Přírodovědecká fakulta Jihočeské Univerzity
v Českých Budějovicích

Kompoziční evoluce genomu strunatců

Radka Symonová

Habilitační práce
2021

Prohlášení

Prohlašuji, že jsem předkládanou habilitační práci s názvem „Komposiční evoluce strunatců“ vypracovala samostatně s použitím uvedených literárních zdrojů a výsledků vlastní vědecké práce. Výsledků vlastní vědecké práce v habilitační práci jsem dosáhla ve spolupráci s kolegy na UHK, či kolegy z jiných pracovišť v ČR a jinde ve světě od doby získání doktorského titulu do sepsání této práce (období 2009 – 2021).

V Týništi nad Orlicí, dne 25. 4. 2021 Mgr. Radka Symonová, Ph.D.

Obsah

Český abstrakt

Symonová R. 2021. Kompoziční evoluce genomu strunatců. Habilitační práce, PřF JU.

Klíčová slova: AT/GC organisace genomu, GC obsah, repetitivní sekvence, chromosomy, transposony, rDNA geny.

Tuto habilitační práci tvoří komentovaná kompilace 19 vědeckých publikací autorky doplněná o dosud nepublikované relevantní výsledky, čímž vytváří ucelený počin ve snaze nastínit mechanismy, které se podílely na kompoziční evoluci genomu strunatců.

DNA kódující proteiny zaujímá jen ca. 1,5 - 5 % jaderného genomu strunatců. Genetický kód se uplatňuje pouze v této malé frakci. Poměr zastoupení jednotlivých nukleotidů A, T, G, C v DNA však plní krom realisace genetického kódu řadu úloh, čímž vytváří další úroveň kompozičního kódu genomu. Po strukturní stránce v DNA rozlišujeme slabé nukleotidové páry A=T resp. T=A tvořící mezi sebou dva vodíkové můstky a silné páry G≡C/C≡G se třemi můstky. Proto sledujeme procento G+C (GC%) vůči A+T, což patří k nejdůležitějším druhově specifickým veličinám při sekvenování genomu. Cytosin (C) je navíc substrát pro metylaci DNA (i RNA), jeden z nejuniversálnějších epigenetických mechanismů regulace genové exprese. GC% asociuje s četnými aspekty biologie genomu: densita genů, načasování replikace DNA, typ a proporce transposonů, proporce CpG a CpG ostrůvků, eu- vs. heterochromatin, rozeznání míst sestřihu exonů, tvorba nukleosomů, intensita rekombinace, intensita exprese genů, počet a délka exonů a intronů na gen, atd. Z cytogenetických a ultracentrifugačních studií víme, že genom savců a ptáků je AT/GC heterogenní, tzn., střídá se v něm DNA bohatší na GC s úseky GC chudšími. Genom ostatních strunatců je AT/GC homogenní a jediné výrazně GC bohaté oblasti jsou 45S rDNA klastry, které mohou nabývat značného počtu kopií a s tím spojené evoluční dynamiky. AT/GC heterogenita savců má zásadní důsledky v lidské cytogenetice a po desetiletí se jí využívá při vyšetření lidského karyotypu. Cytogenetika nižších obratlovců je odkázána na morfologickou podobnost homologních chromosomů a/nebo na techniky fluorescenčního barvení chromosomů. Přesné příčiny a mechanismy vzniku AT/GC heterogenity genomu ptáků a savců zůstávají neznámé a začíná být jasné, že nejde o triviální jev dříve mylně připisovaný jejich vyšší tělesné teplotě. Nejnovější data naznačují, že nejde ani o neadaptivní následek genové konverse (gBGC) při reparativní rekombinaci napodobující selekci, jak se snaží vysvětlit současně nejrozšířenější koncept gBGC. Nejde ani výhradně o důsledky hyper-mutability metylcytosinu extrémně náchylného k deaminaci a konversi. Ve svém výzkumu jsem k vysvětlení těchto otázek přispěla mj. 1. zjištěním, že se savci srovnatelná AT/GC heterogenita existuje i u archaických ryb kostlínů, zatímco jejich nejbližší recentní příbuzný kaproun vykazuje typickou rybí AT/GC homogenitu; 2. jednotlivé linie strunatců zaujímají svou „cytogenomickou niku“, kdy přítomnost mikrochromosomů (paryby, jeseteři, kostlíni, ptáci, a někteří plazi) vede k exponenciálnímu vztahu mezi velikostí chromosomů a jejich GC%, zatímco nepřítomnost mikrochromosomů (savci, ryby, obojživelníci) vede buď k lineárnímu vztahu mezi těmito proměnnými, nebo k absenci vztahu; 3. velikost genomu a chromosomů jsou důležité faktory v kompoziční biologii, ne však pouze ve smyslu gBGC, a jsou determinovány transposony. 4. GC% transposonů ovlivňuje GC% celého genomu, čímž může zapříčinit inversní vztah mezi velikostí a GC%; 5. k vysvětlení původu AT/GC heterogenity savců, ptáků a kostlínů musíme zapojit pluralistický přístup a zahrnout působení selekce; 6. k tomuto úsilí jsme vytvořili *ad hod* bioinformatickou platformu kompoziční biologie strunatců.

Anglický abstrakt

Symonová R. 2021. Compositional genome evolution in chordates. Habilitation thesis, Faculty of Science, University of South Bohemia in České Budějovice

Key words: AT/GC genome organization, GC content, repetitive sequences, chromosomes, transposons, rDNA genes.

This thesis consists of 19 commented research papers of this author and is supplemented with so far unpublished relevant results to produce a comprehensive treatise trying to elucidate mechanisms involved in the compositional genome evolution in chordates.

Proteins coding DNA makes up merely ca. 1.5 – 5 % of the chordate nuclear genome. The genetic code thus works only in this small fraction. The ratio of nucleotides A, T, G, C in DNA is crucial not only for realization of the genetic code, forming a further level of genome composition code. Here, we distinguish weak nucleotide pairs A=T and T=A with two hydrogen bonds between each other and strong pairs G≡C and C≡G with three hydrogen bonds. Hence, the proportion of G+C (GC%) is an important species specific trait accompanying any genome assembly report. Moreover, cytosine (C) is an important substrate for DNA (but also RNA) methylation, one of the most universal epigenetic mechanisms of regulation of gene expression. GC% associates with numerous aspects of genome biology: genes density, DNA replication timing, type and proportion of transposons (TEs), proportion of CpG and CpG islands, eu- vs. heterochromatin, exon splicing site recognition, introns length, etc. Based on cytogenetic and ultracentrifugation studies, we know that mammalian and avian genomes are AT/GC heterogeneous with GC-rich regions alternating AT-rich regions. Genome of other chordates is AT/GC homogeneous and the only GC-rich regions are 45S rDNA clusters, sometimes reaching high copy numbers and evolutionary dynamics. The AT/GC heterogeneity has far-reaching implications in human cytogenetics and has been utilized for clinical karyotyping for decades. This is impossible in lower vertebrates, where karyotyping is based only on morphological similarities of chromosomes and/or fluorescent staining. The exact reasons and mechanisms of the AT/GC heterogeneity in mammals and birds are unknown and it is increasingly clear that they are not caused by a simple factor earlier ascribed to their higher body temperature. The latest results also exclude GC-biased gene conversion (gBGC), a recombination-related mechanisms mimicking selection, as the only one factor involved. Neither any mutational bias nor hypermutability linked to methylcytosines can explain these questions. My own contribution to answer these questions are findings: 1. The mammalian-like AT/GC heterogeneity occurs also in the ancient ray-finned fish lineage gars, whereas their closest related species bowfin shows the typical teleost AT/GC homogeneity; 2. The main chordate lineages occupy their own “cytogenomic niche”, when the presence of microchromosomes (chondrichthyans, sturgeons, gars, birds and some reptiles) means a negative relationship between GC% and chromosome size behaving as if there was no association in large(r) (macro)chromosomes, whereas the absence of microchromosomes (mammals, fish, amphibians) means either a linear or no relationship; 3. Size of genomes and chromosomes are crucial factors of GC%, but not in the terms of the gBGC and are determined by TEs; 4. GC% of TEs shapes GC% of the host genome and can change or invert the relationship between GC% and size; 5. The explanation of the AT/GC heterogeneity in higher vertebrates and gars will require a pluralistic approach and involvement of selection; and finally, 6. We have constructed a bioinformatics platform to solve the above issues.

1. Úvod

1.1. Kompoziční biologie – vymezení oblasti zájmu

Tato habilitační práce se věnuje proporci guaninu a cytosinu, GC%, vůči adeninu a tyminu, AT%, tedy GC kompoziční biologii, na úrovni celého jaderného genomu strunatců. Práce sleduje faktory, které mají vliv na evoluci GC% a zejména na rozdíly mezi teplokrevnými obratlovci a ostatními skupinami. Mezi tyto faktory patří velikost genomu a chromosomů a subchromosomální struktury a oblasti DNA, většinou repetitivní na straně jedné a na ně vzájemně působící mechanismy a fylogenetické vztahy na straně druhé. Celkový obraz evoluce GC% na celogenomové úrovni se snažím ve své dosavadní práci poskládat z dat molekulárně-cytogenetických a genomických a to přímo z dat sekvenačních bioinformaticky a současně meta-analýzou dat o velikosti genomů, chromosomů a jejich GC%. Tento integrativní cytogenomický přístup, kde využívám svou zkušenost z molekulárně-cytogenetické laboratoře a snažím se osvojit a využívat bioinformatiku a statistiku se osvědčil a otevřel nové, dosud neprobádané možnosti moderní biologie. Datům z mitochondriálního event. plastidového genomu eukaryot a genomu prokaryot jsem se dosud aktivně nevěnovala, tudíž nejsou zahrnuta v tomto habilitačním pojednání. Ačkoliv je i GC biologie virů srovnatelně významné a zajímavé téma, tak je evolučně natolik vzdálena, že ho ve své práci rovněž opomíjím, ačkoliv základy GC biologie byly položeny právě při studiu sekvencí virů a prokaryot (vzhledem k jejich velikosti genomu). Totéž platí do určité míry tzv. biologii kódů („Code biology“, *sensu* Barbieri, 2015), která zahrnuje mj. i biologii kodonů. To jsou sice relevantní oblasti biologie, vztahují se však pouze ke kódující frakci jaderného genomu (exomu), jež tvoří jen zlomek jaderné DNA (u člověka 1,15 – 1,17 % Hatje et al., 2019). Protože mi v této fázi mého výzkumu jde o globální pohled na kompoziční evoluci genomu a o faktory formující celogenomovou proporci GC, tak i biologii kodonů (prozatím) opomíjím.

1.2. Význam GC% v genomu

V učebnicích i vědeckých publikacích se běžně uvádí, že tří vodíkové můstky mezi G a C páry (G≡C) jsou důvodem vyšší stability DNA bohaté na GC (např. Snustad & Simmons, 2017; Murray et al., 2006; Vinogradov & Anatskaya, 2017). Ve skutečnosti mají na vyšší stabilitu GC bohaté DNA mnohem větší vliv silné patrové (stacking) interakce způsobené hydrofobními silami mezi nad sebou umístěným G a C (Yakovchuk et al., 2006). Ví se však, že GC bohatá a tedy stabilnější DNA dává vzniku stabilnější RNA a ta i teplotně stabilnějším aminokyselinám (Bernardi & Bernardi, 1986; Costantini & Bernardi, 2008a). GC% DNA také ovlivňuje hydrofobitu kódovaných aminokyselin (Arhondakis et al., 2004). Tím se GC biologie již dostává na pomezí molekulární biofyziky a fyzikální chemie, kde relevantní recentní souhrn poznatků podává např. Vologodskii et al., 2018. Ve struktuře a funkci genomu hraje GC% celou řadu významných rolí: v GC% se odráží densita genů a GC% určuje čas replikace DNA – na geny a GC bohaté oblasti DNA se replikují v časné fázi (Costantini & Bernardi, 2008b), GC% určuje charakter chromatinu a jeho lokalisaci v buněčném jádře – GC bohatá DNA se vyskytuje zejm. ve formě euchromatinu a je umístěna uvnitř jádra, zatímco AT bohatá DNA častěji existuje jako heterochromatin ukotvený v lamině na vnitřním okraji jaderné membrány (Saccone et al., 2002; Federico et al., 2006). Typ chromatinu je dán negativní korelací mezi GC% a potenciálem tvorby nukleosomů (Vinogradov, 2005; Peckham et al., 2007; Jabbari et al., 2019). Jiná situace je u spermií savců, kde regionální variabilita GC% určuje přenos paternální DNA a její modifikace (Vavouri & Lehner, 2011). GC bohatá DNA je u živočichů obecně kvantitativně méně zastoupena – GC% genomu obratlovců nepřesahuje 50%, malé genomy ptáků a čtverzubců jsou GC bohatší než genomy větší, ale nikdy nedosahují 50 % GC (Borůvková et al., 2021). Naproti tomu u bakterií GC% dosahuje rozpětí 25 – 75 % (Sueoka, 1962). GC bohatší geny obsahují méně krátkých intronů a UTR, i GC bohatší exony bývají kratší, i když situace zde je komplexnější a

komplikovanější (Haddrill et al., 2005; Gazave et al., 2007; Zhu et al., 2009; Wang & Yu, 2011; Amit et al., 2012). GC bohatá DNA obsahuje méně transposonů a jejich jiné skupiny, které jsou kratší a GC bohatší (Eyre-Walker & Hurst, 2001). Vychýlení (bias) ve složení nukleotidů přímo ovlivňuje lokální regulační procesy jako je alternativní sestřih (Lemaire et al., 2019). Vzhledem k tomu, že metylace DNA se odehrává zejm. na C, tak GC% určuje množství substrátu pro tuto významnou epigenetickou modifikaci. Vlivem své hypermutability (C→T) metylovaný C, mC, zpětnovazebně ovlivňuje i GC% DNA (Mugal et al., 2015a) a tudíž i transkripční kontrolu transposonů (Zhou et al., 2020). S GC%, mírou transkripce a torsního stresu a s metylací C koreluje i výskyt levotočivé Z formy DNA, jejíž existence je u živých buněk dostatečně podložena ovšem její funkce nikoliv (Vinogradov, 2003; Craig et al., 2021).

Ačkoliv se následující charakteristiky týkají pouze exomu, jsou hodné zřetele v této práci, protože dokreslují význam GC% v té frakci genomu, kde se soustředí selekce: GC% na 3. posici kodonu (GC3) genů ve funkční třídě „metabolismus“ database KOG (Tatusov et al., 2001, 2003) je vyšší než u ostatních genů. Zatímco GC3 genů ve funkční třídě „uchování a zpracování genetické informace“ je signifikantně nižší, geny třídy „buněčné procesy a signalizace“ nevykazují žádný trend. Dále v každé funčkní třídě genů existuje významná positivní korelace mezi GC3 a GC% intronů (GCi), přičemž průměr GC3 je vždy vyšší než průměr GCi korespondujících intronů (D'Onofrio et al., 2007). Navíc AT bohaté geny vykazují vysokou tkáňovou specificitu ve své expresi a krom cílové tkáně je jejich transkripce většinou umlčena. Naopak mezi GC bohaté geny patří tzv. housekeeping geny, které jsou mj. vzhledem k jejich GC% méně metylované a tudíž více transkripnčně aktivní (Vinogradov, 2003; Vinogradov & Anatskay, 2017).

Z technického hlediska je GC% a jeho lokální heterogenitu také potřeba vést v patrnosti při rekonstrukci fylogenetických stromů, při detekci selekce v sekvenci DNA a při odhadu využití kodonů (Romiguier & Roux, 2017). GC% mělo dále vliv i na kvalitu sekvenace genomu, kdy extrémně GC bohaté oblasti DNA nebylo možné prosekvenovat (Peona et al., 2018). S rozvojem další generace sekvenačních technologií lze doufat, že tyto komplikace budou stále méně a méně časté (Amarasinghe et al., 2020), ačkoliv zatím stále pracujeme s genomy vytvořenými předchozími generacemi sekvenování.

1.3. Vývoj v pojetí kompoziční biologie

V průběhu posledních desetiletí se přístup ke kompoziční biologii značně lišil, což bylo dané zejména dostupnými metodami a s tím spojenými daty. Zpočátku šlo o popis vztahů mezi jednotlivými nukleotidy v nukleových kyselinách (odstavec 1.3.1.). V posledních dekádách již dominují snahy o pochopení kompoziční variability uvnitř genomu a mezi genomy teplokrevných a ostatních živočichů. Tyto snahy vyústily v návrh až pěti evolučních hypotes s cílem vysvětlit pozorovanou kompoziční variabilitu. Tyto hypotesy nebo koncepty lze rozdělit na adaptivní (selekční) a neadaptivní (selekčně neutrální) (*sensu* Eyre-Walker & Hurst, 2001) nebo intra- a extra-celulární (*sensu* Berná et al., 2012).

1.3.1. Chargaffova pravidla a „homostabilisační tendence“ genomu

Průkopníkem snah o kvantifikaci nukleotidů v DNA zejm. virů a bakterií byl Erwin Chargaff, který formuloval základní principy, pro jejichž označení se později vžil termín „Chargaffova pravidla“. Ta souhrnně praví, že DNA každého druhu má stechiometrický podíl purinů (A+G) a pyrimidinů (T+C) 1 : 1 (použité zkratky jsou vysvětleny v tabulce Tab. 1):

I. První pravidlo parity je druhově nespecifické a říká, že v dvoušroubovici DNA celého genomu platí A% = T% a C% = G% (Chargaff 1950, 1951; empiricky potvrzeno prací Watson & Crick, 1953).

II. Druhé pravidlo parity je také druhově nespecifické a říká, že v jednom řetězci DNA (ssDNA) celého genomu platí A% ≈ T% a C% ≈ G% (Rudner et al., 1968; empiricky ověřeno konceptem the Symmetry

Principle, Prabhu, 1993). Platí u 4 z 5 typů dvouřetězcových genomů (dsDNA) – eukaryota, bakterie, dsDNA viry a archea, ale neplatí u genomů organel (mtDNA a ptDNA; Mitchell & Bridge, 2006).

III. Klastrové pravidlo je druhově nespecifické a praví, že v ssDNA se nejméně 60% pyrimidinů vyskytuje v klastrech po třech a více v řadě za sebou. Vzhledem ke komplementaritě basí toto platí i pro puriny (Chargaff, 1963; dále rozpracováno v práci Szybalski et al., 1966 a dalších pracech 1. autora).

IV. Pravidlo GC% říká, že v ssDNA (G+C)/(A+T+G+C) neboli S/(W+S) má tendenci být konstantní a je druhově specifické (Chargaff, 1951, 1979).

	Purin (R)	Pyrimidin (Y)
slabé (W)	adenin A	= T tymin/U
silné (S)	guanin G	≡ C cytosin

Tab. 1 Základní rozdělení nukleotidů a jejich vlastnosti relevantní pro tuto práci podle IUPAC (Cornish-Bowden, 1985). Demetylovanou formou T je uracil (U).

Krom prvního Chargaffova pravidla do určité míry (zejm. u vyšších eukaryot) upadla v zapomnění, pravděpodobně proto, že Chargaff předběhl dobu a tématem se zabýval dlouho před rozvojem rutinního sekvenování DNA. Jeho práci však „vzkřísil“ profesor Donald R. Forsdyke jednak svým review (Forsdyke & Mortimer, 2000) a zejména pak svou rozsáhlou monografií Evolutionary Bioinformatics věnovanou právě Chargaffovi (a Susumo Ohnovi; Forsdyke, 2016). Forsdyke na základě starších, jím citovaných studií, vytvořil koncept „homostabilisační tendence“ DNA, který praví, že relativně GC% uniformní oblasti působí jako isolátory bránící homologní rekombinaci mezi duplikovanými geny, čímž umoží jejich sekvenční a funkční diferenciaci a zabrání jejich ztrátě (Forsdyke, 2004, 2016). Forsdyke tak vytvořil koncept, který do určité míry kompensuje nedostatky níže uvedených hypotes (např. bere v potaz GC% okolních sekvencí, flanking regions), avšak vůbec mu není věnována adekvátní pozornost mainstreamových studií (viz níže).

1.3.2. Koncept isochor a komposičních domén

Další etapa GC biologie je neodmyslitelně spjata s úctyhodným dílem profesora Giorgia Bernardiho. Jeho tým nejdříve metodou analytické ultracentrifugace DNA v CsCl gradientu zjistil komposiční heterogenitu DNA eukaryot (Macaya et al., 1976; Thiery et al., 1976; Cuni et al., 1981; Bernardi & Bernardi, 1985). Genom teplokrevných obratlovců označil jako mosaiku isochor – typicky 300 a více kB dlouhých úseků DNA s charakteristickým GC%, kterým se zásadně lišily od sousedních úseků (Cuny et al., 1981). Srozumitelnou technickou definici isochor podává Oota et al., 2010: v porovnání s pouhou distribucí GC% sekvencí isochory charakterisuje jako „plató hodnoty dosažené standardní odchylkou GC% náležících různým rodinám isochor“. Takto by výsledky měly reflektovat prostorové struktury ve smyslu původní definice Bernardiho a kolegů jako nenáhodnou prostorovou variabilitu v GC% celého genomu (Bernardi, 2005; Oota et al., 2010; Cozzi et al., 2015). Isochory Bernardi klasifikoval pro lidský genom do pěti „rodin“ dle jejich GC% na lehké L1 - L2 a těžké H1 - H3 (Bernardi, 2005). V genomu ptáků identifikovali ještě na GC nejbohatší, tedy „nejtěžší“, isochoru H4 (Costantini et al., 2009). V genomech ryb a obojživelníků Bernardi se spolupracovníky našli jednak žádné nebo malé zastoupení H isochor a současně zjistili výskyt pouze dvou sousedících isochor, např. L1 a L2 pro dánio, L2 a H1 pro medaku, H1 a H2 pro koljušku a čtverzubce, z čehož odvodili a zobecnili AT/GC homogenitu rybích genomů (Bernardi, 2008; Costantini et al., 2007, 2009 – tyto a další publikace, které zde pro stručnost již neuvádím, shodně presentují stejné grafy pouze výše uvedených ***čtyř (!)*** druhů ryb, ačkoliv sekvenovaných genomů výrazně přibývalo). AT/GC homogenitu rybích genomů jsme potvrdili na podstatně větším vzorku druhů (Borůvková et al., 2021), avšak situace je komplexnější (např. Symonová et al., 2017a, viz níže). Naopak savčí a ptačí genomy označili Bernardi se spolupracovníky za

AT/GC heterogenní s výskytem jak AT, tak GC bohatých oblastí a s širším rozpětím zastoupených rodin isochor. Přítomnost na GC bohatých isochor připsali adaptaci (tedy následkem selekce) na jejich zvýšenou tělesnou teplotu, tzv. hypotesa termodynamické stability (Bernardi & Bernardi, 1986; Bernardi, 2005; Costantini et al., 2009). Tu se jim nějak dařilo udržovat při životě, protože ptáci mají vyšší tělesnou teplotu a jejich genomy jsou GC nejbohatší (Kadi et al., 1993), zatímco oposum, ježura a ptakopysk mají nižší teplotu a zásadně se liší v GC% - ježura a ptakopysk jsou výrazně GC bohatší než např. člověk (Zhou et al., 2021), ačkoliv tělesná teplota ptakopyska i ježury je nižší než u ostatních savců (Smyth, 1973; Grigg et al., 2003). Naopak oposum má genom GC chudý (Samollow, 2008). Navíc hledali a nacházeli určitou podporu pro svá tvrzení u ryb žijících za různých teplot (Jabbari & Bernardi, 2004; Bucciarelli et al., 2009), ačkoliv jiné práce jejich výsledky interpretovaly zcela odlišně (jen pro příklad Berná et al., 2013). Svá zjištění zobecnili tak, že nárůst enivronmentální nebo tělesné teploty spouští nárůst GC% stabilizující DNA, RNA i jí kódované aminokyseliny (Bernardi, 2005). Současně Bernardi odmítá koncept GC-vychýlené genové konverse (gBGC; Mugal et al., 2015b) jakožto majoritně přijímané a nejrobustnější vysvětlení mechanismu vzniku AT/GC heterogenity savců a ptáků a homogenity nižších obratlovců (viz odstavec 1.3.5). Dále Bernardi systematicky ignoruje skutečnost, že geny pro rRNA (rDNA) představují universálně GC nejbohatší frakci genomu a patří mezi DNA s nejvyšší intensitou rekombinace, tedy s vysokou incidencí gBGC (Escobar et al., 2011; Varriale et al., 2008), což přesně odpovídá konceptu gBGC. S dalším rozvojem sekvenování a bioinformatiky Bernardiho tým hledal podporu pro svůj koncept isochor v genomech jenže počet analysovaných druhů ryb (ani dalších obratlovců) nerozšířil. Obdivuhodný počet publikací na toto téma mohl naznačovat, že se jim jejich úsilí dařilo. Avšak s přibývajícími alternativními přístupy ke GC biologii jim byla stále více vyčítána „subjektivní intervence uživatele“ (Elhaik & Graur, 2014) a snaha najít to, hledali. Ostatní autoři totiž tak uspokojivých výsledků při verifikaci konceptu isochor nedosahovali, což vyústilo v četné disputace prof. Bernardiho s ostatními autory, kteří k této problematice zaujímali na rozdíl od něho kritický a objektivní postoj. Bernardiho práce pak již vesměs citoval (a cituje) on sám nebo jeho spolupracovníci a jeho četné odpovědi autorům kritických článků (např. Bernardi, 2001; Clay & Bernardi, 2005) zůstávaly bez reakcí. Výsledkem je, že část vědecké komunity Bernardiho práci stále zmiňuje jako relevantní (Vinogradov, 2005; Oota et al., 2010; Amit et al., 2012; Lemaire et al., 2019), část novějších publikací jeho práci zmiňuje jako přežitou a již neplatnou (např. Hurst & Merchant, 2001; Cohen et al., 2005; Chojnowski et al., 2007) a část na jeho výsledky navázala a v modifikované formě je nadále cituje (Melodelima & Gautier, 2008; Forsdyke, 2016), mj. i Bernardi sám (Bernardi, 2015; Jabbari & Bernardi, 2017; Bernardi, 2018; Bernardi, 2019). Téměř všechny Bernardiho práce jsou ke stažení na jeho webu http://www.giorgiobernardi.eu/. Z pohledu této habilitační práci a jejich cílů je významná monografie autora „Structural and Evolutionary Genomics: Natural Selection in Genome Evolution“ (Bernardi, 2005; pdf na výše uvedeném linku), neboť přehledně a detailně shrnuje jejich poznatky. Avšak původní (neo)selekcionistický koncept isochor v pojetí prof. Bernadiho (Bernardi, 2007) a jeho spolupracovníků byl překonán a některými autory značně modifikován (např. Hughes et al., 1999; Fortes et al., 2007). K tomuto kroku přispěla i moje dosavadní práce a to nejen prokázáním savčího typu uspořádání isochor u archaických studeokrevných kostlínů (Symonová et al., 2017a). Zásadní modifikací konceptu isochor představuje model komposičních domén (komposičně homo-genní domény a komposičně nehomogenní domény, Elhaik et al., 2010a,b, 2014). U člověka byl subset homogenních domén arbitrárně definován jako ≥300 kb a nazván isochorické domény (*sensu* Bernardi, 2005). Tito autoři uvádí, že 1/3 genomu člověka je tvořena nehomogenními doménami a zbylé 2/3 genomu představují „směs“ četných krátkých homogenních domén a relativně nízkého počtu domén dlouhých. Isochorické domény pak tvoří méně než 1/3 lidského genomu. Podobná situace byla zjištěna i u genomu tura (Elsik

et al., 2009). Další pokračování hledání kompozičních domén však ustalo, stejně jako tomu bylo u isochor a zůstalo tak mnoho nezodpovězených otázek.

1.3.3. GC% a vyšší struktury organisace eukaryotického jádra

Rozvoj moderních metod analysy 3D organisace eukaryotického jádra (ultramikroskopie a sekvenace interagující DNA Hi-C) vedl k dalšímu posunu v chápání funkčních struktur genomu. Nejdříve byla objevena a detailně charakterisována tzv. chromosomová teritoria, která zaujímají jednotlivé chromosomy v interfásním jádře savců (Cremer & Cremer, 2001). Už v této fási se objevují pojítka s GC% a faktory, které s GC% korelují, jako např. densita genů, čas replikace DNA, atd. (Cremer & Cremer, 2010) a s tím související AT/GC heterogenita interfásního jádra napříč fylogenetickým stromem obratlovců, tedy ne pouze u savců a ptáků (Saccone et al., 2002 a Federico et al., 2006). Při hledání jemnějšího členění genomu byly následně definovány dva základní typy domén chromatinu na subchromosomální úrovni: 1. topologicky asociované domény TAD, jejichž DNA mezi sebou interagují víc než s jinými doménami a 2. s laminou asociované domény LAD, které úzce interagují s laminou jaderné membrány (Pope et al., 2014) a jsou AT-bohaté (Jabbari & Bernardi, 2017). I zde nacházíme nemalý vliv prof. Bernardiho a jeho následovníků, kteří se snaží etablovat isochory v jejich tradičním pojetí jako „základní úroveň struktury a organisace genomu“ (Costantini & Musto, 2017) a tvrdí, že i. TAD a LAD domény odpovídají isochorám a podléhají chromatinovým doménám (eu- a hetero-chromatin); ii. TAD a LAD domény existují v genomech savců díky evolučnímu zakonservování isochor; a iii. chromatinové domény korespondující s GC chudými isochorami interagují s ostatními GC chudými doménami i na větší vzdálenost na chromosomech, zatímco domény korespondující GC bohatým isochorám vykazují více lokalisované interakce (Jabbari & Bernardi, 2017). Zde narážíme na výše popsaný rozpor v pojetí isochor a jejich nahrazení konceptem kompozičních domén některými autory (Elhaik et al., 2010a,b, 2014). Jde zejm. o skutečnost, že ačkoli „tradiční isochory“ *sensu* Bernardi, 2005 v genomu dosud studovaných savců zaujímají prokazatelně pouhou 1/3 DNA (Elhaik et al., 2010a,b, 2014), tak Bernardi dokázal „své“ isochory „udat“ v celém genomu. Koncept TAD a LAD se etabloval a je široce přijímán. Souvislosti mezi GC% a vnitřním uspořádáním TAD již nezávisle na Bernardim prokázali Jabbari et al., 2019 (spolupracovník Bernardiho). Nejnovější poznatky o supramolekulárním uspořádání DNA v chromosomech eukaryot i bakterií a archeí jsou shrnuty v práci Birnie & Dekker, 2021. Z dosud publikovaných údajů lze usuzovat, že GC% se přinejmenším do určité míry, přímo či nepřímo, uplatňuje při 3D organisaci interfásního jádra (Di Stefano et al., 2016; Das et al., 2020). Díky stále přibývajícímu množství dostupných genomů dostatečné kvality v této oblasti tak zůstává mnoho prostoru pro další výzkum, kterému se věnuji v rámci mého EU projektu na TUM.

1.3.4. Mutační vychýlení a (hyper)mutabilita metylcytosinu a hydroxymetylcytosinu

Mezi selekčně neutrální, resp. neadaptivní hypotesy komposiční evoluce genomu patří „koncept mutačního vychýlení“ (Sueoka, 1988). Tento koncept si všímá časté skutečnosti, že některé typy mutací se vyskytují častěji, než se předpokládá (Stoltzfus et al., 2009). I zde existuje debata o roli mutačního vychýlení při adaptaci, ačkoliv je tento mechanismus dosud považován za vesměs neutrální (Svensson & Berger, 2019; Storz et al., 2019). Tento koncept se poslední desetiletí stále více překrývá s nejuniversálnější epigenetickou modifikací – DNA metylací (Mugal et al., 2015a; Storz et al., 2019). Cytosin ale hlavně 5-metylcytosin (mC) vykazují vysokou míru spontánní hydrolytické deaminace. mC vytváří v genomu savců mutační hotspoty a produktem jeho deaminace je tymin (T; Lutsenko & Bhagwat, 1999). Reparace takové C→T transice je méně efektivní než reparace transice C→U a tudíž má dalekosáhlé následky a to i evoluční (Fryxel & Zuckerkandl, 2000). Na druhou stranu, oxidativní derivát mC, hydroxymetylcytosin (hmC) mutuje C→G transversí (Supek et al., 2014). Evoluce rozsahu

metylace DNA a (de)metylačních enzymů napříč strunatci by do budoucna mohla být jedním z vodítek k pochopení jejich GC biologie (Iwasaki et al., 2014; He et al., 2015). Dalším pojítkem mezi regulací genové exprese a GC% jsou faktory doprovázející transkripci, které přímo souvisí s GC% - bendabilita a zakřivení (curvature) DNA a potenciál přechodu mezi B a Z konformacemi DNA (Vinogradov, 2003).

Klasické pojetí konceptu mutačního vychýlení se více soustředí na chyby při replikaci DNA během její časné a pozdní fáse, přičemž byla zjištěna tzv. mutační spektra specifická pro každou fázi, která mj. odrážela dostupnost volných nukleotidů (Kenigsberg et al., 2016). Načasování replikace DNA přímo podléhá regionálnímu GC% (Costantini & Bernardi, 2008b). K mutačním hypotesám lze přičlenit i koncept distribuce zlomů v DNA (BPR, DNA breakpoints distribution; Lemaitre et al., 2009). Výsledky velkého projektu podávají robustní podporu myšlenky, že celogenomové složení nukleotidů je silně ovlivněno mutacemi napříč genomem a že usměrňující síly nad mutacemi (selekce a/nebo GC-vychýlená genová konverse, viz níže) rovněž hrají svou nezanedbatelnou roli (Long et al., 2018).

1.3.5. gBGC

Současně nejšířeji přijímaným a nejrobustnějším konceptem vysvětlujícím AT/GC heterogenitu uvnitř a mezi genomy obratlovců je GC-vychýlená genová konverze (gBGC, GC-biased gene conversion; Eyre-Walker, 1993; Meunier & Duret, 2004; Galtier & Duret, 2007; Duret & Galtier, 2009). Jde o selekčně neutrální, neadaptivní, intracelulární mechanismus reparace DNA závislý na meiotické rekombinaci, který připomíná selekci, jsou-li následky selekčně neutrální nebo pouze mírně škodlivé (Galtier et al., 2009). gBGC má tendci zvyšovat odhadované tempo adaptivní substituce (Rousselle et al., 2019). Vyskytne-li se při meiose v heteroduplexu DNA neshoda mezi páry basí, jsou do DNA preferenčně začleněny alely bohatší na GC bez ohledu na to, jaké následky na fitness tato změna má (Kostka et al., 2012). Výše uvedené by se v širším slova smyslu dalo interpretovat jako mutace (Flegr, 2018), ale mechanismus gBGC je považován za tak významný, že je z mutačních konceptů vyčleňován jako samostatně stojící (např. Mugal et al., 2015b). Tento zdánlivě nevýznamný proces však nabývá na síle se zvyšující se populační velikostí (Mugal et al., 2015b). gBGC v sobě integruje populačně-genetické a demografické aspekty, které přičítají nárůst velikosti genomu (např. eukaryot vs. prokaryot, event. u savců) poklesu populačních densit a s tím spojenému sníženému selekčnímu tlaku na co nejmenší velikost genomu (Lefébure et al., 2017). Ruku v ruce s těmito faktory jsou u savců dále uváděné délka trvání životního cyklu, vlastnosti chromosomů a velikost těla (Romiguier et al., 2010). Zatím ani tento koncept nedokáže uspokojivě vysvětlit AT/GC heterogenitu amniot a homogenitu ostatních živočichů (Bohlin et al., 2019; Hurst, 2019). Pro gBGC svědčí široce doložená positivní korelace mezi GC% a intensitou rekombinace, která se v rámci genomu může lišit podobně jako GC% (Stapley et al., 2017; Latrillot, 2013) a fakt, že počty GC → AT mutací se nerovnají počtu AT → GC mutací (Lander et al., 2001; Alvarez-Valin et al., 2002), jiná věc je procento fixace těchto mutací v genomu (Rousselle et al., 2019; Paudel et al., 2020). K nejzásadnějším nedostatkům gBGC konceptu patří zejm. skutečnosti, že frekvence synonymních substitucí je kovariantní s GC%, že existují staré GC bohaté geny na chromosomu Y, jehož rekombinace je zanedbatelná a že parametry modelů molekulární evoluce s gBGC jsou vysoce sensitivní (Eyre-Walker & Hurst, 2001). Variabilita v intensitě gBGC má potenciál vysvětlit nukleotidové substituce na neutrálních místech DNA ve spojení s velikostí těla a znaky karyotypu, avšak pouze u placentálních savců a nikoliv u ptakořitných a vačnatců (Latrillot, 2013). Mechanismus gBGC se může uplatňovat i v akceleraci substitucí a tedy i v celkovém tempu evoluce (Kostka et al., 2012), které by potenciálně mohlo mít rovněž vliv na GC% genomu, avšak dosud mu v tomto ohledu nebyla věnována pozornost. Proces obecné genové konverse (tedy ne té pouze GC-vychýlené) je jevem velmi častým týkajícím se většiny genů, který má výrazně vyšší frekvenci než

samotné crossingovery, které jsou předpokladem gBGC. Proto existují indicie, že se na genové konversi podílí ještě další molekulární mechanismus (Storlazzi et al., 1995). U savců a některých dalších obratlovců (ale ne u ptáků a obojživelníků) lze alespoň část regionální variability rekombinačních hotspotů připsat aktivitě proteinu PRDM9 (Baker et al., 2017). Tento fakt by se do budoucna měl stát dalším zohledněným parametrem studia GC biologie genomu.

1.3.6. Metabolická hypotesa a metabolické aspekty biologie genomu

Metabolická hypotesa má své kořeny v rozsáhlé práci prof. A. E. Vinogradova, kterou následně dále rozpracovala skupina spolupracovníků původně z okruhu prof. G. Bernardiho avšak zcela nezávisle na jeho pojetí GC biologie. Primárním východiskem tohoto konceptu jsou dvě vlastnosti DNA, které signifikantně korelují s GC%: i. DNA bendabilita koreluje positivně (Vinogradov 2001) a ii. potenciál tvorby nukleosomů koreluje negativně (Vinogradov 2005; Peckham et al., 2007). Mj. následkem těchto dvou vlastností je na GC nejbohatší DNA transkripčně nejaktivnější krom toho, že obsahuje také nejvíce genů (Versteeg et al., 2003; Arhondakis et al., 2004). Vyšší GC% sekvence vede k vyšší resistenci vůči torsnímu stresu, která je významná zejm. během transkripce, protože redukuje risiko zlomů DNA (Vinogradov, 2003). Dalším východiskem metabolické hypotesy je, že intensita metabolismu souvisí s molekulárními procesy, jako jsou reparace DNA a intensita mutací (Berná et al., 2013). To se odvíjí od skutenčosti, že metabolismus produkuje volné radikály kyslíku, které jsou vysoce reaktivní díky svým volným elektronům a způsobují oxidativní mutace DNA. Tyto mutace jsou sice průběžně kompenzovány reparačními mechanismy, avšak během reparace mohou vznikat mutace další. Tudíž druhy s intensivním metabolismem by měly mít vyšší tempo substitučních mutací (Gillooly et al., 2001, 2005). Nejdříve byla zjištěna signifikantní positivní korelace mezi GC% a intensitou metabolismu korigovanou na teplotu prostředí Boltzmanovým faktorem u 154 druhů kostnatých ryb z různých habitatů. Bylo zjištěno, že GC% genomu klesá od polárních k tropickým rybám (Uliano et al., 2010). Fakt, že tropické ryby mají nejnižší GC%, tato zjištění odporují Bernardiho hypotese termodynamické stability (Bernardi, 2005) podobně jako jiné studie (např. Vinogradov, 2003; Belle et al., 2020). Další studie ukázala, že tři kategorie genů savců se liší svou hodnotou GC% na 3. posici kodonu (GC3) – nejvyšší GC3 byla popsáno u genů zapojených do metabolických procesů, střední GC3 u genů buněčných procesů a signalisace a nejnižší GC3 u genů pro uchování a zpracování informace (Berná et al., 2012). Tato zjištění jsou autory interpretována jako „footprint of metabolism in the organisation of mammalian genomes“, jak zní název uvedeného článku Berná et al., 2012 a jsou považována za důkaz proti platnosti selekčně neutrálního konceptu gBGC (Berná et al., 2013). Následovala studie z oblasti selekce ve prospěch ekonomičnosti a efektivnosti transkripce, která navrhuje metabolickou hypotesu při hledání důvodů vnitro- a mezigenomových rozdílů v GC% - Charausia et al., 2014 zjistili, že průměrná délka intronů pěti druhů kostnatých ryb se zkracuje s jejich rostoucím GC% a oba tyto údaje positivně korelují s intensitou metabolismu. Ačkoliv autoři výše uvedených studií podporujících metabolickou hypotesu tvrdí, že jejich hypotesa je ta nejlepší a nejuniversálnější, tak se tomuto přístupu v komposiční biologii nedostává takové pozornosti, jak by si asi představovali. Svědčí o tom mj. malý počet jejich citací, navíc fakt, že od r. 2016 (viz níže) jinou práci na toto téma již nepublikovali.

Dostupnost energie pomáhá pochopit rozdíly v evoluci malých genomů prokaryot a velkých genomů eukaryot, která využívají dodatečné zdroje energie z mitochondrií a plastidů (Lane & Martin, 2010). Proto se metabolická hypotesa jeví opodstatněná i při hledání rozdílů mezi AT/GC homogenními genomy studenokrevných a AT/GC heterogenními genomy teplokrevných obratlovců, kteří se energeticky liší významně (Willmer et al., 2000; van der Meer, 2021). K tomuto úsilí jsem se snažila přispět studií, která testuje srdeční index a intensitu dýchání u AT/GC homogenního kaprouna a AT/GC

heterogenních kostlínů (*sensu* Tarallo et al., 2016) a výsledky porovnává s ostatními rybami a teplokrevnými obratlovci (Tarallo et al., *in prep*). Naše dosavadní zjištění však nenasvědčují tomu, že by kostlíni měli efektivnější metabolismus než kaproun a ostatní ryby, což by jim umožnilo udržovat energeticky náročnější GC bohaté oblasti genomu. Navíc existují náznaky, že intensita metabolismu může u obratlovců do určité míry asociovat s velikostí genomu (Gardner et al., 2019). Velikost genomu je jeden ze zásadních parametrů komposiční evoluce, který představuje významnou proměnnou v evoluci genomu (Elliot & Gregory, 2015). I v případě velikosti genomu se uplatňuje energetický aspekt podmíněný dostupností základních stavebních kamenů (dusík, fosfor) DNA v prostředí (Auboeuf, 2020) a vliv ekologických proměnných podobný tomu, který byl výše popsán u ryb v případě GC% (např. Tarallo et al., 2016) a který je znám např. u obojživelníků (Lertzman-Lepofsky et al., 2019), viz např. recentní studie Vieira dos Santos et al., 2021.

1.3.7. Velikost genomu, celogenomové duplikace a transposony

Velikost genomu představuje sama o sobě obsáhlé téma se svými četnými koncepty, hypotesami a nezodpovězenými otázkami (viz recentní review Blommaert, 2020). Platí, že velikost genomu *eukaryot* nekoresponduje s komplexitou organismu, tzv. „C-value enigma“, dříve známé jako „C-value paradox“ (C-value udává velikost haploidního genomu v pikogramech, čili cytologicky stanovenou; Elliot & Gregory, 2015). Explicitně vztah mezi velikostí genomu a GC% zkoumal Vinogradov, 1998 a zabýval se jím Bernardi se svými spolupracovníky, např. Musto et al., 2006, jejichž výsledky následně upřesnil Mitchell, 2007 s tím, že ani u prokaryot neexistuje tak zřetelná korelace, jak oni ukázali, a že situace je komplexnější. První a krom naší práce Borůvková et al., 2021 dosud jedinou kvantitativní studií na toto téma je Li & Du, 2014, která se zaměřila také na prokaryota a z eukaryot analysovala zanedbatelné množství tehdy dostupných druhů, kde popsala negativní vztah mezi GC% a velikostí genomu u živočichů. Vztah mezi velikostí genomu a GC% lze pojmout z několika rovin: i. Nejobecněji existuje nepřímá úměra mezi GC% a velikostí jak genomu a chromosomů, tak např. intronů, exonů, genů, ale i transposonů (např. Charausia et al., 2010 a četné uvedené citace tamtéž, dále např. Borůvková et al., 2021); ii. Velikost genomu spolu s počtem a morfologií chromosomů determinují intensitu rekombinace na megabasi DNA (Stapley et al., 2017) a tudíž sílu gBGC (Mugal et al, 2015b). To je podmíněno výskytem alespoň jednoho crossingoveru na chromosomové raménko při meiose. Pak většinou platí, že čím intensivnější gBGC, tj. na čím menší část DNA připadá rekombinační událost, čili čím menší raménko, tím vyšší GC% (Mugal et al, 2015b); iii. Tento v určitých situacích triviální lineární vztah mohou zásadně narušit až převrátit repetitivní sekvence, zejm. transposony (TE), které hrají významnou úlohu v evoluci genomu obratlovců (Pappalardo et al., 2021). Pak přestává ve výjimečných případech platit bod i., kdy velké genomy s velkými chromosomy mohou být GC bohaté – *Thalassophryne amazonica* (NCBI, 2018), blatňák *Umbra pygmaea* (NCBI, 2018), axolotl *Ambystoma mexicanum* (Smith et al., 2019), lososovité ryby (např. Lien et al., 2016; Christensen et al., 2020) atd. Zde je nutno podotknout, že výše uvedené příklady „velkých genomů“ a „velkých chromosomů“ u *T. amazonica*, *U. pygmaea* a lososovitých se týkají výhradně ryb. U savců jsou tyto velikosti odlišné, viz níže. Axolotl má gigantický genom i chromosomy i v porovnání se savci (např. Borůvková et al. 2021).

Globálně byla velikost genomu strunatců zásadně ovlivněna dvěma celogenomovými duplikacemi (WGD) v linii vedoucí k obratlovcům (např. Sacerdot et al., 2018; Simakov et al., 2020). Ty byly následovány dalšími WGD specifickými již jen pro některé linie (jeseteři, kostnaté ryby a v rámci nich lososovité a kaprovité ryby, atd.; Le Comber & Smith, 2004; Glasauer & Neuhauss, 2014, dále obojživelníci, Mable et al., 2011). Tyto události byly spojeny buď s následnou redukcí genomu (kostnaté ryby), nebo naopak s další expansí podmíněnou reaktivací TEs jako např. u lososovitých ryb (např. Lien

et al., 2016; Christensen et al., 2020), blatňáků rodu *Umbra* (Lehman et al., *in prep*), *Thalassophryne amazonica* (NCBI, 2018), obojživelníků (např. Nowoshilov et al., 2018)). Čím dále se pohybujeme po fylogenetickém stromu strunatců od basálních WGD, tím více přebírají kontrolu na velikostí genomu transposony (TEs) a jejich vlastní GC% vliv na GC% hostitelského genomu. Dobrými příklady jsou ptáci a jejich kompaktní na GC bohatý genom (Kapusta et al., 2017) a letouni, u kterých TEs kompletně ovládly cytogenomickou evoluci, tzn. velikost chromosomů i GC%, a převrátily výše uvedený lineární vztah (Andreas et al., *in prep*). Existenci takových případů je třeba mít na paměti, neboť mohou narušit jinak lineární vztah mezi GC% a velikostí chromosomů/genomu a tím vzbudit mylný dojem, že žádný vztah neexistuje. Podobnou situaci jsme ukázali u lososovitých ryb, jejichž začlenění do dat má stejný efekt při analyse kostnatých ryb globálně (Symonová & Suh, 2019; Gaffaroglu et al., 2020; Borůvková et al., 2021). Velikost genomu do značné míry determinuje velikost jednotlivých chromosomů a tím i GC%, ačkoliv ne vždy je tomu tak a existují výjimky – tři v našich podmínkách patrně nejznámější kaprovité ryby mají všechny C ≈ 1,8 pg (Gregory, 2021) a GC ≈ 37% (NCBI, 2018), avšak dánio pruhované (Cyprinidae, Danioninae *sensu* Nelson et al., 2016) má diploidní počet chromosomů 2n = 48 zatímco kapr obecný a karas zlatý (oba Cyprinidae, Cyprininae *sensu* Nelson et al., 2016) mají 2n = 100. Velikost chromosomů těchto ryb se výrazně liší a je patrná i v cytogenetických publikacích (např. Knytl et al., 2013; Sola & Gornung, 2001). Tento rozdíl lze ji vysvětlit masivní expansí GC chudých transposonů u dánia (Howe et al., 2013) a celogenomovou duplikací se zachovaným, tedy nediploidisovaným, počtem chromosomů u karase a kapra (Li & Guo, 2020). Malé ale na GC chudé chromosomy mají i pláštěnci rodu *Ciona* (NCBI, 2018; Borůvková et al., 2021).

1.4. Cíle této habilitační práce

Z předchozích odstavců by mělo být patrné, jak významným parametrem biologie a funkce genomu GC% je a tudíž o jak zásadní výzkum v našem pochopení evoluce a fungování (nejen) lidského genomu se jedná. Tato habilitační práce by měla k tomuto výzkumu přispět a vytvořit výchozí bod pro další výzkum. Za cíle této práce jsem si stanovila:

i. Definovat přetrvávající otázky v oblasti komposiční evoluce živočichů jakožto předstupeň k širšímu integrativnímu uchopení tohoto tématu globálně napříč skupinami eukaryot (zejm. Úvod)

ii. Kritické zhodnocení dosavadních přístupů k výše uvedeným otázkám, ačkoliv je nutné uznat, že nashromáždit všechny více či méně relevantí dosud publikované studie se v tuto chvíli nemůže stroprocentně povést, vzhledem k jejich obrovskému množství (zejm. Úvod)

iii. Vytvořit souhrn dosud doložených proměnných korelujících přímo či nepřímo s GC% a podat výčet těchto proměnných v širším kontextu reflektující jejich event. podobnosti (tj. ne jako pouhý seznam položek) a tento souhrn v budoucnu nadále rozšiřovat

iv. Výčet důkazů a indicií působení selekce na GC% mimo neoselekcionistickou argumentaci prof. Bernardiho jakožto kompensace početných studií zastánců selekčně neutrálních mechanismů v komposiční biologii, kteří však systematicky opomíjejí ostatní linie výzkumu (zejm. Diskuse, ale i Úvod)

v. Kritické zhodnocení vývoje a aktuálního stavu mého přínosu k těmto otázkám včetně souhrnu dosavadních výsledků mého současného stipendia EU´s Horizon 2020 Research and Innovation Programme v rámci Marie Sklodowska-Curie grantu 754462 v kategorii EuroTechPostdoc Programme Excellence in Science and Technology řešeného na Technické Univerzitě Mnichov a Dánské Technické Univerzitě v Kodani (2019-2021) s názvem „Vertebrates´ GC biology – linking patterns with functions“.

vi. V konečném důsledku má tato práce sloužit jako „state-of-the-art information hub“ pro výše definovanou oblast výzkumu a přispět k dalšímu posunu našich znalostí vpřed. Primárně má sloužit jako „úložiště“ dosud publikovaných informací a studií, které jinak zatím nebyly pojednány dohromady, což vzhledem k nepostradatelnosti interdosctiplinarity komposiční biologie představuje jednu ze současných překážek.

2. Metodika komposiční biologie

Metody komposiční biologie lze rozdělit na kvalitativní (cytogenetické založené na barvení chromosomů AT- nebo GC-specifickými fluorochromy, např. Symonová et al., 2017a) a kvantitativní. Kvantitativní metody mají svůj počátek v ultracentrifugaci DNA v CsCl gradientu (shrnuto v Bernardi, 2005). Rovněž ultracentrifugace však neumožňuje dosáhnout vyššího rozlišení vzhledem k tomu, že experimenty probíhají ve zkumavce (Li & Du, 2014). Další možností kvantitativního stanovení GC% je průtoková cytometrie spojená s použitím fluorochromů a využívaná zejm. u rostlin (Šmarda et al., 2012). Zde je třeba mít na paměti, že hojně využívaný fluorochrom DAPI zkresluje získané výsledky směrem k obsahu GC (Li & Du, 2014). Nejpřesnějším kvantitativním přístupem s nejvyšším možným rozlišením je sekvenace genomu, metoda, na které nejvíce spočívá tato práce a to čistě metaanalyticky s využitím dat z veřejných online úložišť. I sekvenace genomů je zatížena určitými nedostatky – nejdříve to byla nedostupnost nemodelových (tedy drtivé většiny) druhů, následně nekompletnost genomových assembly způsobená různými druhy metodických problémů (např. Peona et al., 2018). Nedostupnost genomů nemodelových druhů dnes nahradila záplava nových genomů na různé úrovni assembly a člověk musí jejich dostupnost pečlivě sledovat. Propastné rozdíly v kvalitě genomových sekvencí se však prohlubují využitím hybridního a/nebo long-read sekvenování (Amarasinghe et al., 2020) u vybraných druhů a tím, že u jiných druhů zůstávají k disposici a jsou stále produkovány genomy za využití sekvenování předchozích generací (viz NCBI, 2018).

Metody, které jsem využila ve zde presentovaných studiích, nejdříve spočívaly na fluorescentní *in situ* hybridisaci rDNA sond za účelem detekce rDNA na chromosomech (např. Symonová et al., 2013; Dion-Côté et al., 2016; Symonová et al., 2017b, 2017c). Takto získaná data jsme v některých případech vyhodnotili statisticky v evolučním kontextu (Dion-Côté et al., 2016). Dále jsme aplikovali sekvenční barvení chromosomů strunatců chromomycinem A (afinita ke GC bohatým oblastem) a DAPI (afinita k AT) za účelem kvalitativní detekce rozložení AT/GC oblastí na chromosomech jednak ve snaze identifikovat homologní resp. homeologní chromosomy, přičemž u ryb krom kostlínů toto barvení visualisuje zejména rDNA, a následně při vlastní komposiční analyse chromosomů napříč strunatci (např. Symonová et al., 2017a; Majtánová et al., 2018). V další fási jsme již pouze zpracovávali data z veřejně dostupných databasí jako jsou NCBI a Ensebml (např. Symonová et al., 2017a, Symonová & Suh, 2019; Borůvková et al., 2021). V současné době využíváme našich *ad hoc* skriptů v jazyce Python pod názvem „EVANGELIST“ (tj. EValuatioN on Genome List; Matoulek et al., 2021) a připravovaný nový nástroj „EVAN“ (Matoulek et al., *in prep*) a v jazyce R (Borůvková et al., 2021). Veškeré podklady k nástroji EVANGELIST a výsledky vznilé jeho využitím jsou k disposici online https://github.com/bioinfohk/evangelist.

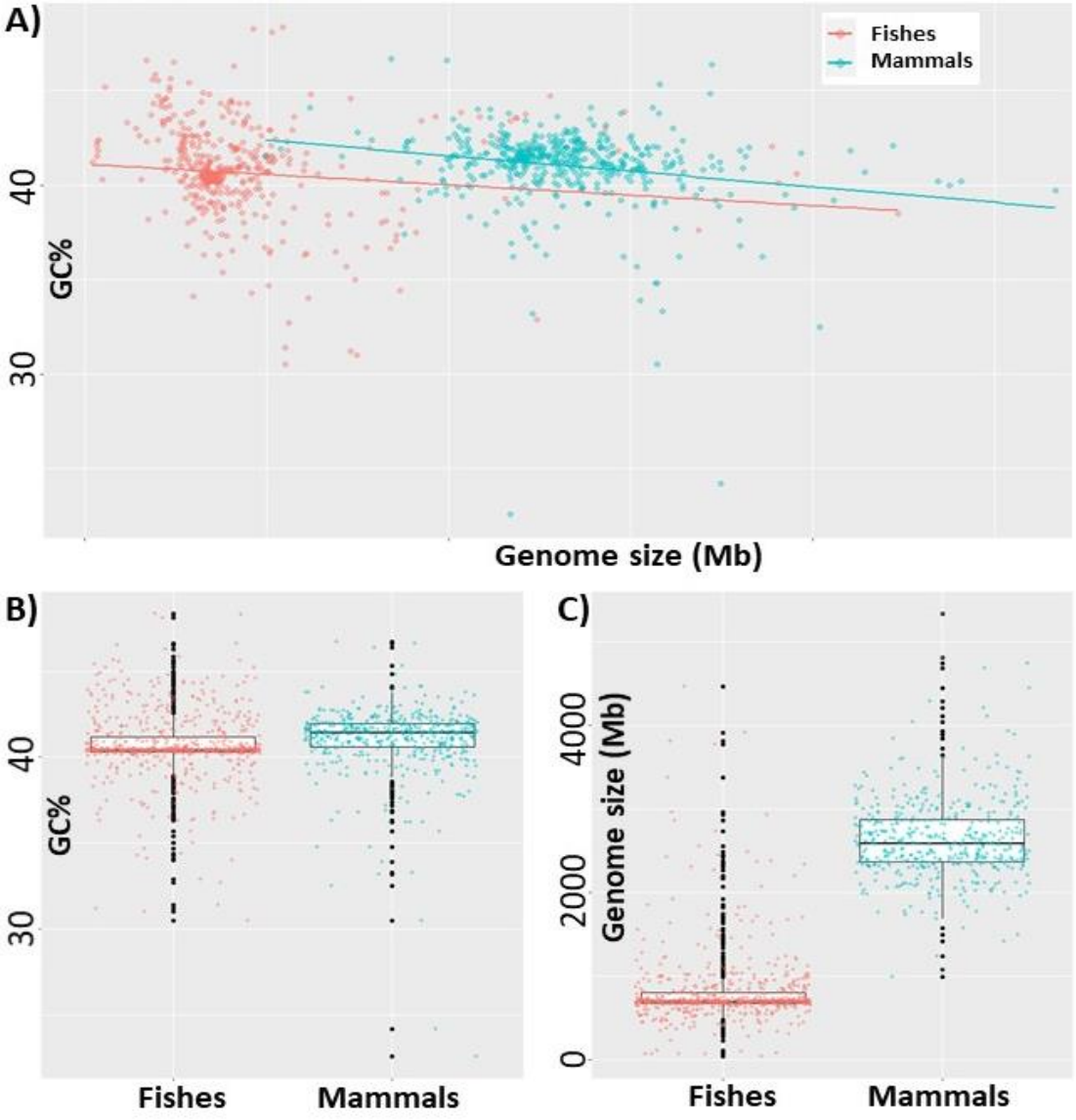
3. Výsledky a diskuse

Většina dosud publikovaných prací zaměřená na kompoziční evoluci strunatců se věnuje vlivu gBGC na kódující sekvence zejm. vyšších obratlovců a modelových skupin: u savců (např. Romiguier et al., 2010; Latrillot, 2013; Shen et al., 2015; Huttener et al., 2019, etc.), u ptáků (např. Weber et al., 2014; Bolivar et al., 2016; Rousselle et al., 2018) a plazů (Matsubara et al., 2012; Figuet et al., 2015). Kódující sekvence (exom) však tvoří u dosud nejlépe probádaného jaderného genomu člověka 1,15 - 1,17 % (Hatje et al., 2019). U ostatních strunatců naše data naznačují proporci exomu ca. 2 - 5% celého genomu (Symonová & Matoulek, nepublikováno), což je v souladu s nedávno publikovaným údajem u *D. rerio* činícím 4,67 % (de Mendoza et al., 2019; pozn.: zcela jinak tomu je u basálních mnohobuněčných, kde exom zaujímá 10 - 25 % genomu; de Mendoza et al., 2019). Tudíž u strunatců jde o zanedbatelnou frakci genomu. Navíc Bernardiho snahy etablovat GC% na 3. posici kodonů (GC3) této frakce jako proxy údaj pro danou isochoru resp. celý genom (Bernardi et al., 1985; Bernardi et al., 1997) nakonec selhaly (Elhaik et al., 2009; Tatarinova et al., 2010). Z těchto důvodů jsme se v našem dosavadním výzkumu zaměřili na *globální* cytogenomickou situaci strunatců a na vliv hlavních faktorů jako je velikost genomu, velikost chromosomů a proporce a uspořádání transposonů a často vysoce repetitivních genů pro jaderné ribosomálni RNA (rRNA).

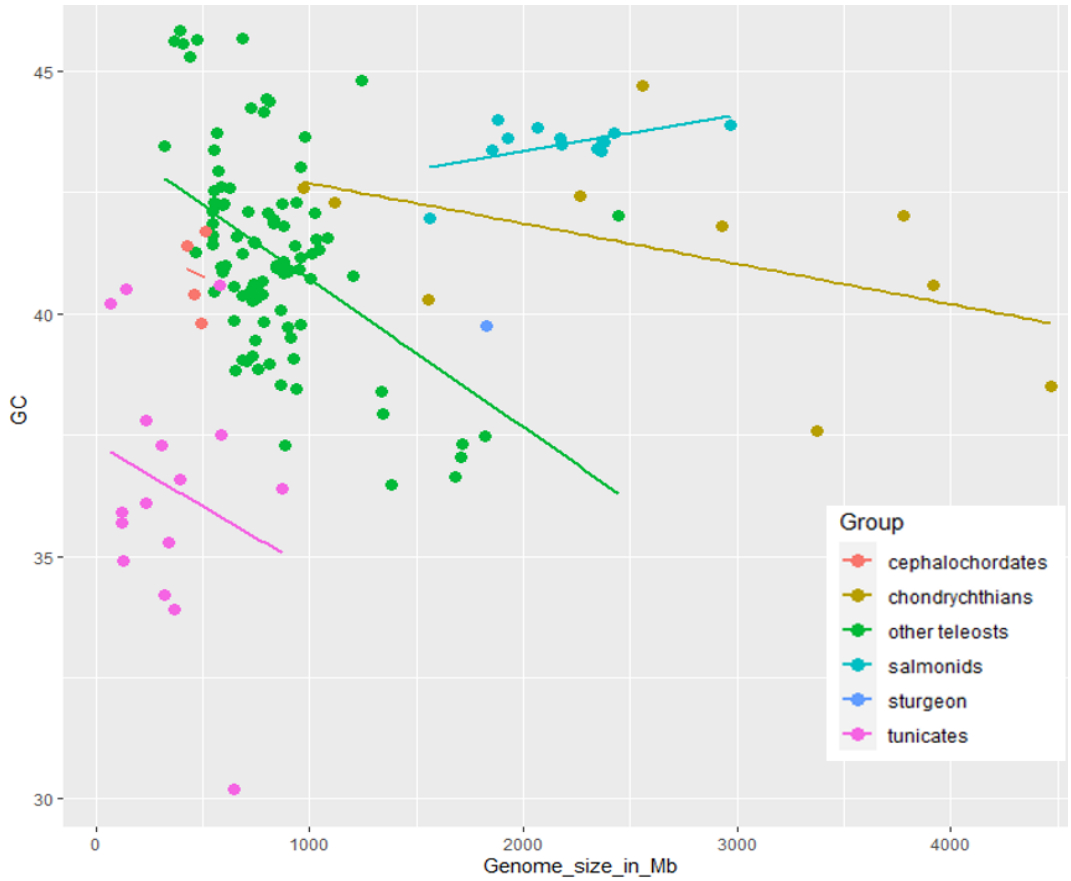
3.1. Vztah mezi GC% a velikostí chromosomů a velikostí genomu

Detailní analysu vztahu mezi velikostí genomu, velikostí chromosomů a GC% napříč fylogenetickým stromem strunatců avšak s důrazem na ryby jsme publikovali v Příloze 1 (Borůvková et al., 2021). V této práci ukazujeme, že každá hlavní skupina strunatců zaujímá svou vlastní více či méně specifickou „cytogenomickou niku“. Nejbasálnější pláštěnci mají chromosomy nejmenší a na GC nejchudší. U kopinatců došlo k zvětšení chromosomů i zvýšení GC%, u mihulí pak pouze k výraznému nárustu GC%, zatímco u paryb k nárustu GC% i zvětšení makrochromosomů. Zajímavá je situace u kostnatých ryb a savců, kteří mezi obratlovci „konvergují“ na absenci mikrochromosomů (na rozdíl od ptáků, některých plazů, jeseterů, kostlínů, paryb a mihulí) a nedosahují extrémních velikostí genomu jako někteří obojživelníci (Smith et al., 2019). Jejich celkové uspořádání genomu je tak v globálním pohledu zdánlivě podobné jakož i jejich GC% (Obr. 1 A-C). Velikost genomu savců však v průměru dva- až třikrát převyšuje ryby i přes rybí celogenomovou duplikaci (WGD) a další WGD zejm. u basálních sladkovodních linií kostnatých ryb (Glasauer & Neuhauss, 2014; detailně analysováno Morim et al., *in prep*). U většiny ryb tak přes jejich toleranci k WGD pozorujeme výraznou tendenci k redukci genomu, která se nejvíce projevuje u fylogeneticky mladých a mořských linií (Morim et al., *in prep*). Výjimkou jsou lososovité ryby, kde přes jedinou dodatečnou WGD (Macqueen & Johnston, 2014) došlo až ke ztrojnásobení velikosti jejich genomu v porovnání s ostatními kostnatými rybami, což se patrně odrazilo i na jejich neobvykle vysokém GC% (pravděpodobně vlivem GC bohatých TEs) při velkém genomu (Obr. 2). U ryb je expanse genomu doprovázena spíše GC ochuzením, čehož jsou typickým příkladem kaprovité ryby (NCBI, 2018), kde však je extrémně variabilní i velikost chromosomů (viz níže). Vedle lososovitých jsou mezi kostnatými rybami známé další dva příklady expanse genomu doprovázené neobvyklým nárůstem GC% - blatňáci rodu *Umbra* (Esociformes, sesterská skupina lososovitých) a druh *Thalassophryne amazonica* (Batrachoidiformes *sensu* Nelson et al., 2016) (NCBI, 2018; Borůvková et al., 2021; obr. 3). U všech uvedených příkladů expanse genomu ryb lze usuzovat na namnožení mj. i na GC bohatých transposonů (detailněji viz níže). Redukce genomu u fylogeneticky mladších linií ryb vedla ke zvýšení GC%, což je nejvíce patrné u řádu čtverzubci (Tetraodontiformes), jejichž genomy jsou nejmenší v rámci obratlovců a GC nejbohatší nejen v rámci ryb (NCBI, 2018). Lze tedy říci, že tyto na GC bohaté, ale stále AT/GC homogenní ryby s malým genomem dosáhly GC

bohatosti jiným mechanismem než lososovití a také než savci. U savců lze předpokládat vliv transposonů na jejich velikost genomu a GC% bez zapřičinění WGD událostí. Reaktivaci a následnou amplifikaci transposonů u ryb je však možné do značné míry připsat právě jejich četným WGD (Lien et al., 2016; Rodriguez & Arkhipova, 2018). Transposony tak přispívají k opačnému trendu v evoluci velikosti genomu ryb, kterým je jinak rychlá rediploidisace po WGD (Symonová et al., 2017b) a obecně zmenšování genomu (Morim et al., *in prep*). V tomto světle však hypotesa gBGC postrádá podporu, neboť malé chromosomy by měly umožňovat intensivnější rekombinaci (*sensu* Stapley et al., 2017), vést k redukci velikosti genomu (*sensu* Nam & Ellegren, 2012) a tudíž by jejich velikost měla vést ke GC obohacení (Mugal et al., 2015b). Obr. 1 ukazuje, že celkový GC% ryb je srovnatelný se situací u savců i přes očekávanou mnohem intensivnější rekombinaci danou zásadními rozdíly ve velikosti genomu. U savců tedy existují zcela odlišné mechanismy evoluce velikosti genomu.

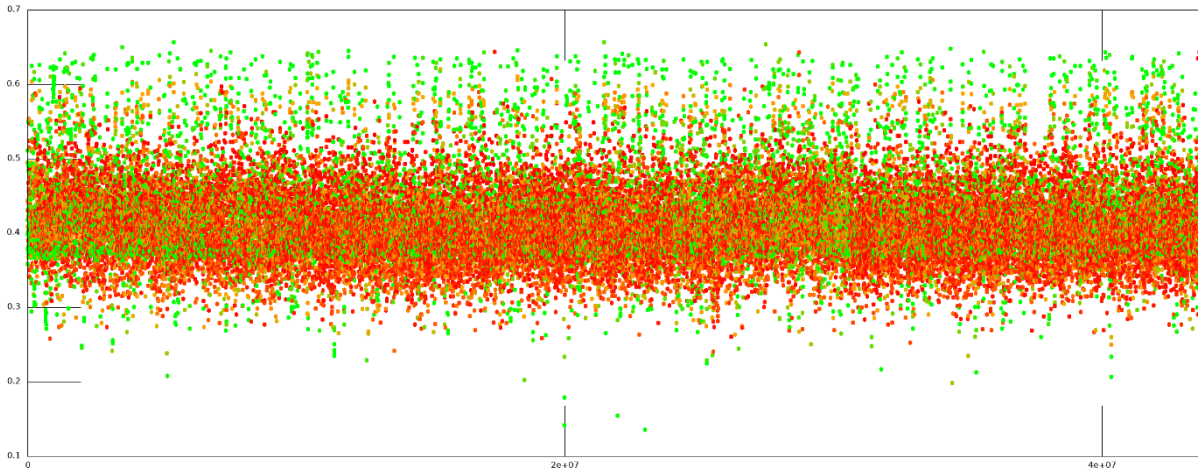


Obr. 1 Globální podobnosti (dané absencí mikrochromosomů) a rozdíly ve vztahu mezi GC% a velikostí genomu u ryb a savců. **A)** Scatterplot znázorňující podobný ačkoliv zanedbatelný vztah mezi GC% a velikostí genomu, pro savce N = 490 a R^2 = 0,0397 a pro ryby N = 601 včetně lososovitých a R^2 = 0,04. Hodnoty R^2 lze vysvětlit signifikantně positivní asociací u lososovitých (viz Obr. 2) a letounů v rámci savců (viz Gaffaroglu et al., 2020; Andreas et al., *in prep*; viz níže); **B-C)** boxploty vyjadřující podobnost v GC% a rozdíl ve velikosti genomu ryb a savců (data z NCBI, 2018; Symonová, nepublikováno).



Obr. 2 Vztah mezi GC% a velikostí genomu u jesetera a basálních skupin strunatců. Kostnaté ryby (zeleně) bez lososovitých (tyrkysově, uprostřed nahoře) vykazují slabší, ale významnou negativní asociaci mezi GC% a velikostí genomu, R^2=0,34, zatímco lososovití slabší positivní R^2=0,2773 (data z NCBI, 2018; Symonová, nepublikováno).

Další významnou kapitolou cytogenomické evoluce strunatců je výskyt mikrochromosomů u paryb, jeseterů, kostlínů, některých plazů a ptáků, který rovněž determinuje GC% těchto skupin (Příloha 1, Borůvková et al., 2021) a současně je úzce spjat s intensitou rekombinace a množstvím TEs, které se navzájem negativně ovlivňují (Kent et al., 2017). GC bohatost (nejen) mikrochromosomů u archaických ryb je patrná již z molekulárně-cytogenetických článků s vhodně použitými fluorochromy (např. Symonová et al., 2017a, 2017b) avšak z hlediska významu pro kompoziční biologii byla opomíjena (např. Ráb et al., 1999). Kompaktnost (mj. vlivem eliminace TEs) a GC bohatost mikrochromosomů napříč fylogenetickým stromem strunatců podporuji koncept gBGC. Vzhledem k následkům, jaké mikrochromosomy na genom mají, by ptáci neměli být sdružováni se savci na základě jejich endotermie jako AT/GC heterogenní, protože zaujímají svou vlastní cytogenomickou niku. Náš koncept cytogenomické niky podpořila nedávná studie, která odhalila „ptačí“, tedy výrazně redukovaný, genom u obojživelníků (u žáby; Lamichhaney et al., 2021), u kterých ocasatí dosahují naopak gigantických genomů mezi tetrapody ale i mezi strunatci (Smith et al., 2019). Ačkoliv podobná miniaturisace genomu u ptáků je připisovaná zejm. zvýšené intensitě rekombinace vlivem mikrochromosomů (Burt, 2002), u této žáby byly zjištěny čtyři faktory spojené s kontrolou TEs jakožto hlavním faktorem redukce velikosti genomu: i. Redukce abundance všech hlavních skupin TEs; ii. Zvýšená delece TEs; iii. Drastická redukce délky intronů (které jsou obývány TEs); a iv. Expanse Piwi genů, které prostřednictvím RNA interference potlačují aktivitu TEs (Lamichhaney et al., 2021). Vidíme tak další mechanismy evoluce velikosti genomu u strunatců, které mají v konečném důsledku i vliv na výsledné GC% genomu: intensita rekombinace a s ní spojená velikost genomu a chromosomů zejm. vlivem eliminace TEs, delece TEs *sensu* Kapusta et al., 2017 a RNA interference (Lamichhaney et al., 2021).



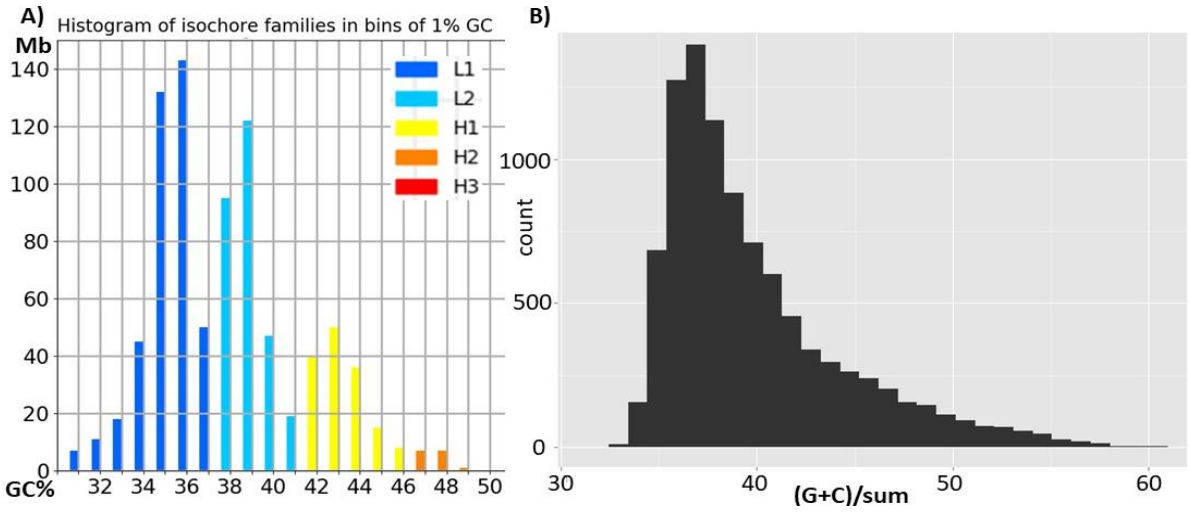
Obr. 3 Parciální profil GC% (osa y) a proporcí unikátní vs. repetitivní DNA (barevná škála, zeleně repetice, červeně unikátní DNA, přechodné odstíny odráží proporci výše uvedené DNA ve sliding window o velikosti 1 kb) u druhu *Thalassophryne amazonica* chromosomu 16, kde je patrná GC bohatost repetitivních sekvencí (TEs) dokonce přesahující 60% GC, kompletní profil spolu s dalšími druhy je k nahlédnutí na naší presentaci na GitHub https://github.com/bioinfohk/evangelist_plots/blob/master/plots_1000/Thalassophryne_amazonica.png (data z Ensembl Rapid Release; Matoulek & Symonová, nepublikováno).

3.2. Vliv transposonů na GC% strunatců

Transposony (TEs) už dávno nejsou považovány pouze za nepotřebné či škodlivé genetické haraburdí, jak tomu bylo dříve (např. Piskurek & Jackson, 2012). Protože intragenomická AT/GC heterogenita se týká jak kódující tak nekódujících DNA sekvencí (Eyre-Walker, 2001), tak TEs jsou významným faktorem s potenciálním vlivem na vznik AT/GC heterogenity teplokrevných obratlovců resp. s vlivem na AT/GC homogenisaci u ostatních živočichů (Symonová & Suh, 2019). Tomuto konceptu nasvědčuje skutečnost, že se liší skupiny TEs u teplokrevných a studenokrevných – DNA transposony dominují u ryb a obojživelníků, zatímco retrotransposony dominují u savců (Canapa et al., 2015). S TEs navíc neodmyslitelně souvisí otázka velikosti genomu. U ptáků, jejichž genom patří k nejmenším mezi amnioty, došlo k zásadní redukci množství i zastoupení skupin TEs (Kapusta et al., 2017). V předchozím odstavci bylo pojednáno o vlivu TEs u ostatních strunatců. TEs je tak nezbytné považovat za významný faktor v kompoziční evoluci genomu. Tato skutečnost byla známa u lidského genomu (např. Eyre-Walker & Hurst, 2001; Bernardi, 2005), avšak u živočichů ji dosud nebyla věnovaná adekvátní pozornost a TEs nebyly chápány jako významný faktor. Až v posledních letech tento stav změnily jednak studie Boissonota (např. Ruggiero & Boissinot, 2020) a následně mého týmu. Nejdříve jsme vyhodnotili vliv TEs na GC% globálně (Symonová & Suh, 2019), následně jsme porovnali distribuci TEs na chromosomech ryb a savců (Matoulek et al., 2021). Zajímavé potenciální následky expanse TEs jsme shledali v genomu lososovitých ryb, kde přes značnou velikost genomu existuje výrazná GC bohatost (NCBI, 2018). Protože u nich TEs, resp. soft-maskované oblasti, dosahují až 60 % genomu (Matoulek et al., 2021), připisujeme tuto nezvyklou GC bohatost TEs (Gaffaroglu et al., 2019), ačkoliv detailní analysa GC% TEs losovitých musí být teprve provedena. Podobnou situaci lze předpokládat i u dalších GC bohatých avšak AT/GC homogenních genomů – vhodným příkladem je dosud isolovaný výskyt takové situace u druhu *Thalassophryne amazonica* (u lososovitých jde o celý řád), jehož genom se na NCBI a současně na webu projektu Genome Ark (https://vertebrategenomesproject.org/) objevil na konci r. 2019 dosud bez udání citace. I zde soft-maskované oblasti dosahují vysokých hodnot GC%, nezřídka více než 60 % (Obr. 3) i přes patrně nedostatečný repeat-masking. Detailní analysa GC% TEs napříč strunatci, zejm. savců, přinese další zajímavé informace, přičemž význam tohoto výzkumu poroste s očekávaným zvyšováním kvality genomových assembly a repeat-maskingu (viz níže).

3.3. GC% strunatců - absolutní GC% x AT/GC heterogenita

V návaznosti na podkapitolu 1.3.2. a naši publikaci o AT/GC heterogenitě genomu kostlínů (Příloha 8, Symonová et al., 2017b) je vhodné zde presentovat dosud nepublikované výsledky isochor strunatců. Obr. 4 A) je prvním počinem, kdy jsme využili volně dostupný nástroj IsoSegmenter v jazyce Python (Cozzi et al., 2015) pocházející z výzkumné skupiny prof. Bernardiho. Již tento první graf potvrzuje naše předchozí výsledky u kostlínů a jejich savcům podobnou AT/GC heterogenitu, ačkoliv zde využitý genom kostlína (Braasch et al., 2016) je výrazně nekompletní. Velikost tohoto assembly je 945.878 Mb (Braasch et al., 2016) zatímco etablovaný online zdroj cytologických dat uvádí haploidní velikost genomu zástupců dvou rodů kostlínů 1,2 – 1,4 pg (Gregory, 2021). Lze tedy usuzovat, že chybějící frakce genomu mohla být GC bohatá (Peona et al., 2018) a výsledná heterogenita tudíž ještě vyšší. Na tomto místě je nutné upozornit na rozdíl mezi celkovým GC% genomu a právě jeho heterogenitou. Uvedený graf je součástí naší publikace *in prep*, kde kriticky hodnotíme potřeby visualisace isochor, které je vytýkána přílišná subjektivní intervence (Elhaik & Graur, 2014) v porovnání s dalšími nástroji (Elhaik et al., 2010a, b) anebo s využitím prostého histogramu GC% u primárních sekvenačních readů nebo s pomocí nepřekrývajícího se sliding-window u již assemblovaných genomů (viz. Obr. 3 B). Na obr. 3 je zjevná diskrepance mezi histogramem GC% a grafem rodin isochor, který výrazně eliminuje GC bohaté oblasti. Proto považujeme za vhodné provést revisi dosavadních výsledků, zaměřit se na analysu GC% bez subjektivní intervence (Obr. 4 B) a současně identifikovat a kvantifikovat přispění GC% TEs (*sensu* Matoulek et al., 2021). Proporce TEs v genomu vykazuje změny podél fylogenetickémho stromu strunatců, ale je podmíněna i kvalitou sekvenování a assembly genomu. S novými versemi dostupných genomů se proporce mobilomu zvyšuje, což souvisí krom sekvenačních technologií také s kvalitou knihoven repetitivních sekvencí, vůči kterým jsou genomové sekvence srovnávány. Tyto knihovny jsou pravidelně aktualisovány a rozšiřovány díky stále novým sekvenačním projektům založených na sekvenaci dlouhých úseků DNA a tudíž mají vyšší výtěžnost i v případě repetitivních oblastí. Lze tedy očekávat stále se zvyšující kvalitu knihoven, kde bude možné analysovat GC% jednotlivých typů sekvencí s vyšší přesností než jak jsme ukázali ve studii Matoulek et al. 2021, kde repeat-masking zejm. ryb je zatím velmi nedostatečný. Tato skutečnost prozatím snižuje efektivitu našeho výzkumu.



Obr. 4 Dvě možnosti analysy zastoupení sekvencí a jejich GC% (obě osy x). **A)** Visualisace a kvantifikace tzv. rodin isochor *sensu* Bernardi, 2005 kostlína (*L. oculatus*) z genomu Braasch et al., 2016 s využitím nástroje IsoSegmenter (Cozzi et al., 2015) (sekvence genomu z Ensembl, Symonová & Sedláková, nepubl.); **B)** Visualisace a kvantifikace pomocí histogramu (sekvence genomu z Ensembl, Symonová & Mořkovský, nepubl.).

3.4. Geny rRNA a jejich význam v cytogenetice studenokrevných obratlovců

rDNA neboli ribosomální DNA kóduje ribosomální RNA (rRNA), nezbytnou složku ribosomů zodpovědných za proteosyntesu každé buňky (Potapova & Gerton, 2019). Z pohledu cytogenetiky ryb rDNA byla a je významnou frakcí genomu a to jak z metodických důvodů, tak z pohledu evoluční dynamiky chromosomů, proto jí je věnovaná nemalá pozornost (např. Symonová et al., 2013; Dion-Côté et al., 2016; Symonová, 2019; Symonová & Howell, 2018) stejně jako u jiných živočichů (Sochorová et al. 2018). Obrovský význam této universální a jedné z nejstarších frakcí genomu sahá i do klinické oblasti a rDNAomu (rDNA subgenomu) se věnuje stále větší množství publikací mj. proto, že hybridní sekvenování nyní umožňuje tyto nepostradatelné vysoce repetitivní sekvence lépe studovat a napravit dosavadní nedostatky v kvalitě jejich sekvencí a také omyly (Hall et al., 2021). Vysoké GC% rDNA a mnohdy monstrosní počet kopií (desítky tisíc např. u ryb Symonová et al., 2017c) činily tyto repetice zajímavým faktorem ve studiu velikosti genomu a GC% genomu. Zásadní vliv rDNA na tyto dva parametry dosud prokázán nebyl (např. Lehman et al., *in prep*), ačkoliv byl popsán zajímavý vztah mezi velikostí genomu a velikostí rDNA subgenomu (Prokopowich et al., 2003), což naznačuje význam počtu zdánlivě redundantních kopií rDNA, tj. těch, které nejsou přepisovány na rRNA. Je možné, že nové výsledky pohled na rDNA ještě změní. Extrémně vysoké GC% 45S rDNA může reflektovat samotný význam a nepostradatelnost této frakce genomu a proto by mohlo být považováno za výsledek selekce podobně jako u genů kódujících proteiny. Vysoké GC% 45S rDNA navíc dle Forsdyka usnadňuje jejich harmonizovanou evoluci (concerted evolution; Forsdyke, 2016), což je zásadní mechanismus homogenisace vysokého počtu kopií rDNA a výsledek selekce.

3.5 Doklady selekce v komposiční biologii

Největším zastáncem selekce v GC evoluci je G. Bernardi a jeho již přežitá (neo)selektcionistická hypotesa termodynamické stability. Zajímavým zjištěním z Bernardiho laboratoře je GC bohatost genů (Bernardi, 2005). Zde existuje obecný konsensus, že na exony působí silná selekce kvůli jejich roli uchování a realisace genetické informace (např. Vinogradov, 2003). Jsou-li exony všeobecně GC bohaté, měla by jejich GC bohatost pramenit také ze selekce. Tomu se věnuje další představitel selekce v GC evoluci, A. E. Vinogradov, se svými vysoce citovanými publikacemi. Ten tvrdí, že působí-li v GC evoluci selekci, jde o selekci znaků spojených s transkripcí DNA, tedy ne primárně GC% jako takové (Vinogradov, 2003). Jeho odůvodnění vzniku GC bohatých úseků u amniot zní logicky:

„Because heavy isochores are known to be gene-rich and show a high level of transcription, it is suggested here that isochores arose not as an adaptation to elevated temperature but because of a certain grade of general organization and correspondingly advanced level of genomic organization, reflected in genome structuring, with physical properties of DNA in the gene-rich regions being optimized for active transcription and in the gene-poor regions for chromatin condensation ('transcription/grade' concept).“ Vinogradov, 2003

Avšak GC bohaté geny mají i ryby (Symonová et al., 2017a; Symonová & Matoulek nepubl., Obr. 5).

Exonů se týká další oblast, kde existuje konsensus nebo alespoň snaha o hledání robustních důkazů, že GC% podléhá selekci, ačkoliv se v případě exomu jedná o tak malou frakci genomu – selekce pro efektivitu a přesnost translace (Gingold & Pilpel, 2021).

„Selection for translation efficiency was shown also in some multicellulars such as C. elegans, D. melanogaster *and* Arabidopsis thaliana*. Yet, as expected, attempts to demonstrate selection for translation efficiency in human, and to further correlate it with expression levels, yield contradictory results. Some studies found no evidence for translational selection in human suggesting that synonymous codons in human are not selected to maximize translation efficiency. Conversely, other*

studies do indicate weak, yet significant, translational selection in human, according to estimates of codon usage adaptation to the global tRNA pool.“ Gingold & Pilpel, 2021

Pokud GC% genů koreluje s GC% oblasti, kde se tyto geny vyskytují (např. Bernardi, 2005), tak musí existovat vztah mezi GC% těchto dvou frakcí a mechanismus zajišťující to, aby byla tato korelace dlouhodobě udržena. Takovou ačkoliv slabou, selekci vůči GC% u intronů našli Pozzoli et al., 2008. Tito autoři konstatovali, že GC evoluce lidského genomu je pod kontrologu jak neutrální gBGC, tak selekce. Selekci směrem k optimální délce ale i GC% lidských intronů popsali Wang & Ju, 2011.

Selekci v kompoziční evoluci transposonů diskutují a dále prokázali Ruggiero & Boissinot, 2020 – mj. považují bohatost na adenin u savčích L1 non-LTR retrotransposonů jako jejich vlastní regulační mechanismus reflektující efektivitu své transposice a potenciální negativní důsledky na svého hostitele. Tito autoři studovali GC% u retrotransposonů v rámci jednoho a mezi různými hostitelskými genomy savců, což je přesně oblast, která je naprosto zásadní i pro ostatní strunatce, ačkoliv jí dosud nebyla věnována žádná pozornost. Jiní autoři navrhují, že existují další dosud neznámé neutrální mechanismy kompoziční evoluce (Storlazzi et al., 1995; Paudel et al., 2020). U bakterií byla selekce směrem ke GC bohatosti již dříve prokázána (Hildebrand et al., 2011). Působení selekce by mělo být zahrnuto v našich snahách o pochopení kompoziční evoluce celého jaderného genomu strunatců.

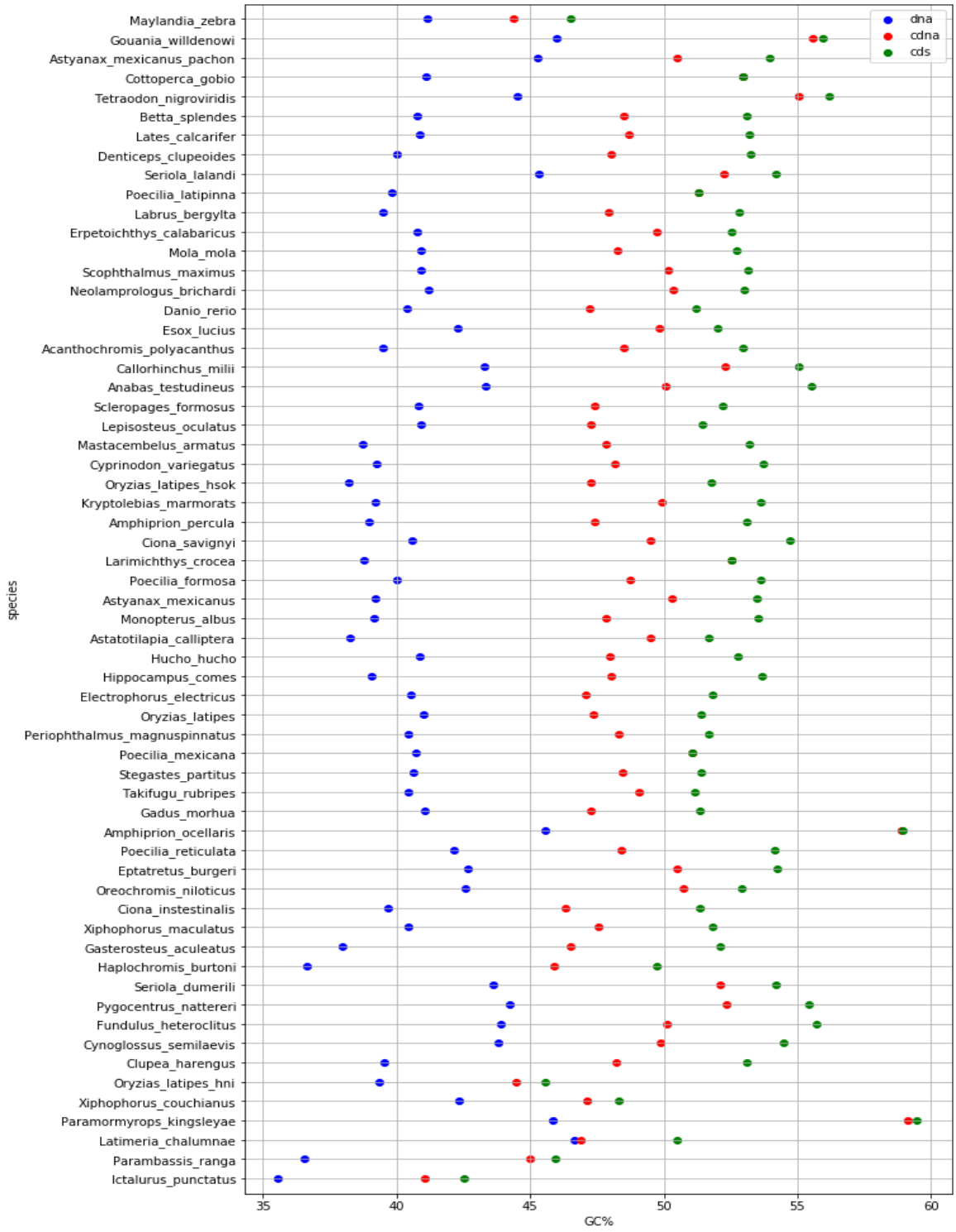
3.6 Souhrn a zhodnocení současného stavu poznání

V současnosti nejpalčivějším problémem ve zde popsaném tématu je skutečnost, že zástupci a zastánci jednotlivých konceptů do značné ignorují koncepty alternativní, které se však nijak nevylučují. Spíše naopak mají potenciál se doplňovat a je jasné, že je nanejvýš žádoucí pluralistický přístup k celé problematice, nezavrhovat „konkurenční“ koncepty ale spíše se snažit o jejich propojení a také mít na paměti další aspekty, které dosud nebyly formulovány jako explicitní hypotesy (např. role TE).

Vlastní prací jsem k tomuto tématu přispěla tím, že 1) můj tým cytogeneticky a následně opakovaně bioinformaticky odhalil kostlíny jako nositele AT/GC heterogenního genomu, což je dalším důvodem pro upuštění od již přežité teorie termodynamické stability z dílny prof. Bernardiho a jeho spolupracovníků; 2) ukázali jsme transposony jako významné hráče v této vysoce komplexní souhře početných faktorů; 3) popsali jsme kvantitativní vztahy mezi velikostí genomu, velikostí chromosomů a GC% genomu napříč fylogenetickým stromem strunatců a označili transposony jako faktor, který může tento primárně triviální lineární vztah narušit až převrátit; 4) vytvořili jsme *ad hoc* bioinformatické nástroje k detailním analysám a visualisacím GC% a proporcí transposonů napříč assemblovanými chromosomy; 5) tím jsme dali vzniku novému, integrativnímu a pluralistickému pojetí kompoziční biologie eukaryot a vytvořili podmínky pro globální výzkum napříč fylogenetickým stromem živočichů s potenciálem expandovat směrem k bezobratlým, houbám, rostlinám a jednobuněčným organismům. Tímto způsobem budeme schopni zachytit GC% celého genomu, ale i exomu a repetitivních sekvencí druhů, jejichž genom bude postupem času dostupný.

3.7 Diskuse

Vznik AT/GC heterogeneity u ptáků a savců dosud nebyl uspokojivě vysvětlen, ačkoliv se tomuto tématu po desetiletí věnovala nemalá pozornost, jak je v základních obrysech a velmi stručně popsáno výše. Tuto skutečnost konstatují zásadní review, které se věnují (mj.) kompoziční evoluci genomu (Eyre-Walker & Hurst, 2001, Duret et al., 2006; Long et al., 2018; Bohlin et al., 2019 a Hurst, 2019). Knihu Bernardi, 2005 nelze považovat za objektivní krom výše uvedeného mj. proto, že ignoruje fakt, že ptáci mají sice vyšší tělesnou teplotu, tudíž jejich vyšší GC% by odpovídala teorie termodynamické stability, ale současně mají početné drobné mikrochromosomy (Burt, 2002; NCBI, 2018), které jednoznačně zvyšují efektivitu rekombinace hovořící ve prospěch gBGC konceptu.



Obr. 5 GC% ryb na úrovni celé DNA (modře), cDNA, tj. DNA komplementární k mRNA, (červeně) a cds (zeleně), data z Ensembl, nepublikováno Matoulek & Symonová.

Srovnatelná situace existuje u basálních savců, kteří sice mají nižší tělesnou teplotu a jejichž nižší GC% by opět odpovídala termodynamické stabilitě. Současně však mají také obrovské a málo početné chromosomy (NCBI, 2018), které výrazně snižují efektivitu rekombinace a tudíž hovoří ve prospěch gBGC. Toto Bernardi zcela opomíjí. Tyto dva příklady demonstrují komplexitu komposiční biologie genomu a hlavně naprostou nezbytnost pluralistického přístupu a kombinace cytogenomických a

molekulárně-biologických dat. Absence takového přístupu do značné míry přispívá k dosavadnímu nedostatečnému pochopení komposiční biologie genomu (Bohlin et al., 2019; Hurst, 2019).

Roli teploty v komposiční evoluci genomu by mohlo být možné spatřovat v jiné souvislosti, než prosazuje Bernardi. Funkce genomu ektotermních živočichovů, jejichž metabolismus musí fungovat při širokém rozmezí teplot, je do značné míry závislá právě na teplotě prostředí, kde se nachází (Willmer et al., 2000). Je známo množství příkladů, kdy teplota prostředí ovlivňuje např. expresi genů (Politis et al., 2017). Tento regulační aspekt teploty u endotermních ptáků a savců odpadá – jejich metabolismus funguje pouze ve velmi úzkém rozmezí teplot, kde není prostor pro regulaci či modulaci transkripce, enzymy buď fungují, nebo nefungují (Willmer et al., 2000). Stojí tedy za úvahu, zda nikoliv vyšší tělesná teplota *sensu* Bernardi, 2005, ale vysoce stabilní a neměnná tělesná teplota ptáků a savců nemohla hrát roli při vzniku jejich komoposičního GC kódu. Mechanismus takové vnitrogenomové signalisace založený na AT/GC heterogenitě lze spatřit v případě vzniku GC bohatých transkripčních elementů v době okolo vzniku amniot (Zhang, eta l., 2004; Khuu et al., 2007; Ho, 2009; Calistri et al., 2011). Další příklad uplatnění GC% při transkripci je vliv na místo sestřihu pre-mRNA, čili vymezení hranice mezi exony a introny (Zhang et al., 2011; Amit et al., 2012). Ačkoliv jde v případě sestřihu pre-mRNA opět pouze o kódující frakci genomu, tak oba zde uvedené příklady (dalších je celá řada) naznačují adaptivní význam GC%, navíc spojený s rozdíly mezi teplokrevnými a studenokrevnými obratlovci. V případě GC bohatých transkripčních elementů dochází k rozšíření pole působnosti GC% z exomu na další oblasti, což zvyšuje jejich dopad společně s kódujícími sekvencemi na globální GC% a AT/GC heterogenitu.

Jaderný genom savců je z desítek procent tvořen repeticemi, u člověka přilibžně 50%, ačkoliv alternativní přístupy naznačují, že 66 - 69 % lidského genomu je tvořeno repeticemi nebo od nich odvozených sekvencí (de Koning et al., 2011). Přestože jen necelá 2 % lidského genomu kódují peptidy event. tRNA či rRNA, tak jsou výrazně GC bohatá a extrémně komplexní. Pokud by i repetitivní sekvence měly při své velikosti být GC bohaté, tak by znamenaly obrovskou zátěž a risiko pro celý genom vzhledem k jejich hypermutabilitě. Naopak, je výhodné repetice udržovat pod kontrolou v podobě vysoce kompaktního heterochromatinu, čemuž právě přispívá AT bohatost. Substituční tlak, alespoň v lidském genomu, je obecně vychýlen ve směru AT (vzhledem k početní převaze repetic a nekódujících sekvencí; Eyer-Walker & Hurst, 2001; Petrov & Hartl, 1999; Paudel et al., 2020), což je v souladu s AT bohatostí repetic. Geny potřebují být přístupné enzymům transkripční mašinerie, což jim zajišťuje jejich GC bohatost snižující potenciál tvorby nukleosomů a zvyšující bendabilitu spolu s odolností vůči torsním tlakům. GC bohatost spolu s rozvolněnou konformací euchromatinu však genům současně zvyšují pravděpodobnost mutací. Proto je snadné si představit evolučně neutrální mechanismus jako je gBGC, který bez ohledu na fitnes jedince „paušálně“ zvyšuje GC% v oblastech bohatých na geny a zajišťuje tak udržitelnost jejich GC bohatosti např. i jako kompensaci hypermutability. Skutečnost, že GC bohaté nebo bohatší jsou i oblasti kolem genů a introny může být vnímána jako vedlejší efekt gBGC, nebo jako „pojištění si“, že GC% bude dostatečné jako signál „pozor, blíží se geny“, anebo kombinace výše uvedeného. Tato GC bohatost genů je však evolučně významná zejm. s ohledem na zvyšující se komplexitu (*sensu* Vinogradov, 2003), ale i velikost genomu a tudíž výsledkem selekce. Výsledkem selekce by mohla být i samotná gBGC, ačkoliv sama působí selekčně neutrálně. U savců a ptáků je tak přítomný „komposiční GC kód“, který jim pomáhá se vypořádat s velkým a komplexním genomem a hlavně s do detailů vyladěnou regulací exprese jeho genů. Daleko větší genomy basálních ryb nebo obojživelníků mohou však být i GC bohatší než savčí, avšak GC bohatost tu zprostředkovávají transposony, které homogenisují GC% na chromosomech. Ve smyslu současných hypotes lze tyto genomy charakterisovat sníženou selekcí na velikost genomu (např. vlivem nízkých populačních hustot

sensu Nam & Ellegren, 2012), která vedla k expansi chromosomů akumulací transposonů a tím zpětnovazebně k dalšímu snížení intensity rekombinace, atd. (Mugal et al., 2015b; Stapley et al., 2017). Otázkou zůstává, zda výkyvy podmínek vnějšího prostředí včetně proměnlivé teploty by mohly představovat dostatečné opodstatnění jejich AT/GC homogenity genomu.

4. Závěry

4.1. Význam výsledků pro vědní obor a možnosti směřování dalšího výzkumu

Příloha 1 Borůvková et al., 2021 využívá dat extrahovaných naším nástrojem EVANGELIST v jazyce Python a analysovaných s využitím knihoven jazyka R. Tento kvantitativní přístup ukázal podobnost v organizaci chromosomů a lokálního GC% u paryb, jeseterů a kostlínů s plazy a ptáky a zároveň rozdíly ve velikosti jejich genomu determinující globální GC%. Dále naše výsledky ukázaly podobný negativní vztah mezi velikosti chromosomů a jejich GC% mezi rybami a savci. Tato asociace je však výsledkem zcela odlišné cytogenomické evoluce – ryby, ačkoliv oproti savcům prošly jednou celogenomovou duplikací navíc, mají 2-3x menší genom při podobném GC%. Naznačují tak na jedné straně konvergentní komposiční evoluci ryb a savců a současně trend ryb k redukci genomu a savců k jeho expansi. Tyto výsledky vrhají nové světlo na koncept gBGC, kde velikost genomu a jednotlivých chromosomů ve vztahu ke GC% hrají významnou roli a ukazují potřebu jeho revise.

Příloha 2 Matoulek et al., 2021 prezentuje výsledky spolupráce profesionálního informatika – programátora a matematičky při zpracování a interpretaci exponenciálně narůstajícího množství genomických dat u ryb a dalších obratlovců. Vytvořili jsme nástroj v jazyce Python, který dosud bezprecedentním způsobem umožňuje integrovat sekvenační data při chromosomální úrovni assembly a data repeat-maskingu a vytvořit profil GC% celé sekvence DNA a současně proporci repetitivních sekvencí podél chromosomu. Tento přístup ukázal rozdíly ve vzájemném uspořádání GC% a repetitivních sekvencí u ryb a savců. Získaná data lze nejen visualisovat ale i extrahovat a následně dále (meta)analysovat. Nástroj je použitelný pro všechny genomy s assembly na úrovni chromosomů https://github.com/bioinfohk/evangelist_plots. Zásadním výsledkem je odhalení homogenisace GC% ryb prostřednictvím transposonů (TE) a jejich GC%, což ukazuje TE jako významný komposiční faktor potenciálně ovlivnitelný selekcí. To by však znamenalo posun od selekčně neutrální gBGC jakožto současně nejšířeji uznávaného faktoru.

Příloha 3 Gaffaroglu et al., 2020 presentuje lososovité ryby mj. jako příklad neobvyklých následků celogenomové duplikace (WGD) obratlovců a jejich komposiční evoluce genomu – opačný trend vztahu velikosti genomu vůči GC% než nacházíme u ostatních ryb ale i dalších obratlovců. Zajímavé je, že lososovité ryby duplikovaly svůj genom jednou, ale mají genom až třikrát větší než většina ostatních kostnatých ryb (Teleostei). Data ze sekvenování genomů lososovitých shodně kvantifikují podíl TE na více než 50 %, což naznačuje, že vlivem WGD došlo u těchto ekonomicky významných ryb k amplifikaci nezvykle GC bohatých transposonů a tedy vedle WGD k další expansi genomu. To je zcela opačný trend, než lze pozorovat u ostatních kostnatých ryb, které jednak dynamicky rediploidisují genom po WGD a dále, pokud dojde k expansi TE, jsou tyto elementy GC chudé, resp. AT bohaté. Podobnou situaci jsme zaznamenali u jedné skupiny letounů (Andreas et al., *in prep*) a je možné, že bude prokázána i jinde mezi obratlovci. Skupiny s podobným inversním vztahem mezi GC% a velikostí genomu mohou dramaticky zkreslit existující asociace mezi těmito dvěmi faktory na vyšší taxonomické úrovni a je nezbytné brát na ně zřetel a zohlednit je.

Příloha 4 Symonová & Suh, 2019 ukazuje roli transposonů (TE) jakožto významného faktoru kompoziční evoluce genomu ryb s potenciálem přinejmenším přispět k objasnění GC biologie kódujících vs. nekódujících oblastí genomu. Tato práce vytvořila důležitý odrazový můstek k jejich detailnějšímu studiu u dalších linií obratlovců (tj. hostitelských genomů TE), protože ukazuje, že GC% se může výrazně lišit mezi jednotlivými skupinami TE ale i mezi liniemi jejich hostitelských genomů. Ukázali jsme dosud nezjištěnou korelaci mezi GC% hostitelského genomu a GC% jeho TE. Tato studie zohlednila skutečnost, že odlišné skupiny TEs, které invadují genomy teplokrevných a studenokrevných obratlovců, mohou mít následek v kompoziční biologii genomu těchto skupin.

Přílohy 5 a 6 Symonová, 2019 a Symonová & Howell, 2018 shrnují význam komplexu genů kódujících ribosomální RNA (rDNA) u ryb a dalších nemodelových organismů. rDNA je nejstarší, nejuniversálnější a GC nejbohatší frakcí genomu a podílí se na endogenním repeatomu, čili repetitivních sekvencí různého počtu kopií. Vedle tradiční detekce rDNA na chromosomech a to jak metodou FISH nebo prostým AT- a GC-specifickým fluorescenčním barvením, v těchto pracech ukazujeme mj. cytogeneticky nedetekovatelné sekvence rDNA pomocí bioinformatiky. Nezastupitelnou roli cytogenetiky u ryb při studiu expanse rDNAomu (rDNA subgenom) lze demonstrovat na příkladu štiky obecná, jejíž genom obsahuje přinejmenším desítky tisíc kopií 5S rDNA, **Příloha 7** (Symonová et al., 2017c).

Přílohy 8 a **9** (Symonová et al., 2017a a Majtánová et al., 2018) jsou důkazem stále nenahraditelné role molekulární cytogenetiky v evolučně-biologickém výzkumu. V první studii jsme prokázali savcům podobný způsob AT/GC heterogenity genomu u archaických kostlínů (Lepisosteiformes, Actinopterygii) a to jak na jejich chromosomech cytogeneticky, tak následně v genomovém assembly bioinformaticky. Autoři studie, která v té době genom kostlína (*L. oculatus*) publikovala, si této skutečnosti nevšimli – nejspíš proto, že GC% chromosomů detailně neanalysovali, neboť je nenapadlo, že takto archaická skupina ryb by mohla vykazovat jakoukoliv AT/GC heterogenitu (Braasch et al., 2016). Ve druhé studii jsme již pouze cytogeneticky ukázali, že nejbližší žijící příbuzný kostlínů a současně jediný recentní zástupce řádu Amiiformes, kaproun obecný (*A. calva*) má „typicky rybí" AT/GC homogenní genom. Bioinformatické pokračování této druhé studie bude následovat velmi brzy, neboť genom kaprouna již čeká na své zveřejnění (zatím preprint: Thompson et al., 2021).

Přílohy 11 (Symonová et al., 2017b), **12** (Symonová et al., 2013) a **13** (Havelka et al., 2016) jsou molekulárně-cytogenetické studie jeseterů a veslonose (Acipenseriformes), které byly zásadní mj. z toho hlediska, že ukázaly cytogeneticky detekovatelný vyšší GC% u početných mikrochromosomů těchto rovněž archaických ryb. Oproti kostlínům však jsou jejich makrochromosomy AT/GC homogenní. To se stalo jedním z východisek naší nejnovější studie (**Příloha 1**), která tato kvalitativní cytogenetická pozorování kvantifikovala, a přispělo k odhalení diversity v organisaci genomu na fylogenetickém stromu obratlovců (**Přílohy 1 a 2**).

Přílohy 14 a **15** (Dion-Coté et al., 2015, 2017) nás přivedly k extrémní cytogenomické dynamice severo-amerických síhů rodu *Coregonus* a významu na GC nejbohatší frakce genomu, rDNA, těchto lososovitých ryb s vícenásobně duplikovaným (polyploidním) genomem. K tomuto výzkumu jsme byli přizváni na základě našeho předchozího výzkumu evropských síhů, který vyústil v mou dosud nejcitovanější práci **Příloha 16** (Symonová et al., 2013). Zde jsme ukázali význam transposonů v evoluční dynamice genomu a chromosomů prostřednictvím rDNA. Současně jsme prokázali význam expanse rDNA subgenomu a transposonů při rekonstrukci akcelerované evoluce ve sladkovodním biocyklu během postglaciálního období.

Přílohy **17**, **18** a **19** (Knytl et al., 2013; Sember et al., 2015; Majtánová et al., 2016) jsou tři molekulárně-cytogenetické studie ryb řádu Cypriniformes, které se týkají různě ploidních ryb s dynamickým rDNA subgenomem, malými chromosomy a výrazně GC chudým genomem. Jde o významný řád ryb, kam patří mj. dánio pruhované (*Danio rerio*) a kaprovité ryby významné pro akvakulturu (kapr, karas, atd.). Svým nízkým GC% a častou dodatečnou polyploidií představují zajímavý protipól rovněž polyploidním avšak GC velmi bohatým lososovitým rybám. Na přípravě těchto článků jsem figurovala mj. jako školitelka disertační práce Z. Majtánové, školitelka-konsultantka disertační práce A. Sembera a tutorka čerstvého postdoc vědce M. Knytla.

3.2. Možnosti směřování dalšího výzkumu

Další směřování výzkumu kompoziční evoluce genomu strunatců ale i obecně eukaryot bude určována dostupností osekvenovaných genomů na chromosomové úrovni a bioinformatických nástrojů k jejich zpracování. V současné době jako nejakutálnější téma považuji využít náš kvantitativní cytogenomický přístup (*sensu* Borůvková et al., 2021) u savců a podrobit savčí transposony podobné studii jako jsme provedli u ryb (Symonová & Suh, 2019). Naše právě probíhající studie na letounech (Andreas et al., *in prep*) poskytla již nyní zajímavé a slibné výsledky a je jasné, že letouni budou z pohledu kompoziční biologie významnou skupinou nejen v rámci savců i do budoucna.

Podobné otázky jako jsou popsané v této práci u strunatců, existují u rostlin a bezobratlých. Další kapitolou kompoziční biologie budou houby a jednobuněční, jejichž genomy jsou také stále více sekvenovány a utěšeně se hromadí ve veřejně dostupných databásích. Tudíž dalším směrem kompoziční biologie bude integrace výsledků ze všech těchto hlavních skupin eukaryot. Lze však předpokládat, že výzkum u jednobuněčných skupin bude ještě notnou řadu let trvat vzhledem k jejich fylogenetické diversitě a početnosti, o to přínosnější by mohl být. Následně by mělo dojít k pochopení důvodů propastných rozdílů mezi prokaryoty a eukaryoty *sensu* Bohlin & Pettersson, 2019.

Dalšími očekávanými milníky v oblasti GC biologie bude jednak genom assemblovaný na úroveň chromosomů kaprouna obecného (*Amia calva*), který v době psaní této práce ještě není zveřejněn, ačkoliv publikace ke genomu je již online ve formě preprintu (Thompson et al., 2021). Po zveřejnění naší práce Borůvková et al., 2021 byl publikován genom bichira (*Polypterus senegalus*) assemblovaný na úroveň chromosomů (Bi et al., 2021), který jsme sice zahrnout nemohli, ale jeho dodatečná cytogenomická analysa nepřináší žádný zásadní posun. Dále jsou zcela nově k disposici genom veslonose (*Polyodon spathula*), genom kaprouna (*Amia calva*) z jiné studie, než je výše uvedená, a genom dalšího kostlína (*Atractosteus spatula*), avšak pouze na úrovni draftu (Bi et al., 2021). Výše popsané dokládá současnou dynamiku sekvenování i velkých genomů nemodelových strunatců. Pro kompoziční evoluci genomu obratlovců bude zásadní buď vylepšená verse genomu kostlína *L. oculatus* nebo lepší genom na úrovni chromosomů jiného druhu kostlína. Současný jediný genom druhu *L. oculatus* (Braasch et al., 2016) je velmi nekompletní, což ale nezabránilo možnosti prokázat jeho savcům podobnou AT/GC heterogenitu (Symonová et al., 2017).

Zcela zásadní bude pluralistický a integrativní přístup ke kompoziční biologii, jak je nastíněn v této práci, protože dosavadní obrovské úsilí byť špičkových laboratoří z celého světa nedospělo k uspokojivému vysvětlení důvodů a mechanismů vnitrogenomových a mezidruhových rozdílů v GC%, zejm. mezi studenokrevnými a teplokrevnými obratlovci.

4.3. Využití dosažených výsledků při výuce

Teoretické vědomosti i praktické zkušenosti získané v průběhu řešení zde presentované problematiky byly kromě jiného využity při tvorbě nového studijního předmětu „Evoluce genomu“ na Univerzitě Hradec Králové (UHK) a podobného předmětu „Genomevolution“ na Universitě Innsbruck (UIBK). Zde presentovaná problematika se stala těžištěm výzkumného zaměření Bioinformatického centra HK v rámci projektu Erasmus+, v jehož rámci nyní finišují 3 diplomové a 1 bakalářská práce. Od zimního semestru 2021/22 plánujeme 2 disertační práce na již aplikovaná (biomedicínská) bioinformatická témata. Další vysokoškolské kvalifikační práce jakéhokoliv stupně zaměřené na komposiční biologii nejen strunatců lze kdykoliv nabídnout případným zájemcům.

K dosažení presentovaných výsledků přispěli studenti doktorského studia Mgr. Veronika Borůvková a Ing. Dominik Matoulek na UHK a jedna studentka na PřF UK Mgr. Zuzana Majtánová, Ph.D. (již po úspěšné obhajobě doktorské práce). Na výzkumu na UHK dále participovaly dvě studentky navazujícího magisterského stupně studia Bc. Kateřina Rodrová a Bc. Kristýna Patočková a jeden student bakalářského stupně studia Bc. Lukas Aichler na UIBK. Všichni výše jmenovaní studenti a studentky pracovali nebo pracují pod mým vedením.

4.4. Využití dosažených výsledků v praxi

Výsledy této habilitační práce dosud v praxi využity nebyly a v současné době a podobě ani nemají ambici se v praxi uplatnit. V této fási jde o čistě základní výzkum avšak s potenciálem přesahu do výzkumu aplikovaného, zejm. biomedicínského. Pro pochopení fungování lidského genomu ve zdraví i nemoci je zcela zásadní porozumět, jaké mechanismy komplexitu savčího genomu utvářely a jaké byly předstupně při jeho vzniku. V biomedicíně se nejvíce uplatňuje rDNAomika, přičemž v současné době dochází k upřesňování dosud získaných informací o vlivu počtu kopií rDNA při rozvoji různých chorob vzhledem ke stále se zvyšující dostupnosti hybridního a long-read sekvenování. Samotná GC biologie má v současné době patrně největší uplatnění v genomice a evoluční biologii virů včetně SARS-Cov-2. U virů GC% přímo ovlivňuje jejich tzv. „codon adaptation index“ využívající vysoce exprimované lidské geny jako referenční hodnotu, čímž se může teoreticky predikovat efektivita exprese virových genů v lidských buňkách (Auwarakul, 2005, doi: 10.1016/j.virusres.2004.10.004). Jak ukazuje významná česká stopa v současné komposiční biologii virů nejen u SARS-Cov-2 (Matyášek & Kovařík, 2020, doi: 10.3390/genes11070761), tak GC biologie virů je důležitou součástí moderní biologie.

5. Seznam použité literatury

Alvarez-Valin F, Lamolle G, Bernardi G. 2002. Isochores, GC3 and mutation biases in the human genome. Gene 300(1-2): 161-168 10.1016/S0378-1119(02)01043-0

Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Gouil Q. 2020. Opportunities and challenges in long-read sequencing data analysis. Genome Biology 21, 30 10.1186/s13059-020-1935-5

Amit M, Donyo M, Hollander D, Goren A, Kim E, Gelfman S, Lev-Maor G, Burstein D, Schwartz S, Postolsky B, Pupko T, Ast G. 2012. Differential GC content between exons and introns establishes distinct strategies of splice-site recognition. Cell Rep 1(5):543-56 10.1016/j.celrep.2012.03.013

Andreas M, Borůvková V, Matoulek D, Liu H, Horáček I, Symonová R. *in prep*. Transposons drive chiropteran quantitative cytogenomic evolution and follow bats phylogeny

Arhondakis S, Auletta F, Torelli G, D'Onofrio G. 2004. Base composition and expression level of human genes. Gene 325:165-9 10.1016/j.gene.2003.10.009

Auboeuf D. 2020. Physicochemical Foundations of Life that Direct Evolution: Chance and Natural Selection are not Evolutionary Driving Forces. Life (Basel) 10(2):7 10.3390/life10020007

Belle EM, Smith N, Eyre-Walker A. 2002. Analysis of the phylogenetic distribution of isochores in vertebrates and a test of the thermal stability hypothesis. J Mol Evol 55(3):356-63 10.1007/s00239-002-2333-1

Berná L, Chaurasia A, Angelini C et al. 2012. The footprint of metabolism in the organization of mammalian genomes. BMC Genomics 13, 174 https://doi.org/10.1186/1471-2164-13-174

Berná L, Chaurasia A, Tarallo A, Agnisola C, D´Onofrio G. 2013. The shifting and the transition mode of vertebrate genome evolution in the light of the metabolic rate hypothesis: a Review. In: Advances in Zoology Research, vol. 5, chapter 3, pp. 65-93, Nova Science Publishers, Inc.

Baker Z, Schumer M, Haba Y, Bashkirova L, Holland C, Rosenthal GG, et al. 2017. Repeated losses of PRDM9-directed recombination despite the conservation of PRDM9 across vertebrates. eLife 6: e24133 10.7554/eLife.24133

Barbieri M. 2015. Code Biology. A New Science of Life. Springer, Cham, pp. 248

Bernardi G. 2001. Misunderstandings about isochores. Part 1. Gene 276(1-2):3-13 10.1016/s0378-1119(01)00644-8

Bernardi G. 2005. Structural and Evolutionary Genomics, Natural Selection in Genome Evolution, vol. 37, 1st Edition, Elsevier, pp 458

Bernardi G. 2007. The neoselectionist theory of genome evolution. PNAS 104(20): 8385– 8390

Bernardi G. 2008. Fish genomics: A mini-review on some structural and evolutionary issues. Marine Genomics 1:3-7

Bernardi G. 2015. Chromosome Architecture and Genome Organization. PLoS One 11(10) e0143739

Bernardi G. 2018. The formation of chromatin domains involves a primary step based on the 3-D structure of DNA. Scientific Reports 8(1):17821 doi: 10.1038/s41598-018-35851-0

Bernardi G. 2019. The Genomic Code: A Pervasive Encoding/Molding of Chromatin Structures and a Solution of the "Non-Coding DNA" Mystery. BioEssays 41(12): 1900106

Bernardi G, Bernardi G. 1985. Codon usage and genome composition. J Mol Evol 22: 363–365

Bernardi G, Bernardi G. 1986. Compositional constraints and genome evolution. J Mol Evol 24: 1-11

Bernardi G, Olofsson B, Filipski J, Zerial M, Salinas J, Cuny G, Meunier-Rotival M, Rodier F. 1985. The mosaic genome of warm-blooded vertebrates. Science 228: 953-958

Bernardi G, Hughes S, Mouchiroud D. 1997. The major compositional transitions in the vertebrate genome, J Mol Evol 44 (1):S44-S51

Bi X, Wang K, Yang L, Pan H, Jiang H, Wei Q, et al. 2021. Tracing the genetic footprints of vertebrate landing in non-teleost ray-finned fishes. Cell 184(5):1377-1391.e14 10.1016/j.cell.2021.01.04

Birnie A & Dekker C. 2021. Genome-in-a-Box: Building a Chromosome from the Bottom Up. ACS Nano. 15(1):111-124 10.1021/acsnano.0c07397

Blommaert J. 2020. Genome size evolution: towards new model systems for old questions. Proc Biol Sci. 287(1933):20201441. doi: 10.1098/rspb.2020.1441

Bohlin J & JH-O Pettersson. 2019. Evolution of Genomic Base Composition: From Single Cell Microbes to Multicellular Animals. Computational and Structural Biotechnology Journal, 17: 362-370

Borůvková V, Howell WM, Matoulek D, Symonová R. 2021. Quantitative approach to fish cytogenetics in the context of vertebrate genome evolution. Genes 12(2) 312 10.3390/genes12020312

Braasch et al., 2016. The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons. Nature Genetics

Bucciarelli G, Di Filippo M, Costagliola D, Alvarez-Valin F, Bernardi G, Bernardi G. 2009. Environmental Genomics: A Tale of Two Fishes. Mol Biol Evol 26(6):1235-43 10.1093/molbev/msp041

Burt DW. 2002. Origin and evolution of avian microchromosomes. Cytogenet Genome Res 96(1-4):97-112 10.1159/000063018

Calistri E, Livi R, Buiatti M. 2011. Evolutionary trends of GC/AT distribution patterns in promoters. Mol Phylogenet Evol 60(2):228-35 10.1016/j.ympev.2011.04.015

Canapa A, Barucca M, Biscotti MA, Forconi M, Olmo E. 2015. Transposons, Genome Size, and Evolutionary Insights in Animals. Cytogenet. Genome Res 147, 217–239

Chargaff E. 1950. Chemical specificity of nucleic acids and mechanism of their enzymic degradation. Experientia 6: 201-209

Chargaff E. 1951. Structure and function of nucleic acids as cell constituents. Fed. Proc. 10: 654-659

Chargaff E. 1963. Essays on Nucleic Acids, Elsevier, Amsterdam

Chargaff E. 1979. How genetics got a chemical education. Ann. NY Acad. Sci. 325:345-360

Charausia A, Tarallo A, Bernà L, Yagi M, Agnisola C, D'Onofrio G. 2014. Length and GC content variability of introns among teleostean genomes in the light of the metabolic rate hypothesis. Plos ONE 9(8): e103889 10.1371/journal.pone.0103889

Chojnowski JL, Franklin J, Katsu Y, Iguchi T, Guillette LJ Jr, Kimball RT, Braun EL. 2007. Patterns of Vertebrate Isochore Evolution Revealed by Comparison of Expressed Mammalian, Avian, and Crocodilian Genes. Journal of Molecular Evolution 65: 259-266 10.1007/s00239-007-9003-2

Christensen KA, Rondeau EB, Minkley DR, Sakhrani D, Biagi CA, Flores AM, Withler RE, et al. 2020. The sockeye salmon genome, transcriptome, and analyses identifying population defining regions of the genome. PLoS One 15(10):e0240935 10.1371/journal.pone.0240935

Clay O, Bernardi G. 2005. How not to search for isochores: a reply to Cohen et Al. Mol Biol Evol 22(12): 2315-7 10.1093/molbev/msi231

Cohen N, Dagan T, Stone L, Graur D. 2005. GC Composition of the Human Genome: In Search of Isochores. Molecular Biology and Evolution 22(5): 1260–1272

Costantini M, Auletta F, Bernardi G. 2007. Isochore patterns and gene distributions in fish genomes. Genomics 90: 364-371

Costantini M, Bernardi G. 2008a. The short-sequence designs of isochores from the human genome, Proc Natl Acad Sci USA 10: 13971-13976

Costantini M, Bernardi G. 2008b. Replication timing, chromosomal bands and isochores. Proc Natl Acad Sci USA 105(9):3433–3437

Costantini M, Cammarano R, Bernardi G. 2009. The evolution of isochore patterns in vertebrate genomes. BMC Genomics 10:146

Costantini M, Musto H. 2017. The Isochores as a Fundamental Level of Genome Structure and Organization: A General Overview. Journal of Molecular Evolution 84:93–103

Cornish-Bowden A. 1985. Nomenclature for incompletely specified bases in nucleic acid sequences: recommendations 1984. Nucleic Acids Research 13(9): 3021–3030 10.1093/nar/13.9.3021

Cozzi P, Milanesi L, Bernardi G. 2015. Segmenting the Human Genome into Isochores. Evol Bioinform Online 11:253-61. 10.4137/EBO.S27693

Craig N, Green R, Greider C, Storz G, C Wolberger. 2021. Molecular Biology, Principles of Genome Function 3rd Edition, Oxford University Press, ISBN 9780198788652, pp. 981

Cremer T & C Cremer. 2001. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. Nat Rev Genet 2, 292–301 10.1038/35066075

Cremer T & M Cremer. 2010. Chromosome territories. Cold Spring Harb Perspect Biol. 2(3): a003889 doi: 10.1101/cshperspect.a003889

Cuny G, Soriano P, Macaya G, Bernardi G. 1981. The major components of the mouse and human genomes: Preparation, basic properties and compositional heterogeneity. European Journal of Biochemistry 115: 227–233

Das P, Shen T, McCord RP. 2020. Inferring chromosome radial organization from Hi-C data. BMC Bioinformatics 21, 511 10.1186/s12859-020-03841-7

de Koning AP, Gu W, Castoe TA, Batzer MA, Pollock DD. 2011. Repetitive elements may comprise over two-thirds of the human genome. PLoS Genetics 7(12):e1002384 10.1371/journal.pgen.1002384

de Mendoza A, A, Hatleberg WL, Pang K, Leininger S, Bogdanovic O, Pflueger J, Buckberry S, Technau U, Hejnol A, Adamska M, Degnan BM, Degnan SM, Lister R. 2019. Convergent evolution of a vertebrate-like methylome in a marine sponge. Nature Ecology & Evolution 3:146-1476

Dion-Côté AM, Symonová R, Lamaze FC, Pelikánová Š, Ráb P, Bernatchez L. 2016. Standing chromosomal variation in Lake Whitefish species pairs: the role of historical contingency and relevance for speciation. Mol Ecol 26(1):1 78-192 10.1111/mec.13816

Di Stefano M, Paulsen J, Lien T et al. 2016. Hi-C-constrained physical models of human chromosomes recover functionally-related properties of genome organization. Sci Rep 6, 35985 10.1038/srep35985

D'Onofrio G, Ghosh TC, Saccone S. 2007. Different functional classes of genes are characterized by different compositional properties. FEBS Lett 581(30):5819-24 10.1016/j.febslet.2007

Duret L, Eyre-Walker A, Galtier N. 2006. A new perspective on isochore evolution. Gene 385: 71-4 10.1016/j.gene.2006.04.030

Duret L & N Galtier. 2009. Biased gene conversion and the evolution of mammalian genomic landscapes. Annu Rev Genomics Human Genet 10:285-311

Elhaik E, Landan G, Graur D. 2009. Can GC Content at Third- Codon Positions Be Used as a Proxy for Isochore Composition? Molecular Biology and Evolution 26 (8): 1829–1833 10.1093/molbev/msp100

Elhaik E, Graur D, Josić K. 2010a. Comparative testing of DNA segmentation algorithms using benchmark simulations. Mol Biol Evol 27: 1015–1024.

Elhaik E, Graur D, Josić K, Landan G. 2010b. Identifying compositionally homogeneous and nonhomogeneous domains within the human genome using a novel segmentation algorithm. Nucleic Acids Res 38: e158.

Elhaik & Graur, 2014. A Comparative Study and a Phylogenetic Exploration of the Compositional Architectures of Mammalian Nuclear Genomes. PLoS Computational Biology 10(11): e1003925 10.1371/journal.pcbi.1003925

Elliot TA & Gregory TR. 2015. What's in a genome? The C-value enigma and the evolution of eukaryotic genome content. Philos Trans R Soc Lond B Biol Sci. 370(1678):20140331 10.1098/rstb.2014.0331

Elsik CG, Tellam RL, Worley KC, Gibbs RA, Muzny DM, et al. 2009. The genome sequence of taurine cattle: a window to ruminant biology and evolution. Science 324: 522–528

Escobar JS, Glémin S, Galtier N. 2011. GC-biased gene conversion impacts ribosomal DNA evolution in vertebrates, angiosperms, and other eukaryotes. Mol Biol Evol28(9):2561-75

Eyre-Walker A. 1993. Recombination and mammalian genome evolution. Proceecings of the Royal Society Biological Sciences 252:237-243

Eyre-Walker A & LD Hurst. 2001. The evolution of isochores. Nature Review Genetics 2(7): 549-55 10.1038/35080577

Federico C, Scavo C, Cantarella CD, Motta S, Saccone S, Bernardi G. 2006. Gene-rich and gene-poor chromosomal regions have different locations in the interphase nuclei of cold-blooded vertebrates. Chromosoma 115:123-128

Figuet E, Ballenghien M, Romiguier J, Galtier, N. 2015. Biased Gene Conversion and GC-Content Evolution in the Coding Sequences of Reptiles and Vertebrates. Genome Biology and Evolution 7, 240–250 10.1093/gbe/evu277.

Flegr J. 2018. Evoluční biologie. Academia, Praha, pp. 570

Forsdyke DR. 2004. Regions of relative GC% uniformity are recombinational isolators. Journal of Biological Systems 12(3): 261-271

Forsdyke DR. 2016. Evolutionary Bioinformatics, 3rd Ed, Springer, pp 471

Forsdyke DR & JR Mortimer. 2000. Chargaff´s legacy. Gene 261: 127-137

Fortes GG, Bouza C, Martínez P, Sánchez L. 2007. Diversity in isochore structure among cold-blooded vertebrates based on GC content of coding and non-coding sequences. Genetica 129(3): 281-9 10.1007/s10709-006-0009-2

Fryxell KJ & Zuckerkandl E. 2000. Cytosine deamination plays a primary role in the evolution of mammalian isochores. Mol Biol Evol 17(9):1371-83 10.1093/oxfordjournals.molbev.a026420

Gaffaroglu M, Majtánová Z, Symonová R, Pelikánová Š, Unal S, Lajbner Z, Ráb P. 2020. Present and Future Salmonid Cytogenetics. Genes 11(12), 1462 10.3390/genes11121462

Galtier N & Duret L. 2007. Adaptation or biased gene coversion? Extending the null hypothesis of molecular evolution. TRENDS in Genetics 23(6): 273-276

Galtier N, Duret L, Glémin S, Ranwez V. 2009. GC-biased gene conversion promotes the fixation of deleterious amino acid changes in primates, Trends Genet 25 1:1-5

Gardner JD, Laurin M, Organ ChL. 2019. The relationship between genome size and metabolic rate in extant vertebrates. Philosophical Transactions of the Royal Society B. 375:20190146

Gazave E, Marques-Bonet T, Fernando O, Charlesworth B, Navarro A. 2007. Patterns and rates of intron divergence between humans and chimpanzees. Genome Biol 8 (2): R21-10.1186/gb-2007-8-2-r21

Gillooly JF, Brown JH, West GB, Savage VM, Charnov EL. 2001. Effects of size and temperature on metabolic rate. Science 293(5538):2248-51 10.1126/science.1061967

Gillooly JF, Allen AP, West GB, Brown JH. 2005. The rate of DNA evolution: effects of body size and temperature on the molecular clock. Proc Natl Acad Sci USA 102(1):140-5 10.1073/pnas.0407735101

Gingold H & Y Pilpel. 2021. Determinants of translation efficiency and accuracy. Molecular Systems Biology 7: 481 10.1038/msb.2011.14

Glasauer SMK & SCF Neuhauss. 2014. Whole-genome duplication in teleost fishes and its evolutionary consequences. Molecular Genetics and Genomics 289: 1045-1060

Gregory TR. 2021. Animal Genome Size Database. http://www.genomesize.com

Grigg GC, Beard LA et al. 2003. Body temperature in captive long-beaked echidnas (*Zaglossus bartoni*). Comp Biochem Physiol A Mol Integr Physiol 136(4):911-6. 10.1016/j.cbpb.2003.09.004

Haddrill PR, Charlesworth B, Halligan DL, Andolfatto P. 2005. Patterns of intron sequence evolution in *Drosophila* are dependent upon length and GC content. Genome Biol 6(8): R67-10.1186/gb-200568r67

Hall AN, Turner TN, Queitsch C. 2021. Thousands of high-quality sequencing samples fail to show meaningful correlation between 5S and 45S ribosomal DNA arrays in humans. Sci Rep 11(1):449 10.1038/s41598-020-80049-y

Hatje K, Mühlhausen S, Simm D, Kollmar M. 2019. The Protein-Coding Human Genome: Annotating High-Hanging Fruits. BioEssays 41 (11): e1900066: 1 -14

He X, Tillo D, Vierstra J, Syed KS et al. 2015. Methylated Cytosines Mutate to Transcription Factor Binding Sites that Drive Tetrapod Evolution. Genome Biol Evol 7(11):3155-69 10.1093/gbe/evv205

Hildebrand F, Meyer A, Eyre-Walker A. 2011. Evidence of Selection upon Genomic GC-Content in Bacteria. PLoS Genet 6(9): e1001107 10.1371/journal.pgen.1001107

Ho PS. 2009. Thermogenomics: thermodynamic-based approaches to genomic analyses of DNA structure. Methods 47(3):159-67 10.1016/j.ymeth.2008.09.007

Howe K, Clark MD, Torroja CF, Torrance J et al. 2013. The zebrafish reference genome sequence and its relationship to the human genome. Nature 496(7446): 498-503 10.1038/nature12111

Hughes S, Zelus D, Mouchiroud D. 1999. Warm-blooded isochore structure in Nile crocodile and turtle. Mol Biol Evol 16(11): 1521-1527

Hurst LD. 2019. A century of bias in genetics and evolution. Heredity 123:33-43

Hurst LD & AR Merchant. 2001. High guanine-cytosine content is not an adaptation to high temperature: a comparative analysis amongst prokaryotes. Proc. Biol. Sci. 268 (1466): 493–497 10.1098/rspb.2000.1397

Huttener R, Thorrez L, Veld T et al. 2019. GC Content of Vertebrate Exome Landscapes Reveal Areas of Accelerated Protein Evolution. BMC Evolutionary Biology 19 10.1186/s12862-019-1469-1

Iwasaki Y, Abe T, Okada N, Wada K, Wada Y, Ikemura T. 2014. Evolutionary changes in vertebrate genome signatures with special focus on coelacanth. DNA Res 21(5):459-67 10.1093/dnares/dsu012

Jabbari K & Bernardi G. 2004. Body temperature and evolutionary genomics of vertebrates: a lesson from the genomes of *Takifugu rubripes* and *Tetraodon nigroviridis*. Gene 333:179-81

Jabbari K & Bernardi G. 2017. An Isochore Framework Underlies Chromatin Architecture. PLoS One 12(1): e0168023

Jabbari K, Chakraborty M, Wiehe T. 2019. DNA sequence-dependent chromatin architecture and nuclear hubs formation. Sci Rep 9(1):14646 10.1038/s41598-019-51036-9

Kadi F, Mouchiroud D, Sabeur G, Bernardi G. 1993. The compositional patterns of the avian genomes and their evolutionary impications. Journal of Molecular Evolution 37:544-551

Kapusta A, Suh A, Feschotte C. 2017. Dynamics of genome size evolution in birds and mammals. PNAS 114(8):E1460-E1469 10.1073/pnas.1616702114

Kenigsberg E, Yehuda Y, Marjavaara L, Keszthelyi A, Chabes A, Tanay A, Simon I. 2016. The mutation spectrum in genomic late replication domains shapes mammalian GC content. Nucleic Acids Res 44(9): 4222-32 10.1093/nar/gkw268

Kent TV, Uzunović J, Wright SI. 2017. Coevolution between transposable elements and recombination. Philos Trans R Soc Lond B Biol Sci 372(1736):20160458 10.1098/rstb.2016.0458

Kostka D, Hubisz MJ, Siepel A, Pollard KS. 2012. The role of GC-biased gene conversion in shaping the fastest evolving regions of the human genome. Mol Biol Evol 29(3):1047-57 10.1093/molbev/msr279

Khuu P, Sandor M, DeYoung J, Ho PS. 2007. Phylogenomic analysis of the emergence of GC-rich transcription elements. PNAS 104(42):16528-33 10.1073/pnas.0707203104

Lamichhaney S, Catullo R, Keogh JS, Clulow S, Edwards SV, Ezaz T. 2021. A bird-like genome from a frog: Mechanisms of genome size reduction in the ornate burrowing frog, *Platyplectrum ornatum*. PNAS 118 (11) e2011649118 10.1073/pnas.2011649118

Lander et al., 2001. Initial sequencing and analysis of the human genome. Nature 409(6822):860-921 10.1038/35057062

Lane N & Martin W. 2010. The energetics of genome complexity. Nature 467(7318): 929-34 10.1038/nature09486

Lartillot N. 2013. Phylogenetic patterns of GC-biased gene conversion in placental mammals and the evolutionary dynamics of recombination landscapes. Mol Biol Evol 30(3):489-502 10.1093/molbev/mss239

Le Comber SC & Smith C. 2004. Polyploidy in fishes: patterns and processes. Biological Journal of the Linnean Society 32:431-442

Lefébure T, Morvan C, Malard F et al., 2017. Less effective selection leads to larger genomes. Genome Res 27(6):1016-1028 10.1101/gr.212589.116

Lehman R, Kovařík A, Ocalewicz K, Kirtiklis L, Tegner JN, Wanzenböck J, Bernatchez L, Lamatsch DK, Symonová R. *submitted.* DNA transposon expansion is associated with genome size increase in mudminnows.

Lemaire S, Fontrodona N, Aubé F, Claude JB et al. 2019. Characterizing the interplay between gene nucleotide composition bias and splicing. Genome Biology 20(1):259 10.1186/s13059-019-1869-y

Lemaitre C, Zaghloul L, Sagot MF, Gautier C, Arneodo A, Tannier E, Audit B. 2009. Analysis of fine-scale mammalian evolutionary breakpoints provides new insight into their relation to genome organisation. BMC Genomics 10:335 10.1186/1471-2164-10-335

Lertzman-Lepofsky G, Mooers AØ, Greenberg DA. 2019. Ecological constraints associated with genome size across salamander lineages. Proc. R. Soc. B 286:20191780 10.1098/rspb.2019.1780

Li XQ & Du D. 2014. Variation, evolution, and correlation analysis of C+G content and genome or chromosome size in different kingdoms and phyla. PLoS One 9(2):e88339 10.1371/journal.pone.0088339

Li X & B Guo. 2020. Substantially adaptive potential in polyploid cyprinid fishes: evidence from biogeographic, phylogenetic and genomic studies. Proc Biol Sci 287(1920): 20193008 10.1098/rspb.2019.3008

Lien S, Koop BF, Sandve SR, Miller JR, Kent MP, Nome T, Hvidsten TR, et al. 2016. The Atlantic salmon genome provides insights into rediploidization. Nature 533(7602):200-5 10.1038/nature17164

Long H, Sung W, Kucukyildirim S, Williams E, Miller SF, Guo W, et al. 2018. Evolutionary determinants of genome-wide nucleotide composition. Nat Ecol Evol 2(2):237-240 10.1038/s41559-017-0425-y

Lutsenko E, Bhagwat AS. 1999. Principal causes of hot spots for cytosine to thymine mutations at sites of cytosine methylation in growing cells. Mutation Research/Reviews in Mutation Research 437:11–20

Mable BK, Alexandrou MA, Taylor MI. 2011. Genome duplication in amphibians and fish: an extended synthesis. Journal of Zoology 284(3): 151-182 10.1111/j.1469-7998.2011.00829.x

Macaya G, Thiery JP, Bernardi G. 1976. An approach to the organization of eukaryotic genomes at a macromolecular level. J Mol Biol 108(1):237-54 10.1016/s0022-2836(76)80105-2

Macqueen DJ, Johnston IA. 2014. A well-constrained estimate for the timing of the salmonid whole genome duplication reveals major decoupling from species diversification. Proc Biol Sci 281(1778): 20132881 10.1098/rspb.2013.2881

Matoulek D, Borůvková V, Ocalewicz K, Symonová R. 2021. GC and Repeats Profiling along Chromosomes — The Future of Fish Compositional Cytogenomics. Genes 12 (1): 50 10.3390/genes12010050

Matsubara K, Kuraku S, Tarui H et al. 2012. Intra-Genomic GC Heterogeneity in Sauropsids: Evolutionary Insights from CDNA Mapping and GC3 Profiling in Snake. BMC Genomics 13, 604 10.1186/1471-2164-13-604

Melodelima Ch & Ch Gautier. 2008. The GC-heterogeneity of teleost fishes. BMC Genomics 9: 632 10.1186/1471-2164-9-632

Meunier J & Duret L. 2004. Recombination drives the evolution of GC-content in the human genome. Mol Biol Evol 21(6):984-90 10.1093/molbev/msh070

Mitchell D & R Bridge, 2006. A test of Chargaff's second rule". Biochem Biophys Res Com 340(1):90–94

Mitchell D. 2007. GC content and genome length in Chargaff compliant genomes. Biochem Biophys Res Commun. 2007 353(1):207-10 10.1016/j.bbrc.2006.12.008

Morim T, Smaers J, Symonová R. *in prep*. Trends and patterns in cytogenomic evolution along the fish phylogenetic tree

Mugal CF, Arndt PF, Holm L, Ellegren H. 2015a. Evolutionary consequences of DNA methylation on the GC content in vertebrate genomes. G3 (Bethesda) 5(3):441-7 10.1534/g3.114.015545

Mugal CF, Weber CC, Ellegren H. 2015b. GC-biased gene conversion links the recombination landscape and demography to genomic base composition: GC-biased gene conversion drives genomic base composition across a wide range of species. Bioessays 37(12):1317-26

Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil APOD. 2006. Harper's Illustrated Biochemistry; 28th ed., The McGraw-Hill, pp. 1271

Musto H, Naya H, Zavala A, Romero H,et al. 2006. Genomic GC level, optimal growth temperature, and genome size in prokaryotes. Biochem Biophys Res Commun 347(1):1-310.1016/j.bbrc.2006.06.054

Nam K, Ellegren H. 2012. Recombination drives vertebrate genome contraction. PLoS Genet 8(5): e1002680 10.1371/journal.pgen.1002680

NCBI Resource Coordinators. 2018. Database resources of the National Center for Biotechnology Information. Nucleic Acids Research 46(D1):D8-D13 10.1093/nar/gkx1095

Nowoshilow S, Schloissnig S, Fei JF, Dahl A et al. 2018. The axolotl genome and the evolution of key tissue formation regulators. Nature 554(7690):50-55 10.1038/nature25458

Oota S, Kawamura K, Kawai Y, Saitou N. 2010. A new framework for studying the isochore evolution: estimation of the equilibrium GC content based on the temporal mutation rate model. Genome Biol Evol 2:558-71 10.1093/gbe/evq041

Pappalardo AM, Ferrito V, Biscotti MA, Canapa A, Capriglione T. 2021. Transposable Elements and Stress in Vertebrates: An Overview. Int J Mol Sci 22(4):1970 10.3390/ijms22041970

Paudel R, Fedorova et al. 2020. Adapting Biased Gene Conversion theory to account for intensive GC-content deterioration in the human genome by novel mutations. PLOS ONE 15(4): e0232167

Peckham HE, Thurman RE, Fu Y, Stamatoyannopoulos JA, Noble WS, Struhl K, Weng Z. 2007. Nucleosome positioning signals in genomic DNA. Genome Res 17(8):1170-7 10.1101/gr.6101007

Petrov DA & Hartl DL. 1999. Patterns of nucleotide substitution in Drosophila and mammalian genomes. PNAS 96: 14751479

Peona V, Weissentsteiner M, Suh A. 2018. How complete are "complete" genome assemblies? - An avian perspective. Mol Ecol Resour 18:1188–1195

Piskurek O & Jackson JJ. 2012. Transposable Elements: From DNA Parasites to Architects of Metazoan Evolution. Genes 3(3): 409-422 10.3390/genes3030409

Politis SN, Mazurais D et al. 2017. Temperature effects on gene expression and morphological development of European eel, *Anguilla anguilla* larvae. PLoS One 12(8):e0182726 10.1371/journal.pone.0182726

Potapova TA & JL Gerton. 2019 Ribosomal DNA and the nucleolus in the context of genome organization. Chromosome Res 27(1-2):109-127 10.1007/s10577-018-9600-5

Pozzoli U, Menozzi G, Fumagalli M, Cereda M et al. 2008. Both selective and neutral processes drive GC content evolution in the human genome. BMC Evol Biol 8:99. doi: 10.1186/1471-2148-8-99

Pope B, Ryba T, Dileep V. et al. 2014. Topologically associating domains are stable units of replication-timing regulation. Nature 515: 402–405 10.1038/nature13986

Prabhu VV. 1993. Symmetry observation in long nucleotide sequences. Nucleic Acids Research 21 (12): 2797–2800

Prokopowich CD, Gregory TR, Crease TJ. The correlation between rDNA copy number and genome size in eukaryotes. Genome 46(1): 48-50 10.1139/g02-103

Ráb P, Rábová M, Reed KM, Phillips RB. 1999. Chromosomal characteristics of ribosomal DNA in the primitive semionotiform fish, longnose gar *Lepisosteus osseus*. Chromosome Res 7(6): 475-80 10.1023/a:1009202030456

Rodriguez F & Arkhipova IR. 2018. Transposable elements and polyploid evolution in animals. Curr Opin Genet Dev 49:115-123 10.1016/j.gde.2018.04.003

Romiguier J, Ranwez V, Douzery EJP, Galtier N. 2010. Contrasting GC-Content Dynamics across 33 Mammalian Genomes: Relationship with Life-History Traits and Chromosome Sizes. Genome Research 20, 1001–1009 10.1101/gr.104372.109

Romiguier J, Roux C. 2017. Analytical Biases Associated with GC-Content in Molecular Evolution. Front Genet 8:16 10.3389/fgene.2017.00016

Rousselle M, Laverré A et al. 2019. Influence of Recombination and GC-biased Gene Conversion on the Adaptive and Nonadaptive Substitution Rate in Mammals versus Birds. Mol Biol Evol 36(3):458-471 10.1093/molbev/msy243

Rudner R, Karkas JD, Chargaff E. 1968. Separation of *B. subtilus* DNA into complementary strands. III. Proc. Natl. Acad. Sci. USA 60: 921-922

Ruggiero RP & Boissinot S. 2020. Variation in Base Composition Underlies Functional and Evolutionary Divergence in Non-LTR Retrotransposons. Mobile DNA 2020, 11 10.1186/s13100-020-00209-9

Saccone F, Federico C, Bernardi G. 2002. Localization of the gene-richest and the gene-poorest isochores in the interphase nuclei of mammals and birds. Gene 300:169-178

Sacerdot C, Louis A, Bon C, Berthelot C, Roest Crollius H. 2018. Chromosome evolution at the origin of the ancestral vertebrate genome. Genome Biol 19(1):166 10.1186/s13059-018-1559-1

Samollow PB. 2008. The opossum genome: Insights and opportunities from an alternative mammal. Genome Research 18:1199-1215 10.1101/gr.065326.107

Shen W, Wang D, Ye B, Shi M, Ma L, Zhang Y, Zhao, Z. 2015. GC3-Biased Gene Domains in Mammalian Genomes: Fig. 1. Bioinformatics 31, 3081–3084 10.1093/bioinformatics/btv329

Simakov O, Marlétaz F, Yue JX, O'Connell B et al. 2020. Deeply conserved synteny resolves early events in vertebrate evolution. Nat Ecol Evol 4(6):820-830 10.1038/s41559-020-1156-z

Smith JJ, Timoshevskaya N, Timoshevskiy VA, Keinath MC, Hardy D, Voss SR. 2019. A chromosome-scale assembly of the axolotl genome. Genome Res 29(2):317-324 10.1101/gr.241901.118

Smyth DM. 1973. Temperature regulation in the platypus, *Ornithorhynchus anatinus*. Comparative Biochemistry and Physiology Part A: Physiology 45(3): 705-715

Snustad DP & MJ Simmons. 2017. Genetika. Masarykova univerzita, 2. aktualizované vydání

Sochorová J, Garcia S, Gálvez F, Symonová R, Kovařík A. 2018. Evolutionary trends in animal ribosomal DNA loci:introduction to a new online database. Chromosoma 127(1):141-50 10.1007/s00412-017-0651-8

Sola L & Gornung E. 2001. Classical and molecular cytogenetics of the zebrafish, *Danio rerio* (Cyprinidae, Cypriniformes): an overview. Genetica 111(1-3):397-412 10.1023/a:1013776323077

Stapley J, Feulner PGD et al. 2017. Variation in recombination frequency and distribution across eukaryotes: patterns and processes. Phil Trans R Soc B 372: 20160455 10.1098/rstb.2016.0455

Stoltzfus A, Yampolsky LY. 2009. Climbing mount probable: mutation as a cause of nonrandomness in evolution. J. Heredity 100: 637–647 doi:10.1093/jhered/esp048

Storz JF, Natarajan C et al. 2019 The role of mutation bias in adaptive molecular evolution: insights from convergent changes in protein function. Phil. Trans R Soc B 374: 20180238 10.1098/rstb.2018.0238

Storlazzi A, Xu L, Cao L, Kleckner N. 1995. Crossover and noncrossover recombination during meiosis: timing and pathway relationships. PNAS 92(18):8512-6 10.1073/pnas.92.18.8512

Sueoka N. 1962. On the genetic basis of variation and heterogeneity of DNA base composition, PNAS 48(4): 582-592

Sueoka N. 1988. Directional mutation pressure and neutral molecular evolution. PNAS 85(8):2653-57

Supek F, Lehner B, Hajkova P, Warnecke T. 2014. Hydroxymethylated cytosines are associated with elevated C to G transversion rates. PLoS Genet. 10(9):e1004585 10.1371/journal.pgen.1004585

Svensson EI & D Berger. 2019 The Role of Mutation Bias in Adaptive Evolution. Trends Ecol Evol 34(5): 422-434 10.1016/j.tree.2019.01.015

Symonová R 2019. Integrative rDNAomics - Importance of the oldest repetitive fraction of the eukaryote genome Genes 10 (5) 345

Symonová R, Majtánová Z, Arias-Rodriguez L, Mořkovský L, Kořínková T, Cavin L, Pokorná MJ, 2017a. Genome Compositional Organization in Gars Shows More Similarities to Mammals than to Other Ray-Finned Fish. J Exp Zool B Mol Dev Evol 328(7): 607-619 10.1002/jez.b.22719

Symonová R, Havelka M, Amemiya CT, Howell WM, Kořínková T, Flajšhans M, Gela D, Ráb P. 2017b. Molecular cytogenetic differentiation of paralogs of Hox paralogs in duplicated and re-diploidized genome of the North American paddlefish (*Polyodon spathula*). BMC Genetics 18(1): 19 10.1186/s12863-017-0484-8

Symonová R, Ocalewicz K, Kirtiklis L, Delmastro GB, Pelikánová Š, Garcia S, Kovařík A. 2017c. Higher-order organisation of extremely amplified, potentially functional and massively methylated 5S rDNA in European pikes (*Esox* sp.). BMC Genomics 18(1):391 10.1186/s12864-017-3774-7

Symonová R & WM Howell. 2018. Vertebrate Genome Evolution in the Light of Fish Cytogenomics and rDNAomics Genes 9 (2) 96

Symonová R & A Suh. 2019. Nucleotide composition of transposable elements likely contributes to AT/GC compositional homogeneity of teleost fish genomes. Mobile DNA 10: 1–8

Szybalski W, Kubinski H, Sheldrick O. 1966. Pyrimidine clusters on the transcribing strand of DNA and their possible role in the initiation of RNA synthesis. Cold Spring Harb Symp Quant Biol. 31:123–127

Šmarda P, Bureš P, Šmerda J, Horová L. 2012. Measurements of genomic GC content in plant genomes with flow cytometry: A test for reliability. New Phytol 193: 513–521 10.1111/j.1469-8137.2011.03942.x

Tarallo A, Angelini C, Sanges R, Yagi M, Agnisola C, D'Onofrio G. 2016. On the genome base composition of teleosts: the effect of environment and lifestyle. BMC Genomics 17:173 10.1186/s12864-016-25371

Tarallo A, Howell WM, Sammons S, Symonová R. *in prep*. Metabolic rate does not explain the conundrum of the AT/GC heterogeneity of gars and the homogeneity of bowfin

Tatarinova TV, Alexandrov NN, Bouck JB, Feldmann KA. 2010. GC3 biology in corn, rice, sorghum and other grasses. BMC Genomics 11:308 10.1186/1471-2164-11-308

Tatusov RL, Natale DA, Garkatsev IV et al. 2001. The COG database: new developments in phylogenetic classification of proteins from complete genomes. Nucleic Acids Res 29(1): 22–28

Tatusov RL, Fedorova ND, Jacskon JD, Jacobs AR et al. 2003. The COG database: an updated version includes eukaryotes. BMC Bioinformatics 4, 41.

Thiery JP, Macaya G, Bernardi G. 1976. An analysis of eukaryotic genomes by density gradient centrifugation. J Mol Biol 108: 219–235

Thompson A, Hawkins M, et al. 2021. The genome of the bowfin (*Amia calva*) illuminates the developmental evolution of ray-finned fishes. https://www.researchsquare.com/article/rs-92055/v1

Uliano E, Chaurasia A, Bernà L, Agnisola C, D'Onofrio G. 2010. Metabolic rate and genomic GC: what we can learn from teleost fish. Mar Genomics 3(1):29-34 10.1016/j.margen.2010.02.001

van der Meer J. 2021. Production efficiency differences between poikilotherms and homeotherms have little to do with metabolic rate. Ecol Lett 24(2):219-226 10.1111/ele.13633

Varriale A, Torelli G, Bernardi G. 2008. Compositional properties and thermal adaptation of 18S rRNA in vertebrates. RNA 14(8):1492-500. doi: 10.1261/rna.957108

Vavouri T, Lehner B. 2011. Chromatin organization in sperm may be the major functional consequence of base composition variation in the human genome. PLoS Genet 7(4):e1002036 10.1371/journal.pgen.1002036

Versteeg R, van Schaik BD, van Batenburg MF, Roos M, Monajemi R, et al. 2003. The human transcriptome map reveals extremes in gene density, intron length, GC content, and repeat pattern for domains of highly and weakly expressed genes. Genome Res 13(9):1998-2004 10.1101/gr.1649303

Vieira dos Santos E, Martinez PA, Souza G et al. 2021 Genome size drives ecological breadth in Pomacentridae reef fishes. J Experiment Mar Biol Ecol 540: 151544 10.1016/j.jembe.2021.151544

Vinogradov AE. 1998. Genome size and GC-percent in vertebrates as determined by flow cytometry: The triangular relationship. Cytometry 31:100-109 10.1002/(sici)1097-0320(19980201)31:2<100::aid-cyto5>3.0.co;2-q

Vinogradov AE. 2001. Bendable genes of warm-blooded vertebrates. Mol Biol Evol 18 (12): 2195-2200 10.1093/oxfordjournals.molbev.a003766.

Vinogradov AE. 2003. DNA helix: the importance of being GC-rich. Nucleic Acid Research 31(7): 1838-1844 10.1093/nar/gkg296

Vinogradov AE. 2005. Noncoding DNA, isochores and gene expression: nucleosome formation potential. Nucleic Acids Res 33 (2): 559-563. 10.1093/nar/gki184

Vinogradov AE & OV Anatskaya. 2017. DNA helix: the importance of being AT-rich. Mammalian Genome 28: 455-464

Vologodskii A & MD Frank-Kamenetskii. 2018. DNA melting and energetics of the double helix. Phys Life Rev 25:1-21 10.1016/j.plrev.2017.11.012

Wang D & J Yu. 2011. Both Size and GC-Content of Minimal Introns Are Selected in Human Populations. PLoS ONE 6(3): e17945 10.1371/journal.pone.0017945

Watson JD & FHC Crick. 1953. Genetical implications of the structure of deoxyribonucleic acid. Nature 171:964-967

Weber CC, Boussau B, Romiguier J, Jarvis ED, Ellegren H. 2014. Evidence for GC-Biased Gene Conversion as a Driver of between-Lineage Differences in Avian Base Composition. Genome Biology 15 10.1186/s13059-014-0549-1

Willmer P, Stone G, Johnston IA. 2000. Environmental physiology of animals. Blackwell Science, London. 644 pages, ISBN 0-632-03517-X

Yakovchuk P, Protozanova E et al. 2006. Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. Nucleic Acids Res 34(2):564-74 10.1093/nar/gkj454

Zhang L, Kasif S, Cantor CR, Broude NE. 2004. GC/AT-content spikes as genomic punctuation marks. Proc Natl Acad Sci USA 101(48):16855-60 10.1073/pnas.0407821101

Zhang J, Kuo CC, Chen L. 2011. GC content around splice sites affects splicing through pre-mRNA secondary structures. BMC Genomics 12:90 10.1186/1471-2164-12-90

Zhou W, Liang G, Molloy PL, Jones PA. 2020. DNA methylation enables transposable element-driven genome expansion. PNAS 117(32): 19359-19366. 10.1073/pnas.1921719117

Zhou Y, Shearwin-Whyatt L, Li J, Song Z, Hayakawa T, et al. 2021. Platypus and echidna genomes reveal mammalian biology and evolution. Nature. 2021 Jan 6. doi: 10.1038/s41586-020-03039-0

Zhu L, Zhang Y, Zhang W et al. 2009. Patterns of exon-intron architecture variation of genes in eukaryotic genomes. BMC Genomics 10, 47 10.1186/1471-2164-10-47

6. Poděkování

Ráda bych poděkovala své rodině za podporu v průběhu mého studia a vědecké práce. Poděkování patří všem mým kolegům i studentům z tuzemských i zahraničních institucí, kde jsem působila, a všem, se kterými jsem spolupracovala a měla možnost se od nich učit novému.

Tato práce vznikla za finanční podpory následujících projektů od různých poskytovatelů autorce této práce jakožto hlavní řešitelce:

GAČR, post-doktorský projekt „Allopolyploidisace jako klíčový faktor v evoluci genomu basálních linií obratlovců č. P506/11/P596, 2011-2013 (hodnocen jako vynikající)

Bilaterální projekt AV ČR – CONACYT Mexiko „Molekulárně-cytogenetická analýza genomu jeseterů, veslonosů a kostlínů – žijících svědků pradávných evolučních experimentů s genomem praobratlovců"

EMBO fellowship at the Arctic University of Norway, Tromso ASTF 390-2014

Projekt na Universitě Insbruck NWF 2015-2016 „Epigenetika ryb jako nástroj v genomice"

Tyrolský fond na podporu vědy „Evoluční význam a dynamika multiplikovaných rDNA v genomu ryb", UNI-0404/2015

Interní projekty na PřF UHK: 2018 „Cytogenomika transposonů ryb" a 2019 „Role transposonů v evoluci architektury genomu obratlovců"

Projekt Erasmus+ kategorie Strategická partnerství ve VŠ s názvem „Bioinformatické centrum Hradec Králové", 2019-1-CZ01-KA203-061433 řešen v letech 2019-2022

Horizon 2020 EuroTech Postdoc „GC biology of vertebrates" na Technické Universitě Mnichov (TUM) a Dánské Technické Universitě (DTU) Marie Skłodowska-Curie č. 754462, 2019-2021 – zde bych ráda poděkovala jmenovitě profesoru **Dmitriji Frishmanovi** z TUM a **Francesce Bertolini**, PhD z DTU, že mi na základě jediného emailu umožnili vypracovat a následně řešit tento projekt.

Dotace na individuální účel Krajského úřadu Královéhradeckého kraje s názvem „Bioinformatika – cesta k rozvoji vědy v Královéhradeckém kraji" č. 19RGI02

Mimořádné poděkování patří ing. **Dominiku Matoulkovi** a mgr. **Veronice Borůvkové** za jejich odvahu pustit se se mnou na pole bioinformatiky a biostatistiky nemodelových organismů. Bez jejich velkého přispění a entusiasmu by můj posun v této oblasti nebyl tak hladký a dynamický. Dominik ačkoliv softwarový inženýr biologií nedotčený mému výzkumu propůjčil své rozsáhlé IT know-how a vyvinul *ad hoc* nástroj EVANGELIST na analýzu a visualisaci DNA sekvencí v jazyce Python a trpělivě mě provázel a provází v mých prvních krocích při využívání jazyka Python. Veronika neváhala a pomohla mi proklestit se nástrahami statistiky a programování v jazyce R, čímž mi pomohla otevřít cestu *našemu* „kvantitativnímu přístupu v cytogenomice". Náš tým dokázal využít veřejně dostupná data a na mě bylo zasadit je do kontextu zde popsané komposiční biologie a využít jich ve snaze o pochopení dosavadních bílých míst. Takto si představuji interdisciplinární spolupráci a jsem za ni vděčná. Jinak snad evoluční cytogenomika nemůže v současné záplavě dat fungovat. Dále bych ráda poděkovala **Anastázii Sedlákové**, PhD za pomoc s „reaktivací" Python nástroje IsoSegmenter a další podporu v bioinformatické práci a jejímu manželovi **Filipu Sedlákovi**, PhD za podporu během mých prvních kroků na MetaCentru, tj. práce v systému Linux. Stejně tak patří můj velký dík **Liboru Mořkovskému**, PhD, který se mnou měl trpělivost v době, kdy jsme cytogeneticky zjistili AT/GC heterogenitu u kostlínů a já se teprve snažila zorientovat v nekonečných možnostech, které nabízí bioinformatika. Za naši spolupráci a přečtení a okomentování tohoto textu bych chtěla poděkovat RNDr. **Aleši Kovaříkovi**, CSc.

7. Přílohy

Seznam publikovaných článků, které jsou součástí habilitační práce

1. Borůvková V, Howell WM, Matoulek D, **Symonová R**. 2021. Quantitative Approach to Fish Cytogenetics in the Context of Vertebrate Genome Evolution. Genes (Basel) 12(2): 312 10.3390/genes12020312

2. Matoulek D, Borůvková V, Ocalewicz K, **Symonová R**. 2021. GC and Repeats Profiling along Chromosomes-The Future of Fish Compositional Cytogenomics. Genes (Basel) 12(1): 50 10.3390/genes12010050

3. Gaffaroglu M, Majtánová Z, **Symonová R**@, Pelikánová Š, Unal S, Lajbner Z, Ráb P.2020. Present and Future Salmonid Cytogenetics. Genes (Basel) 11(12):1462 10.3390/genes11121462

4. **Symonová R** & A Suh. 2019. Nucleotide composition of transposable elements likely contributes to AT/GC compositional homogeneity of teleost fish genomes. Mobile DNA 10: 49 10.1186/s13100-019-0195-y

5. **Symonová R**. 2019. Integrative rDNAomics-Importance of the Oldest Repetitive Fraction of the Eukaryote Genome. Genes (Basel) 10(5):345. doi: 10.3390/genes10050345

6. Sochorová J, Garcia S, Gálvez F, **Symonová R**, Kovařík A. 2018. Evolutionary trends in animal ribosomal DNA loci: introduction to a new online database. Chromosoma 127(1): 141-150 10.1007/s00412-017-0651-8

7. **Symonová R** & WM Howell. 2018. Vertebrate Genome Evolution in the Light of Fish Cytogenomics and rDNAomics. Genes (Basel) 9(2): 96 10.3390/genes9020096

8. **Symonová R**, Majtánová Z, Arias-Rodriguez L, Mořkovský L, Kořínková T, Cavin L, Pokorná MJ, Doležálková M, Flajšhans M, Normandeau E, Ráb P, Meyer A, Bernatchez L. 2017. Genome Compositional Organization in Gars Shows More Similarities to Mammals than to Other Ray-Finned Fish. J Exp Zool B Mol Dev Evol 328(7): 607-619 10.1002/jez.b.22719

9. Majtánová Z#, **Symonová R#**, Arias-Rodriguez L, Sallan L, Ráb P. 2017. "Holostei versus Halecostomi" Problem: Insight from Cytogenetics of Ancient Nonteleost Actinopterygian Fish, Bowfin *Amia calva*. J Exp Zool B Mol Dev Evol 328(7): 620-628 10.1002/jez.b.22720 #equal contribution

10. **Symonová R**, Ocalewicz K, Kirtiklis L, Delmastro GB, Pelikánová Š, Garcia S, Kovařík A. 2017. Higher-order organisation of extremely amplified, potentially functional and massively methylated 5S rDNA in European pikes (*Esox* sp.). BMC Genomics 18(1):391 10.1186/s12864-017-3774-7

11. **Symonová R**, Havelka M, Amemiya CT, Howell WM, Kořínková T, Flajšhans M, Gela D, Ráb P. 2017. Molecular cytogenetic differentiation of paralogs of Hox paralogs in duplicated and re-diploidized genome of the North American paddlefish (*Polyodon spathula*). BMC Genetics 18(1): 19 10.1186/s12863-017-0484-8

12. **Symonová R**, Flajšhans M, Sember A, Havelka M, Gela D, Kořínková T, Rodina M, Rábová M, Ráb P. 2013. Molecular cytogenetics in artificial hybrid and highly polyploid sturgeons: An evolutionary story narrated by repetitive sequences. Cytogenet Genome Res 141(2-3):153-62 10.1159/000354882

13. Havelka M, Bytyutskyy D, **Symonová R**, Ráb P, Flajšhans M. 2016. The second highest chromosome count among vertebrates is observed in cultured sturgeon and is associated with genome plasticity. Genet Sel Evol 48: 12 10.1186/s12711-016-0194-0

14. Dion-Côté AM, **Symonová R**, Lamaze FC, Pelikánová Š, Ráb P, Bernatchez L. 2016. Standing chromosomal variation in Lake Whitefish species pairs: the role of historical contingency and relevance for speciation. Mol Ecol 26(1):1 78-192 10.1111/mec.13816

15. Dion-Côté AM, **Symonová R**, Ráb P, Bernatchez L. 2015.Reproductive isolation in a nascent species pair is associated with aneuploidy in hybrid offspring. Proc Biol Sci 282(1802):20142862 10.1098/rspb.2014.2862

16. **Symonová R**, Majtánová Z, Sember A, Staaks GB, Bohlen J, Freyhof J, Rábová M, Ráb P. 2013b. Genome differentiation in a species pair of coregonine fishes: an extremely rapid speciation driven by stress-activated retrotransposons mediating extensive ribosomal DNA multiplications. BMC Evol Biol 13:42 10.1186/1471-2148-13-42

17. Knytl M, Kalous L, Symonová R, Rylková K, Ráb P. 2013. Chromosome studies of European cyprinid fishes: cross-species painting reveals natural allotetraploid origin of a Carassius female with 206 chromosomes. Cytogenet Genome Res 139(4):276-83 10.1159/000350689

18. Sember A, Bohlen J, Šlechtová V, Altmanová M, **Symonová R**, Ráb P. 2015. Karyotype differentiation in 19 species of river loach fishes (Nemacheilidae, Teleostei): extensive variability associated with rDNA and heterochromatin distribution and its phylogenetic and ecological interpretation. BMC Evol Biol 15:251 10.1186/s12862-015-0532-9

19. Majtánová Z, Choleva L, **Symonová R**, Ráb P, Kotusz J, -Pekárik L, Janko K. 2016. Asexual Reproduction Does Not Apparently Increase the Rate of Chromosomal Evolution: Karyotype Stability in Diploid and Triploid Clonal Hybrid Fish (*Cobitis*, Cypriniformes, Teleostei). PLoS One 11(1):e0146872 10.1371/journal.pone.0146872

8. Vědecké publikace s IF nezahrnuté v této habilitační práci

Symonová R, Fuková I, Lamatsch DK, Matzke-Karasz R, Paar J, Schmit O, Müller S. 2018. Karyotype variability and inter-population genomic differences in freshwater ostracods (Crustacea) showing geographical parthenogenesis. Genes (Basel) 9(3), 150 10.3390/genes9030150

Matzke-Karasz R, Neil JV, Smith RJ, **Symonová R**, Mořkovský L, Archer M, Hand, SJ; Cloetens, P; Tafforeau, P. 2014. Subcellular preservation in giant ostracod sperm from an early Miocene cave deposit in Australia. Proceedings of the Royal Society B – Biological Sciences 281(1786): 20140394

Bruvo R, Adolfsson S, **Symonová R**, Lamatsch DK, Schön, I, Jokela J, Butlin RK, Müller S. 2011. Few parasites, and no evidence for Wolbachia infections, in a freshwater ostracod inhabiting temporary ponds. Biological Journal of the Linnean Society 102(1): 208-216 10.1111/j.1095-8312.2010.01556.x

Matzke-Karasz R, Smith RJ, **Symonová R**, Miller CG, Tafforeau P. 2009. Sexual Intercourse Involving Giant Sperm in Cretaceous Ostracode. Science 324(5934):1535-1535 10.1126/science.1173898

Symonová R, Smrž J. 2009. First record of hemocytes and oenocytes in freshwater ostracodes. Journal of Crustacean Biology 29(1): 18-25 10.1651/08-3003.1

Symonová R. 2007. Ultrastructure of hepatopancreas and its possible role as a hematopoietic organ in non-marine cypridoidean ostracods (Crustacea). Hydrobiologia 585: 213-223 10.1007/s10750-007-0639-0

genes

MDPI

Article

Quantitative Approach to Fish Cytogenetics in the Context of Vertebrate Genome Evolution

Veronika Borůvková [1], W. Mike Howell [2], Dominik Matoulek [1] and Radka Symonová [3,*]

[1] Faculty of Science, University of Hradec Kralove, Hradec Kralove, 500 03 Czech Republic; veronika.boruvkova@uhk.cz (V.B.); dominik.matoulek@uhk.cz (D.M.)
[2] Department of Biological and Environmental Sciences, Samford University, Birmingham, AL, 35226 USA; wmhowell@samford.edu
[3] Department of Bioinformatics, Wissenschaftszentrum Weihenstephan, Technische Universität München, Freising 85354, Germany
* Correspondence: radka.symonova@gmail.com

Abstract: Our novel Python-based tool EVANGELIST allows the visualization of GC and repeats percentages along chromosomes in sequenced genomes and has enabled us to perform quantitative large-scale analyses on the chromosome level in fish and other vertebrates. This is a different approach from the prevailing analyses, i.e., analyses of GC% in the coding sequences that make up not more than 2% in human. We identified GC content (GC%) elevations in microchromosomes in ancient fish lineages similar to avian microchromosomes and a large variability in the relationship between the chromosome size and their GC% across fish lineages. This raises the question as to what extent does the chromosome size drive GC% as posited by the currently accepted explanation based on the recombination rate. We ascribe the differences found across fishes to varying GC% of repetitive sequences. Generally, our results suggest that the GC% of repeats and proportion of repeats are independent of the chromosome size. This leaves an open space for another mechanism driving the GC evolution in vertebrates.

Keywords: GC content; chromosome size; linkage group; microchromosomes; GC-biased gene conversion

Citation: Borůvková, V.; Howell, M.; Matoulek, D.; Symonová, R. Quantitative Approach to Fish Cytogenetics in the Context of Vertebrate Genome Evolution. *Genes* **2021**, *12*, 312. https://doi.org/10.3390/genes12020312

Academic Editor: Rossi Anna Rita and Samuel Martin

Received: 23 November 2020
Accepted: 17 February 2021
Published: 22 February 2021

1. Introduction

The quantitative approach in fish cytogenetics, or more precisely in cytogenomics, has so far been rather neglected and the same can be said for vertebrates in general. This is due to the lack of available data. The traditional cytogenetics prevailing during the last decades has mostly used a qualitative approach, which was indispensable for the exploration of chromosomal traits generally. With the still accelerating progress of genome sequencing, a large body of evidence on base composition, i.e., the proportion of guanines and cytosines in DNA (GC%), became available also in non-model vertebrates. In parallel, the Animal Genome Size Database [1] is another valuable resource of information useful particularly to estimate the completeness of genome assemblies. Only this recent development enabled us to tackle compositional cytogenomics of vertebrates from a quantitative viewpoint.

There were attempts to quantify results of the mostly qualitative (molecular) cytogenetics during all the phases of its development. With the first human karyotypes being presented during the early 1960s, it was not surprising to see an explosion of karyotypes being rapidly published in other living groups of organisms, especially vertebrates. However, this was before chromosome banding techniques (G-, Q-, C-, R-, AgNOR, etc.), karyotypes were simply grouped according to their size and the position of the centromere. Quantitatively, one could still determine a number of chromosomal

features: 1. The chromosome number of a species; 2. the position of the centromere; 3. often the position and number of the secondary constriction(s) = nucleolus organizer regions; 4. a measurement of the chromosome long arm length; 5. measurement of the short arm length; 6. measurement of the total chromosome length (TCL); 7. the long arm/short arm ratio (this was important as it pinpointed the position of the centromere along the chromosome) [2]; 8. the total haploid chromosome length (THCL); 9. the fundamental number (FN) of a karyotype (determined by counting the chromosome arms). Therefore, the FN ≤ 2 × 2 n, with the difference depending on the number of mono-armed chromosomes present (telocentric or acrocentric); 10. measurement of the centromeric index (CI), which is the length of the short arm divided by the total length of the chromosome × 100. The CI proved to be a very valuable quantitative measure for describing the shape of a chromosome, especially before chromosome banding methods [2,3].

Since the chromosomes of the vast majority of fish species could not be banded, especially G-banded, most of the authors of early fish karyotype papers included from one to several of the above counts and measurements in efforts to quantify their findings. The difficulty in obtaining chromosome banding patterns in fishes is likely due to their small size relative to that of higher vertebrates, and that, unlike birds and mammals, whose genomes can be divided into multiple GC-rich and AT-rich segments, fishes have been suggested to have little compartmentalization of their genomes by base composition [4–7], etc. Other quantitative traits were developed and have often been investigated in fish separately: Chromosome numbers, particularly in a phylogenetic context [8]; genome size expressed as the C-value based on the Animal Genome Size Database [1], as related to the population size [9], and to the nucleotypic effect [10,11].

The base composition at the level of fish chromosomes has so far been addressed in two model species in relation to isochores [12]. In two more fish species, the base composition was analysed in the entire genome [13,14]. However, these quantitative cytogenomic traits, particularly at the chromosome level, are crucial inputs in the analyses of mechanisms driving the regional GC% and the AT/GC heterogeneity in mammals and birds vs. the AT/GC homogeneity in lower vertebrates. Namely, one of the currently most accepted concepts trying to elucidate the aforementioned questions is the GC-biased gene conversion, gBGC [15–17]. This concept explains the increased GC% and its heterogeneous organisation in mammals and birds by the recombination rate (RR) per megabase pair (Mb; recently reviewed by [18]). Here, the chromosome or chromosome arm size plays an important role as there is at least one crossing-over (i.e., one recombination event) per one chromosome arm in bi-armed chromosomes and per chromosome in mono-armed chromosomes [18,19]. Among vertebrates, the efficiency of gBGC on the GC% has been investigated in great detail mostly in the coding (i.e., exonic) regions and above all in mammals [20–22] and birds [23,24], but also in reptiles [25,26]. The same applies for the online available GCevobase, an evolution-based database for the GC content in eukaryotic genomes displaying GC contents for all the annotated coding sequences from Ensembl [27]. However, the coding sequences make up not more than 2% of the human genome and similarly small fractions in other vertebrates (Matoulek et Symonová, unpublished data). Therefore, these highly detailed and sophisticated analyses referenced above could not catch the compositional evolution along large genome fractions as, e.g., the far more abundant repetitive fraction. Another approach represented by Frenkel et al., employed a far higher resolution beyond the chromosome, thus omitting their importance. They also investigated a limited number of species available at that time [28]. A more relevant study by Li and Du already focused on the chromosome level, however, dividing animals into mammals and non-mammals blurred the lineage specific traits together with the low number of analysed species available at that time [29]. Later, fully irrespective of GC% and its organisation, fishes were demonstrated to have the highest recombination rate among vertebrates [19] even without including the information on chromosome arms (FN) numbers, which would further

increase the acquired values. This shows an urgent need to revisit this issue from another, so far unexplored viewpoint.

GC% is linked to the genome size [30] and hence, chromosome counts are related to the genome size and chromosome size as well as their morphology (bi-armed vs. mono-armed; [19]). The genome size shows a clear positive association with the proportion of repetitive elements, particularly of transposable elements (TEs [31]; etc.).

In terms of genome composition, we need to distinguish the overall genomic GC-richness or GC content, i.e., the percentage of G + C (GC%) from the avian or mammalian situation recognized as the AT/GC heterogeneity [32], known also in non-teleost gars [33] and further confirmed by [34] in this special issue. The latter situation is characterized by an alternation of GC-rich and GC-poor regions along chromosomes, thus forming banding patterns upon an AT- and GC-specific staining (for more details, see [34] in this special issue). This AT/GC heterogeneity had been ascribed to a higher diversity of the isochores and their families distinguished according to their GC% [35]. In the AT/GC heterogeneous (mammalian, avian, and gar) genomes, the overall GC% can, however, be even lower than it is in the AT/GC homogeneous (fish) genomes. Considering that all of the currently available vertebrate genome assemblies contain gaps due to either repeat-rich or GC-rich regions [36], fish with GC-rich genomes might actually be even more GC-rich than estimated, and potentially even more GC-rich than mammalian and avian genomes. However, this bias should be comparable for all the genomes along the vertebrate phylogenetic tree. The only difference can be expected in newer improved versions of genome assemblies employing the hybrid approach of sequencing as, e.g., [37] recently in bats.

Our goal in this study is to utilize the increasingly available vertebrate genomes assembled to the chromosome level to assess the role of chromosome size in the overall GC content and in the AT/GC genomic heterogeneity across vertebrates. To do so, we employed our novel Python-based tool EVANGELIST published in this special issue [34]. This tool uses the sliding window approach to visualize and quantify the percentage of repeats (rep%) and GC percentage (GC%) in both repeats and non-repetitive DNA simultaneously along the chromosomes [34]. Employing this tool revealed several interesting quantitative traits at the chromosomal level across the vertebrates. These inputs together with a large-scale meta-analysis of genomic data serve as a test of the concept of GC-biased gene conversion in fishes and in other vertebrates.

2. Materials and Methods

We meta-analysed NCBI/genome records (accessed in January 2021) on the genome size, chromosome size, and GC percentage (GC%) of entire genomes and their chromosomes (in the genome assemblies called linkage groups, LGs).

2.1. Data Acquisition, Filtering, and Manual Curation

At the NCBI webpage (NCBI, 2016) using the available genome selection filters, we selected chordates assembled to the chromosome level. We manually checked the obtained datasets for multiple records (particularly human, dog, primates, and other model species), incomplete assemblies and hybrids. The online filtering yielded 157 fishes, 10 amphibians, 13 reptiles, 66 birds, and 327 mammals. However, upon the manual inspections, the numbers of species usable in our study decreased to 85 mammals, 55 birds, 12 reptiles, nine amphibians, and 118 fishes. Among fishes, there were four chondrichthyans, three lampreys, one lungfish with an incomplete dataset, one sturgeon, one gar, one bichir, and 107 teleosts (lists of species are in the Appendix). These data were used for Figures 1–4 and Figures S2 and S3. These figures were prepared using R [38] with the library scales included in the basic R and further with packages ggplot2, forcats, and ggpmisc, all belonging to tidyverse [39]. There was a bias in mammalian lineages represented among the species assembled to the chromosome level. Namely, some lineages were not represented at all (Afrotheria, Cingulata, Scandentia, Dermoptera,

Eulipotyphla, and Pholidota), some were underrepresented (Chiroptera, the second largest mammalian order with only three species available), and some were overrepresented (primates, rodents, and artiodactyls). More details are available in the Supplementary Methods. Therefore, we have selected a subset of 26 mammals trying to compensate for this bias and to avoid the overrepresented species skewing the results. Similarly, in birds, the most abundant order available were Passeriformes and to a less extent also Anseriformes. Hence, we again selected a subset of 18 avian species to avoid any bias towards these orders. However, some avian species, despite being filtered, turned out not to have the required chromosome level assembly available. Finally, we performed a comparable species selection in fish and produced a subset of 49 species representing the orders, as much as possible, and not overrepresenting the better explored ones.

2.2. Repeats Analyses and Genome Size Data

To employ the Python-based tool EVANGELIST [34], we have used repeat-masked data from Ensembl to analyse a potential relationship between the GC% and repeats proportion (rep%). This approach was utilized also to address the issue of a potential effect of the decreasing chromosome size (particularly in microchromosomes) on the GC% and rep%. These data were used for Figure 5 and Figures S4–S7.

Genome size data obtained from the Animal Genome Size Database [1] was utilized during the manual curation of the completeness of genomic data from NCBI and to the comparison of genome size and GC% between mammals and fish (Figure S1).

3. Results

Initially, we analysed the potential relationships between the chromosome size and its GC% across ray-finned fishes and compared them to one tunicate, one cephalochordate, all the three available lampreys, all the three available chondrichthyans, and to other vertebrates (amphibians, reptiles, birds, and mammals; Figure 1a, each point represents a single chromosome; graphs for each group are shown separately in Figure 1b–j). This shows that most fishes occupy the area with the lowest GC% and intermediate chromosome sizes (blue points). The single tunicate represented, the sea squirt (*Ciona intestinalis*, brown points in Figure 1b), occupies the actual lowest values of GC% and chromosome sizes and shows no association between these measures. The single cephalochordate represented, the Florida lancelet (*Branchiostoma floridae*, pink points in Figure 1c) is embedded approximately in the middle of the "fish area" surrounded by fish and shows a significant positive association between the GC% and chromosome size (better details in Figure 2a). In lampreys, the GC% and chromosome sizes are significantly positively associated (black points, Figure 1d, Figure 2b in better details). Chondrichthyans (black points, Figure 1e) but also sturgeon and gar, show a negative association between the GC% and chromosomes forming a curve resembling a hyperbole similarly to avian chromosomes (more details in Figure 3). In fishes (Figure 1a,f, blue), the linear approximation is mostly a suitable model representing their relationship between the chromosome size and GC%, however, all negative, positive, and no associations occur (more details in Figure 2 and below). Only eight of the nine currently available amphibian species are visualized here (Figure 1a,g, red) due to the extremely huge chromosome size in axolotl (*Ambystoma mexicanum*). All the amphibians are visualized in Figure 4 in proper details showing different levels of the negative association between the GC% and chromosome size not apparent at the resolution in Figure 1. Reptiles (Figure 1a,h, violet) with macro- and microchromosomes show a similar association of GC% with the chromosome size to chondrichthyans, basal fish lineages (sturgeon and gar), and birds. This means a negative relationship of GC% vs. chromosome size, behaving as if there was no association in large(r) (macro)chromosomes and a steep descent in GC% in microchromosomes. Birds (Figure 1a,i, orange) show a clear hyperbolic curve. Mammals (Figure 1a,j, green) show a negative relationship between GC% and the chromosome size with a varying strength similar to fishes. Violin plots of compositional differences between the selected fishes and mammals are in Figure S2.

3.1. Variability in Relationships between the Chromosome Size and GC% in Ray-Finned Fishes

Figure 2 shows representative species with diverse trends in their relationship between the chromosome size and GC%. The cephalochordate outgroup of all vertebrates, the lancelet (*B. floridae*), shows a positive correlation between GC% and the chromosome size that fits the linear model well (Figure 2a). Its haploid genome assembly size 513.461 Mb corresponds to the smallest teleosts and the size of its nineteen haploid chromosomes ranging between 17.12 and 35.34 Mb. The jawless outgroup of other vertebrates, the lamprey (*Lethenteron resissneri*) shows a negative correlation between GC% and the chromosomes size fitting the linear model well (Figure 2b). Its haploid genome assembly size is 1063.01 Mb and corresponds to the average teleosts, however, the size of its 72 haploid chromosomes ranges between 4.5 and 26 Mb. The remaining fish species depicted show diverse associations of diverse strength or no association at all. The non-teleost reedfish (*Erpetoichthys calabaricus,* Figure 2c) has a huge (haploid) genome of 3209 Mb separated into only eighteen chromosomes ranging in their size between 88.37 and 350.1 Mb (the largest chromosome is almost as large as the entire genome of pufferfishes, which have a reduced genome size). Its regression line is clearly decreasing, however, due to the extremely different scales of the X and Y axes, the line slope is close to zero. The relationship fits the linear model well (Figure 2c). Following are two teleosts belonging to the superorder Protacanthopterygii [40] but showing different trends and evolutionary history. Namely, in the Northern pike (*Esox Lucius*, Figure 2d), GC% decreases almost linearly with the increasing chromosome size, thus fitting the linear model well. This species has a typical teleost haploid genome size and chromosome number (940 Mb and 25 chromosomes ranging between 22.6 and 55.4 Mb). Whereas in the Atlantic salmon (*Salmo salar*, Figure 2e), that underwent the salmonid specific whole genome duplication points in the graph, form a cloud that does not fit the linear model and with no obvious simple relationship between the variables. This species has the haploid genome size 2966.89 Mb in 29 chromosomes of size between 40 and 160 Mb. The same situation exists in two other salmonids in NCBI assembled to the chromosome level (*Salmo*, *Oncorhynchus*, but not in *Thymallus* and *Salvelinus* that both show a slightly negative association between GC% and the chromosome size, not shown). The next species analysed is the model organism zebrafish (*Danio rerio*, Figure 2f) of the basal teleost lineage Ostariophysi [40], known to have a rather larger [1,41] and AT-enriched genome, e.g., [33]. Its haploid genome assembly of about 1408 Mb is divided into 25 chromosomes (2 n = 50) ranging in size between 37.5 and 59.58 Mb. Excluding the extreme, rightmost point of the graph the GC% oscillates around the mean independently of the chromosome size and the relationship does not fit a linear model. Other cypriniform fishes analysed (*Cyprinus* and *Carassius*, not shown) show similar trends, i.e., no association between the chromosome size and GC%. Similarly, in the Japanese rice fish (*Oryzias latipes*, Figure 2g), Beloniformes, the lower GC% appears to occur in larger chromosomes, but the relationship cannot be described in simple terms and does not fit the linear model well. So far, a unique situation among teleosts occurs in toadfish (*Thalassophryne amazonica*, Figure 2h) of the order Batrachoidiformes [40]. This species shows a significantly positive association between GC% and the chromosome size, similarly to the cephalochordate and a large haploid genome for a teleost, 2446.59 Mb with 23 large chromosomes ranging in size between 36.25 and 175.46 Mb. These traits together with its GC-richness (42.02%) resemble salmonids. Furthermore, the Southern platyfish (*Xiphophorus maculatus*, Figure 2i) of the order Cyprinodontiformes shows a tight linear relationship between GC% and the chromosome size ($R^2 = 0.7896$) with points distributed visibly along the decreasing line. Finally, we selected pufferfish (*Takifugu rubripes*, Figure 2j) representing the order Tetraodontiformes with the smallest vertebrate genomes. It might appear that the shape of the line is influenced mainly by the two outer points. However, the decreasing tendency holds (although not that strong) without them and their negative association fits the linear model well.

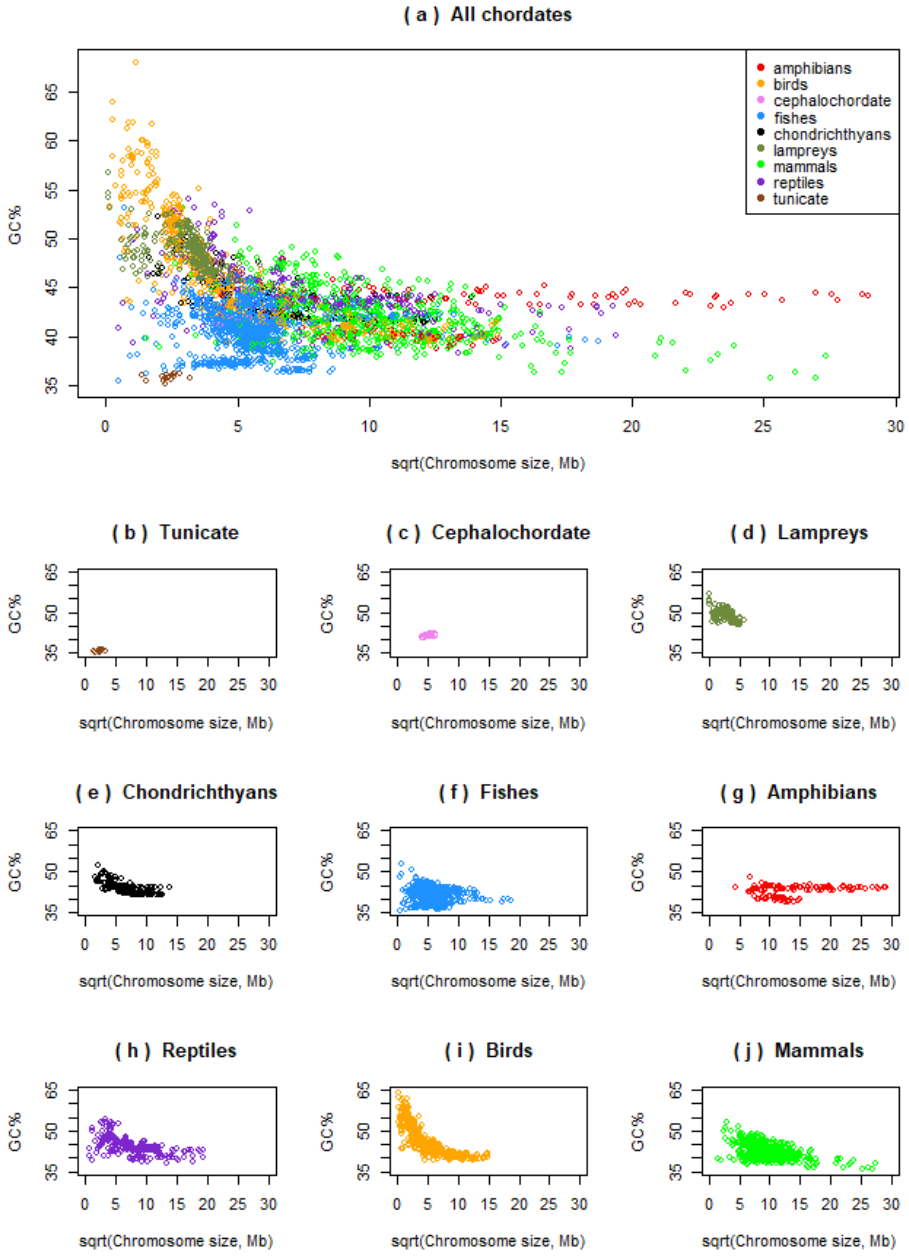


Figure 1. Large-scale overview of the relationships between chromosome size (*x*-axis, square root transformed for a better visualization) and chromosome GC% (*y*-axis) across selected chordates on equidistant axes x and y. (**a**) All chordates together; (**b**–**j**) major chordate lineages separately ordered phylogenetically. The graph (**f**), fishes, includes two basal ray-finned species (sturgeon and gar) and teleosts. In (**g**), the largest genome of axolotl is not visualized, see Figure 4 for amphibians. Each point represents a single chromosome of a species within the given group.

3.2. Basal Fish Lineages Show a Similar Relationship between the Chromosome Size and GC% to Birds and Some Reptiles

In the two non-teleost ray-finned fishes and three chondrichthyans with available genomes assembled to the chromosome level (four shown here, Figure 3a–d), the GC% is inversely related to their chromosome size as in some reptiles (a snake and a turtle shown here, Figure 3e–f) and in birds (four shown here, Figure 3g–j). This relationship results in a hyperbolic curve in all the species analysed here. The curve is "smoother" for birds than

in these basal fish lineages and the selected reptiles, where points representing single chromosomes are more scattered. In chondrichthyans, reptiles, and birds the chromosome size around 50 Mb represents a breakpoint behind the fact that the GC% becomes independent of the chromosome size. Sterlet and gar have this breakpoint around 20 Mb.

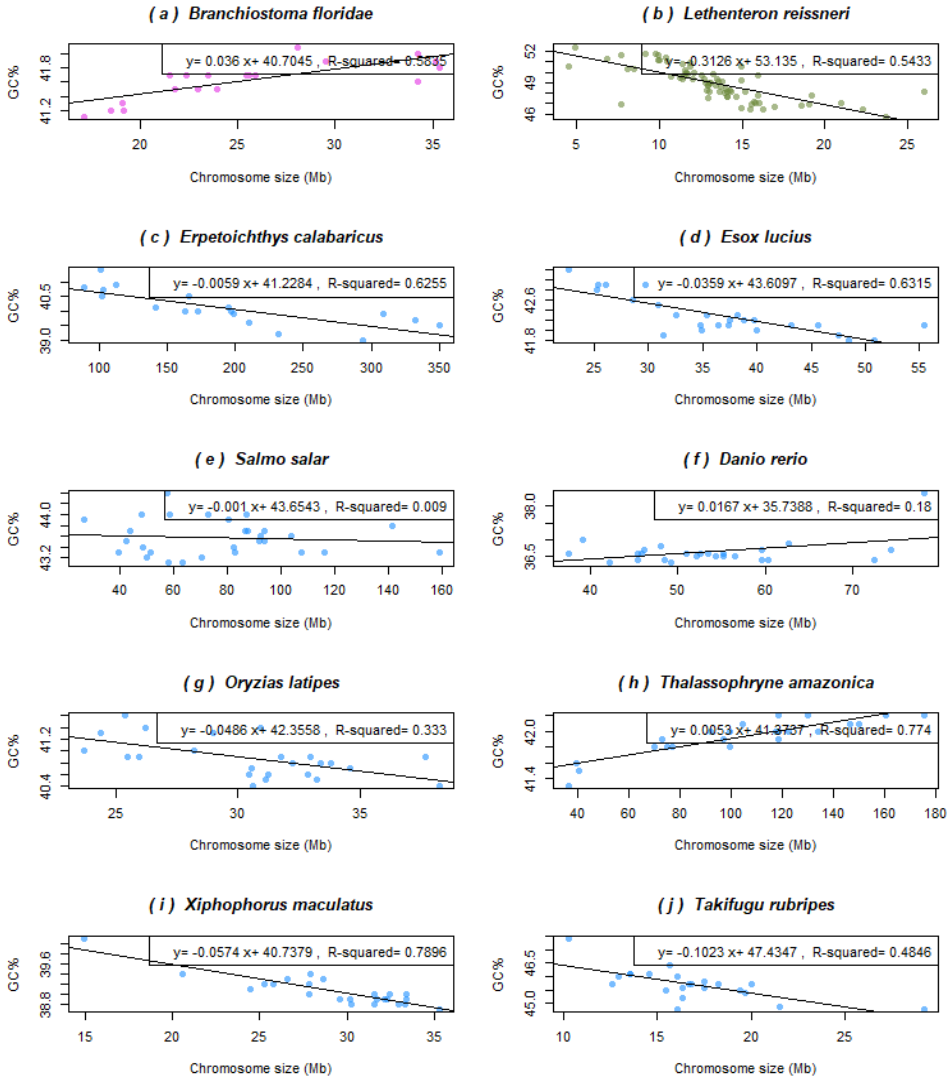


Figure 2. Relationships between GC% and the chromosome size in Mb in selected fish and fish-like species representing different patterns of this association. (**a**) Lancelet as the basal chordate considered the ancestral condition shows a significantly positive association between the chromosome size and GC%; (**b**) lamprey, the extremely GC-rich jawless fish with numerous mostly tiny chromosomes, and reedfish (**c**) with its giant non-teleost genome and extremely large chromosomes both show a significantly negative association; (**d**) the Northern Pike, an important pre-duplication model species for salmonids, shows a significantly negative association, as well; (**e**) in the Atlantic salmon, there is no association between these two measures; (**f**) zebrafish, although the mostly investigated fish model with a larger teleost genome, shows a rather weak positive association; (**g**) medaka, another important model species, shows a weaker negative association; (**h**) toadfish with a surprisingly large teleost genome shows a significantly positive association and the largest teleost chromosome, although no whole genome duplication is known here as, e.g., in salmonids with comparable genome and chromosome size but higher chromosome numbers; (**i**) the Southern platyfish shows the most significant negative association and the typical teleost chromosome sizes; (**j**) fugu represents the compact teleost and vertebrate genomes shows a slightly less prominent but still significantly negative association. The colour code corresponds to Figure 1.

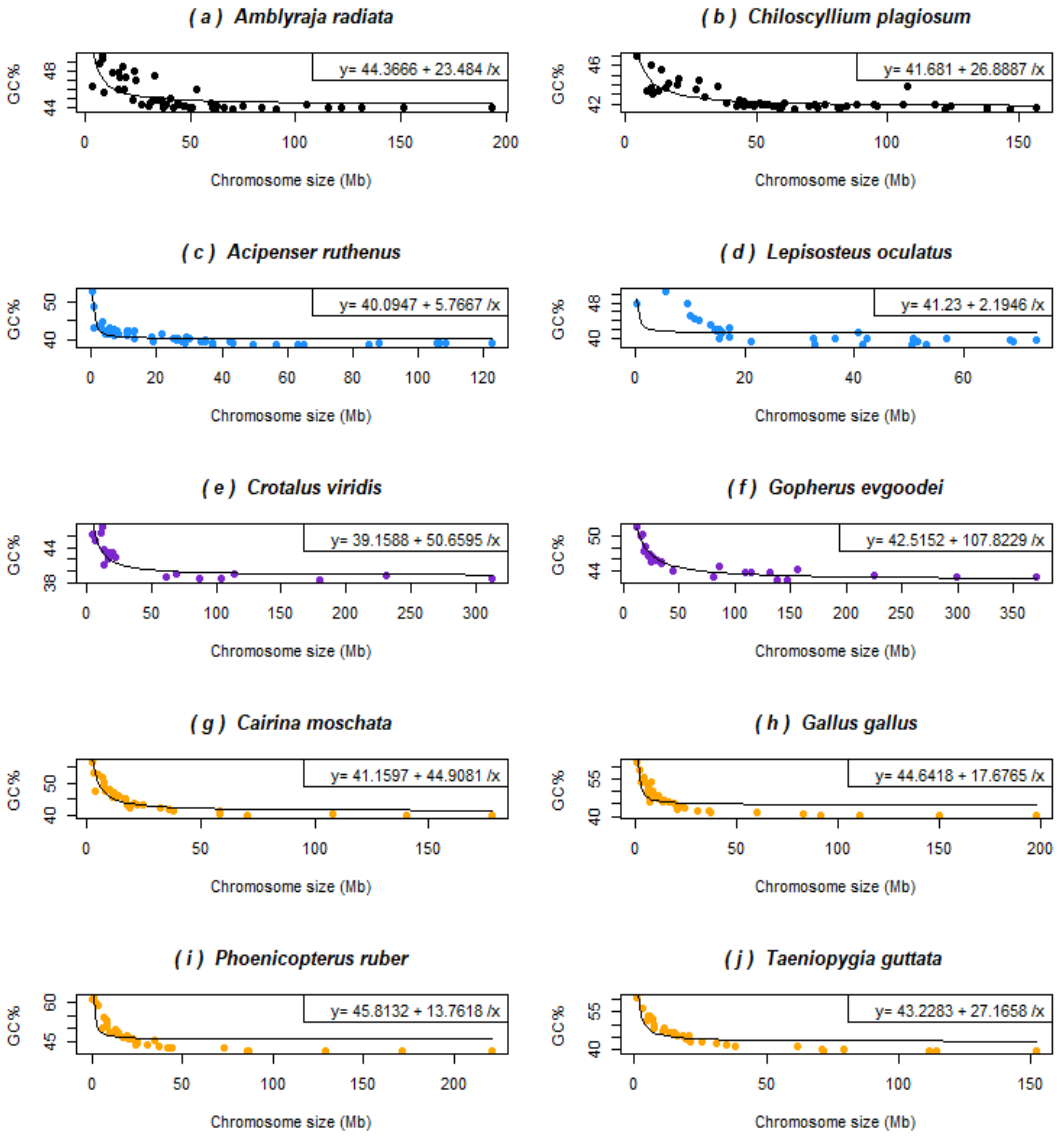


Figure 3. Relationships between GC% and the chromosome size (Mb) in basal fish lineages that all possess microchromosomes show an inverse relationship as do birds and reptiles with microchromosomes. (**a**) The thorny skate, representing the superorder Batoidea, is the GC-richest species in this group; (**b**) the whitespotted bamboo shark represents the superorder Selachimorpha; (**c**) sterlet, one of the only two sequenced sturgeons, with the GC-poorest genome; (**d**) the spotted gar belongs to the infraclass Holostei and is known for its mammalian-like AT/GC compartmentalization; (**e**) the rattlesnake represents a smaller and GC-poorer reptilian genome; (**f**) gopher tortoise represents a larger and more GC-rich reptilian genome; (**g**) the Muscovy duck is the first avian genome compared here; (**h**) chicken represents one of the best assembled avian genomes; (**i**) the American flamingo represents large wading birds; finally, (**j**) the zebra finch with its fifteen genome assemblies available represents songbirds. Violin plots of these species are available in the Appendix, Figure S3. The colour code corresponds to Figure 1.

3.3. Genome Size Drives GC% in Amphibians but Inversely Than Expected

Amphibians greatly differ in their genome size and salamanders are known to reach the second highest vertebrate genome sizes (116 and 118 Gb [42]). This fact highly influences the availability and also quality of amphibian genome assemblies in the NCBI database [43], whereas cytological data (C-values is pg) are very rich and provide robust evidence on the genome size in anurans and urodels [1]. The mean C-value of anurans and urodels is 3.59 pg and 35.21, respectively [1]. Our results show interesting consequences

of this dynamic genome size evolution in amphibians for their GC%. Here, even such a low sample size (nine species) enabled to distinguish three distinct groups, partly phylogenetically independent, according to their GC%, chromosome size, and associations thereof (Figure 4):

1. A single salamander species with its giant and extremely GC-rich chromosomes (and genome) showing a weak negative association between GC% and the chromosome size (R^2 = 0.32; genome size 32,396.4 Mb, GC ~ 46.5%).
2. Three caecilians and two frogs (pelobatid and bufonid) with an intermediate GC% and chromosome sizes (and intermediate genome sizes between 3779.43 and 5319.24 Mb, GC ~ 43%–44%) showing no association in two caecilians (R^2 = 0.066 in *Rhinatrema*, R^2 = 0.021 in *Geotrypetes*) to a significant negative association in the remaining caecilian *Microcaecilia* and two anurans (R^2 = 0.63–0.72) between GC% and the chromosome size.
3. Three remaining frogs (two pipids and a pyxicephalid) with a significantly (R^2 = 0.5837, 0.6689, and 0.7824) negative association between GC% and the chromosome size (the smallest genomes between 1451.3 and 2718.43 Mb, GC ~ 39%–40.5%).

This indicates that GC% increases together with the genome size and the genome size appears to be the determinant of this peculiar situation.

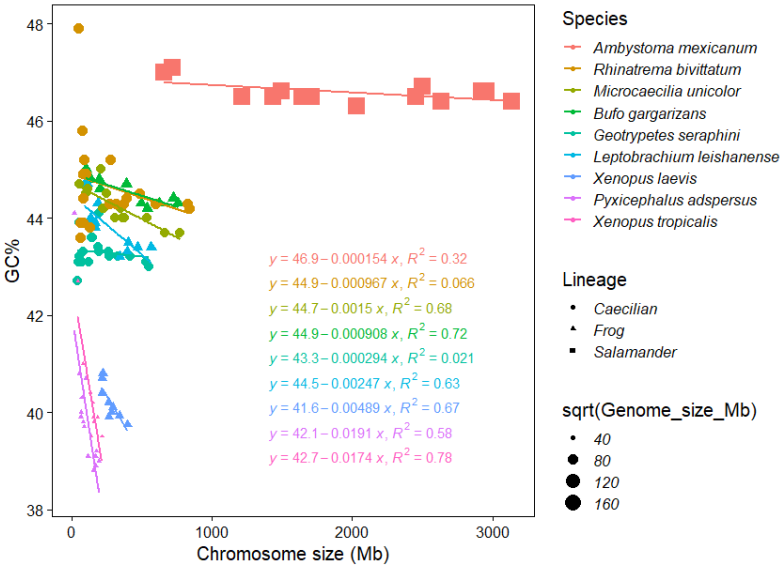


Figure 4. The chromosome size and GC% of chromosomes in currently available amphibians. Nine species ordered according to the size of their entire genomes in the legend and equations, which is also reflected in the point size. Point shapes reflect the main amphibian lineages. Patterns of the associations between the chromosome size and GC% forms three distinct groups partly corresponding to the main lineages and fully corresponding to genome sizes.

3.4. GC% vs. Repeats Proportion (%) and Chromosome Size

To explore any potential relationship between the chromosome size and their corresponding GC%, rep%, and GC% of repeats, we calculated these measures for each chromosome separately in 43 fish and fish-like species and in eleven mammals. These data were extracted from our tool EVANGELIST [34] and represent individual sliding windows along chromosomes and their GC% and rep% values. The results are pictured as scatter plots for each chromosome organized as tiles in Supplementary online data S4-S7 and on the GitHub online repository of our tool EVANGELIST (https://github.com/bioinfohk/evangelist_plots/tree/master/rep%25_vs_GC%25). These plots show several patterns in the relationships between GC% and repeats%, but no obvious association of GC%,

rep%, and GC% of repeats with the chromosome size. The distribution of repeats percentage is asymmetric for each chromosome, while typically unchanging throughout the species and independent of the chromosome size. This can be seen from the position of the median of rep% (vertical line) in each tile. The same holds for the position of regression lines (similar in shape and position within the species). We also analysed the relationship between the repeats proportion at the level of entire chromosomes extracting the median rep% for each chromosome and it size (Figure 5). This is first shown in three cyprinids (goldfish, carp, and zebrafish, Figure 5a) differing in chromosome counts but similar genome sizes, where a significant ($R^2 = 0.46$) negative association exists in goldfish, a very slight positive association in zebrafish, and no association at all in carp. Two compact tetraodontiform genomes show a moderate positive association in tetraodon and no association in fugu (Figure 5b). Another complex situation is shown in the Northern pike and three salmonids, where whole-genome duplication might have played an important role (Figure 5c). Interestingly, there is no clear association between the chromosome size and rep% even in microchromosomes (Figure 5d), where GC% is otherwise negatively associated with the chromosome size. Here, however, the availability of repeats libraries and hence, the quality of the repeat-masking process needs to be considered particularly in basal fish lineages (sturgeon and gar).

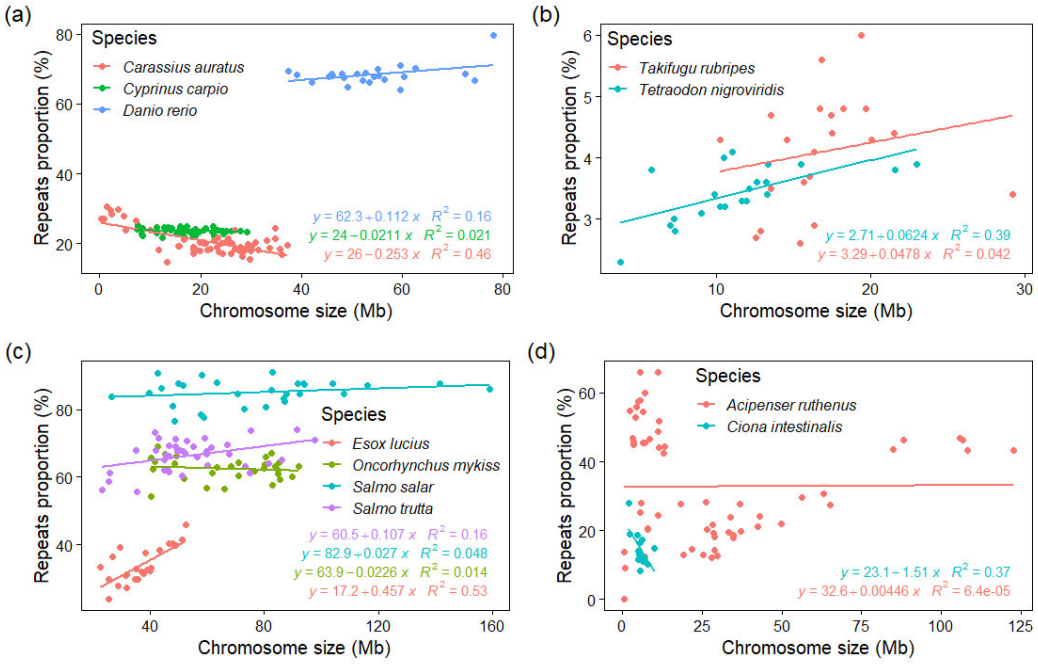


Figure 5. Relationships between the repeats proportion and chromosome size in selected fish species with relevant traits—genome size, chromosome numbers and sizes, and whole-genome duplication or genome compaction underwent. (**a**) In three cyprinids with highly different cytogenomic traits; (**b**) two of the smallest teleost (and vertebrate) genomes show different associations; (**c**) the Northern pike as the pre-duplication ancestor of salmonids shows a clear positive association, while three salmonids show no association with a substantial repeats enrichment; (**d**) the tunicate sea squirt has a chromosome size comparable with microchromosomes of sturgeon, however, differs greatly in proportions of repeats showing a negative association.

4. Discussion

Chromosome sizes greatly differ among currently sequenced chordates. In parallel to chromosome sizes, genome sizes and proportions of repeat also differ considerably. Despite some mostly superficial similarities, each major lineage of vertebrates occupies its own specific cytogenomic "niche" composed of a unique combination of cytogenomic traits.

4.1. Genome Size and Chromosome Numbers Do Not Entirely Explain the Difference in GC Evolution between Fish and Mammals

The meta-analysis of currently available data on genome size and GC% in vertebrates [1,43] shows that fish genomes are, despite their additional teleost-specific genome duplication [44], about two to three times smaller than mammalian genomes (Figure S1). However, both fish and mammals have converged to very similar diploid chromosome numbers (2 n) despite this difference in their genome evolution. Namely, the means and the medians of 2 n are 48–50 for fish [8] and 46–47 for mammals [1]. With these chromosome counts and fish genome size two to three times smaller than in mammals, this would mean roughly double the recombination rate per megabase (Mb) of DNA in fish in comparison with mammals [19]. This should result in higher GC contents in fish according to the concept of GC-biased gene conversion (gBGC) suggesting that the GC-richness of the mammalian genome is positively associated with higher recombination rates [18]. However, this is not the case—fish genomes are GC-poorer than the mammalian ones (Figure S1). Moreover, fish genomes are AT/GC homogenous in comparison with AT/GC heterogeneity in mammals as we show here in Appendix B, in our previous study in this special issue [34], as well as in [35]. This implies that the basic quantitative traits (genome size and 2 n) are not the only clues to the differences in nucleotide composition between fish and mammals and that another component must be involved. This component might be the repeatome, i.e., all repeated DNA sequences including also the simple tandem repeats, etc. [45] or specifically the mobilome, i.e., the sum of all mobile elements *sensu* [46]. The large-scale view of associations between the chromosome size vs. GC is similar in both mammals and fishes (Figure 1).

4.2. The Role of Chromosome Size Combines with the Influence of Repeats

The situation in chondrichthyans, sturgeon, and gar demonstrated here in comparison with birds indeed indicates an important role of the chromosome size on GC% of the entire genome. This is impacted by the presence of microchromosomes in all these species. The larger genome size in chondrichthyans and sturgeon can explain, in line with the concept of gBGC, their overall lower GC% (Figure 3). Despite this similar karyotype organization (numerous tiny microchromosomes and clearly larger macrochromosomes), the AT/GC heterogeneity of macrochromosomes has evolved only in gars [33] and not in sturgeons [47] and paddlefishes [48]. The closest living relatives of the AT/GC heterogeneous gars, the bowfin, possess the typical teleost-like karyotype in terms of AT/GC homogeneity and the absence of the microchromosomes [49]. On the other hand, both these ancient fish and birds show the independence of GC% from the chromosome size in their macrochromosomes (Figure 3), although the GC% of the mammalian chromosomes larger than 50–100 Mb is still negatively associated with the respective chromosome size (not shown here). This finding in both basal fish lineages and birds is contradictory to the concept of GC-biased gene conversion (gBGC) [18].

Salmonids and amphibians may provide another clue to understanding the relationship between the genome size and GC%. As shown here at the chromosomal level (Figure 5 c), in another paper in this special issue [50] and earlier [30] at the genome level, there is no simple association between these two measures in salmonids. This can be at least partly explained by their extreme repeat content: 58%–60% in the Atlantic salmon [51], ~52% transposable elements in *Coregonus* sp. "Balchen" from the Lake Thun [52], 57.1% of the rainbow trout genome [53], RepeatMasker associated 53.07% of the genome with interspersed repeats, and masked 56.48% of the genome as repeat-derived in the Chinook salmon [54], etc. These high proportions of repeats belong to the highest ones among fishes and found in any vertebrate [51,55], similar to the only salamander species with a sequenced genome to data, the axolotl (*A. mexicanum*) with its 65.6% of repetitive sequences [56]. The impact of repeats on the GC% of the entire genome might be similar to that in zebrafish and some other cyprinids, where repeats occur in high proportions as

well, 52.2% in Zv9 [41]. However, repeats in most cyprinids are GC-poor [30], whereas those in salmonids might be GC-rich, a possibility that needs to be explored. Both Class I and II transposons are almost equally represented in the sequenced salmonids. Therefore, it is currently impossible to ascribe the potential GC-enrichment to one of them as the dominant one *sensu* [30]. In axolotl, however, distinct long terminal repeat (LTR) retroelement classes and endogenous retroviruses made up the largest portion of the repetitive sequences and included elements of more than 10 kb in length [56]. This shows the broad repertoire in repeats composition in lineages with giant genomes and the need to explore them. The entire genomics of polyploid animals is still in its infancy and hence, important new facts can be expected soon [57] since new genomes have been sequenced in the meanwhile and others are on the way.

4.3. Nucleotide Composition Investigations in Fish

The currently most plausible findings link the genome GC content with the chromosome size and with life history traits, where the effective population size (Ne) is the most important indicator of the strength of the gBGC [18,20]. For fish, there is no systematic analysis of GC% based on purely genomic data as are available for mammals [20–22], birds [23,24], and reptiles [25,26]. One study assessing GC% based on non-genomic determination showed a slightly higher GC% in marine and migrating marine fish species [58]. Hence, we can merely indirectly assess the above mentioned factors with correlatives with GC% known for fish. Genome size is one of them and it is negatively correlated with the genomic GC% in fish except for salmonids [30]. Genome size was shown to negatively correlate also with the Ne and this association was demonstrated to have resulted in a larger genome in freshwater ray-finned fish [59]. This is in line with the aforementioned results on the effect of environment and lifestyle in fish [58]. However, the reduced Ne presumably resulting in larger and more complex genomes in freshwater fish [59] are contradictory to the situation in chondrichthyans, which are marine [40] and possess large genomes [43], but were not included in the study [59]. Another recent study indicated that the fish species distributed at higher latitude might have a small long-term Ne [60]. A link between Ne and the transposon evolution/transposition intensity was also indicated [61,62]. Moreover, there is a highly complex and ambiguous relationship between the transposons insertion and accumulation and the recombination rate (summarized in [63]).

Crucial results come from the compositional biology of mammalian transposons. There are indications that the base composition is evolving under selection and may be reflective of the long-term co-evolution between non-LTR retrotransposons and their host [64]. These authors also hypothesise that the coexistence of elements with drastically different base compositions suggests that these elements may be using different strategies to persist and multiply in the genome of their host. Similarly, other authors proposed that in addition to gBGC, there may be additional, still uncharacterized molecular mechanisms that either preserve genomic regions with biased nucleotide compositions from mutational degradation or fail to degrade such inhomogeneities in specific chromosomal regions [65].

5. Conclusions

To sum up, fishes show an immense diversity in their genome organisation and in the mutual relationships between their major genomic traits (GC% and genome size). On the other hand, their diploid chromosome numbers are relatively conserved [8], which means that their chromosome sizes need to decrease with the genome compaction observed in modern fish lineages. This resulted in an increased recombination rate as evidenced in the literature [19] and in a GC-enrichment, however, without AT/GC heterogeneity. Hence, modern fish lineages reached the smallest genome sizes among vertebrates despite their teleost-specific whole genome duplication [44]. Birds increased their absolute GC% and acquired or retained their AT/GC heterogeneity and substantially decreased their genome size, at least partly thanks to their microchromosomes (they increase the

recombination rate that leads to genome compaction [66]). In contrast, basal fish lineages with microchromosomes do show a certain GC-enrichment, but they retained larger genome sizes despite the increased recombination rate mediated by microchromosomes that should have led to genome compaction similarly as in birds. Finally, mammals substantially expanded their genomes without any additional whole genome duplication and acquired or retained the AT/GC heterogeneity without any significant increase in their GC%. Therefore, mammals rather redistributed their genomes to GC-rich and GC-poor regions and accumulated them separately based on their GC%. The exact role and extent of GC-biased gene conversion in shaping the genome composition across vertebrates needs to be clarified and further investigated with the newly available resources. The quantitative approach to this issue, presented here, opens new possibilities to address related questions.

Supplementary Materials: The following are available online at www.mdpi.com/2073-4425/12/2/312/s1. Figure S1: Meta-analysis of GC% (A) and genome size (B) in fish (red) and mammals (green). Data from NCBI/genome and www.genomesize.com. Each point represents an entire genome; Figure S2: GC% of individual chromosomes in one lancelet (left, pink) and six ray-finned fishes (six blue violin plots) with diversified associations between the chromosome size and GC% compared with seven mammals (green violin plots). The overall comparison of GC% and genome size in fish and mammals can be found in Figure 1. The colour-code corresponds to Figure 1; Figure S3: GC% of individual chromosomes in lineages with microchromosomes—four ancient fish and in reptiles and birds. The left four violin plots are two chondrichthyans (thorny skate and bamboo shark in black), two basal ray-finned fishes (sterlet and spotted gar in blue), two violin plots in the middle are reptiles (rattlesnake and tortoise purple), and four right violin plots are birds (the Muscovy duck, chicken, the American flamingo, and zebra finch in yellow). The colour code corresponds to Figure 1; Figure S4: *Clupea harengus* GC% vs. repeats% in each chromosome. The chromosome size decreases from 33 to 12.4 Mb. Neither substantial changes in repeats% nor GC% of repeats occur with the decreasing chromosome size as predicted by the gBGC concept. A similar pattern was identified in all the three salmonids analysed here. The vertical line represents the median of rep%; Figure S5: *Danio rerio* GC% vs. repeats% in each chromosome. The chromosome size decreases from 78 to 37.5 Mb. Such a pattern has so far been identified only in this species. The vertical line represents the median of rep%; Figure S6: *Lepisosteus oculatus* GC% vs. repeats% in each chromosome. The chromosome size decreases from 73.2 to 5.4 Mb (we excluded the smallest chromosome with size 294 kb). Such a pattern occurs in numerous species; Figure S7: *Erpetoichthys calabaricus* GC% vs. repeats% in each chromosome. The chromosome size decreases from 350 to 88 Mb. The vertical line represents the median of rep%.

Author Contributions: Conceptualization, R.S., W.M.H., and V.B.; methodology, R.S., V.B., and D.M.; software, D.M.; validation, V.B.; resources, R.S.; data curation, D.M.; writing—original draft preparation, R.S. and W.M.H.; writing—review and editing, R.S., W.M.H., and V.B.; visualization, V.B. and R.S.; project administration, R.S.; funding acquisition, R.S. All authors have read and agreed to the published version of the manuscript.

Funding: This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 754462. The APC was funded by the Erasmus + programme of the European Union with contract no. 2019-1-CZ01-KA203-061433.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable

Data Availability Statement: https://github.com/bioinfohk/evangelist_plots

Acknowledgments: Computational resources were supplied by the project "e-Infrastruktura CZ" (e-INFRA LM2018140) provided within the program Projects of Large Research, Development, and Innovations Infrastructures.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Gregory, T.R. Animal Genome Size Database. Available online: http://www.genomesize.com (accessed on 31 January, 2021).
2. Levan, A.; Fredga, K.; Sandberg, A.A. Nomenclature for Centromeric Position on Chromosomes. *Hereditas* **1964**, *52*, 201–220, doi:10.1111/j.1601-5223.1964.tb01953.x.
3. Mark, H.F.; Mark, R.; Pan, T.; Mark, Y. Centromere Index Derivation by a Novel and Convenient Approach. *Ann. Clin. Lab. Sci.* **1993**, *23*, 267–274.
4. Comings, D.E. Mechanisms of Chromosome Banding and Implications for Chromosome Structure. *Annu. Rev. Genet.* **1978**, *12*, 25–46, doi:10.1146/annurev.ge.12.120178.000325.
5. Luo, C. Multiple Chromosomal Banding in Grass Carp, Ctenopharyngodon Idellus. *Heredity* **1998**, *81*, 481–485, doi:10.1046/j.1365-2540.1998.00323.x.
6. Medrano, L.; Bernardi, G.; Couturier, J.; Dutrillaux, B.; Bernardi, G. Chromosome Banding and Genome Compartmentalization in Fishes. *Chromosoma* **1988**, *96*, 178–183, doi:10.1007/BF00331050.
7. Mayr, B.; Kalat, M.; Ráb, P.; Lambrou, M. Band Karyotypes and Specific Types of Heterochromatins in Several Species of European Percid Fishes (Percidea, Pisces). *Genetica* **1987**, *75*, 199–205, doi:10.1007/BF00123574.
8. Mank, J.E.; Avise, J.C. Phylogenetic Conservation of Chromosome Numbers in Actinopterygiian Fishes. *Genetica* **2006**, *127*, 321–327, doi:10.1007/s10709-005-5248-0.
9. Gregory, T.R.; Witt, J.D.S. Population Size and Genome Size in Fishes: A Closer Look. *Genome* **2008**, *51*, 309–313, doi:10.1139/G08-003.
10. Hardie, D.C.; Hebert, P.D.N. The Nucleotypic Effects of Cellular DNA Content in Cartilaginous and Ray-Finned Fishes. *Genome* **2003**, *46*, 683–706, doi:10.1139/g03-040.
11. Hardie, D.C.; Hebert, P.D. Genome-Size Evolution in Fishes. *Can. J. Fish. Aquat. Sci.* **2004**, *61*, 1636–1646, doi:10.1139/f04-106.
12. Melodelima, C.; Gautier, C. The GC-Heterogeneity of Teleost Fishes. *BMC Genom.* **2008**, *9*, 632, doi:10.1186/1471-2164-9-632.
13. Costantini, M.; Auletta, F.; Bernardi, G. Isochore Patterns and Gene Distributions in Fish Genomes. *Genomics* **2007**, *90*, 364–371, doi:10.1016/j.ygeno.2007.05.006.
14. Costantini, M.; Cammarano, R.; Bernardi, G. The Evolution of Isochore Patterns in Vertebrate Genomes. *BMC Genom.* **2009**, *10*, 146, doi:10.1186/1471-2164-10-146.
15. Eyre-Walker, A. Recombination and Mammalian Genome Evolution. *Proc. R. Soc. Lond. B Biol. Sci.* **1993**, *252*, 237–243, doi:10.1098/rspb.1993.0071.
16. Fullerton, S.M.; Bernardo Carvalho, A.; Clark, A.G. Local Rates of Recombination Are Positively Correlated with GC Content in the Human Genome. *Mol. Biol. Evol.* **2001**, *18*, 1139–1142, doi:10.1093/oxfordjournals.molbev.a003886.
17. Montoya-Burgos, J.I.; Boursot, P.; Galtier, N. Recombination Explains Isochores in Mammalian Genomes. *Trends Genet.* **2003**, *19*, 128–130, doi:10.1016/S0168-9525(03)00021-0.
18. Mugal, C.F.; Weber, C.C.; Ellegren, H. GC-Biased Gene Conversion Links the Recombination Landscape and Demography to Genomic Base Composition: GC-Biased Gene Conversion Drives Genomic Base Composition across a Wide Range of Species. *BioEssays* **2015**, *37*, 1317–1326, doi:10.1002/bies.201500058.
19. Stapley, J.; Feulner, P.G.D.; Johnston, S.E.; Santure, A.W.; Smadja, C.M. Variation in Recombination Frequency and Distribution across Eukaryotes: Patterns and Processes. *Philos. Trans. R. Soc. B Biol. Sci.* **2017**, *372*, 20160455, doi:10.1098/rstb.2016.0455.
20. Romiguier, J.; Ranwez, V.; Douzery, E.J.P.; Galtier, N. Contrasting GC-Content Dynamics across 33 Mammalian Genomes: Relationship with Life-History Traits and Chromosome Sizes. *Genome Res.* **2010**, *20*, 1001–1009, doi:10.1101/gr.104372.109.
21. Huttener, R.; Thorrez, L.; In't Veld, T.; Granvik, M.; Snoeck, L.; Van Lommel, L.; Schuit, F. GC Content of Vertebrate Exome Landscapes Reveal Areas of Accelerated Protein Evolution. *Bmc Evol. Biol.* **2019**, *19*, doi:10.1186/s12862-019-1469-1.
22. Shen, W.; Wang, D.; Ye, B.; Shi, M.; Ma, L.; Zhang, Y.; Zhao, Z. GC3-Biased Gene Domains in Mammalian Genomes. *Bioinformatics* **2015**, *31*, 3081–3084, doi:10.1093/bioinformatics/btv329.
23. Weber, C.C.; Boussau, B.; Romiguier, J.; Jarvis, E.D.; Ellegren, H. Evidence for GC-Biased Gene Conversion as a Driver of between-Lineage Differences in Avian Base Composition. *Genome Biol.* **2014**, *15*, doi:10.1186/s13059-014-0549-1.
24. Bolívar, P.; Mugal, C.F.; Nater, A.; Ellegren, H. Recombination Rate Variation Modulates Gene Sequence Evolution Mainly via GC-Biased Gene Conversion, Not Hill–Robertson Interference, in an Avian System. *Mol. Biol. Evol.* **2016**, *33*, 216–227, doi:10.1093/molbev/msv214.
25. Matsubara, K.; Kuraku, S.; Tarui, H.; Nishimura, O.; Nishida, C.; Agata, K.; Kumazawa, Y.; Matsuda, Y. Intra-Genomic GC Heterogeneity in Sauropsids: Evolutionary Insights from CDNA Mapping and GC3 Profiling in Snake. *BMC Genom.* **2012**, *13*, 604, doi:10.1186/1471-2164-13-604.
26. Figuet, E.; Ballenghien, M.; Romiguier, J.; Galtier, N. Biased Gene Conversion and GC-Content Evolution in the Coding Sequences of Reptiles and Vertebrates. *Genome Biol. Evol.* **2015**, *7*, 240–250, doi:10.1093/gbe/evu277.
27. Wang, D. GCevobase: An Evolution-Based Database for GC Content in Eukaryotic Genomes. *Bioinformatics* **2018**, *34*, 2129–2131, doi:10.1093/bioinformatics/bty068.
28. Frenkel, S.; Kirzhner, V.; Korol, A. Organizational Heterogeneity of Vertebrate Genomes. *PLoS ONE* **2012**, *7*, e32076, doi:10.1371/journal.pone.0032076.
29. Li, X.-Q.; Du, D. Variation, Evolution, and Correlation Analysis of C+G Content and Genome or Chromosome Size in Different Kingdoms and Phyla. *PLoS ONE* **2014**, *9*, e88339, doi:10.1371/journal.pone.0088339.

30. Symonová, R.; Suh, A. Nucleotide Composition of Transposable Elements Likely Contributes to AT/GC Compositional Homogeneity of Teleost Fish Genomes. *Mob. DNA* **2019**, *10*, doi:10.1186/s13100-019-0195-y.
31. Carducci, F.; Barucca, M.; Canapa, A.; Carotti, E.; Biscotti, M.A. Mobile Elements in Ray-Finned Fish Genomes. *Life* **2020**, *10*, 221, doi:10.3390/life10100221.
32. Bernardi, G. The Vertebrate Genome: Isochores and Evolution. *Mol. Biol. Evol.* **1993**, doi:10.1093/oxfordjournals.molbev.a039994.
33. Symonová, R.; Majtánová, Z.; Arias-Rodriguez, L.; Mořkovský, L.; Kořínková, T.; Cavin, L.; Pokorná, M.J.; Doležálková, M.; Flajšhans, M.; Normandeau, E.; et al. Genome Compositional Organization in Gars Shows More Similarities to Mammals than to Other Ray-Finned Fish: Cytogenomics of Gars. *J. Exp. Zool. Part B Mol. Dev. Evol.* **2017**, *328*, 607–619, doi:10.1002/jez.b.22719.
34. Matoulek, D.; Borůvková, V.; Ocalewicz, K.; Symonová, R. GC and Repeats Profiling along Chromosomes—The Future of Fish Compositional Cytogenomics. *Genes* **2021**, *12*, 50, doi:10.3390/genes12010050.
35. Bernardi, G. *Structural and Evolutionary Genomics Natural Selection in Genome Evolution*; Elsevier: Amsterdam, The Netherlands, 2005.
36. Peona, V.; Weissensteiner, M.H.; Suh, A. How Complete Are "Complete" Genome Assemblies?—An Avian Perspective. *Mol. Ecol. Resour.* **2018**, *18*, 1188–1195, doi:10.1111/1755-0998.12933.
37. Jebb, D.; Huang, Z.; Pippel, M.; Hughes, G.M.; Lavrichenko, K.; Devanna, P.; Winkler, S.; Jermiin, L.S.; Skirmuntt, E.C.; Katzourakis, A.; et al. Six Reference-Quality Genomes Reveal Evolution of Bat Adaptations. *Nature* **2020**, *583*, 578–584, doi:10.1038/s41586-020-2486-3.
38. R Core Team. *R: A Language and Environment for Statistical Computing*, Version 2.6.2; R Core Team: Vienna, Austria, 2013.
39. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis; Use R!*, 2nd ed.; Springer International Publishing: Cham, Switzerland, 2016; ISBN 978-3-319-24277-4.
40. Nelson, J.S.; Grande, T.; Wilson, M.V.H. *Fishes of the World*, 5th ed.; John Wiley & Sons: Hoboken, NJ, USA, 2016; ISBN 978-1-118-34233-6.
41. Howe, K.; Clark, M.D.; Torroja, C.F.; Torrance, J.; Berthelot, C.; Muffato, M.; Collins, J.E.; Humphray, S.; McLaren, K.; Matthews, L.; et al. The Zebrafish Reference Genome Sequence and Its Relationship to the Human Genome. *Nature* **2013**, *496*, 498–503, doi:10.1038/nature12111.
42. Hidalgo, O.; Pellicer, J.; Christenhusz, M.; Schneider, H.; Leitch, A.R.; Leitch, I.J. Is There an Upper Limit to Genome Size? *Trends Plant Sci.* **2017**, *22*, 567–573, doi:10.1016/j.tplants.2017.04.005.
43. NCBI Genome Browser. Available online: https://www.ncbi.nlm.nih.gov/genome/browse (accessed on 31 January, 2021).
44. Meyer, A.; Schartl, M. Gene and Genome Duplications in Vertebrates: The One-to-Four (-to-Eight in Fish) Rule and the Evolution of Novel Gene Functions. *Curr. Opin. Cell Biol.* **1999**, *11*, 699–704.
45. Hannan, A.J. Tandem Repeats and Repeatomes: Delving Deeper into the 'Dark Matter' of Genomes. *EBioMedicine* **2018**, *31*, 3–4, doi:10.1016/j.ebiom.2018.04.004.
46. Arkhipova, I.R.; Yushenova, I.A. Giant Transposons in Eukaryotes: Is Bigger Better? *Genome Biol. Evol.* **2019**, *11*, 906–918, doi:10.1093/gbe/evz041.
47. Fontana, F.; Bruch, R.M.; Binkowski, F.P.; Lanfredi, M.; Chicca, M.; Beltrami, N.; Congiu, L. Karyotype Characterization of the Lake Sturgeon, *Acipenser fulvescens* (Rafinesque 1817) by Chromosome Banding and Fluorescent in Situ Hybridization. *Genome* **2004**, *47*, 742–746, doi:10.1139/g04-028.
48. Symonová, R.; Havelka, M.; Amemiya, C.T.; Howell, W.M.; Kořínková, T.; Flajšhans, M.; Gela, D.; Ráb, P. Molecular Cytogenetic Differentiation of Paralogs of Hox Paralogs in Duplicated and Re-Diploidized Genome of the North American Paddlefish (*Polyodon spathula*). *BMC Genet.* **2017**, *18*, doi:10.1186/s12863-017-0484-8.
49. Majtánová, Z.; Symonová, R.; Arias-Rodriguez, L.; Sallan, L.; Ráb, P. "Holostei versus Halecostomi" Problem: Insight from Cytogenetics of Ancient Nonteleost Actinopterygian Fish, Bowfin *Amia calva*: Molecular Cytogenetics of *Amia calva*. *J. Exp. Zool. Part B Mol. Dev. Evol.* **2017**, *328*, 620–628, doi:10.1002/jez.b.22720.
50. Gaffaroglu, M.; Majtánová, Z.; Symonová, R.; Pelikánová, Š.; Unal, S.; Lajbner, Z.; Ráb, P. Present and Future Salmonid Cytogenetics. *Genes* **2020**, *11*, 1462.
51. Lien, S.; Koop, B.F.; Sandve, S.R.; Miller, J.R.; Kent, M.P.; Nome, T.; Hvidsten, T.R.; Leong, J.S.; Minkley, D.R.; Zimin, A.; et al. The Atlantic Salmon Genome Provides Insights into Rediploidization. *Nature* **2016**, *533*, 200–205, doi:10.1038/nature17164.
52. De-Kayne, R.; Zoller, S.; Feulner, P.G.D. A de Novo Chromosome-level Genome Assembly of *Coregonus* sp. "*Balchen*": One Representative of the Swiss Alpine Whitefish Radiation. *Mol. Ecol. Resour.* **2020**, *20*, 1093–1109, doi:10.1111/1755-0998.13187.
53. Pearse, D.E.; Barson, N.J.; Nome, T.; Gao, G.; Campbell, M.A.; Abadía-Cardoso, A.; Anderson, E.C.; Rundio, D.E.; Williams, T.H.; Naish, K.A.; et al. Sex-Dependent Dominance Maintains Migration Supergene in Rainbow Trout. *Nat. Ecol. Evol.* **2019**, *3*, 1731–1742, doi:10.1038/s41559-019-1044-6.
54. Christensen, K.A.; Leong, J.S.; Sakhrani, D.; Biagi, C.A.; Minkley, D.R.; Withler, R.E.; Rondeau, E.B.; Koop, B.F.; Devlin, R.H. Chinook Salmon (*Oncorhynchus tshawytscha*) Genome and Transcriptome. *PLoS ONE* **2018**, *13*, e0195461, doi:10.1371/journal.pone.0195461.
55. Canapa, A.; Barucca, M.; Biscotti, M.A.; Forconi, M.; Olmo, E. Transposons, Genome Size, and Evolutionary Insights in Animals. *Cytogenet. Genome Res.* **2015**, *147*, 217–239, doi:10.1159/000444429.
56. Nowoshilow, S.; Schloissnig, S.; Fei, J.-F.; Dahl, A.; Pang, A.W.C.; Pippel, M.; Winkler, S.; Hastie, A.R.; Young, G.; Roscito, J.G.; et al. The Axolotl Genome and the Evolution of Key Tissue Formation Regulators. *Nature* **2018**, *554*, 50–55, doi:10.1038/nature25458.

57. Rodriguez, F.; Arkhipova, I.R. Transposable Elements and Polyploid Evolution in Animals. *Curr. Opin. Genet. Dev.* **2018**, *49*, 115–123, doi:10.1016/j.gde.2018.04.003.
58. Tarallo, A.; Angelini, C.; Sanges, R.; Yagi, M.; Agnisola, C.; D'Onofrio, G. On the Genome Base Composition of Teleosts: The Effect of Environment and Lifestyle. *Bmc Genom.* **2016**, *17*, doi:10.1186/s12864-016-2537-1.
59. Yi, S.; Streelman, J.T. Genome Size Is Negatively Correlated with Effective Population Size in Ray-Finned Fish. *Trends Genet.* **2005**, *21*, 643–646, doi:10.1016/j.tig.2005.09.003.
60. Rolland, J.; Schluter, D.; Romiguier, J. Vulnerability to Fishing and Life History Traits Correlate with the Load of Deleterious Mutations in Teleosts. *Mol. Biol. Evol.* **2020**, *37*, 2192–2196, doi:10.1093/molbev/msaa067.
61. Tollis, M.; Boissinot, S. The Evolutionary Dynamics of Transposable Elements in Eukaryote Genomes. In *Genome Dynamics*; Garrido-Ramos, M.A., Ed.; S. KARGER AG: Basel, Switzerland, 2012; Volume 7, pp. 68–91, ISBN 978-3-318-02149-3.
62. Bourgeois, Y.; Boissinot, S. On the Population Dynamics of Junk: A Review on the Population Genomics of Transposable Elements. *Genes* **2019**, *10*, 419, doi:10.3390/genes10060419.
63. Kent, T.V.; Uzunović, J.; Wright, S.I. Coevolution between Transposable Elements and Recombination. *Philos. Trans. R. Soc. B Biol. Sci.* **2017**, *372*, 20160458, doi:10.1098/rstb.2016.0458.
64. Ruggiero, R.P.; Boissinot, S. Variation in Base Composition Underlies Functional and Evolutionary Divergence in Non-LTR Retrotransposons. *Mob. DNA* **2020**, *11*, doi:10.1186/s13100-020-00209-9.
65. Paudel, R.; Fedorova, L.; Fedorov, A. Adapting Biased Gene Conversion Theory to Account for Intensive GC-Content Deterioration in the Human Genome by Novel Mutations. *PLoS ONE* **2020**, *15*, e0232167, doi:10.1371/journal.pone.0232167.
66. Nam, K.; Ellegren, H. Recombination Drives Vertebrate Genome Contraction. *PLoS Genet.* **2012**, *8*, e1002680, doi:10.1371/journal.pgen.1002680.

genes

MDPI

Article

GC and Repeats Profiling along Chromosomes—The Future of Fish Compositional Cytogenomics

Dominik Matoulek [1], Veronika Borůvková [1], Konrad Ocalewicz [2] and Radka Symonová [3,*]

[1] Faculty of Science, University of Hradec Kralove, 500 03 Hradec Králové, Czech Republic; dominik.matoulek@uhk.cz (D.M.); veronika.boruvkova@uhk.cz (V.B.)
[2] Department of Marine Biology and Ecology, Institute of Oceanography, Faculty of Oceanography and Geography, University of Gdansk, 80-309 Gdansk, Poland; konrad.ocalewicz@ug.edu.pl
[3] Department of Bioinformatics, Wissenschaftszentrum Weihenstephan, Technische Universität München, 80333 Freising, Germany
* Correspondence: radka.symonova@gmail.com

Abstract: The study of fish cytogenetics has been impeded by the inability to produce G-bands that could assign chromosomes to their homologous pairs. Thus, the majority of karyotypes published have been estimated based on morphological similarities of chromosomes. The reason why chromosome G-banding does not work in fish remains elusive. However, the recent increase in the number of fish genomes assembled to the chromosome level provides a way to analyse this issue. We have developed a Python tool to visualize and quantify GC percentage (GC%) of both repeats and unique DNA along chromosomes using a non-overlapping sliding window approach. Our tool profiles GC% and simultaneously plots the proportion of repeats (rep%) in a color scale (or vice versa). Hence, it is possible to assess the contribution of repeats to the total GC%. The main differences are the GC% of repeats homogenizing the overall GC% along fish chromosomes and a greater range of GC% scattered along fish chromosomes. This may explain the inability to produce G-banding in fish. We also show an occasional banding pattern along the chromosomes in some fish that probably cannot be detected with traditional qualitative cytogenetic methods.

Keywords: AT/GC heterogeneity; chromosome banding; fish cytogenetics; GC-profile; repeats organization

Citation: Matoulek, D.; Borůvková, V.; Ocalewicz, K.; Symonová, R. GC and Repeats Profiling along Chromosomes—The Future of Fish Compositional Cytogenomics. *Genes* **2021**, *12*, 50. https://doi.org/10.3390/genes12010050

Received: 2 December 2020
Accepted: 29 December 2020
Published: 31 December 2020

1. Introduction

Classical chromosome banding methods such as G- (Giemsa), R- (reverse) and Q- (quinacrine) banding allow for routine chromosome analysis in higher vertebrates, including human clinical cytogenetics [1–3], and many more. A recent review of these heterogeneous chromosomal bands and sequence features is available [4]. A fully different and incomparable situation exists in lower vertebrates, particularly in fishes. Compared with other vertebrates, fish have smaller chromosomes and a narrower range of GC% values in entire genomes [5,6]. Despite numerous attempts, e.g., [7–9], the chromosome banding methods mentioned above do not yield usable patterns in fish. The research performed up to now was summarized concluding that C-banding [10] and silver-staining [11] in fishes provide reasonably good results, whereas very little success has been achieved using G-bands [12]. The only way to produce a reliable pattern on fish chromosomes was the application of replication labelling, as different regions of genome replicate at different moments during S phase of the cell cycle [13]. The replication banding utilizes the incorporation of a thymidine analogue, 5-bromo-2′-deoxiuridine (BrdU), into nuclear DNA during the S-phase of DNA replication. Then regions with BrdU are visualized by detection on metaphase chromosomes. Bands with incorporated BrdU may be revealed, for example, by Hoechst 33,258 fluorescence, acridine orange fluorescence, or fluorochrome-photolysis-Giemsa staining (FPG), among others [14]. It has been shown

that heterochromatic, AT-rich G-bands and C-bands are late replicating, while euchromatic, GC-rich R-bands replicate early during the S-phase [2]. Despite its high resolution in mammals, replication banding patterns have been produced in a limited number of fish species so far. Application of FPG enabled the identification of early and late replicating chromosomal regions with high resolution banding patterns in salmonids [15,16], white sturgeon [16], and eels [17,18]. Less clear patterns after BrdU incorporation were observed on chromosomes of cyprinids [19,20], anastomids [21], ictalurids [22], flatfish [23], pufferfish [24] and characids [25]. The interspecies differences in the resolution of the replication banding may result from different genome composition and the size of chromosomes. Salmonid genomes are of polyploid origin and have relatively large chromosomes that are favourable for the distinct and clear replication banding pattern [16]. However, even this laborious procedure applied in fish did not always produce results comparable with those in mammalian and avian cytogenetics [12]. The presence of very small microchromosomes along with larger macrochromosomes in some basal fish lineages (chondrichthyans, sturgeons, gars) resembling those in birds and some reptiles complicate fish cytogenetics even more because of their indistinguishable chromosome morphology.

Values of GC% are associated with numerous traits including gene density, chromatin structure, the proportion and types of transposable elements, DNA replication timing, nucleosome formation potential etc. [26]. To test whether GC content differences might explain the lack of G-bands in fish, we investigated the fine-scale AT/GC organization in fish. Thanks to the increasing availability of fish genomes assembled to the chromosome level and at the same time of their soft-masking, i.e., labelling repetitive elements as the lower case in the otherwise upper case represented DNA sequence, it is possible to produce a virtual banding pattern of GC% and repeats percentage (rep%) along chromosomes. The recently published genomes of sterlet sturgeon [27] and reedfish [28] were important milestones for fish compositional cytogenomics sensu [29] together with the immense body of evidence accumulated by traditional cytogenetics. In the traditional (i.e., qualitative) fish cytogenetics, there are two mutually non-exclusive ways to visualize GC% and rep% even on the same metaphase. These are the CDD-staining combining AT- and GC-specific fluorochromes to the same metaphase [30,31] for GC% and fluorescence in situ hybridization (FISH) with a repetitive DNA fraction, e.g., cot-1, as a probe [32] or a more destructive visualization of constitutive heterochromatin using C-banding [10,33] for rep%. However, application of these methods is limited in fish due to the small size of their chromosomes. Moreover, it requires time-consuming laboratory processing including chromosome preparation from living fish. On the other hand, cytogenetic methods including C-banding and DAPI-staining usually enable identification of the centromeres, which is not yet possible in most of fish genomes.

To the best of our knowledge, there is no such specialized bioinformatics tool available to integrate and plot both GC% and rep% into a single image. There are some tools producing GC-profiles along chromosomes, e.g., [34,35], or tools integrated, e.g., in Bioconductor plotting diversified features along chromosomes [36] but never plotting simultaneously the proportions of repetitive DNA together with the GC% of non-soft-masked (non-repetitive) and soft-masked (repetitive) DNA.

Our aims were: (1) to assess differences in compositional organization (GC and repeats proportions) of chromosomes at multiple levels of resolution (i.e., with different sliding window sizes) among vertebrates with a focus on fishes; (2) to utilize the increasingly available genomic data on the chromosome level and their constantly increasing quality; (3) to virtualize the traditional qualitative molecular cytogenetic methods in silico; (4) to assess the role of transposons and other repetitive elements on the entire AT/GC composition along chromosomes; and (5) to produce a publicly available tool visualizing and quantifying these two major features (GC and repeats proportions) along chromosomes assembled to the chromosome level.

Producing two types of plots, combining a color scale with percentage values along chromosomes with a customized non-overlapping sliding window size helped to resolve

the conundrum of unavailability of banding patterns in fish cytogenetics. Namely, the fine-scale organization of repeats and their own GC content homogenize the overall GC% along fish chromosomes, preventing the formation of larger regions with an elevated GC% separated by sharp borders.

2. Materials and Methods

2.1. Data Acquisition and Processing

Altogether, we utilized genome assemblies of 41 fish and one tunicate species (Table A1) assembled to the chromosome level available in the database Ensembl (37 species with already available soft-masking; Release 100; [37]) and in NCBI six species which genomes had to be processed with soft-masking software [28], e.g., using the online tool RepeatMasker version 4 [38]). These species include one tunicate (*Ciona intestinalis*), three chondrichthyan species, three non-teleost ray-finned fish, i.e., one reedfish (*Erpetoichthys calabaricus*), one sturgeon (*Acipenser ruthenus*) and one gar (*Lepisosteus oculatus*), and 35 teleosts. To compare fish GC% and repeats organization along chromosomes with mammals, we further utilized genome assemblies of gorilla, cat, little brown bat, and greater horseshoe bat, also available already soft-masked in Ensembl. We compared three different non-overlapping sliding window sizes with 1 kbp as default. Furthermore, we tested non-overlapping sliding window sizes 3 kbp and 10 kbp in selected species. This is highly relevant for polyploid (e.g., salmonids) or (extremely) large (reedfish, zebrafish) fish genomes. The sliding window size 3 kbp reflects the fact that mammalian genomes are about three times larger than fish genomes, while both converge on approximately 2n = 46–50. This enabled us to compare fish and mammalian chromosomes at a corresponding scale.

2.2. DNA Profiling Tool

The tool called EVANGELIST (=EVAluatioN on GEnome LIST) utilizes the non-overlapping sliding window (referred to as sliding window below) approach to quantify and visualize the percentage of repeats and GC percentage (GC%) in both repeats and non-repetitive DNA simultaneously. It includes the following Python components: DNA_puller, gnuplot_generator and a set of Jupyter notebooks. To run this tool, it is necessary to have the BioPython [39] library installed. The tool performs four basic steps to produce the presented results:

1. Data download from a database such as Ensembl or NCBI, where they are accessible by the FTP. The tool saves data for every requested species into its own folder and unzips them.
2. Data analysis by the sliding window approach is performed for each FASTA file separately with "DNA_puller", a component provided on GitHub. Each window position yields the number of occurrences of each letter (i.e., ATGC), discerning the upper and the lowercase ones.
3. The raw data are processed as a preparation for charts, giving GC% and the ratio between soft-masked (identified repeats) and non-soft-masked (non-repetitive DNA or not identified repeats) DNA in a CSV file for each chromosome. Such a file has three columns (index, i.e., position in DNA, GC%, and ratio) and a generally high number of rows, each of which will present a point in the chart. For instance, for a chromosome with 10 Mbp and a sliding window of size 1kbp, the result file has 10 Mbp/1 kbp = 1000 rows hence 1000 points in the chart.
4. Generation of the definition files and rendering charts is a two-step process performed with the tool GNUplot, version 5.2. The former is executed with our component "gnuplot_generator". During this step, the CSV files are sorted by the number of lines counted by the wc ('word count') program in Linux. Finally, the charts are rendered.

2.3. Plotting Large-Scale Profiles and Statistical Analyses

Plotting extremely large chromosomes presented a crucial issue. The size of "normal" (macro)chromosomes ranges from 15 to 150 Mb. To prevent information loss, our tool

produces plots with a tailored size according to chromosome sizes in each species separately. This ensures that each set of chromosomes is plotted as large as possible, which is crucial because of the requirements to visualize an extreme number of points: e.g., the largest chromosome in Northern pike (average C-value 1.1 pg [40], and average assembly size 921 Mbp; GenBank) is the linkage group (LG) 11 with size 55.41 Mbp, meaning that 55,410 points have to be visualized for this single chromosome (each point represents 1000 bp or 1 kbp). The complete set of chromosomes in this species is 10,000 × 25,000 pixels large and the file size is about 10 MB. On the other hand, the scale differs in each species.

We have tested the obtained results for GC% and repeats% for the linear relationship and correlation between these two measures in all species under study using BioPython [39].

Icons made by https://www.flaticon.com/authors/freepik.

The tool is available on GitHub https://github.com/bioinfohk/evangelist and the complete collection of all profiles produced in the framework of this study and in full resolution is available on the link https://github.com/bioinfohk/evangelist_plots.

3. Results

In the default setting, our tool plots GC% along chromosomes as points representing each consecutive 1000 bp (1 kbp) with 0–100% of GC on the y axis (Figure 1). The percentage of repeats (rep%) is plotted as a color gradient of these points, where green represents 1 kbp of soft-masked DNA, i.e., 100% of repeats, and red represents 1 kbp of non-soft-masked DNA, i.e., no repeats detected within the range of these 1 kbp (Figure 2). Our efforts to produce graphs as informative as possible resulted in very large plots. We have chosen this setting as the primary one because of its higher information value. This pattern of GC% values and colors can be easily swapped so that the scale of GC% can actually mimic the CDD-staining on chromosomes, where GC-rich regions are red and AT-rich regions are green and the rep% is on the y axis (Figure 3).

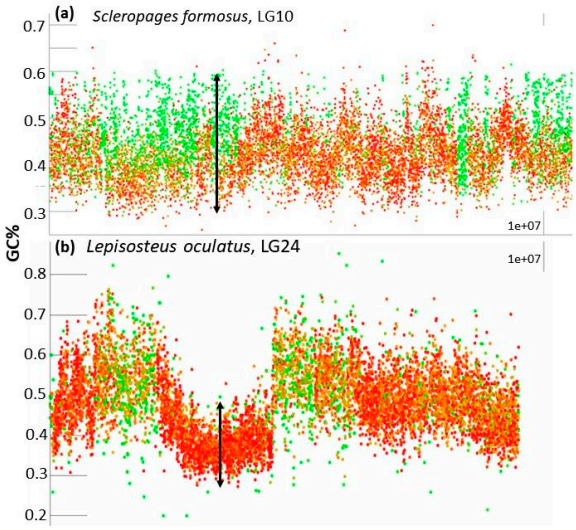


Figure 1. Details of parts of chromosomes of two representative fish species produced with the default setting of the non-overlapping sliding window size 1 kbp. (**a**) Asian arowana (*Scleropages formosus*), where the soft-masked DNA, i.e., repeats (green) attain high GC% whereby they homogenize the overall GC content to form a flattened upper bound of GC%. Here, the wide range of GC% values in repeats is also apparent and shows the importance of the small window size used here as default; (**b**) a different situation exists in the spotted gar (*Lepisosteus oculatus*) with one GC-poorer non-soft-masked (i.e., non-repetitive DNA, red) region surrounded by regions with a sharply elevated GC%. Each dot represents a single sliding window value of GC% (y axis) and soft-masked (repetitive) DNA percentage (red no repeats, green 100% repetitive, orange approx. 50% of repetitive DNA). The arrows in both images indicate a greater range of values of GC% in Asian arowana (**a**) than in the selected region of the spotted gar (**b**).

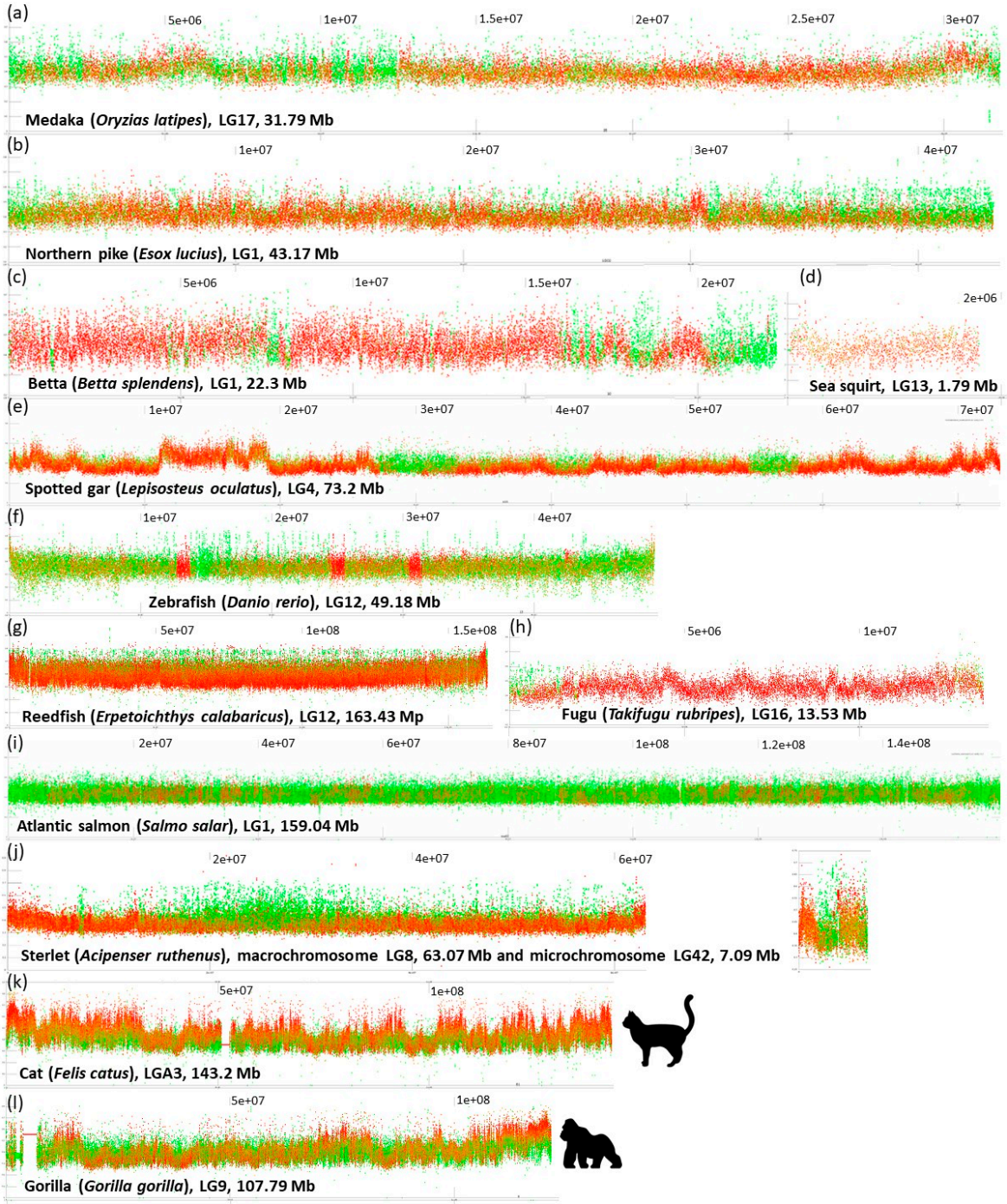


Figure 2. Graphs of mostly middle-sized chromosomes (unless otherwise indicated) with the default setting of the sliding window size 1 kbp. (**a**) Medaka shows repeats intermingled with unique sequences resulting in an overall orange coloration alternating with prevailing repeats (green) and unique (red) regions; (**b**) northern pike, with all acrocentric chromosomes, the largest chromosome shown; (**c**) betta with repeats localized in interstitial blocks and at a single end of the chromosome resulting in an overall red coloration; (**d**) sea squirt with homogeneous GC-poor DNA, the smallest chromosome shown; (**e**) spotted gar, the only fish so far known with the AT/GC heterogeneity, the largest chromosomes shown; (**f**) zebrafish, an example of an extremely GC-depleted fish genome with almost no fluctuations; (**g**) Reedfish with extremely large chromosomes without any prominent fluctuations in GC%; (**h**) fugu, a short linkage group (LG) with an extremely reduced amount of repeats; (**i**) Salmon, a polyploid AT-rich genome, the largest chromosome shown; (**j**) Sterlet, another polyploid fish with AT-rich(er) macro- and GC-rich(er) microchromosomes; (**k**) cat and gorilla (**l**) are mammalian outgroups with GC- and gene-rich peaks and rather AT-rich repeats. Complete plots of all analysed species are available at our online repository https://github.com/bioinfohk/evangelist.

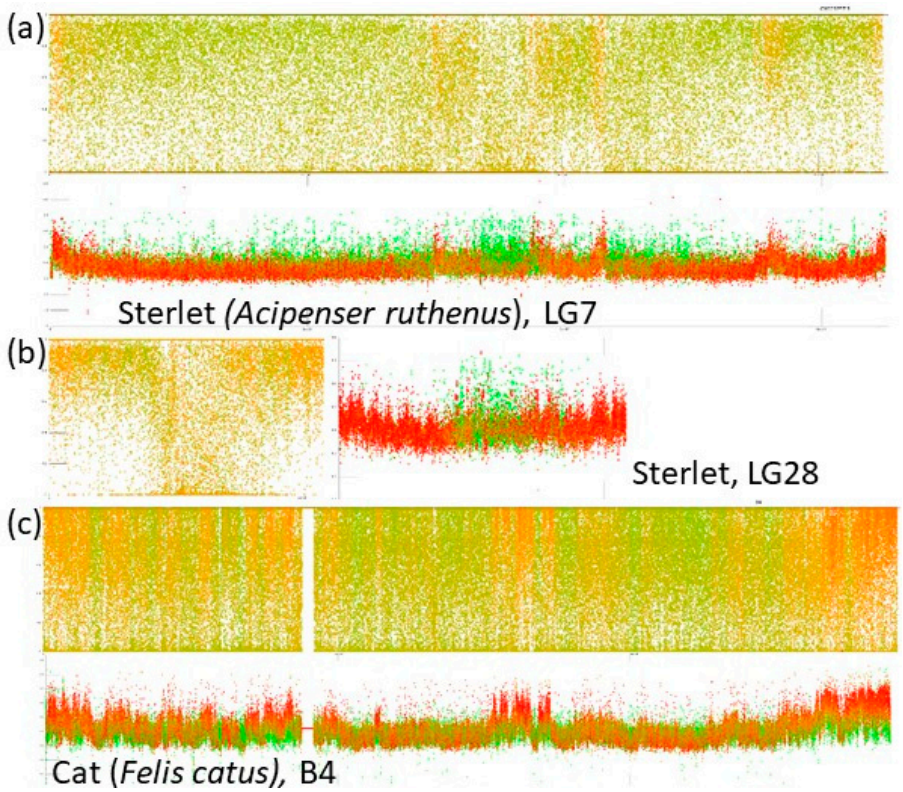


Figure 3. Comparison of the two major options of setting of GC% and rep% visualization with our tool with the default sliding window size 1 kbp. (**a**) One macrochromosome of sterlet, where GC% is represented by the color-scale mimicking the CMA_3-fluorescence staining in the upper panel (GC-rich in red, AT-rich in green) compared with the swapped setting, where GC% is plotted as a profile and rep% as the color-scale in the lower panel; (**b**) the same for one microchromosome of sterlet; (**c**) B4 of cat as an example of a mammalian LG.

3.1. GC-Profiles in Fish

Regarding the GC% values, the sliding window size 1 kbp proved to yield the best resolution and the fish species analysed so far produced the following patterns:

1. The entire chromosome is formed by a generally flattened range of points with GC% between the minimal values around 35% and the maximal values around 55% (*Oryzias latipes*, Figure 2) or sometimes 30–60% (*Betta splendens*, Figure 2) with only rare or occasional slight departures from this pattern. Whereas some species show a narrower GC% range with almost no fluctuations/departures, e.g., in the Blunt-snouted Clingfish (*Gouania willdenowi*), some other species show an even broader range of GC% 30–65% with some more prominent local elevations or depletions of GC% (*Scleropages formosus*). Occasional slight elevations in GC% occur at the ends of chromosomes.
2. No prominent pattern occurs in the basal chordate (tunicate) sea squirt (*Ciona intestinalis*). This pattern can be ascribed to an extremely low amount of DNA in the chromosomes (4.5–10 Mb). The majority of points occur in the range 30–40% of GC with only very rare and narrow peaks or isolated points reaching 50% of GC.
3. So far, the only known fish species with heterogeneous AT/GC organization along LGs is the spotted gar (*Lepisosteus oculatus*, Figure 2). Here, a rather narrow "baseline" of densely organized points of GC% between 30–50% alters with sharp and compact peaks reaching over 60% of GC%.
4. Another extreme situation exists in the reedfish (*Erpetoichthys calabaricus*, Figure 2) with a dense organization, however, resulting in a flat range of values between 30–55% GC. This flattened appearance can be ascribed to the exceptionally large size of

chromosomes (88.37–350.1 Mb) that are even larger than mammalian chromosomes (gorilla 32.72–219.76 Mb).

5. More fluctuating GC% values exist in tetraodontid fish with reduced genome size (*Tetraodon nigroviridis*, *Takifugu rubripes*, Figure 2; [41–43]) and to some extent in other species with reduced genomes e.g., the three-spined stickleback (*Gasterosteus aculeatus*).
6. A combination of a flattened range of GC% values in large(r) chromosomes (i.e., macrochromosomes) and more or less clear GC% elevations in smaller chromosomes (i.e., microchromosomes) exists in the sterlet (*Acipenser ruthenus*, Figure 2j) and all three chondrichthyan species analysed (*Amblyraja radiata*, *Chiloscyllium plagiosum*, and *Pristis pectinata*). Here, with the decreasing chromosome size, elevations in GC% firstly appear at the ends of chromosomes. In smaller chromosomes, internal GC% fluctuations occur.

3.2. Repeats Content and Organization in Fish

The default sliding window size of 1 kbp proved to yield the best resolution relative to repeat distribution along chromosomes. The following patterns and their mutual combinations have so far been observed:

1. Blocks of repeats prevailing over the non-repetitive DNA at both ends of chromosomes. This pattern is particularly prominent in species with all acrocentric chromosomes (e.g., *Esox lucius* (Figure 2; [44]), *Oreochromis niloticus* [45], *Sparus aurata*, etc.). The size of these blocks of repeats varies within and among species.
2. Interstitial, clearly delineated small blocks of almost exclusively repetitive DNA. (e.g., *Betta splendens*, Figure 2, *Ictalurus punctatus*, *Scleropages formosus*, *Oryzias latipes*).
3. Dispersed and intermingled repeats occurring mostly in fish species with larg(er) genomes (e.g., *Danio rerio*, *Astyanax mexicanus*, and pseudotetraploid salmonids *Oncorhynchus mykiss* and *Salmo salar*). Here, either completely green or orange regions of varying size are interrupted with small blocks of non-repetitive DNA.
4. Limited extent of repeats proportion caused by reduced genome size through repeats elimination (*Tetraodon nigroviridis*, *Takifugu rubripes*, Figure 2, *Gasterosteus aculeatus*) or through insufficient repeat-masking (*Oryzias javanicus*, *Scophthalmus maximus*, etc.).

These patterns of repeats distribution can combine and co-occur in a single fish species. However, it is necessary to stress that these patterns depend on the quality of soft-masking that is linked to the genome assembly quality. Hence, the obtained patterns cannot be considered ultimate in genomes, where soft-masking revealed only a smaller fraction of repeats.

Interestingly, in regions, where GC% decreases in the non-repetitive fractions, the GC% of repeats increases and thus compensates for this decrease, keeping the overall GC% values with a flattened upper bound, e.g., in Figure 1b in Asian arowana, Figure 2a medaka, Figure 2b the Northern pike or Figure 2c betta. More fish species showing this phenomenon can be seen on our GitHub repository. In regions, where non-repetitive DNA becomes fully absent, the repetitive DNA follows the GC% of the non-repetitive fraction from the surrounding regions. This prevents the formation of peaks with a higher GC% and of sharper borders in GC%.

The inverted representation of GC% and rep% shown in Figure 3 was produced to enable a direct comparison with cytogenetic CMA_3 staining. This helps to understand why this AT/GC-based CMA_3 staining does not work in fish–the GC-rich regions are too small and less prominent to be recognizable on small fish chromosomes.

3.3. GC- and Repeat-Content in Selected Mammals and Comparison with Fish

A fully different picture exists in the four representatives of mammals (gorilla, cat, little brown bat, and greater horseshoe bat). Here, the flat "baseline" is formed by a mixture of repeats and non-repetitive DNA (orange points), whereas the highly GC-enriched genomic fractions are formed by clearly gene-rich DNA and the GC-depleted fractions mostly by repeats. The gene- and GC-rich regions form sharp borders and clearly delineated peaks along the chromosomes. There are some repeats with a higher GC%, however they hardly

reach the GC% of gene-rich DNA and never form peaks as the gene-rich DNA does. Hence, there are no regions of GC-rich(er) repeats as described above in fish.

3.4. Different Sliding Window Sizes in Fish and Mammals

Since fish genomes are mostly up to three-times smaller than the mammalian ones but both groups converge on approximately 2n = 46–50 chromosomes, mammalian chromosomes are larger. Similarly, genomes of polyploid fish are substantially larger. This is reflected in our tool by the possibility to select one of three currently available sliding window sizes (1 kbp, 3 kbp, and 10 kbp). Examples of results with these three different sliding window sizes are shown in the Figure 4. Following species are compared: one fish with a typical teleost haploid genome size around 1 pg, the Northern pike, one polyploid fish with the genome size around 3 pg, the Atlantic salmon and one mammal with genome size 3.5–4 pg, the gorilla. The sliding window size 1 kbp appears the best suitable for teleosts and other species with a comparable genome size. The sliding window size 3 kbp appears suitable for polyploid fish and mammals and better enables downsizing of resulting plots. The sliding window size 10 kbp can be used in the best way when an extreme downsizing of the plots is required or in species with (extremely) large genomes (e.g., amphibians, reedfish, mammals or other organisms including highly polyploid plants).

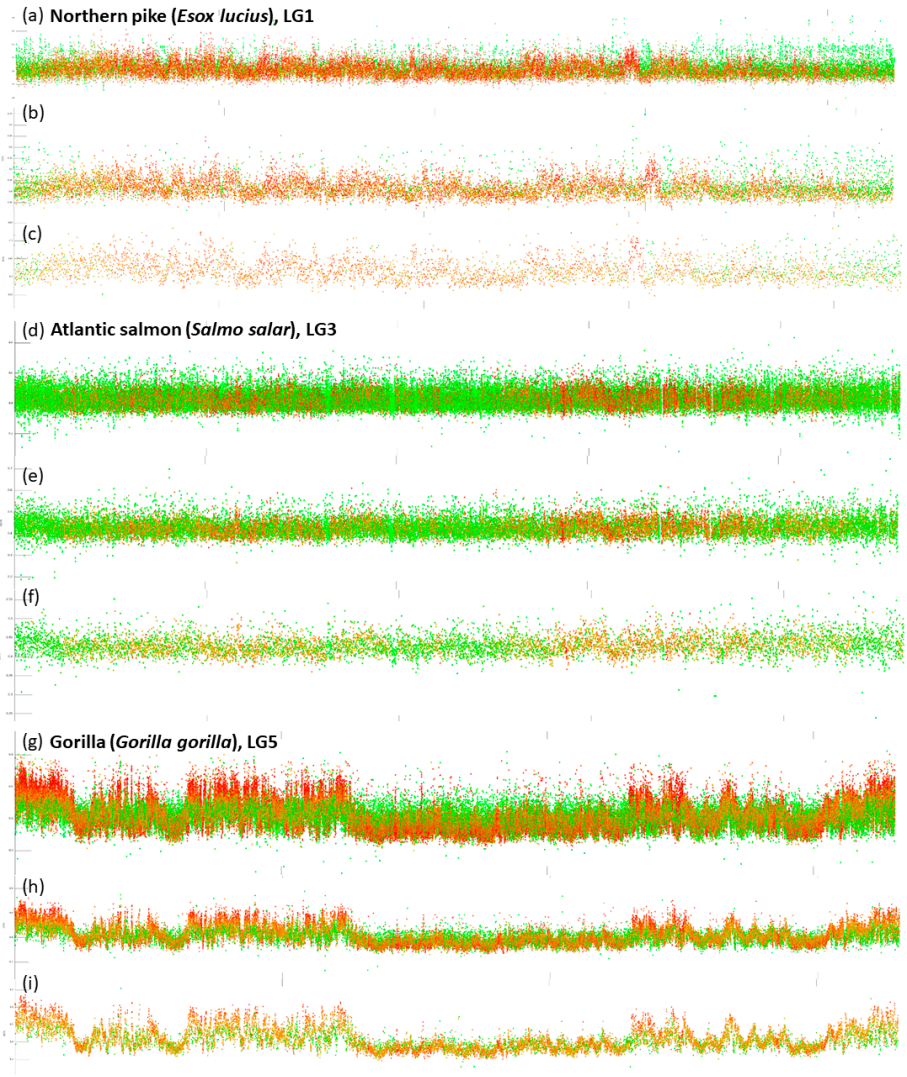


Figure 4. Comparison of three consecutive sliding window sizes, i.e., 1, 3 and 10 kbp in vertebrates with substantially different genome size. The Northern pike (**a**–**c**); the Atlantic salmon (**d**–**f**); gorilla (**g**–**i**).

3.5. Relationship between GC% and Repeats Percentage in Fishes and Mammals

Our tool enables a fast extraction of the values of GC% and rep% for each sliding window analysed (represented as a dot in the plots), makes scatterplots of these two measures and calculates Pearson's correlation coefficient (r). Separately, we tested for the linear relationship and correlation between these two measures in all species under study. This analysis shows a weak but significant positive correlation (r = 0.1–0.225, $p = 10^{-16}$) between GC% and rep% in nineteen of the 42 fish or fish-like species with the exception of *Amphiprion percula*, where r = −0.172. In the remaining fish species, r < 0.1 and in eight of them r = −0.082–−0.029, $p = 10^{-16}$–10^{-6}). These nineteen fish species show now phylogenetic relatedness. In the four mammals tested, there was a weak but significant negative correlation (r = −0.226–−0.046, $p = 10^{-16}$) between GC% and rep%. Data quality (either soft-masking or genome assembly) was insufficient for the following four species (*C. plagiosum*, *A. radiata*, *G. morhua*, *P. pectinata*). It is necessary to say that this analysis is highly dependent on the repeat masking quality and its accuracy will be increasing in the future.

Scatterplots including the r values for each species are available at our GitHub repository https://github.com/bioinfohk/evangelist_plots/tree/master/rep%25_vs_GC%25.

3.6. Functionality of the Tool

What makes this tool useful is the fully automated approach to data analysis. All steps are performed by a computer without any need of user´s input. The user only provides the names of species and waits for some time that depends on the bandwidth and the provided computer.

4. Discussion

4.1. Technical Requirements and Limitations

The presented plots shown here were created using a Linux server (64 GB RAM) however, the tool can run on a standard desktop computer only with a longer waiting time. The tool is fully dependent on the quality of the input data. This is the genome assembly quality and the quality of the repeat-(soft)masking (RM) procedure. RM can be redone in older genome assemblies against any up-to-date and/or custom repeat libraries in a separate step using, e.g., RepeatMasker tool [38]. We assume that the newly available genome assemblies will have increasingly better RM quality because of the rapid development in the masking strategies and the number of repeats newly identified. Currently, it is always necessary to bear in mind what might be the available level of RM of each species and hence until what extent the RM was sufficient, e.g., the very low rep% in *Tetraodon nigroviridis* might be indeed ascribed to its extremely streamlined genome with eliminated TEs [43]. Similarly, the high rep% in salmonids or zebrafish can be ascribed to their large genomes full of TEs [42,43]. On the other hand, the rep% in *Oryzias javanicus* is far more reduced in comparison with its much more explored congener *O. latipes* (Figure 1a) or another well explored model species *A. mexicanus* [43]. This means that the genome assembly and/or RM quality in *O. javanicus* is substantially lower than in other species.

There are several types of resulting plots based on the resolution of these considerable datasets: (1) A3-format; (2) large-scale plots; (3) crops; and (4) a combination of the previous ones.

Linking of chromosomes with their corresponding linkage groups (LGs) from genome assemblies is available only for a few fish species and this appears to be another limitation of LG profiling in practice. This means that in fish, it is mostly impossible to deduce the chromosome morphology (meta- vs. acrocentric, etc.) from the GC and repeats profiles at this stage. So far, we depend on the comparison of size-sorted LGs with the subjective size of chromosomes from cytogenetic studies and/or on the usage of genome browsers (e.g., the recently released NCBI Genome Data Viewer) to identify potential centromeres along LGs. This means that we can only estimate the position of centromeres after the comparison with chromosome size and morphology. How we could proceed with the identification of centromeres further depends on the quality of genome assemblies that is however

increasingly better, particularly thanks to long-read sequencing and its combination with the more accurate short-read sequencing (the hybrid approach). Another possibility is to localize genes for nuclear ribosomal RNA in the genome browser and on chromosomes.

4.2. GC- and Repeats-Profiling and Chromosome Banding in Fish

Replication banding has been used in fish to assign chromosomes to their homologous pairs [46], to identify sex chromosomes [21,47], and to describe chromosome rearrangements and polymorphisms [15,46]. It worked well on large salmonid chromosomes [16,46], but it is less applicable to small cyprinid or poecilid chromosomes [48]. On the other hand, the application of replication banding may be limited not only by the chromosome size, the degree of their spiralization but also by the genomic composition. A distinct and quite clear replication banding pattern has been observed in salmonids, whose repetitive DNA accounts for up to 60% of the genome [49]. On the contrary, a reduced number of replication bands was recognized along pufferfish chromosomes [24], whose genomes contain less than 10% repetitive elements due to their compaction [50,51]. Comparison of the replication banding pattern on the chromosomes of rainbow trout or masu salmon [16] and pufferfish clearly shows that salmonid chromosomes exhibit many early and late replicating bands alternating along their chromosomes [16], while pufferfish chromosomes are mostly composed of large early replicating bands sometimes covering almost entire chromosomal arms and small late replicating bands restricted to centromeric regions [24]. Genomes of salmonids and pufferfish underwent different (opposite) evolution, namely, whole genome duplication and genome compaction, respectively, that affected AT/GC composition in these fishes. This can be clearly observed in the GC-profiles of LGs studied in these species in the present research (Figure 2). In rainbow trout and salmon, repetitive DNA is equally distributed in the genome and interrupted with small blocks of non-repetitive DNAs while, in pufferfish most of the genome is composed of non-repetitive DNAs (Figure 2) given that repeat masking was of comparable quality in these species. This shows that the reduction of repetitive genomic elements during evolution decreases the resolution (and efficiency) of chromosomal banding based on the different phases of replication. GC% and repetitive DNAs profiling described here may indeed become an efficient tool in approaching "computational cytogenetics" in the future because this compensates for the small sizes of teleost chromosomes. Hence this approach might be complementary to the replication banding in species with suitable genomes/chromosomes.

Our results are consistent with previous findings that the GC% of the repetitive (soft-masked) genomic fraction is mostly higher than the genome-wide GC% in fish [52]. Namely, our plots in fish show that the repetitive fraction homogenizes GC% (compensates for the decrease in GC% of the non-repetitive fraction) and even increases the regional GC% values. This was not the case in the four mammalian genomes analysed. Since there is still no consensus about the origin of the AT/GC heterogeneity in vertebrates and the evolutionary mechanisms responsible, which may be varied [53], we assess our results in fish following the three main concepts discussed in [53]. First, the currently best supported view is that GC-biased gene conversion (gBGC) increases GC% at selectively neutral or weakly selected sites. Here, we can speculate that the small size of fish chromosomes might have resulted in a more effective gBGC through a higher rate of crossing over per Mbp [54,55] and led to GC-richness even in repeats. This should have, however, resulted in GC-richer genomes in fish than in mammals, which is not the case. Second, the high proportion of transposons in genomes results in a high rate of DNA methylation [56], and methylated cytosines are hypermutable and highly susceptible to spontaneous oxidative deamination [57,58], leading to a reduction in genomic GC% [59]. This could explain the observed homogeneous base composition of fish genomes. Moreover, the compact pufferfish genome, with low repeat and transposon density is GC-rich and heterogeneous. Finally, the role of selection in the GC evolution of the host genome [26,60] has largely been abandoned [53]. However, selection may play a role in the evolution of GC% of transposons and in their compositional interactions with host genomes. Here, it will be

necessary to assess GC% first in functional and degraded transposons and in their different classes. The first results in this field show a higher GC% in the Class II transposons than in the Class I [52]. More importantly, there are indications that the base composition of human non-LTR retrotransposons is indeed evolving under selection and may be reflective of the long-term co-evolution between non-LTR retrotransposons and the host genome [61]. This study summarizes current knowledge on the base composition of transposons in mammals and its impact.

4.3. Towards Understanding the AT/GC Homogeneity of Fish Genomes

The inability to achieve G-banding in fish has been largely ascribed to their AT/GC homogeneity [29], and our detailed analyses of sequence data support this, albeit in only a small fraction of fish species (Table A1) covering 27 fish orders/groups (of the total 85; [62]).

There are no substantial differences among the here analysed teleosts indicating any so far hidden AT/GC heterogeneity, up to the role of genome size and repeats proportion in tetraodontiform fishes. On the other hand, a very special case is gars (Lepisosteiformes). These last survivors of an ancient lineage [62] were discovered to have a rather mammalian way of AT/GC heterogeneity [34]. In contrast, their most closely related, the last surviving species of Amiiformes, the bowfin (*Amia calva*, [62]), has the typical teleost-like AT/GC homogeneity [63]. These two fish groups still represent a puzzle that will persist at least until the genome assembly of bowfin will be available, which should be soon (Braasch, *pers. comm.*). At this stage, we can describe traits related to chromosome organization in the spotted gar–the only one gar species with a genome assembly available, luckily at the chromosome level [64]. Even more luckily, despite a high degree of incompleteness of the spotted gar's genome assembly (945.878 Mb versus approx. C = 1.4 pg [40]), its GC-profile still clearly shows the mammalian type of AT/GC heterogeneity. The above-mentioned study on gars further compared CMA_3-stained (i.e., GC-rich, red, AT-rich, green) chromosomes of selected vertebrate groups including the starry sturgeon (*Acipenser stellatus*). They show that the small-sized microchromosomes are red or reddish in this sturgeon, whereas macrochromosomes are homogenously green with reddish centromeres (Figure E in [34]). This corresponds to the results presented here (Figures 2 and 3) in sterlet (*A. ruthenus*), where microchromosomes are GC-richer. This is an interesting result regarding the fact that numerous microchromosomes were presented with C-bands visualizing the constitutive heterochromatin in sturgeon hybrids [65]. On the other hand, these authors further present the results of their comparative genomic hybridization and genomic in situ hybridization showing the hybridization signals mostly on microchromosomes [65]. This might be alternatively interpreted that microchromosomes bear mostly coding regions that retain more sequence similarity among the compared species than the DNA on macrochromosomes that contain more repeats. Hence, clearly, this topic deserves further attention from both molecular cytogenetics and genomics to elucidate the potential differences between micro- and macrochromosomes. The importance of combining cytogenetics with genomics is evidenced by the fact that during sequencing, the first sturgeon microdissection of metaphase chromosomes assisted in proper genome assembly [27]. We address the quantitative traits/aspects of GC% in fish and across vertebrates in our other study published in this special issue [66].

Our results further show the GC-richness of small-size (micro)chromosomes also in three chondrichthyans, although the soft-masking did not work properly in the two of them (*P. pectinata* and *A. radiata*). There can be seen a great potential in comparisons with cytogenetic studies using CMA_3-staining, e.g., [67] published an impressive AT/GC pattern in two *Scleropages* species (*S. jardinii* and *S. leichardti*), while the only species with an available genome (and processed here), *S. formosus*, appears to have the typical teleost AT/GC banding pattern [67]. This shows that the question of GC biology in fish and generally in vertebrates is still far from being solved satisfactorily.

Author Contributions: Conceptualization, R.S.; methodology, R.S., D.M. and V.B.; software, D.M.; validation, V.B.; data curation, D.M.; writing—original draft preparation, R.S., V.B. and K.O.; writing—review and editing, R.S. and K.O.; visualization, D.M. and V.B.; supervision, R.S. and K.O.; project administration, R.S.; funding acquisition, R.S. and V.B. All authors have read and agreed to the published version of the manuscript.

Funding: This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 754462. This project was further funded by the Erasmus+ programme of the European Union with contract Nr. 2019-1-CZ01-KA203-061433. The APC was funded by the Faculty of Science, University of Hradec Králové.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available datasets were analysed in this study. This data can be found at: https://github.com/bioinfohk/evangelist_plots.

Acknowledgments: We would like to acknowledge W. Mike Howell for revision of this manuscript. Computational resources were supplied by the project "e-Infrastruktura CZ" (e-INFRA LM2018140) provided within the program Projects of Large Research, Development and Innovations Infrastructures.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Summarizing the fish species analysed in this study.

Species	Order	2n [1]	Genome Size (pg) [2]	GC%
Acipenser ruthenus	Acipenseriformes	120	1.8	39.8
Amblyraja radiata	Rajiformes	98	2.17	44.6
Amphiprion percula	Ovalentaria	48	0.9	39.5
Astatotilapia calliptera	Cichliformes	46	NA	41.1
Astyanax mexicanus	Characiformes	50	~1.5	38.4
Betta splendens	Anabantiformes	42	0.64	45.2
Carassius auratus	Cypriniformes	50	1.8	37.5
Chiloscyllium plagiosum	Orectolobiformes	102	~4.56	42
Ciona intestinalis	Tunicata	28	0.2	36
Clupea harengus	Clupeiformes	54	~0.9	44.2
Cottoperca gobio	Perciformes	48	NA	41
Cynoglossus semilaevis	Pleuronectiformes	44	0.62	41.3
Cyprinus carpio	Cypriniformes	100	1.8	37.1
Danio rerio	Cypriniformes	50	1.95	36.7
Denticeps clupeoides	Clupeiformes	40	NA	43.7
Echeneis naucrates	Carangiformes	48	0.7	41.4
Erpetoichthys calabaricus	Polypteriformes	36	4.7	40.1
Esox lucius	Esociformes	50	1.1	42.2
Gadus morhua	Gadiformes	46	0.65	46.3
Gasterosteus aculeatus	Gasterosteiformes	42	0.65	44.6
Gouania willdenowi	Gobiesociformes	48	NA	38.4
Ictalurus punctatus	Siluriformes	58	1	39.7
Larimichthys crocea	Perciformes	48	NA	41.4
Lepisosteus oculatus	Lepisosteiformes	58	1.4	40.1
Maylandia zebra	Cichliformes	46	NA	41.1
Myripristis murdjan	Beryciformes	48	~0.9	41.8
Oncorhynchus mykiss	Salmoniformes	58	2.7	43.4
Oreochromis niloticus	Cichliformes	46	1	39.9
Oryzias javanicus	Beloniformes	48	0.9	39
Oryzias latipes	Beloniformes	48	1	40.8
Parambassis ranga	Ovalentaria	48	NA	42.5
Poecilia reticulata	Cyprinodontiformes	46	0.88	40.3

Table A1. *Cont.*

Species	Order	2n [1]	Genome Size (pg) [2]	GC%
Pristis pectinata	Pristiformes	92	2.8	42.6
Salarias fasciatus	Blenniformes	46	0.83	44.4
Salmo salar	Salmoniformes	60	3.15	43.9
Scleropages formosus	Osteoglossiformes	50	NA	44.1
Scophthalmus maximus	Pleuronectiformes	44	0.75	43.4
Sparus aurata	Perciformes	48	0.95	41.7
Sphaeramia orbicularis	Kurtiformes	48	NA	37.8
Takifugu rubripes	Tetraodontiformes	44	0.4	45.8
Tetraodon nigroviridis	Tetraodontiformes	42	0.43	46.6
Xiphophorus maculatus	Cypridontiformes	48	0.9	39.8

[1.] Based on data in NCBI or Arai, 2011; [2.] Based on www.genomesize.com; NA, not available.

References

1. Holmquist, G.P. Evolution of chromosome bands: Molecular ecology of noncoding DNA. *J. Mol. Evol.* **1989**, *28*, 469–486. [CrossRef]
2. Bickmore, W.; Craig, J. Chromosome bands: Patterns in the genome; Molecular Biology Intelligence Unit. In *Chapman & Hall*; Landes Bioscience: New York, NY, USA; Austin, TX, USA, 1997; ISBN 978-1-57059-393-2.
3. Holmquist, G.P. Chromosome bands, their chromatin flavors, and their functional features. *Am. J. Hum. Genet.* **1992**, *51*, 17–37.
4. Holmquist, G.P. Chromosomal Bands and Sequence Features. In *Encyclopedia of Life Sciences*; John Wiley & Sons, Ltd.: Chichester, UK, 2005; ISBN 978-0-470-01617-6.
5. Costantini, M.; Auletta, F.; Bernardi, G. Isochore patterns and gene distributions in fish genomes. *Genomics* **2007**, *90*, 364–371. [CrossRef]
6. Melodelima, C.; Gautier, C. The GC-heterogeneity of teleost fishes. *BMC Genom.* **2008**, *9*, 632. [CrossRef] [PubMed]
7. Blaxhall, P.C. Chromosome karyotyping of fish using conventional and G-banding methods. *J. Fish Biol.* **1983**, *22*, 417–424. [CrossRef]
8. Schmid, M.; Guttenbach, M. Evolutionary diversity of reverse (R) fluorescent chromosome bands in vertebrates. *Chromosoma* **1988**, *97*, 101–114. [CrossRef] [PubMed]
9. Medrano, L.; Bernardi, G.; Couturier, J.; Dutrillaux, B.; Bernardi, G. Chromosome banding and genome compartmentalization in fishes. *Chromosoma* **1988**, *96*, 178–183. [CrossRef]
10. Arrighi, F.E.; Hsu, T.C. Localization of heterochromatin in human chromosomes. *Cytogenet. Genome Res.* **1971**, *10*, 81–86. [CrossRef] [PubMed]
11. Howell, W.M.; Black, D.A. Controlled silver-staining of nucleolus organizer regions with a protective colloidal developer: A 1-step method. *Experientia* **1980**, *36*, 1014–1015. [CrossRef]
12. Sharma, O.P.; Tripathi, N.K.; Sharma, K.K. A Review of Chromosome Banding in Fishes. In *Some Aspects of Chromosome Structure and Functions*; Springer: Dordrecht, The Netherlands, 2002; pp. 109–122. ISBN 978-0-7923-7057-4.
13. Toledo, A.; Viegas-Péquignot, E.; Foresti, F.; Filho, T.; Dutrillaux, B. BrdU replication patterns demonstrating chromosome homoeologies in two fish species, genus. *Eig. Cytogenet. Genome Res.* **1988**, *48*, 117–120. [CrossRef]
14. Lemieux, N.; Drouin, R.; Richer, C.-L. High-resolution dynamic and morphological G-bandings (GBG and GTG): A comparative study. *Hum. Genet.* **1990**, *85*, 261–266. [CrossRef] [PubMed]
15. Jankun, M.; Ocalewicz, K.; Woznicki, P. Replication C- and Fluorescent Chromosome Banding Patterns in European Whitefish, *Coregonus lavaretus* L. *Hereditas* **2004**, *128*, 195–199. [CrossRef]
16. Fujiwara, A.; Nishida-Umehara, C.; Sakamoto, T.; Okamoto, N.; Nakayama, I.; Abe, S. Improved fish lymphocyte culture for chromosome preparation. *Genetica* **2001**, *111*, 77–89. [CrossRef] [PubMed]
17. Salvadori, S.; Coluccia, E.; Cannas, R.; Cau, A.; Deiana, A.M. Replication Banding in two Mediterranean Moray eels: Chromosomal Characterization and Comparison. *Genetica* **2003**, *119*, 253–258. [CrossRef] [PubMed]
18. Salvadori, S.; Deiana, A.M.; Deidda, F.; Lobina, C.; Mulas, A.; Coluccia, E. XX/XY sex chromosome system and chromosome markers in the snake eel *Ophisurus serpens* (Anguilliformes: Ophichtidae). *Mar. Biol. Res.* **2018**, *14*, 158–164. [CrossRef]
19. Hellmer, A.; Voiculescu, I.; Schempp, W. Replication banding studies in two cyprinid fishes. *Chromosoma* **1991**, *100*, 524–531. [CrossRef]
20. Daga, R.R.; Thode, G.; Amores, A. Chromosome complement, C-banding, Ag-NOR and replication banding in the zebrafish *Danio rerio*. *Chromosome Res.* **1996**, *4*, 29–32. [CrossRef]
21. Molina, W.F.; Galetti, P.M. Early replication banding in *Leporinus* species (Osteichthyes, Characiformes) bearing differentiated sex chromosomes (ZW). *Genetica* **2007**, *130*, 153–160. [CrossRef]
22. Zhang, Q.; Wolters, W.; Tiersch, T. Brief communication. Replication banding and sister-chromatid exchange of chromosomes of channel catfish (*Ictalurus punctatus*). *J. Hered.* **1998**, *89*, 348–353. [CrossRef]
23. Fujiwara, A.; Fujiwara, M.; Nishida-Umehara, C.; Abe, S.; Masaoka, T. Characterization of Japanese flounder karyotype by chromosome bandings and fluorescence in situ hybridization with DNA markers. *Genetica* **2007**, *131*, 267–274. [CrossRef]
24. Grützner, F.; Lütjens, G.; Rovira, C.; Barnes, D.W.; Ropers, H.; Haaf, T. Classical and molecular cytogenetics of the pufferfish (*Tetraodon nigroviridis*). *Chromosome Res.* **1999**, *7*, 655–662. [CrossRef] [PubMed]

25. Schemczssen-Graeff, Z.; Barbosa, P.; Castro, J.P.; da Silva, M.; de Almeida, M.C.; Moreira-Filho, O.; Artoni, R.F. Dynamics of Replication and Nuclear Localization of the B Chromosome in Kidney Tissue Cells in *Astyanax scabripinnis* (Teleostei: Characidae). *Zebrafish* **2020**, *17*, 147–152. [CrossRef] [PubMed]
26. Bernardi, G. *Structural and Evolutionary Genomics: Natural Selection in Genome Evolution*; Elsevier: Amsterdam, The Netherlands, 2005.
27. Du, K.; Stöck, M.; Kneitz, S.; Klopp, C.; Woltering, J.M.; Adolfi, M.C.; Feron, R.; Prokopov, D.; Makunin, A.; Kichigin, I.; et al. The sterlet sturgeon genome sequence and the mechanisms of segmental rediploidization. *Nat. Ecol. Evol.* **2020**, *4*, 841–852. [CrossRef] [PubMed]
28. NCBI Genome Browser. Available online: https://www.ncbi.nlm.nih.gov/genome/browse (accessed on 30 September 2020).
29. Symonová, R.; Howell, W. Vertebrate Genome Evolution in the Light of Fish Cytogenomics and rDNAomics. *Genes* **2018**, *9*, 96. [CrossRef] [PubMed]
30. Schweizer, D. Simultaneous fluorescent staining of R bands and specific heterochromatic regions (DA-DAPI bands) in human chromosomes. *Cytogenet. Genome Res.* **1980**, *27*, 190–193. [CrossRef] [PubMed]
31. Schweizer, D. Counterstain-enhanced chromosome banding. *Hum. Genet.* **1981**, *57*, 1–14. [CrossRef]
32. Wang, Y.; Minoshima, S.; Shimizu, N. Cot-1 banding of human chromosomes using fluorescence in situ hybridization with Cy3 labeling. *Jpn. J. Hum. Genet.* **1995**, *40*, 243–252. [CrossRef]
33. Sumner, A.T.; Evans, H.J.; Buckland, R.A. New Technique for Distinguishing between Human Chromosomes. *Nat. New Biol.* **1971**, *232*, 31–32. [CrossRef]
34. Symonová, R.; Majtánová, Z.; Arias-Rodriguez, L.; Mořkovský, L.; Kořínková, T.; Cavin, L.; Pokorná, M.J.; Doležálková, M.; Flajšhans, M.; Normandeau, E.; et al. Genome Compositional Organization in Gars Shows More Similarities to Mammals than to Other Ray-Finned Fish. *J. Exp. Zool. Part B Mol. Dev. Evol.* **2017**, *328*, 607–619. [CrossRef]
35. Varadharajan, S.; Rastas, P.; Löytynoja, A.; Matschiner, M.; Calboli, F.C.F.; Guo, B.; Nederbragt, A.J.; Jakobsen, K.S.; Merilä, J. A high-quality assembly of the nine-spined stickleback (*Pungitius pungitius*) genome. *Genome Biol. Evol.* **2019**, *11*, 3291–3308. [CrossRef]
36. Verdugo, R.A.; Orostica, K.Y. Global Visualization Tool of Genomic Data. *Bioinformatics* **2016**, *32*, 2366–2368. [CrossRef]
37. Hunt, S.E.; McLaren, W.; Gil, L.; Thormann, A.; Schuilenburg, H.; Sheppard, D.; Parton, A.; Armean, I.M.; Trevanion, S.J.; Flicek, P.; et al. Ensembl variation resources. *Database* **2018**, *2018*. [CrossRef] [PubMed]
38. Smit, A.F.A.; Hubley, R.; Green, P. RepeatMasker Open-4.0. Available online: http://www.repeatmasker.org2015 (accessed on 30 September 2020).
39. Cock, P.J.A.; Antao, T.; Chang, J.T.; Chapman, B.A.; Cox, C.J.; Dalke, A.; Friedberg, I.; Hamelryck, T.; Kauff, F.; Wilczynski, B.; et al. Biopython: Freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **2009**, *25*, 1422–1423. [CrossRef]
40. Gregory, T.R. Animal Genome Size Database. Available online: http://www.genomesize.com (accessed on 30 September 2020).
41. Carducci, F.; Barucca, M.; Canapa, A.; Carotti, E.; Biscotti, M.A. Mobile Elements in Ray-Finned Fish Genomes. *Life* **2020**, *10*, 221. [CrossRef]
42. Gao, B.; Shen, D.; Xue, S.; Chen, C.; Cui, H.; Song, C. The contribution of transposable elements to size variations between four teleost genomes. *Mob. DNA* **2016**, *7*, 4. [CrossRef] [PubMed]
43. Shao, F.; Han, M.; Peng, Z. Evolution and diversity of transposable elements in fish genomes. *Sci. Rep.* **2019**, *9*, 1–8. [CrossRef] [PubMed]
44. Symonová, R.; Ocalewicz, K.; Kirtiklis, L.; Delmastro, G.B.; Pelikánová, Š.; Garcia, S.; Kovařík, A. Higher-order organisation of extremely amplified, potentially functional and massively methylated 5S rDNA in European pikes (*Esox* sp.). *BMC Genom.* **2017**, *18*, 391. [CrossRef] [PubMed]
45. Supiwong, W.; Tanomtong, A.; Supanuam, P.; Seetapan, K.; Khakhong, S.; Sanoamuang, L. Chromosomal Characteristic of Nile Tilapia (*Oreochromis niloticus*) from Mitotic and Meiotic Cell Division by T-Lymphocyte Cell Culture. *Cytologia* **2013**, *78*, 9–14. [CrossRef]
46. Jankun, M.; Woznicki, P.; Furgala-Selezniow, G. Chromosomal evolution in the three species of Holarctic fish of the Genus *Coregonus* (Salmoniformes). *Adv. Limnol.* **2005**, *60*, 25–37.
47. Bertollo, L.A.C.; Fontes, M.S.; Fenocchio, A.S.; Cano, J. The X1X2Y sex chromosome system in the fish *Hoplias malabaricus*. I. G-, C- and chromosome replication banding. *Chromosome Res.* **1997**, *5*, 493–499. [CrossRef]
48. Ocalewicz, K. Identification of Early and Late Replicating Heterochromatic Regions on Platyfish (*Xiphophorus maculatus*) Chromosomes. *Folia Biol.* **2005**, *53*, 149–153. [CrossRef] [PubMed]
49. Lien, S.; Koop, B.F.; Sandve, S.R.; Miller, J.R.; Kent, M.P.; Nome, T.; Hvidsten, T.R.; Leong, J.S.; Minkley, D.R.; Zimin, A.; et al. The Atlantic salmon genome provides insights into rediploidization. *Nature* **2016**, *533*, 200–205. [CrossRef] [PubMed]
50. Aparicio, S. Whole-Genome Shotgun Assembly and Analysis of the Genome *of Fugu rubripes*. *Science* **2002**, *297*, 1301–1310. [CrossRef] [PubMed]
51. Jaillon, O.; Aury, J.-M.; Brunet, F.; Petit, J.-L.; Stange-Thomann, N.; Mauceli, E.; Bouneau, L.; Fischer, C.; Ozouf-Costaz, C.; Bernot, A.; et al. Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature* **2004**, *431*, 946–957. [CrossRef]
52. Symonová, R.; Suh, A. Nucleotide composition of transposable elements likely contributes to AT/GC compositional homogeneity of teleost fish genomes. *Mob. DNA* **2019**, *10*, 1–8. [CrossRef]

53. Mugal, C.F.; Weber, C.C.; Ellegren, H. GC-biased gene conversion links the recombination landscape and demography to genomic base composition: GC-biased gene conversion drives genomic base composition across a wide range of species. *BioEssays* **2015**, *37*, 1317–1326. [CrossRef]
54. Montoya-Burgos, J.I.; Boursot, P.; Galtier, N. Recombination explains isochores in mammalian genomes. *Trends Genet.* **2003**, *19*, 128–130. [CrossRef]
55. Eyre-Walker, A. Recombination and mammalian genome evolution. *Proc. R. Soc. Lond. Ser. B Biol. Sci.* **1993**, *252*, 237–243. [CrossRef]
56. de Mendoza, A.; Hatleberg, W.L.; Pang, K.; Leininger, S.; Bogdanovic, O.; Pflueger, J.; Buckberry, S.; Technau, U.; Hejnol, A.; Adamska, M.; et al. Convergent evolution of a vertebrate-like methylome in a marine sponge. *Nat. Ecol. Evol.* **2019**, *3*, 1464–1473. [CrossRef]
57. Fryxell, K.J.; Zuckerkandl, E. Cytosine Deamination Plays a Primary Role in the Evolution of Mammalian Isochores. *Mol. Biol. Evol.* **2000**, *17*, 1371–1383. [CrossRef]
58. Wang, R.Y.-H.; Kuo, K.C.; Gehrke, C.W.; Huang, L.-H.; Ehrlich, M. Heat- and alkali-induced deamination of 5-methylcytosine and cytosine residues in DNA. *Biochim. Biophys. Acta (BBA) Gene Struct. Expr.* **1982**, *697*, 371–377. [CrossRef]
59. Mugal, C.F.; Arndt, P.F.; Holm, L.; Ellegren, H. Evolutionary Consequences of DNA Methylation on the GC Content in Vertebrate Genomes. *G3 Genes Genomes Genet.* **2015**, *5*, 441–447. [CrossRef] [PubMed]
60. Bernardi, G. The neoselectionist theory of genome evolution. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 8385–8390. [CrossRef] [PubMed]
61. Ruggiero, R.P.; Boissinot, S. Variation in base composition underlies functional and evolutionary divergence in non-LTR retrotransposons. *Mob. DNA* **2020**, *11*, 1–18. [CrossRef] [PubMed]
62. Nelson, J.S.; Grande, T.; Wilson, M.V.H. *Fishes of the World*, 5th ed.; John Wiley & Sons: Hoboken, NJ, USA, 2016; ISBN 978-1-118-34233-6.
63. Majtánová, Z.; Symonová, R.; Arias-Rodriguez, L.; Sallan, L.; Ráb, P. "Holostei versus Halecostomi" Problem: Insight from Cytogenetics of Ancient Nonteleost Actinopterygian Fish, Bowfin *Amia calva*. *J. Exp. Zool. B Mol. Dev. Evol.* **2017**, *328*, 620–628. [CrossRef] [PubMed]
64. Braasch, I.; Gehrke, A.R.; Smith, J.J.; Kawasaki, K.; Manousaki, T.; Pasquier, J.; Amores, A.; Desvignes, T.; Batzel, P.; Catchen, J.; et al. The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons. *Nat. Genet.* **2016**, *48*, 427–437. [CrossRef]
65. Symonová, R.; Flajšhans, M.; Sember, A.; Havelka, M.; Gela, D.; Kořínková, T.; Rodina, M.; Rábová, M.; Ráb, P. Molecular Cytogenetics in Artificial Hybrid and Highly Polyploid Sturgeons: An Evolutionary Story Narrated by Repetitive Sequences. *Cytogenet. Genome Res.* **2013**, *141*, 153–162. [CrossRef]
66. Borůvková, V.; Howell, W.M.; Matoulek, D.; Symonová, R. Quantitative approach to fish cytogenetics in the context of vertebrate genome evolution. *Genes* **2021**. (submitted).
67. de Bello Cioffi, M.; Ráb, P.; Ezaz, T.; Antonio Carlos Bertollo, L.; Lavoué, S.; Aquiar de Oliveira, E.; Sember, A.; Molina, F.; Henrique Santos de Souza, F.; Majtánová, Z.; et al. Deciphering the Evolutionary History of Arowana Fishes (Teleostei, Osteoglossiformes, Osteoglossidae): Insight from Comparative Cytogenomics. *Int. J. Mol. Sci.* **2019**, *20*, 4296. [CrossRef]

genes

MDPI

Article

Present and Future Salmonid Cytogenetics

Muhammet Gaffaroglu [1], Zuzana Majtánová [2], Radka Symonová [3,*], Šárka Pelikánová [2], Sevgi Unal [4], Zdeněk Lajbner [5] and Petr Ráb [2]

[1] Department of Molecular Biology and Genetics, Faculty of Science, University of Ahi Evran, Kirsehir 40200, Turkey; mgaffaroglu@yahoo.com

[2] Laboratory of Fish Genetics, Institute of Animal Physiology and Genetics, Czech Academy of Sciences, 27721 Liběchov, Czech Republic; majtanova@iapg.cas.cz (Z.M.); pelikanova@iapg.cas.cz (Š.P.); rab@iapg.cas.cz (P.R.)

[3] Department of Bioinformatics, Wissenschaftszentrum Weihenstephan, Technische Universität München, 85354 Freising, Germany

[4] Department of Molecular Biology and Genetics, Faculty of Science, Bartin University, Bartin 74000, Turkey; sunal@bartin.edu.tr

[5] Physics and Biology Unit, Okinawa Institute of Science and Technology, Graduate University, Onna, Okinawa 904 0495, Japan; lajbner@oist.jp

* Correspondence: radka.symonova@tum.de

Received: 6 November 2020; Accepted: 2 December 2020; Published: 6 December 2020

Abstract: Salmonids are extremely important economically and scientifically; therefore, dynamic developments in their research have occurred and will continue occurring in the future. At the same time, their complex phylogeny and taxonomy are challenging for traditional approaches in research. Here, we first provide discoveries regarding the hitherto completely unknown cytogenetic characteristics of the Anatolian endemic flathead trout, *Salmo platycephalus*, and summarize the presently known, albeit highly complicated, situation in the genus *Salmo*. Secondly, by outlining future directions of salmonid cytogenomics, we have produced a prototypical virtual karyotype of *Salmo trutta*, the closest relative of *S. platycephalus*. This production is now possible thanks to the high-quality genome assembled to the chromosome level in *S. trutta* via soft-masking, including a direct labelling of repetitive sequences along the chromosome sequence. Repetitive sequences were crucial for traditional fish cytogenetics and hence should also be utilized in fish cytogenomics. As such virtual karyotypes become increasingly available in the very near future, it is necessary to integrate both present and future approaches to maximize their respective benefits. Finally, we show how the presumably repetitive sequences in salmonids can change the understanding of the overall relationship between genome size and G+C content, creating another outstanding question in salmonid cytogenomics waiting to be resolved.

Keywords: chromosome banding; cytotaxonomy of trout; FISH; NOR phenotype; rDNA, *Salmo platycephalus*

1. Introduction

The taxonomic species diversity of peri-Mediterranean and Near East brown trout is still not well understood, and new species are expected to be discovered and/or resurrected from the westernmost tip of the trout distribution area in Morocco [1,2], across the Iberian peninsula [3], Italy, the Balkans (reviewed in Kottelat and Freyhof [4]) and Greece [5]. Similarly, new trout species were recently described in the territory of Turkey in the Mediterranean Sea, Black Sea and Persian Gulf river drainages [6–11], presently encompassing 12 species. However, several authors have already recognized the taxonomic diversity of this region's brown trout ([9,12,13] and references therein).

Among these authors, Behnke [14] even erected a new subgenus *Platysalmo* within the genus *Salmo* and described species *P. platycephalus* for morphologically distinct trout found in the Zamanti River in the upper parts of the Seyhan River system in southeastern Turkey. The separate taxonomic status of this species was later confirmed by analyses of mtDNA and nuclear molecular markers, which nested the flathead trout within the Adriatic phylogeographic lineage of the brown *S. trutta* complex and advocated for its separate taxonomic status [13]. However, the exact position of *P. platycephalus* within the Adriatic cluster remained unclear. The researchers concluded that classification of the flathead trout as a genus and/or subgenus of *Salmo* is not supported by their data, although their taxonomic construction is generally accepted in all subsequent studies (e.g., Turan et al. [8]). The flathead trout's morphological and life history characteristics were addressed by Kara et al. [15,16]. Recently, flathead trout populations are critically endangered by habitat loss and stockings of non-native trout [17].

In spite of numerous cytogenetic studies of the brown trout [18], available data for trout of the peri-Mediterranean as well as the southeastern distribution range remains highly limited (Table 1). In this study, we described for the first time the karyotype and other chromosomal characteristics of the Anatolian endemic flathead trout, *Salmo platycephalus* Behnke, 1968, as revealed by conventional (Ag-impregnation, CMA_3 fluorescence) and molecular (FISH with 5S and 18S rDNA as well as telomeric probes) techniques. Such a detailed cytogenetic analysis of this species has been missing since the publication of the influential and so far most comprehensive overview of salmonid chromosome evolution [18]. To compare our results with other literature records, we also reviewed available cytotaxonomic data for Eurasian species of the genus *Salmo*. In so doing, we have updated and extensively summarized the present cytogenetics of the salmonid genus *Salmo*.

Current fish cytogenetics has been largely shaped by the huge sequencing effort worldwide, and there are trends to integrate cytogenetics with genomics in fish (e.g., Mazzuchelli et al. [19]; de Oliveira et al. [20]). Salmonids are economically and especially scientifically important [21]; Hence, their genomes have been increasingly sequenced, despite the sizeable obstacles of their genome size [22] and substantial repeats content [23,24] represent particularly to genome assembling. The NCBI/genome currently lists 12 salmonid genomes, of which six species have been assembled to the chromosome level (November 2020). The latest Release 101 of the Ensembl genome browser (August 2020) lists five salmonid species all assembled to the chromosome level [25]. These resources open up new directions for cytogenomic investigations in fish that are particularly relevant for salmonids. Namely, the genome assemblies available in Ensembl can be utilized to produce plots visualizing proportions of repetitive and non-repetitive fractions and their G+C content (GC%) simultaneously with a novel Python tool [26]. Hence, it is now also possible to use this tool for several salmonids and to produce a prototypical virtual karyotype for this group. Actually, the very first plots of *S. salar* are already available by Matoulek et al. [26], in three different resolutions, i.e., different sliding window sizes, (https://github.com/bioinfohk/evangelist_plots). However, *S. salar* belongs to the karyotype category B' *sensu* Phillips and Ráb [17], i.e., salmonids with 2n = ~60 (54–58) and chromosome arm number NF = 72–74. Hence, *Salmo trutta* Linnaeus, 1758, with a karyotype more similar to *S. platycephalus*, is more desirable for cytogenomic comparisons. The first results of virtual karyotyping of *S. salar* show that the soft-masked genome (i.e., repetitive fraction) appears surprisingly GC-rich (even richer in GC than the non-repetitive fraction [26].

Repetitive sequences that are generally highly important for fish cytogenetics are represented in salmonid genomes in thus far unprecedented proportions of up to 60%, among the highest proportions established for any vertebrate [27,28]. Repetitive sequences in salmonids were recently suggested to have a different relationship between salmonid genome GC% and genome size than that of other teleosts [29]. This is in line with the aforementioned results of virtual karyotyping. Now, thanks to the fast development in fish genomics, new teleost genomes including several salmonid species have become available. Therefore, it is desirable to address this still outstanding question of GC% of salmonid repeatome representing another direction of future research—namely the quantitative approach described in more details as related to fish in this special issue by Borůvková

et al. [30]. Hence, we outline future research directions not only of salmonid cytogenetics but of vertebrates' cytogenetics in general.

2. Materials and Methods

2.1. Studied Material

Five males and six females of flathead trout were collected by electrofishing in Karagoz Creek, Zamanti River Basin, 38.7350000 N, 36.4864000 E. The individuals were dissected both for direct chromosome preparation in field conditions as well as for other analyses and thus were not deposited in collection as vouchers. Valid Animal Use Protocol was enforced during study in IAPG CAS (No. CZ 02386).

2.2. Chromosome Preparation and Staining

Standard procedures for chromosome preparation followed those laid out in Ráb and Roth [31]. Chromosomal preparations from all individuals were stained with conventional Giemsa solution (5%, 10 min) to confirm the number and morphology of their chromosomes. Fluorescent staining with chromomycin A_3 (CMA_3) specific for GC-rich regions was applied, counterstained with DAPI, with a higher affinity for AT-rich regions [32]. Silver (Ag-) staining for detection of nucleolar organizer regions (NORs) followed Howell and Black [33]. The sequence of staining followed the protocol of Rábová et al. [34].

2.3. *Fluorescence* in situ *Hybridization (FISH) with Telomeric and rRNA Genes Probes*

Probes for FISH experiments were produced by PCR with the primer pairs and thermal cycling conditions according to Komiya and Takemura [35] for 5S rDNA and White et al. [36] for 28S rDNA. The PCR reactions were carried out in a final volume of 25 μL consisting of 100 ng genomic DNA, 12.5 μL PPP master mix, 0.01 mM of each primer and PCR water to complete the volume (all reagents from TopBio, Prague, Czech Republic). Cycling conditions were as follows: (a) 28S: 2 min at 95 °C; 35 cycles of 1 min at 95 °C, 40 s at 55 °C and 2 min at 72 °C; 5 min at 72 °C; (b) 5S rDNA: 5 min at 94 °C; two cycles of 1 min at 95 °C, 30 s at 61 °C, and 45 s at 72 °C; two cycles of 1 min at 95 °C, 30 s at 59 °C and 45 s at 72 °C; two cycles of 1 min at 95 °C, 30 s at 57 °C and 45 s at 72 °C; 25 cycles of 1 min at 95 °C, 30 s at 61 °C and 45 s at 72 °C; 7 min at 72 °C. The amplified fragments were sequenced at the ABI 3700 sequencer prior FISH experiments. Probes were indirectly labelled with biotin-16-dUTP (Roche, Mannheim, Germany) and digoxigenin-11-dUTP (Roche) through PCR reamplification of previously sequenced PCR products. Reamplification was carried out under the same condition as the previous PCR reaction. Labelled PCR products were precipitated. A hybridization mixture was made consisting of hybridization buffer [37], sonicated salmon sperm blocking DNA (15 μg/slide; Sigma-Aldrich, St. Louis, MO, USA) and differently labelled PCR products of both genes. The hybridization and detection procedure were carried out under conditions described by Symonová et al. [37]. The biotin-dUTP-labelled probes were detected by either the Invitrogen CyTM3-Streptavidin (Invitrogen, San Diego, CA, USA; cat. no. 43-4315) or by the FITC-Streptavidin (cat. no. 43-4311). The digoxigenin-dUTP-labelled probes were detected either by the Roche Anti-Digoxogenin-Fluorescein (cat. no. 11207741910) or by the Anti-Digoxigenin-Rhodamin (cat. no. 11207750910). The chromosomes were counterstained with Vectashield/DAPI (1.5 mg/mL) (Vector, Burlingame, CA, USA).

2.4. Microscopy and Image Analyses

Chromosomal preparations were examined by an Olympus Provis AX 70 epifluorescence microscope (Olympus, Tokyo, Japan). Images of metaphase chromosomes were recorded with a cooled Olympus DP30BW CCD camera (Olympus, Tokyo, Japan). The IKAROS and ISIS imaging programs (Metasystems, Altlussheim, Germany) were used to analyse grey-scale images. The captured digital images from FISH experiments were pseudocoloured (red for Anti-Digoxigenin-Rhodamine, green for Invitrogen FITC-Streptavidin) and superimposed using Adobe Photoshop

software, version CS5. Karyotypes from Giemsa-stained chromosomes were arranged in Ikaros (Metasystems) software. In the case of CMA_3/DAPI staining, the CMA_3 signal was inverted into the red channel while the DAPI signal went into the green channel to enhance the contrast between these two signal types. At least 25 metaphases (of the highest possible quality) per individual and method were analysed, some of them sequentially. Chromosomes were classified according to Levan et al. [38], but modified as m = metacentric, st = subtelocentric and a = acrocentric, where st and a chromosomes were scored as uni-armed to calculate the NF value (Nombre Fondamental, number of chromosome arms *sensu* Matthey [39].

2.5. Cytogenomic Analyses

First, we reviewed current genomic resources (Ensembl and NCBI/genome) for fish and especially salmonid genome assemblies. Second, we applied the novel Python tool EVANGELIST (= EVAluatioN on GEnome LIST) based on the non-overlapping sliding window to visualize and quantify percentage of repeats and GC% in both repeats and non-repetitive DNA simultaneously, introduced by Matoulek et al. [24]. With this tool, we produced the prototypical virtual karyotype for a salmonid (*S. trutta*). Third, we extracted and manually curated data on genome size and GC% from currently available fish genomes assembled to the chromosome level (the best genome quality) using the NCBI online tool https://www.ncbi.nlm.nih.gov/genome/browse#!/overview/. Finally, we processed these data with R [40] and compared them with our previous results [29].

3. Results

3.1. Karyotypes and Molecular Cytogenetic Traits

The diploid chromosome number was 2n = 80 and the karyotype was composed of 7 pairs of metacentric, 5 pairs of subtelocentric, 2 pairs of distinctly large acrocentric and 26 pairs of moderate sized acrocentric chromosomes, decreasing gradually in size (Figures 1 and 2a, b). The NF value equalled 96 (Figure 1). DAPI/CMA_3 fluorescence showed CMA_3-positive, i.e., highly GC-enriched, signals in p arms of the largest subtelocentric chromosome pair (Figure 2c). FISH with the 28S rDNA probe clearly visualized signals in the same position as CMA_3, while 5S rDNA sites were located in pericentromeric regions of one middle-sized metacentric chromosome pair (Figure 2e). FISH with the telomeric probe labelled the terminal regions of all chromosomes and did not reveal any interstitial signals (Figure 2d). Finally, Ag-NOR impregnation marked the same CMA_3 and FISH positive region, i.e., the p arms of the largest sub-telocentric chromosome pairs (Figure 2f).

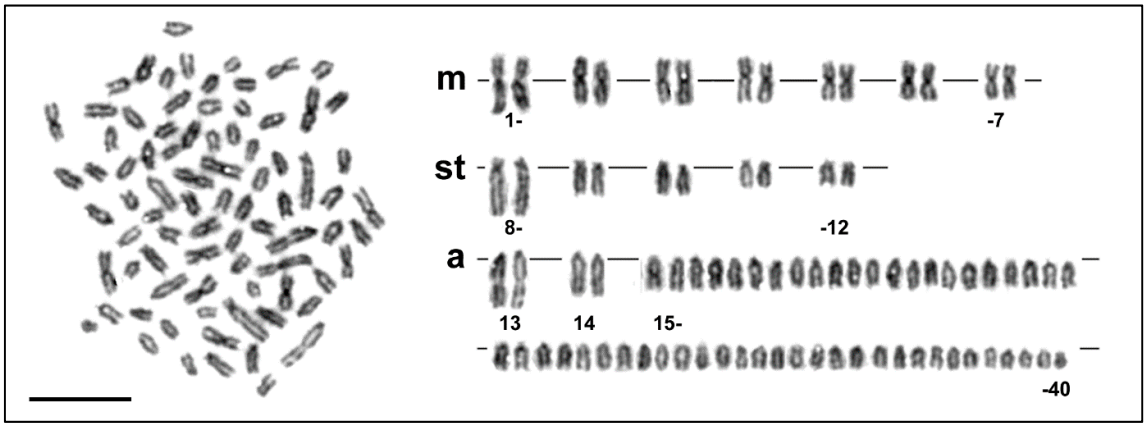


Figure 1. Giemsa stained metaphase plate and the corresponding karyotype of *Salmo platycephalus*. m, metacentric; st, subtelocentric; a, acrocentric chromosomes. Bar equals 10 μm.

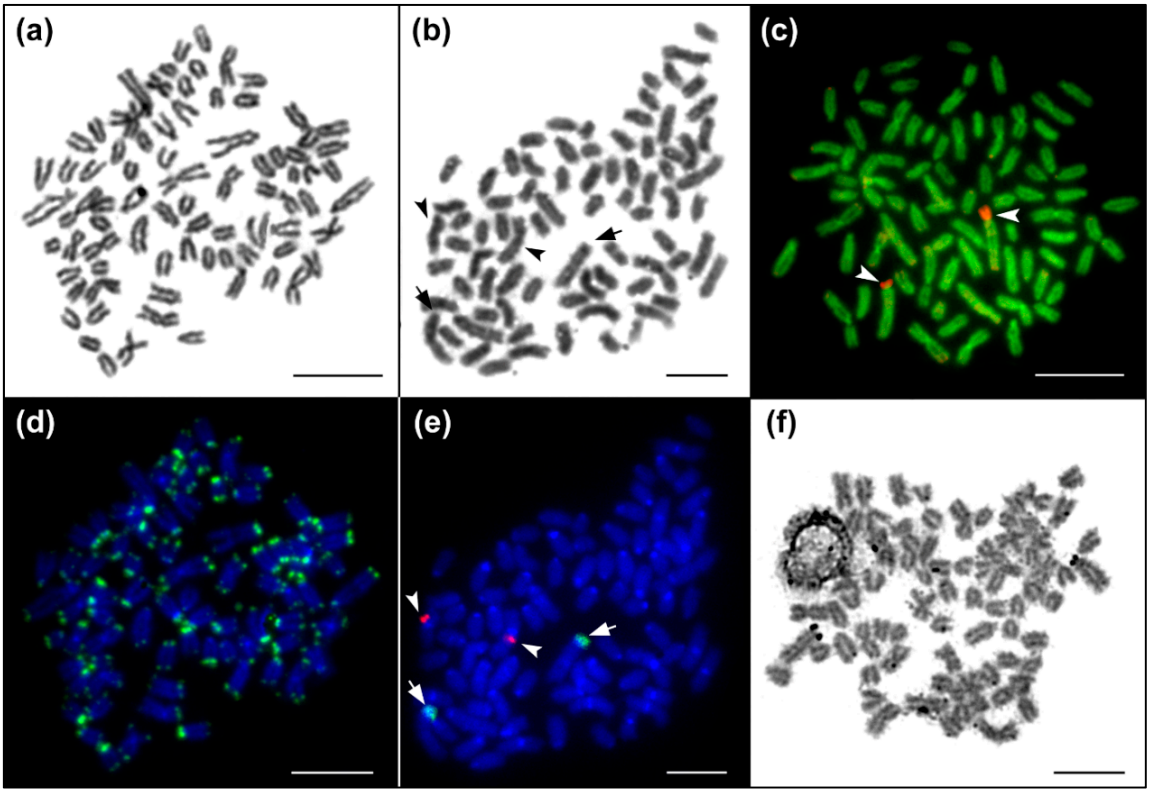


Figure 2. Chromosome analyses of *Salmo platycephalus*. (**a**, **b**) Giemsa-stained metaphases corresponding to (**d**, **e**) panels; (**c**) DAPI/CMA_3 fluorescence, DAPI stained chromosomes (green), CMA_3 signals of GC-rich regions (red); (**d**) DAPI stained chromosomes (blue), telomere repeat hybridization signals (green); (**e**) DAPI stained chromosomes (blue), 28S rDNA (green, indicated by arrows), 5S rDNA hybridization signals (red, indicated by arrowheads); (**f**) Ag-NOR impregnation showing the active major rDNA unit corresponding to the 28S rDNA sites. Bar equals 10 μm.

3.2. Virtual Karyotype and Cytogenomics in Salmonids

Thanks to the close relatedness between the *S. platycephalus* analysed here and the species *S. trutta* [41], which is among the best cytogenetically analysed salmonid fishes [18], it is relevant to compare their karyotypes. Utilizing the EVANGELIST Python tool, we produced the first virtual karyotype for the latter species (Figure 3). The virtual karyotype of *S. trutta* was confronted with its cytogenetics-based congeneric karyotypes. Both virtual karyotypes of the genus *Salmo* show a homogenization in GC% along chromosomes and repetitive as well as non-repetitive fractions reaching 50–60%. The comparison with actual karyotypes enabled an assignment of just the three largest chromosomes (Figure 3) at this stage and shows the need to improve the virtual analysis by including an option to visualise the two nuclear ribosomal fractions in the next step. Visualization of 5S rDNA can already be performed in Ensembl, showing that chromosome No. 1 bears the majority of 5S rDNA sequences (Figure 3, labelled with a blue arrowhead). Moreover, one more 5S rDNA site was identified on chromosome No. 20 (Figure 3, orange arrowhead). This site is probably below the detection range of FISH; however, it shows the potential of virtual karyotyping to visualize DNA sequences otherwise hidden for FISH.

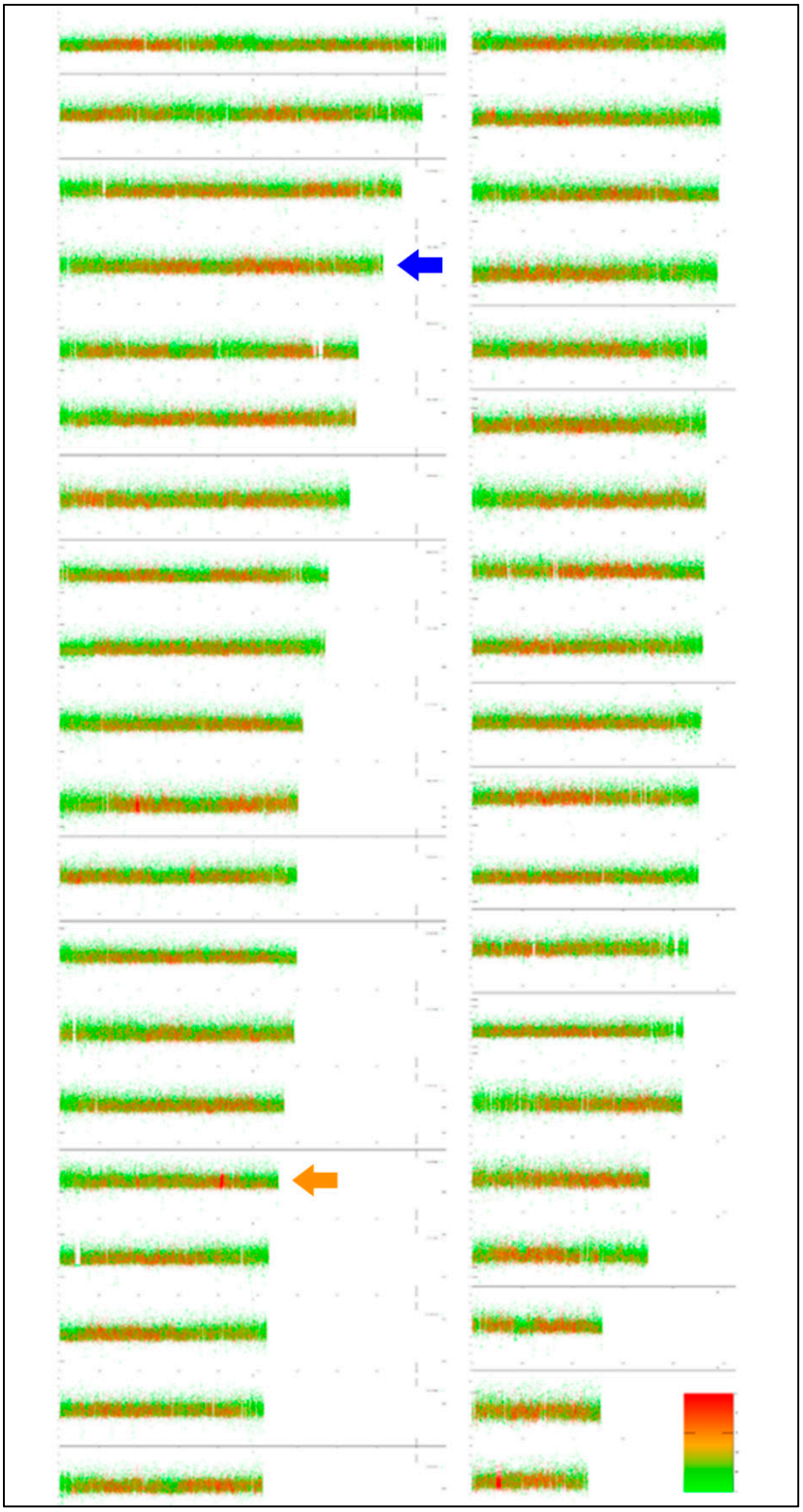


Figure 3. Virtual karyotype of *Salmo trutta* shows the haploid set of size-sorted chromosomes. The colour scale represents the proportion of repetitive (green) and non-repetitive (red) sequences. The *y* axis of each chromosome represents the scale of GC%. The karyotype based on cytogenetics in *S. trutta* enables us to roughly identify only the first three chromosomes according to their size—the largest acrocentric, the largest sub-telocentric and probably the largest metacentric chromosome. According to Ensembl, the 5S rDNA bearing chromosomes are chromosome No. 1 (the main site visualized also by FISH, blue arrow), i.e., the fourth largest chromosome, and chromosome No. 20 (orange arrow), which has a single 5S rDNA sequence.

Finally, we have taken advantage of the increasingly available data on genomic features (GC% and genome size) among teleost fishes with a special focus on salmonids. Including more species than in the previous analysis [29] confirmed earlier results that genome size negatively correlates with the genomic GC% in fish excluding salmonids. Moreover, the results revealed an inverse relationship between these two measures in salmonids in comparison with other teleosts (Figure 4).

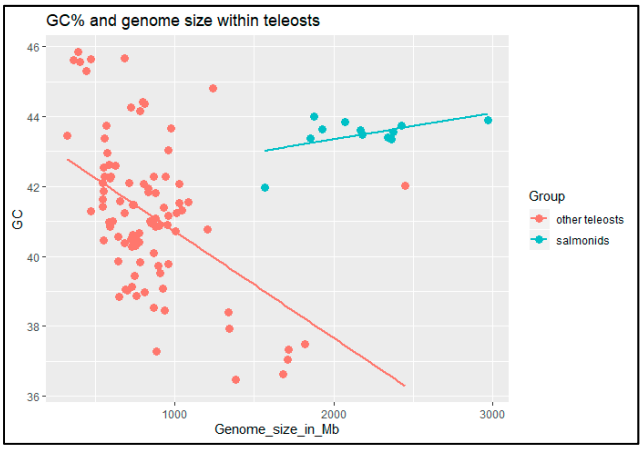


Figure 4. Scatter plot showing the relationship between GC% and genome size in salmonids and other teleosts. Data from https://www.ncbi.nlm.nih.gov/genome/browse#!/overview/.

4. Discussion

The salmonid genus *Salmo* (Linnaeus, 1758) represents freshwater anadromous fishes that are originally widely distributed from the North Atlantic Basin, i.e., Northeastern North America and Europe (including European Arctic) to the upper parts of Amu-Darya R. in Central Asia [42]. The genus contains two sister lineages—a primarily anadromous Atlantic salmon, *S. salar*, and primarily freshwater resident fishes collectively known as brown trout (*Salmo* spp.), although both lineages include numerous anadromous and freshwater populations. However, these salmonids have been introduced and/or stocked outside their native range virtually around the world, mainly as objects of recreational fishery. As a result, several countries report an adverse ecological impact after their introduction [43–47]. Although brown trout are of limited interest in production aquaculture [48] (except the commercially important Atlantic salmon), brown trout have been and still are objects of intense investigations in various types of studies [12]. Similarly, chromosomes of different species and forms of brown trout were already extensively studied by numerous authors by the end of the 19th century (see review of Gas [49]). To compare results of our cytogenetic analysis of flathead trout, we summarized all available chromosome data in brown trout (Table 1). Summing up these studies, we excluded those without reliable locality data, without descriptions of cytogenetic methods used, without the number of individuals used and/or simply reports from other non-referenced sources. In some cases, we reinterpreted specific status of material examined under the name *S. trutta* (i.e., *S. oxianus, S. cenerinus, S. farioides, S. lourosensis, S. peristericus*) since the published locality data clearly pointed to species different from true *S. trutta*. We also included older data based on analyses of anaphase chromosomes from embryo squashes for species/forms not analysed afterwards (i.e., *S. carpio, S. letnica, S. labrax, S. caspius*) but with a sufficient chromosome quality to reliably infer 2n and karyotypes. On the other hand, we are aware that all of these summarized studies had significant flaws. First, none of these studies clearly claimed that examined fishes were deposited in any collection to enable later taxonomic identification of the material analysed [50]. Second, the results of some cytogenetic studies, especially in peri-Mediterranean and central European populations, could certainly have been affected by the stocking of non-autochthonous individuals and their subsequent genetic admixture (e.g., Kohout et al. [51], Leitwein et al. [52]). Nevertheless, our review clearly shows that cytogenetic and/or cytotaxonomic characteristics of flathead trout are nearly or even invariably the same as in other species/forms of brown trout. To explain this conclusion, we further examine in detail the data regarding 2n, karyotype composition and other chromosomal characteristics.

Table 1. Review of reported cytogenetic data for members of Palearctic trout of the genus *Salmo*.

Species/Form	Locality	Country	Examined Individuals	2n	Karyotype Composition		NF	Ref.	Notes
					m/sm	st/a			
Aral Sea Basin									
S. oxianus	Kyzylsu R. (Amu Darya basin)	KAZ	3	80	18	62	98	[53]	1; 2
S. oxianus	Alamedin R. (Chu basin)	KG	5	80	18	62	98	[53]	1; 2
S. oxianus	Bech–Tach (Talas basin)	KG	11	80	18	62	98	[53]	1; 2
Balkans and Mediterranean Sea Basin									
S. carpio	Garda L.	IT	embryos	80	20	60	100	[54]	1
S. cenerinus	Monti Sibillini	IT	57 †	80	14/8	58	102	[55]	2; 5
S. farioides	Drosopigi R.	GR	/	80	20	60	100	[56]	1; 2
S. lourosensis	Louros R.	GR	/	80	20	60	100	[56]	1; 2
S. letnica	Ochrid L.	MK	embryos	80			104	[57]	1
S. marmoratus	Socha R.	SI	/	80	22	58	102	[58]	1
S. marmoratus	Socha R.	SI	1	80	22	58	102	[59]	1
S. marmoratus	Idrijca R.	SI	2	80	22	58	102	[59]	1
S. marmoratus	Friuli–Venezia	IT	57 †	80	14/8	58	102	[55]	2; 5
S. obtusirostris	Buna R. (Neretva R. basin)	BIH	/	82	12	70	94	[60]	1
S. peristericus	Aigos Germanos	GR	/	80	20	60	100	[56]	1; 2
S. trutta *	Buni, Krupica, Bistrica R.	RU	17	80	18–20	62–60	100	[61]	1; 2
S. trutta *	Klinje L.	BIH	17	80	20	60	100	[61]	1; 2
S. trutta *	Pschata R.	SI	/	80	20	60	100	[58]	1
S. trutta *	Tripotamos R.	GR	/	76	16	60	92	[56]	1; 2
Baltic Sea Basin									
S. trutta	Ropsha	RU	/	78	20	58	98	[62]	3
S. trutta (anadromous)	Vistula R.	PL	23	80	14/6	60	100	[63]	
S. trutta (anadromous)	Vistula R.	PL	21	80	22	58	102	[64]	
S. trutta	Vistula R.	PL	18	80	22	58	102	[65]	
S. trutta (lacustrine)	Wdzydze L.	PL	13	80	22	58	102	[65]	
S. trutta	Gawrych Ruda Hatchery	PL	21	80	22	58	102	[66]	
Black Sea Basin									
S. labrax	Local hatchery	GE	embryos	80	18	62	98	[67]	1
S. labrax	Local hatchery	GE	6	80	22	58	102	[68]	
S. trutta *	Black R.	GE	8	80–82	20–22	60	100–104	[68]	3
S. trutta *	Bzyb R.	GE	9	82	22	60	104	[68]	1
S. trutta *	Gumista R.	GE	9	82	22	60	104	[68]	1
S. trutta *	Kodori R.	GE	8	80–82	20–22	60	100–104	[68]	1; 2
S. trutta *	Bicaz, Prejmer, Azuga	RO	/	80	24	56	104	[69]	1; 2
S. trutta *	Western–Middle Carpathians	RO	/	80	24	56	104	[69]	1; 2
Caspian Sea Basin									
S. caspius			embryos	80	18	62	98	[70]	1

S. caspius	Kura R.	AZ	2	82	20	62	102	[68]	1; 2
S. ischchan "winter ischchan"	Sevan L.	AR	11	80	16	64	96	[67,71]	1; 2
S. ischchan "gegarkuni"	Sevan L.	AR	17	80	18	62	98	[67,71]	1; 2
S. ischchan "summer ischchan"	Sevan L.	AR	23	82	18	64	100	[67,71]	1; 2
S. ischchan "bodjak"	Sevan L.	AR	7	82	16	66	98	[67,71]	1; 2
S. trutta "alabalach" *	Argichi R.	AR	8	80	16	64	96	[72]	1
S. trutta	Marmarik R.	AR	/	82	16	66	98	[73]	1
S. trutta	Vedi R.	AR	/	78	20	58	98	[73]	1
S. trutta *	Azat R.	AR	8	78	20	58	98	[68]	1
S. trutta *	Arindg R.		7	80	18	62	98	[68]	1
S. trutta *	Vedi R.	AR	18	78	20	58	98	[68]	1
S. trutta *	Korotan R.		15	80	20	60	100	[68]	1
S. trutta *	Dzeoraget R.	AR	8	80	20	60	100	[68]	1
S. trutta *	Kcia R.		3	82	20	62	102	[68]	1
S. trutta *	Kyuretchai R.		9	84	16	68	100	[68]	1
S. trutta *	Marmarik R.	AR	7	82	16	66	98	[68]	1
S. trutta *	Ochtchi R.		8	82	20	62	102	[68]	1
S. trutta *	Chatchen R.		7	80	20	60	100	[68]	1
S. trutta *	Tchaki R.		8	82	18	64	100	[68]	1
S. trutta *	Goygol L.	AZ	7	80	20	60	100	[68]	1
S. trutta *	Tabackuri L.	GE	15	80	20	60	100	[68]	1
Northern Sea Basin, European Atlantic coast									
S. trutta	Cares R.	ES	49	80	22–23	57–58	102–103	[74]	1
S. trutta	Pyrenees hatchery	ES	44	81	22–24	57–59	103–105	[74]	1
S. trutta (anadromous)	Galicia	ES	14	80	20	60	100	[75]	4
S. trutta (local hatchery strain)	Galicia	ES	19	80	20	60	100	[75]	4
S. trutta	Pšovka Cr.	CZ	10	80	14/4	62	98	[76]	4
S. trutta	Navia, Tambre, Umia, Mino R.	ES	133	78–80	20	58–60	98–100	[77,78]	2; 4; 5
S. trutta	Galicia	ES	15	80	20	60	100	[79]	3
S. trutta	Hatchery stock AT lineage	IT	20	80	14/8	58	102	[55]	2; 5
S. trutta	Loch Lomond	SCT	6	79–80	21–22	58–59	100–102	[80]	3
S. trutta	Norway (migratory)	NO	/	80	14	66	94	[81]	3
S. trutta	Germany	DE	6	78–82	20–26	52–62	102–104	[82]	2
S. trutta	10 localities across all Sweden	SW	14	80	20	60	100	[83]	1

Notes: 1: Giemsa-stained chromosomes only; 2: Robertsonian polymorphism detected; 3: Replication banding pattern discovered cytotype variants in some chromosomes; 4: Ag-, CMA_3- and/or C-banding, cytotype polymorphisms; 5: Ag-, CMA_3- and rDNA ISH and/or FISH; *—Material was analysed under the name *S. trutta* but evidently out of the known autochthonous range of *S. trutta s. str.*, thus likely representing another species of the genus, the species name was determined based on locality data in a given study according to geographical distribution of trout taxa in Kottelat and Freyhof [4]; †—counts reported both trout taxa without distinguishing between them. Studies with incomplete information (without data reflecting karyotype composition, geographic origin, number of examined individuals or methodically problematic studies; all mostly reviewed in Gas [49]) were excluded from this review; symbol "-" in chromosome counts represents observed range, symbol "/"in chromosome counts shows that both categories (m and sm or st and a) were determined.

The karyotype of flathead trout undoubtedly belongs to category A sensu Phillips and Ráb [18], i.e., salmonids with 2n = ~80 and chromosome arm number NF = ~100. The 2n = 80 found in flathead trout has been reported in a majority of studies (e.g., [16,37,48]). Differences from this value are mostly caused by centric fusions of acrocentric chromosomes and/or fissions of metacentric chromosomes as reported in nearly all studies so far. Some reports of different 2n were caused by i) lower quality of metaphases examined and/or ii) low number of analysed individuals (e.g., Kaidanova [62], Karakousis et al. [56]). On the other hand, some studies pointed definitively to different 2n such as 2n = 82 in S. obtusirostris [60], 2n = 84 in S. trutta "alabalach" [72], 2n = 82 in some forms of S. ischchan [71] and 2n = 78 to 82 in some taxonomically unidentified Transcaucasian trout [68]. Regardless, such variation in 2n, frequently documented in other lineages of salmonids with A type karyotype, could be explained despite minor chromosome rearrangements such as pericentromeric inversions that can convert acrocentric chromosomes into sub-telocentric ones [18].

The 2n, karyotypes, and hence NF of the examined species/forms of brown trout are remarkably similar (Table 1). Nevertheless, differences caused by a chromosome classification bias among individual reports exist. Most authors categorise uni-armed and bi-armed chromosomes according to Levan [38] but NF was originally designed to quantify the centric translocations or fissions of the Robertsonian type [39] only. However, some authors scored sub-telocentric chromosomes as bi-armed. Differences in the NF reported for the same form/species thus usually resulted from a difference in the scoring rather than from any real variation. In other words, most of the studies provide the number of metacentric and submetacentric chromosomes together, while a minority of them distinguish these categories, as was done in our study. Another problem in comparing reports on karyotype structures in brown trout is that most of the summarized studies analysed Giemsa-stained chromosomes, published karyotypes and/or metaphase plates of lower quality to infer karyotype structure in more details. The studies using conventional and/or molecular cytogenetic protocols [55,64–66,74–80] revealed very similar or even identical karyotypes as we found in the flathead trout for this study. We can therefore conclude that the karyotype of brown trout typically consist of seven pairs of metacentric, five to six pairs of visibly sub-telocentric chromosomes and all remaining are acrocentric elements of gradually decreasing size. The brown trout's karyotype also contains several distinct chromosome markers—the first two pairs of acrocentric chromosomes distinctly larger from other acrocentric ones and the largest sub-telocentric pair, which is also the largest one in the complement. The short (p) arm of this marker chromosome pair bears the major rDNA sites, as revealed by FISH with 28S rDNA probe, corresponding to positive Ag- and CMA_3-stainings [55,64–66,75–78,84]. Intraspecific variation in the locations and sizes of the chromosomal nucleolar organizer regions (NORs), i.e., major rDNA sites, have been frequently documented [85] but available data for this marker consistently document the same karyotype location across brown trout diversity including flathead trout. However, as in other cases, some intraspecific variability has been observed [64–66]. In our study, we observed the variability in the size of the NOR-bearing p arm of this marker chromosome corresponding to the 28S rDNA signal, similar to Caputo et al. [28] in S. marmoratus. The intraspecific variability of the 5S rDNA cytotaxonomic marker is quite well known [85–87]. Surprisingly, the location of 5S rDNA genes in the genus Salmo was examined in two studies only. Pendás et al. [84] found multichromosomal sites of these genes in brown trout from northwestern Spain, while Caputo et al. [55] observed these sites in telomeres of one middle-sized metacentric pair only. Our study also detected this gene cluster in the pericentromeric region of one middle-sized metacentric pair only. Whether this 5S rDNA bearing chromosome pair is homologous remains to be demonstrated by a cross-species painting protocol (e.g., Ráb et al. [88]). We can therefore conclude that the 2n, and structure as well as number and position of NORs, i.e., the active 28S rDNA sites, of the endemic flathead trout karyotype entirely correspond to those found in other brown trout taxa

4.1. Cytotaxonomy and Diversity of Eurasian Trouts

The species *S. trutta* has long been considered a single but highly polymorphic species broadly distributed in the European ichthyo-geographic region (see Bañarescu [42]) forming three ecotypes—

marine migratory, lacustrine and brook/riverine [4]. In line with this, several subspecies or even distinct species have been described but most of them are simply considered as interindividual and/or interpopulation variability. Though even nominal subgenera of the genus *Salmo* have been described, i.e., *Acantholingua* (for *A. ohridanus*), *Salmothymus* (for *S. obtusirostris*) and *Platysalmo* (for *P. platycephalus*), collectively called ´archaic trout´, they are closely related to the *S. trutta* species complex at the molecular level (e.g., Sušnik et al. [89], Phillips et al. [90]). However, recent detailed investigations of brown trout life histories, biology, distribution and taxonomy suggest that the biological and hence taxonomic diversity of the Eurasian genus *Salmo* is considerably greater than the taxonomy that was accepted up until the 1990s would suggest [4,91], a situation similar for freshwater trout of the genus *Oncorhynchus* [92]. Recently, FishBase [93] lists 50 formally described *Salmo* species. However, many molecular phylogeneticists and phylogeographers question this biological species concept of taxonomic diversity of the genus *Salmo* by pointing to negligible and/or weak genetic differentiation among some of those populations/taxa (to cite from numerous ones e.g., [13,89,94–99]). Yet, other colleagues detected significantly larger genetic differences (e.g., [100–104]). How can the cytotaxonomy of the genus *Salmo* contribute to this debate? Our results of the cytogenetic analysis of flathead trout compared with available cytogenetic data for other trout populations and/or taxa (Table 1) clearly demonstrate that 2n, karyotype structures and other chromosomal markers, especially the position of major rDNA sites, are rather stable or even invariable across trout diversity as described for several lineages of salmonid fishes with A type karyotypes [18,105]. At first glance, this conclusion would support/conform to the view of molecular-based studies. However, the stability of 2n and similar and/or even identical chromosomal characteristics were observed, i.e., karyotype stasis is widely documented in a group of taxonomically different species and/or even lineages. Such uniform stasis has been discovered in groups as diverse as plants [106–108], amphibians [109,110] and birds [111]. Among teleost fishes, multiple groups display such apparent karyotype stasis persisting in significantly long stages of lineage divergences, e.g., pikes of the genus *Esox* [112,113], fishes of the family Leuciscidae ([88,114–117] and references therein), Gobionidae [34], Xenocyprinidae [118] and especially many percomorph groups [119–124]. The underlying evolutionary mechanisms for this mode of karyotype (non) differentiation have not been identified so far but they may be at least partially linked with the functional arrangement of chromatin within the interphase nucleus and the degree of tolerance to its change [125,126]. We therefore conclude that, from the cytotaxonomic point of view, apparent karyotype stasis found in trout of the genus *Salmo* does not challenge their existing and evident taxonomic diversity.

4.2. Cytogenomics in Salmonids

The cytogenomic approach represents a logical continuation of the traditional molecular cytogenetics, which was crucial for understanding fish genome evolution. Cytogenomics effectively integrates the huge body of evidence generated by karyological and cytogenetic research with the genomic approach based on currently extensive genome sequencing [127]. The sequencing effort of fish genomes is still accelerating and highly ambitious; hence, with about 32,000 fish species [128], fish cytogenomics has a good chance of fast becoming as equally crucial as molecular cytogenetics despite the small fraction of genomes that had been sequenced so far in comparison with the number of species already analysed cytogenetically. Already, virtual karyotyping has taken another step forward with the potential to visualize more details with better resolutions than through the use of microscopes for most small-sized fish chromosomes.

At this initial stage, our tool for virtual karyotyping utilizes masking of repeats in the DNA sequence via soft-masking, i.e., identified repetitive sequences become lower-case, whereas the remaining sequences retain their upper-case. It means that the quality of the input assembly and its soft-masking is crucial and cannot be influenced by the tool itself. This tool has been introduced in this special issue to outline the potential future of fish cytogenomics, and so far its functionality has been utilized to address general questions on GC% and repeats evolution not only in fish but also across vertebrates. This means that the tool has not yet been used systematically in cytogenetically analysed fish species and results of both approaches have not yet been compared.

The inverse relationship between the GC% and genome size had been initially ascribed to the extremely dynamic and often highly amplified ribosomal genes [105,129–131] that represent the GC-richest genome fraction [132,133]. However, further molecular cytogenetic results based on FISH with rDNA probes in further salmonids (continuously summarized by the database by Sochorová et al. [134]) as well as the results obtained here do not support these initial assumptions. The genomic approach is less useful here, because the rDNA is mostly disregarded and/or even discarded in the genome assemblies.

Author Contributions: Conceptualization, R.S. and P.R.; Methodology, R.S. and Š.P.; Formal analysis, Z.M.; resources, M.G. and Z.L.; Writing—original draft preparation, P.R.; Writing—review and editing, M.G., Z.M., R.S., S.U., P.R.; Visualization, Z.M.; Project administration, P.R.; Funding acquisition, Z.M., R.S. and P.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Czech Academy of Sciences, the Project EXCELLENCE CZ.02.1.01/0.0/0.0/15_003/0000460 OP RDE, RVO: 67985904 and PPLZ L200451951. This project has received funding from the European Union´s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 754462 (R.S.).

Acknowledgments: The authors would like to express their gratitude to Dominik Matoulek for production of the *S. trutta* prototypical virtual karyotype and to Pei Chen King for language correction.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Delling, B.; Doario, I. Systematics of the trouts endemic to Moroccan lakes, with description of a new species (Teleostei: Salmonidae). *Ichthyol. Explor. Freshw.***2005**, *16*, 49–64.
2. Doadrio, I.; Perea, S.; Yahyaoui, A. Two new species of atlantic trout (Actynopterygii, Salmonidae) from Morocco. *Graellsia* **2015**, *71*, e031.
3. Doadrio, I. *Ictiofauna Continental Española: Bases Para Su Seguimiento*; Ministerio de Medio Ambiente y Medio Rural y Marino, Centro de Publicaciones: Madrid, Spain; 2011; ISBN 84-491-1158-7.
4. Kottelat, M.; Freyhof, J. *Handbook of European Freshwater Fishes*; Publications Kottelat: Cornol, Switzerland; 2007; ISBN 2-8399-0298-2.
5. Delling, B. Diversity of western and southern Balkan trouts, with the description of a new species from the Louros River, Greece (Teleostei: Salmonidae). *Ichthyol. Explor. Freshw.* **2010**, *21*, 331.
6. Turan, D.; Kottelat, M.; Engin, S. Two new species of trouts, resident and migratory, sympatric in streams of northern Anatolia (Salmoniformes: Salmonidae). *Ichthyol. Explor. Freshw.* **2009**, *20*, 333–364.
7. Turan, D.; Kottelat, M.; Bektas, Y. *Salmo tigridis*, a new species of trout from the Tigris River, Turkey (Teleostei: Salmonidae). *Zootaxa* **2011**, *2993*, 23–33.
8. Turan, D.; Kottelat, M.; Engin, S. The trouts of the Mediterranean drainages of southern Anatolia, Turkey, with description of three new species (Teleostei: Salmonidae). *Ichthyol. Explor. Freshw.* **2012**, *23*, 219.
9. Turan, D.; Doğan, E.; Kaya, C.; Kanyılmaz, M. *Salmo kottelati*, a new species of trout from Alakır Stream, draining to the Mediterranean in southern Anatolia, Turkey (Teleostei, Salmonidae). *ZooKeys* **2014**, 135–151, doi: 10.3897/zookeys.462.8177.
10. Turan, D.; Kottelat, M.; Kaya, C. *Salmo munzuricus*, a new species of trout from the Euphrates River drainage, Turkey (Teleostei: Salmonidae). *Ichthyol. Explor. Freshw.* **2017**, *28*, 55–63.
11. Çiçek, E.; Fricke, R.; Sungur, S.; Eagderi, S. Endemic freshwater fishes of Turkey. *FishTaxa* **2018**, *3*, 1–39.
12. Lobón-Cerviá, J.; Sanz, N. *Brown trout: Biology, ecology and management*; John Wiley & Sons, 2017; ISBN 1-119-26831-1.
13. Lobón-Cerviá, J.; Esteve, M.; Berrebi, P.; Duchi, A.; Lorenzoni, M.; Young, K.A. Trout and char of central and Southern Europe and Northern Africa. *Trout and char of the world. Bethesda, Maryland: American Fisheries Society* **2019**.
14. Behnke, R.J. A new subgenus and species of trout, *Salmo* (*Platysalmo*) *platycephalus*, from southcentral Turkey, with comments on the classification of the subfamily Salmoninae. *Mitt. Hamburg. Zool. Mus. Inst.* **1968**, *66*, 1–15.

15. Kara, C.; Alp, A.; Gürlek, M.E. Morphological variations of the trouts (*Salmo trutta* and *Salmo platycephalus*) in the rivers of Ceyhan, Seyhan and Euphrates, Turkey. *Turk. J. Fish. Aquat. Sci.* **2011**, *11*, 77–85.
16. Kara, C.; Alp, A.; Can, M.F. Growth and reproductive properties of flathead trout (*Salmo platycephalus* Bhenke, 1968) population from Zamantı Stream, Seyhan River, Turkey. *Turk. J. Fish. Aquat. Sci* **2011**, *11*, 367–375.
17. Tarkan, A.N.; Tarkan, A.S.; Bilge, G.; Gaygusuz, Ö.; Gürsoy, Ç. Threatened fishes of the world: *Salmo platycephalus* Behnke, 1968 (Salmonidae). *Environ. Biol. Fishes* **2008**, *81*, 371–372.
18. Phillips, R.; Ráb, P. Chromosome evolution in the Salmonidae (Pisces): An update. *Biol. Rev. Camb. Philos. Soc.* **2001**, *76*, 1–25.
19. Mazzuchelli, J.; Kocher, T.D.; Yang, F.; Martins, C. Integrating cytogenetics and genomics in comparative evolutionary studies of cichlid fish. *BMC Genom.* **2012**, *13*, 463, doi:10.1186/1471-2164-13-463.
20. De Oliveira, E.A.; Bertollo, L.A.C.; Ráb, P.; Ezaz, T.; Yano, C.F.; Hatanaka, T.; Jegede, O.I.; Tanomtong, A.; Liehr, T.; Sember, A.; et al. Cytogenetics, genomics and biodiversity of the South American and African Arapaimidae fish family (Teleostei, Osteoglossiformes). *PLoS ONE* **2019**, *14*, e0214225, doi:10.1371/journal.pone.0214225.
21. Pennell, W.; Prouzet, K. Salmonid fish: Bology, conservation status, and economic importance of wild and cultured stocks. *Fish. Aquac. P Safran Enciclopedia Life Suport Syst.* **2009**, *3*, 42–65.
22. Gregory, T.R. Animal Genome Size Database. 2020. Available online: http://www. genomesize.com
23. Koop, B.F.; Davidson, W.S. Genomics and the genome duplication in Salmonids. In *Fisheries for Global Welfare and Environment*; Tsukamoto K., Kawamura R., Takeuchi T., Beard T.D. Jr, Kaiser M.J., Eds; 5th World Fisheries Congress: Tokyo, Japan; 2008; p. 77–86.
24. Davidson, W.S.; Koop, B.F.; Jones, S.J.; Iturra, P.; Vidal, R.; Maass, A.; Jonassen, I.; Lien, S.; Omholt, S.W. Sequencing the genome of the Atlantic salmon (*Salmo salar*). *Genome Biol.* **2010**, *11*, 403, doi:10.1186/gb-2010-11-9-403.
25. Hunt, S.E.; McLaren, W.; Gil, L.; Thormann, A.; Schuilenburg, H.; Sheppard, D.; Parton, A.; Armean, I.M.; Trevanion, S.J.; Flicek, P.; et al. Ensembl variation resources. *Database* **2018**, *2018*, doi:10.1093/database/bay119.
26. Matoulek, D.; Borůvková, V.; Ocalewicz, K.; Symonová, R. GC and repeats profiling along chromosomes – the future of fish compositional cytogenomics. Genes 2020, 11; this special issue.
27. Canapa, A.; Barucca, M.; Biscotti, M.A.; Forconi, M.; Olmo, E. Transposons, genome size, and evolutionary insights in animals. *Cytogenet. Genome Res.* **2015**, *147*, 217–239, doi:10.1159/000444429.
28. Lien, S.; Koop, B.F.; Sandve, S.R.; Miller, J.R.; Kent, M.P.; Nome, T.; Hvidsten, T.R.; Leong, J.S.; Minkley, D.R.; Zimin, A.; et al. The Atlantic salmon genome provides insights into rediploidization. *Nature* **2016**, *533*, 200–205, doi:10.1038/nature17164.
29. Symonová, R.; Suh, A. Nucleotide composition of transposable elements likely contributes to AT/GC compositional homogeneity of teleost fish genomes. *Mobile DNA* **2019**, *10*, 49, doi:10.1186/s13100-019-0195-y.
30. Borůvková, V.; Howell, W.; Matoulek, D.; Symonová, R. Quantitative approach to fish cytogenetics in the context of vertebrate genome evolution. Genes 2020, 11; this special issue.
31. Ráb, P.; Roth, P. Cold-blooded vertebrates. In *Methods of Chromosome Analysis*; Balíček, P., Forejt, J., Rubeš, J., Eds.; Cytogenetic Section of Czechoslovakian Biological Society Publishers: Brno, Czech Republic, 1988; pp. 115–124.
32. Sola, L.; Rossi, A.R.; Iaselli, V.; Rasch, E.M.; Monaco, P.J. Cytogenetics of bisexual/unisexual species of *Poecilia*. II. Analysis of heterochromatin and nucleolar organizer regions in *Poecilia mexicana mexicana* by C-banding and DAPI, quinacrine, chromomycin A3, and silver staining. *Cytogen. Cell Genet.* **1992**, *60*, 229–235.
33. Howell, W.M.; Black, D.A. Controlled silver-staining of nucleolus organizer regions with a protective colloidal developer: A 1-step method. *Experientia* **1980**, *36*, 1014–1015, doi:10.1007/BF01953855.
34. Rábová, M.; Völker, M.; Pelikánová, Š.; Ráb, P. Sequential chromosome banding in fishes. In *Fish Cytogenetic Techniques*; Ozouf-Costaz, C., Pisano, E., Foresti, F., de Almeida Toledo, L.F., Eds.; CRC Press: Boca Raton, FL, USA, 2015; pp. 92–102. ISBN 978-1-4822-1198-6.
35. Komiya, H.; Takemura, S. Nucleotide sequence of 5S ribosomal RNA from rainbow trout (*Salmo gairdnerii*) liver. *J. Biochem.* **1979**, *86*, 1067–1080.

36. White, T.J.; Bruns, T.; Lee, S.; Taylor, J. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In *PCR Protocols: A Guide to Methods and Applications*; Innis, M., Gelfand, J., Sninsky, J., White, T., Eds.; Academic Press: Orlando, FL, USA, 1990; pp. 315–322.
37. Symonová, R.; Sember, A.; Majtánová, Z.; Ráb, P. Characterization of fish genomes by GISH and CGH. In *Fish Cytogenetic Techniques*; Ozouf-Costaz, C., Pisano, E., Foresti, F., de Almeida, L., Eds.; CRC Press: Boca Raton, FL, USA, 2015; pp. 118–131. ISBN 978-1-4822-1198-6.
38. Levan, A.; Fredga, K.; Sandberg, A.A. Nomenclature for centromeric position on chromosomes. *Hereditas* **1964**, *52*, 201–220, doi:10.1111/j.1601-5223.1964.tb01953.x.
39. Matthey, R. L'évolution de la formule chromosomiale chez les vertébrés. *Experientia* **1945**, *1*, 78–86.
40. Team, R.C. *R: A Language and Environment for Statistical Computing*, version 2.6.2; R Foundation for Statistical Computing: Vienna, Austria, 2013.
41. Sušnik, S.; Schöffmann, J.; Snoj, A. Phylogenetic position of *Salmo* (*Platysalmo*) *platycephalus* Behnke 1968 from south-central Turkey, evidenced by genetic data. *J. Fish. Biol.* **2004**, *64*, 947–960.
42. Bănărescu, P. *Zoogeography of Fresh Waters. Volume 2: Distribution and Dispersal of Freshwater Animals in North America and Eurasia*; Aula-Verlag: Wiebelsheim, Germany, 1991; ISBN 3-89104-482-8.
43. Budy, P.; Gaeta, J.W. Brown trout as an invader: A Synthesis of problems and perspectives in North America. In *Brown trout: Biology, ecology, and management*; John Wiley and Sons Ltd.: Hoboken, NJ, USA, 2018; 525–534.
44. Jones, P.; Closs, G. The introduction of brown trout to New Zealand and their impact on native fish communities. *In Brown Trout: Biology, Ecology, and Management*; John Wiley and Sons Ltd.: Hoboken, NJ, USA, 2018; pp. 545–567.
45. Jellyman, P.G.; McHugh, P.A.; Simon, K.S.; Thompson, R.M.; McIntosh, A.R. The effects of brown trout on the trophic webs of New Zealand streams. In *Brown Trout: Biology, Ecology and Management*; John Wiley and Sons Ltd.: Hoboken, NJ, USA, 2017; pp. 569–598.
46. Casalinuovo, M.A.; Alonso, M.F.; Macchi, P.J.; Kuroda, J.A. Brown trout in Argentina: History, interactions, and perspectives. *Brown Trout: Biology, Ecology, and Management*; John Wiley and Sons Ltd.: Hoboken, NJ, USA, 2018; pp. 599–622.
47. Weyl, O.L.; Ellender, B.R.; Ivey, P.; Jackson, M.C.; Tweddle, D.; Wasserman, R.J.; Woodford, D.J.; Zengeya, T.A. Africa: Brown trout introductions, establishment, current status, impacts and conflicts. *Brown trout: Biology, ecology, and management*; John Wiley and Sons Ltd.: Hoboken, NJ, USA, 2018; 623–639.
48. FAO Fisheries and Aquaculture—Statistics—Introduction. Available online: http://www.fao.org/fishery/statistics/en (accessed on 6 November 2020).
49. GAs, M. Revue bibliographique sur la caryologie des Téléostéens. Etude critique des méthodes employées et des résultats obtenus. *Biologie Médical* **1970**, *54*, 54–81.
50. Dettai, A.; Pruvost, P. Storage of Karyotyped Voucher Specimens and their Molecular Identification. In *Fish Cytogenetic Techniques: Ray-Fin Fishes and Chondrichthyans*; CRC Press Taylor and Francis Group: London, UK, 2015; p. 11.
51. Kohout, J.; Jašková, I.; Papoušek, I.; Šedivá, A.; Šlechta, V. Effects of stocking on the genetic structure of brown trout, *Salmo trutta*, in Central Europe inferred from mitochondrial and nuclear DNA markers. *Fish Manag. Ecol.* **2012**, *19*, 252–263.
52. Leitwein, M.; Gagnaire, P.A.; Desmarais, E.; Guendouz, S.; Rohmer, M.; Berrebi, P.; Guinand, B. Genome-wide nucleotide diversity of hatchery-reared Atlantic and Mediterranean strains of brown trout *Salmo trutta* compared to wild M editerranean populations. *J. Fish. Biol.* **2016**, *89*, 2717–2734.
53. Mazik, E.J.; Toktosunov, A.T. Karyotype of the Amu Darya trout *Salmo trutta oxianus* (Salmoniformes, Salmonidae) from Kyzylsu River. *J. Zool.* **1986**, *65*, 1582–1586.
54. Merlo, S. Osservazioni cariologiche su *Salmo carpio*: (Con 1 tavola fuori testo). *Ital. J. Zool.* **1957**, *24*, 253–258.
55. Caputo, V.; Giovannotti, M.; Cerioni, P.N.; Splendiani, A.; Olmo, E. Chromosomal study of native and hatchery trouts from Italy (*Salmo trutta* complex, Salmonidae): Conventional and FISH analysis. *Cytogen. Gen. Res.* **2009**, *124*, 51–62.
56. Karakousis, Y.; Paschos, J.; Triantaphyllidis, C. Chromosomal studies in brown trout (*Salmo trutta* L.) populations. *Cytobios* **1992**, *72*, 117–124.
57. Dimovska, A. Chromosome complement of Ochrid trout (*Salmo letnica* Karaman). *Godisen zb. Prirodno-matem. Fak. Univ. Skopje* **1959**, *12*, 115–135.

58. Al-Sabti, K. Karyotypical studies in three Salmonidaae in Slovenia using leukocyte culture technique. *Ichthyologia* **1983**, *15*, 41–46.
59. Al-Sabti, K. Chromosomal studies by blood leukocyte culture technique on three salmonids from Yugoslavian waters. *J. Fish. Biol.* **1985**, *26*, 5–12, doi:10.1111/j.1095-8649.1985.tb04234.x.
60. Berberovic, L.; Curic, M.; Hadziselimovic, R.; Sofradzija, A. Chromosome complement of *Salmothymus obtusirostris oxyrhynchus* (Steindachner). *Acta Biol. Jug. Genet.* **1970**, *2*, 55–63.
61. Sofradzija, A. The chromosomes of the trout *Salmo trutta m. fario* and *Salmo gairdneri*. *Godisnajak Bioloskog Instituta Univerziteta u Sarajevu* **1982**, *35*, 117–128.
62. Kaidanova, T.I. Karyotype study of brown trout *Salmo trutta morpha fario* from Ropsha population. *Voprosy Ichtiologii* **1975**, *15*, 1124–1128.
63. Jankun, M. Standard karyotype of sea trout (*Salmo trutta morpha trutta*) based on replication banding patterns. *Cytobios* **2000**, *103*, 79–89.
64. Woznicki, P.; Jankun, M.; Kucharczyk, D.; Boron, A.; Luczynski, M. Cytogenetic characterization of sea trout (*Salmo trutta*) from Poland. *Copeia* **1999**, *1999*, 501–505.
65. Woznicki, P.; Sanchez, L.; Martinez, P.; Pardo, B.G.; Jankun, M. A population analysis of the structure and variability of NOR in *Salmo trutta* by Ag, CMA3 and ISH. *Genetica* **2000**, *108*, 113–118.
66. Woznicki, P.; Jankun, M.; Luczynski, M. Chromosome studies in brown trout (*Salmo trutta m. fario*) from Poland: Hypothetical evolution of the 11th, 12th and 14th chromosome pairs in the Salmo karyotype. *Cytobios* **1997**, *91*, 2017–214.
67. Dorofeeva, E.A.; Ruhkjan, R.G. Divergence of *Salmo ischchan* Kessler in light of karyological and morphological characteristics. *Voprosy Ichtiologii* **1982**, *22*, 36–48.
68. Ruhkjan, R.G. *Karyology and Origin of the Transcaucasian Trouts*; Academy of Sciences Armenian SSR Press: Yerevan, Armenia, 1989; p. 166. (In Russian)
69. Raicu, P.; Taisescu, E. Cytogenetic study in *Salmo irideus* and *S. trutta fario*. *Cytologia* **1977**, *42*, 311–314.
70. Dorofeeva, E.A. Karyology and systematic status of Caspian and Black Sea salmons (*Salmo trutta caspius* Kessler, *Salmo trutta labrax* Pallas). *Voprosy Ichtiologii* **1965**, *5*, 28–45.
71. Ruhkjan, R.G. A comparative analysis of the karyotypes of the Sevan trout *Salmo ischchan* Kessler. *Citologia* **1982**, *24*, 66–77.
72. Ruhkjan, R.G. On the origin and species identity of alabalach trout (genus *Salmo*, Salmonidae) based on its karyological characteristics). *Voprosy Ichtiologi* **1984**, *23*, 368–373.
73. Ruhkjan, R.G. Karyotypes of brown trouts of Armenia *Salmo trutta m. fario*. *Biol. J. Armenii* **1981**, *34*, 412–417.
74. Mořan, P.; Pendas, A.M.; García-Vázquez, E.; Linde, A.R. Chromosomal and morphological analysis of two populations of *Salmo trutta sbp. fario* employed in repopulation. *J. Fish Biol.***1989**, *35*, 839–843.
75. Martinez, P.; Vinas, A.; Bouza, C.; Arias, J.; Amaro, R.; Sánchez, L. Cytogenetical characterization of hatchery stocks and natural populations of sea and brown trout from northwestern Spain. *Heredity* **1991**, *66*, 9–17.
76. Mayr, B.; Ráb, P.; Kalat, M. Localisation of NORs and counterstain-enhanced fluorescence studies in *Salmo gairdneri* and *Salmo trutta* (Pisces, Salmonidae). *Theoret. Appl. Genet.* **1986**, *71*, 703–707, doi:10.1007/BF00263267.
77. Castro, J.; Rodríguez, S.; Arias, J.; Sánchez, L.; Martínez, P. A population analysis of Robertsonian and Ag-NOR polymorphisms in brown trout (*Salmo trutta*). *Theoret. Appl. Genetics* **1994**, *89*, 105–111, doi:10.1007/BF00226990.
78. Castro, J.; Viñas, A.; Sánchez, L.; Martínez, P. Characterization of an atypical NOR site polymorphism in brown trout (*Salmo trutta*) with Ag- and CMA3-staining, and fluorescent in situ hybridization. *Cytogenet. Genome Res.* **1996**, *75*, 234–239, doi:10.1159/000134491.
79. Sánchez, L.; Martínez, P.; Bouza, C.; Viñas, A. Chromosomal heterochromatin differentiation in *Salmo trutta* with restriction enzymes. *Heredity* **1991**, *66*, 241–249, doi:10.1038/hdy.1991.30.
80. Hartley, S.E.; Horne, M.T. Chromosome relationships in the genus *Salmo*. *Chromosoma* **1984**, *90*, 229–237, doi:10.1007/BF00292401.
81. Gjedrem, T.; Eggum, Å.; Refstie, T. Chromosomes of some salmonids and salmonid hybrids. *Aquaculture* **1977**, *11*, 335–348, doi:10.1016/0044-8486(77)90083-7.
82. Zenzes, M.T.; Voiculescu, I. C-banding patterns in *Salmo trutta*, a species of tetraploid origin. *Genetica* **1975**, *45*, 531–536, doi:10.1007/BF01772875.

83. Nygren, A.; Nilsson, B.; Jahnke, M. Cytological studies in *Salmo trutta* and *Salmo alpinus*. *Hereditas* **1971**, *67*, 259–267, doi:10.1111/j.1601-5223.1971.tb02378.x.
84. Pendás, A.M.; Morán, P.; García-Vázquez, E. Multi-chromosomal location of ribosomal RNA genes and heterochromatin association in brown trout. *Chromosome Res.* **1993**, *1*, 63–67, doi:10.1007/BF00710608.
85. Gornung, E. Twenty Years of Physical Mapping of Major Ribosomal RNA Genes across the Teleosts: A Review of Research. *Cytogenet. Genome Res.* **2013**, doi:10.1159/000354832.
86. Sember, A.; Bohlen, J.; Šlechtová, V.; Altmanová, M.; Symonová, R.; Ráb, P. Karyotype differentiation in 19 species of river loach fishes (Nemacheilidae, Teleostei): Extensive variability associated with rDNA and heterochromatin distribution and its phylogenetic and ecological interpretation. *BMC Evol. Biol.* **2015**, *15*, 251, doi:10.1186/s12862-015-0532-9.
87. Majtánová, Z.; Unmack, P.J.; Prasongmaneerut, T.; Shams, F.; Srikulnath, K.; Ráb, P.; Ezaz, T. Evidence of Interspecific Chromosomal Diversification in Rainbowfishes (Melanotaeniidae, Teleostei). *Genes* **2020**, *11*, 818, doi:10.3390/genes11070818.
88. Ráb, P.; Rábová, M.; Pereira, C.S.; Collares-Pereira, M.J.; Pelikánová, Š. Chromosome studies of European cyprinid fishes: Interspecific homology of leuciscine cytotaxonomic marker—the largest subtelocentric chromosome pair as revealed by cross-species painting. *Chromosome Res.* **2008**, *16*, 863–873, doi:10.1007/s10577-008-1245-3.
89. Sušnik, S.; Snoj, A.; Wilson, I.F.; Mrdak, D.; Weiss, S. Historical demography of brown trout (*Salmo trutta*) in the Adriatic drainage including the putative *S. letnica* endemic to Lake Ohrid. *Mol. Phylogen. Evol.* **2007**, *44*, 63–76, doi:10.1016/j.ympev.2006.08.021.
90. Phillips, R.B.; Matsuoka, M.P.; Konon, I.; Reed, K.M.; McEachran, M. Phylogenetic Analysis of Mitochondrial and Nuclear Sequences Supports Inclusion of Acantholingua ohridana in the Genus *Salmo*. *Copeia* **2000**, *2000*, 546–550, doi:10.1643/0045-8511(2000)000[0546:PAOMAN]2.0.CO;2.
91. Kottelat, M. Freshwater fishes of western and central Europe. *Biologia* **1997**, *52*, 1–271.
92. Behnke, R.J. Native trout of western North America. *Am. Fish. Soc. Monogr. USA* **1992**, *6*, 233-256.
93. FishBase. A Global Information System on Fishes Available online: https://www.fishbase.se/home.htm (accessed on 6 November 2020).
94. Cortey, M.; Pla, C.; García-Marín, J.-L. Historical biogeography of Mediterranean trout. *Mol. Phylogenetics Evol.* **2004**, *33*, 831–844, doi:10.1016/j.ympev.2004.08.012.
95. Marić, S.; Sušnik, S.; Simonović, P.; Snoj, A. Phylogeographic study of brown trout from Serbia, based on mitochondrial DNA control region analysis. *Genet. Sel. Evol.* **2006**, *38*, 411, doi:10.1186/1297-9686-38-4-411.
96. McKeown, N.J.; Hynes, R.A.; Duguid, R.A.; Ferguson, A.; Prodöhl, P.A. Phylogeographic structure of brown trout *Salmo trutta* in Britain and Ireland: Glacial refugia, postglacial colonization and origins of sympatric populations. *J. Fish. Biol.* **2010**, *76*, 319–347, doi:10.1111/j.1095-8649.2009.02490.x.
97. Snoj, A.; Marić, S.; Bajec, S.S.; Berrebi, P.; Janjani, S.; Schöffmann, J. Phylogeographic structure and demographic patterns of brown trout in North-West Africa. *Mol. Phylogenetics Evol.* **2011**, *61*, 203–211, doi:10.1016/j.ympev.2011.05.011.
98. Ninua, L.; Tarkhnishvili, D.; Gvazava, E. Phylogeography and taxonomic status of trout and salmon from the Ponto-Caspian drainages, with inferences on European Brown Trout evolution and taxonomy. *Ecol. Evol.* **2018**, *8*, 2645–2658, doi:10.1002/ece3.3884.
99. Tougard, C.; Justy, F.; Guinand, B.; Douzery, E.J.P.; Berrebi, P. Salmo macrostigma (Teleostei, Salmonidae): Nothing more than a brown trout (*S. trutta*) lineage? *J. Fish. Biol.* **2018**, *93*, 302–310, doi:10.1111/jfb.13751.
100. Schöffmann, J.; Sušnik, S.; Snoj, A. Phylogenetic origin of *Salmo trutta* L. 1758 from Sicily, based on mitochondrial and nuclear DNA analyses. *Hydrobiologia* **2007**, *575*, 51–55, doi:10.1007/s10750-006-0281-2.
101. Apostolidis, A.P.; Loukovitis, D.; Tsigenopoulos, C.S. Genetic characterization of brown trout (*Salmo trutta*) populations from the Southern Balkans using mtDNA sequencing and RFLP analysis. *Hydrobiologia* **2008**, *600*, 169–176, doi:10.1007/s10750-007-9229-4.
102. Gratton, P.; Allegrucci, G.; Gandolfi, A.; Sbordoni, V. Genetic differentiation and hybridization in two naturally occurring sympatric trout *Salmo* spp. forms from a small karstic lake. *J. Fish. Biol.* **2013**, *82*, 637–657, doi:10.1111/jfb.12022.
103. Berrebi, P.; Retif, X.; Bouhbouh, S. Genetic evidence of unisexual reproduction in the Moroccan hexaploid barbel *Labeobarbus fritschi*. *Folia Zoo.* **2013**, *62*, 257–263, doi:10.25225/fozo.v62.i4.a2.2013.

104. Delling, B.; Sabatini, A.; Muracciole, S.; Tougard, C.; Berrebi, P. Morphologic and genetic characterisation of Corsican and Sardinian trout with comments on *Salmo* taxonomy. *Knowl. Manag. Aquat. Ecosyst.* **2020**, 21, doi:10.1051/kmae/2020013.
105. Dion-Côté, A.-M.; Symonová, R.; Lamaze, F.C.; Pelikánová, Š.; Ráb, P.; Bernatchez, L. Standing chromosomal variation in Lake Whitefish species pairs: The role of historical contingency and relevance for speciation. *Mol. Ecol.* **2017**, *26*, 178–192, doi:10.1111/mec.13816.
106. Mandáková, T.; Heenan, P.B.; Lysak, M.A. Island species radiation and karyotypic stasis in *Pachycladon* allopolyploids. *BMC Evol. Biol.* **2010**, *10*, 367, doi:10.1186/1471-2148-10-367.
107. Bomfleur, B.; McLoughlin, S.; Vajda, V. Fossilized nuclei and chromosomes reveal 180 million years of genomic stasis in royal ferns. *Science* **2014**, *343*, 1376–1377, doi:10.1126/science.1249884.
108. Samad, N.A.; Dagher-Kharrat, M.B.; Hidalgo, O.; Zein, R.E.; Douaihy, B.; Siljak-Yakovlev, S. Unlocking the karyological and cytogenetic diversity of *Iris* from Lebanon: Oncocyclus section shows a distinctive profile and relative stasis during its continental radiation. *PLoS ONE* **2016**, *11*, e0160816, doi:10.1371/journal.pone.0160816.
109. Sessions, S.K.; Kezer, J. Chapter 5—Evolutionary cytogenetics of *Bolitoglossine* Salamanders (Family Plethodontidae). In *Amphibian Cytogenetics and Evolution*; Green, D.M., Sessions, S.K., Eds.; Academic Press: San Diego, CA, USA, 1991; pp. 89–130. ISBN 978-0-12-297880-7.
110. Odierna, G.; Aprea, G.; Capriglion, T.; Castellano, S.; Balletto, E. Evidence for chromosome and Pst I satellite DNA family evolutionary stasis in the *Bufo viridis* group (Amphibia, Anura). *Chromosome Res.* **2004**, *12*, 671–681, doi:10.1023/B:CHRO.0000045746.59805.58.
111. Ellegren, H. Evolutionary stasis: The stable chromosomes of birds. *Trends Ecol. Evol.* **2010**, *25*, 283–291, doi:10.1016/j.tree.2009.12.004.
112. Ráb, P.; Crossman, E.J. Chromosomal NOR phenotypes in North American pikes and pickerels, genus *Esox*, with notes on the *Umbridae* (Euteleostei: Esocae). *Canadian J. Zool.* **2011**, doi:10.1139/z94-265.
113. Symonová, R.; Ocalewicz, K.; Kirtiklis, L.; Delmastro, G.B.; Pelikánová, Š.; Garcia, S.; Kovařík, A. Higher-order organisation of extremely amplified, potentially functional and massively methylated 5S rDNA in European pikes (*Esox* sp.). *BMC Genomics* **2017**, *18*, 391, doi:10.1186/s12864-017-3774-7.
114. Ráb, P.; Collares-Pereira, M.J. Chromosomes of European cyprinid fishes (Cyprinidae, Cypriniformes) Review. *Folia Zoologica* **1995**, *44*, 193–214.
115. Bianco, P.G.; Aprea, G.; Balletto, E.; Capriglione, T.; Fulgione, D.; Odierna, G. The karyology of the cyprinid genera *Scardinius* and *Rutilus* in southern Europe. *Ichthyol. Res.* **2004**, *51*, 274–278, doi:10.1007/s10228-004-0221-y.
116. Pereira, C.S.A.; Ráb, P.; Collares-Pereira, M.J. Chromosomes of European cyprinid fishes: Comparative cytogenetics and chromosomal characteristics of ribosomal DNAs in nine Iberian chondrostomine species (Leuciscinae). *Genetica* **2012**, *140*, 485–495, doi:10.1007/s10709-013-9697-6.
117. Tan, M.; Armbruster, J.W. Phylogenetic classification of extant genera of fishes of the order Cypriniformes (Teleostei: Ostariophysi). *Zootaxa* **2018**, *4476*, 6–39, doi:10.11646/zootaxa.4476.1.4.
118. Sember, A.; Pelikánová, Š.; de Bello Cioffi, M.; Šlechtová, V.; Hatanaka, T.; Do Doan, H.; Knytl, M.; Ráb, P. Taxonomic diversity not associated with gross karyotype differentiation: The case of bighead carps, Genus *Hypophthalmichthys* (Teleostei, Cypriniformes, Xenocyprididae). *Genes* **2020**, *11*, doi:10.3390/genes11050479.
119. Molina, W.F. Chromosomal changes and stasis in marine fish groups. In *Fish Cytogenetics*; Pisano, E., Ozouf-Costaz, C., Foresti, F., Kapoor, B.G., Eds.; Science Publishers: Enfield, NH, USA, 2007; pp. 69–110. ISBN 1-57808-330-3.
120. Motta Neto, C.C.; Cioffi, M.D.B.; Costa, G.W.W.F.; Amorim, K.D.J.; Bertollo, L.A.C.; Artoni, R.F.; Molina, W.F. Overview on karyotype stasis in atlantic grunts (Eupercaria, Haemulidae) and the Evolutionary Extensions for Other Marine Fish Groups. *Front. Mar. Sci.* **2019**, *6*, doi:10.3389/fmars.2019.00628.
121. Neto, C.C.M.; Cioffi, M.B.; Bertollo, L.A.C.; Molina, W.F. Extensive chromosomal homologies and evidence of karyotypic stasis in Atlantic grunts of the genus *Haemulon* (Perciformes). *J. Exp. Mar. Biol. Ecol.* **2011**, *401*, 75–79, doi:10.1016/j.jembe.2011.02.044.
122. Neto, C.C.M.; Cioffi, M.B.; Bertollo, L.A.C.; Molina, W.F. Molecular cytogenetic analysis of Haemulidae fish (Perciformes): Evidence of evolutionary conservation. *J. Exp. Mar. Biol. Ecol.* **2011**, *407*, 97–100, doi:10.1016/j.jembe.2011.07.014.

123. Costa, G.W.W.F. da; Cioffi, M. de B.; Bertollo, L.A.C.; Molina, W.F. The Evolutionary Dynamics of ribosomal genes, histone H3, and transposable Rex elements in the genome of Atlantic snappers. *J. Hered.* **2016**, *107*, 173–180, doi:10.1093/jhered/esv136.
124. Majtánová, Z.; Indermaur, A.; Bitja Nyom, A.R.; Ráb, P.; Musilová, Z. Adaptive radiation from a chromosomal perspective: Evidence of chromosome set stability in cichlid fishes (Cichlidae: Teleostei) from the Barombi Mbo Lake, Cameroon. *Int. J. Mol. Sci.* **2019**, *20*, 4994.
125. Razin, S.V.; Gavrilov, A.A.; Vassetzky, Y.S.; Ulianov, S.V. Topologically-associating domains: Gene warehouses adapted to serve transcriptional regulation. *Transcription* **2016**, *7*, 84–90, doi:10.1080/21541264.2016.1181489.
126. Rosin, L.F.; Crocker, O.; Isenhart, R.L.; Nguyen, S.C.; Xu, Z.; Joyce, E.F. Chromosome territory formation attenuates the translocation potential of cells. *eLife* **2019**, *8*, e49553, doi:10.7554/eLife.49553.
127. Symonová, R.; Howell, W.M. Vertebrate genome evolution in the light of fish cytogenomics and rDNAomics. *Genes* **2018**, *9*, 96, doi:10.3390/genes9020096.
128. Nelson, J.S.; Grande, T.C.; Wilson, M.V.H. *Fishes of the World*, 5 ed.; Wiley: Hoboken, NJ, USA, 2016; ISBN 978-1-118-34233-6.
129. Fujiwara, A.; Abe, S.; Yamaha, E.; Yamazaki, F.; Yoshida, M.C. Chromosomal localization and heterochromatin association of ribosomal RNA gene loci and silver-stained nucleolar organizer regions in salmonid fishes. *Chrom. Res.* **1998**, *6*, 463–471.
130. Symonová, R.; Majtánová, Z.; Sember, A.; Staaks, G.B.O.; Bohlen, J.; Freyhof, J.; Rábová, M.; Ráb, P. Genome differentiation in a species pair of coregonine fishes: An extremely rapid speciation driven by stress-activated retrotransposons mediating extensive ribosomal DNA multiplications. *BMC Evol. Biol.* **2013**, *13*, 42, doi:10.1186/1471-2148-13-42.
131. Dion-Côté, A.-M.; Symonová, R.; Ráb, P.; Bernatchez, L. Reproductive isolation in a nascent species pair is associated with aneuploidy in hybrid offspring. *Proc. Royal Soc. B* **2015**, *282*, 20142862, doi:10.1098/rspb.2014.2862.
132. Bernardi, G. *Structural and Evolutionary Genomics: Natural Selection in Genome Evolution*; Elsevier: Amsterdam, The Netherlands, 2005; ISBN 978-0-08-046187-8.
133. Symonová, R. Integrative rDNAomics—Importance of the oldest repetitive fraction of the Eukaryote genome. *Genes* **2019**, *10*, 345, doi:10.3390/genes10050345.
134. Sochorová, J.; Garcia, S.; Gálvez, F.; Symonová, R.; Kovařík, A. Evolutionary trends in animal ribosomal DNA loci: Introduction to a new online database. *Chromosoma* **2018**, *127*, 141–150, doi:10.1007/s00412-017-0651-8.

Symonová and Suh *Mobile DNA* (2019) 10:49
https://doi.org/10.1186/s13100-019-0195-y

Mobile DNA

COMMENTARY Open Access

Nucleotide composition of transposable elements likely contributes to AT/GC compositional homogeneity of teleost fish genomes

Radka Symonová[1*] and Alexander Suh[2,3]

Abstract

Background: Teleost fish genome size has been repeatedly demonstrated to positively correlate with the proportion of transposable elements (TEs). This finding might have far-reaching implications for our understanding of the evolution of nucleotide composition across vertebrates. Genomes of fish and amphibians are GC homogenous, with non-teleost gars being the single exception identified to date, whereas birds and mammals are AT/GC heterogeneous. The exact reason for this phenomenon remains controversial. Since TEs make up significant proportions of genomes and can quickly accumulate across genomes, they can potentially influence the host genome with their own GC content (GC%). However, the GC% of fish TEs has so far been neglected.

Results: The genomic proportion of TEs indeed correlates with genome size, although not as linearly as previously shown with fewer genomes, and GC% negatively correlates with genome size in the 33 fish genome assemblies analysed here (excluding salmonids). GC% of fish TE consensus sequences positively correlates with the corresponding genomic GC% in 29 species tested. Likewise, the GC contents of the entire repetitive vs. non-repetitive genomic fractions correlate positively in 54 fish species in Ensembl. However, among these fish species, there is also a wide variation in GC% between the main groups of TEs. Class II DNA transposons, predominant TEs in fish genomes, are significantly GC-poorer than Class I retrotransposons. The AT/GC heterogeneous gar genome contains fewer Class II TEs, a situation similar to fugu with its extremely compact and also GC-enriched but AT/GC homogenous genome.

Conclusion: Our results reveal a previously overlooked correlation between GC% of fish genomes and their TEs. This applies to both TE consensus sequences as well as the entire repetitive genomic fraction. On the other hand, there is a wide variation in GC% across fish TE groups. These results raise the question whether GC% of TEs evolves independently of GC% of the host genome or whether it is driven by TE localization in the host genome. Answering these questions will help to understand how genomic GC% is shaped over time. Long-term accumulation of GC-poor(er) Class II DNA transposons might indeed have influenced AT/GC homogenization of fish genomes and requires further investigation.

Keywords: Teleost fish, Transposon, GC content, Genome evolution, Nucleotide composition

* Correspondence: radka.symonova@gmail.com
[1]Department of Biology, Faculty of Science, University of Hradec Králové, Hradec Králové, Czech Republic
Full list of author information is available at the end of the article

Background

Nucleotide composition is a fundamental property of genomes with a strong influence on gene function and regulation [1]. Hence, GC content of a genome (GC_G), i.e., the molar ratio of guanine (G) and cytosine (C) in DNA, is one of the main parameters used to describe nucleotide composition and is frequently related to genome size [1]. For practical reasons, genomes can be segmented in five types of regions called isochores according to their GC percentage (GC%). Two "light" isochores with the lowest GC%, i.e., L1 with approx. 34–36% of GC and L2 approx. 37–40% of GC; as well as three "heavy" isochores, i.e., H1 with approx. 41–45% of GC, H2 46–52% and the "heaviest" H3 with > 53% of GC [2]. In this regard, fish and amphibian genomes are overall AT/GC homogenous because they contain only the GC-poor(er) isochores with a substantially narrower range of GC%, i.e., usually only two neighbouring ones such as L1 and L2 or L2 and H1. On the other hand, avian and mammalian genomes contain all five isochores and their broad range of GC% results in overall GC heterogeneity [2].

An increasing number of recent studies in fish has shown a clear positive correlation between genome size and percentage of TEs, and that TEs are ubiquitous and present in large numbers, e.g., refs. [3–6]. One of these studies [7] documented a surprisingly linear correlation between genome size and TE content in four teleost fish species. A clear but not strictly linear correlation between the percentage of TEs and genome size was identified in a larger dataset of 19 ray-finned and two lobe-finned fish species ([3]; including the four genomes analysed by ref. [7]). The so far most extensive (but still unpublished) study on fish TEs by ref. [5] using in silico explorations of TE activity, diversity and abundance across 74 teleost fish genomes showed that the total genomic TE abundances reflect variation in their host genome size.

Moreover, TEs can be very different in copy numbers and composition [3, 4, 8, 9], which would imply that accumulation or turnover of TE numbers/composition could change genomic GC content (GC_G) because of the TEs' own GC content (GC_{TE}). There are major quantitative and qualitative differences in TEs among vertebrates: Class II DNA transposons are the most abundant group in fish genomes, whereas in avian and mammalian genomes Class I retrotransposons are the most abundant group while DNA transposons are substantially less numerous [3–5, 8, 9]. Hence, the GC_{TE} of different mobilomes, i.e., the sum of TEs within a genome, may potentially result in different overall GC_G organization in fish when compared with birds and mammals. However, the characteristics of GC_{TE} remains understudied in general, particularly in fish. This is despite the fact that TEs make up 6–55% of the total base pairs of fish genomes, and that TEs are clearly depleted in compact and GC-rich genomes (*Takifugu flavidus* [9, 10], *Tetraodon nigroviridis* [11, 12]) while they are massively represented in large and GC-poor genomes such zebrafish (*Danio rerio* [13]) and cod (*Gadus morhua* [14]).

The currently known main features of fish mobilomes can be summarized as follows: i. DNA transposons are the predominant group of TEs in fish; ii. the diversity of TE families is generally high in fish; iii. many TEs show recent activity in fish genomes; and iv. the total genomic abundances of TEs reflect the variation in genome size [3–5, 15]. Since the dynamics of genome size variation can be largely explained by TEs in many eukaryotes [16, 17] and GC_G is negatively linked to genome size in some organisms [1], these findings in fish raise crucial questions about potential roles of TEs in shaping GC_G: i. Do TEs have a different GC% than the non-TE regions of the host genome? ii. Do new TE insertions lead to a decrease in GC% in adjacent regions of the host genome because of TE silencing through cytosine methylation? Methylcytosine frequently undergoes spontaneous deamination resulting in point mutation to thymine [18]. iii. Do TEs change local recombination rates (negatively if TEs are heterochromatinized or positively if they contain motifs attracting the recombination machinery [19, 20]) and hence influence the GC_G as discussed below? These factors all may contribute to the overall nucleotide compositional landscape, i.e., the heterogeneous organization in birds and mammals in comparison with the homogeneous organization in fish and amphibians. Such manifold effects of TEs might be particularly pronounced in species where TEs comprise a substantial genomic fraction, e.g., zebrafish (*D. rerio*) [13].

Both the local GC_G as well as TE density are linked to the local recombination rate. Evidence to date suggests that TE densities correlate negatively with recombination rate, but the strength of this correlation varies across TE types [20]. At the same time, the currently most plausible explanation of the AT/GC heterogeneity in avian and mammalian genomes is a non-adaptive process called GC-biased gene conversion (gBGC), whereby increased GC% is tightly related to an increased recombination rate (recently extensively reviewed by ref. [19]). In mammals and some other vertebrates (but not birds), at least a part of the regional variation in the location of recombination hotspots can be ascribed to the activity of the protein PRDM9 [21].

One may expect that TEs contribute to the length and GC% of noncoding sequences, and continue to do so even long after they are no longer recognizable as TEs. While TE insertions are a major factor in the expansion or turnover of noncoding regions (both introns and

intergenic sequences [17, 22]), the potential influence of the GC_{TE} on the host regional GC_G has only been comprehensively assessed for the human genome. Around 42% of the human genome is made up of retrotransposons, whereas DNA transposons only account for about 2–3%, and the insertion or accumulation of TEs depends on the isochore region involved [23]. For instance, *Alu* (the most abundant TE in human) and L1 insertions contribute to the AT/GC heterogeneity of the human genome due to their differential accumulation: *Alu* SINEs (approx. 50% GC_{TE} in their consensus sequence) reside preferentially in GC-rich regions, whereas L1 LINEs (approx. 37% GC_{TE} in their consensus sequence) reside preferentially in GC-poor regions [24]. Recognizable *Alu* elements make up 20% of GC-rich regions and 7% of GC-poor regions, whereas recognizable L1 elements make up 5% of GC-rich regions and 20% of GC-poor regions [25]. For fish, a single study briefly investigated the potential correlation between TEs and GC% along *T. nigroviridis* and *D. rerio* genomes [26]. However, they did not observe any effect of TEs on GC_G in *T. nigroviridis* and *D. rerio*. Three studies investigated in detail some unusual examples of GC-rich TEs in crabs [27–29] and reported different GC% between DNA transposons of marine and continental species. A bit more is known from plants and their TEs, e.g., Pack-MULEs elements in grasses specifically acquire and amplify GC-rich gene fragments [30].

In this study, we aim to bring a novel viewpoint on the vertebrate nucleotide compositional evolution by analysing the GC_{TE} of fish TEs and assessing their potential contribution to the GC_G and the overall nucleotide compositional landscape of their host genomes.

Results

Genome size positively correlates with the genomic density of TEs in fish

To summarize the previously reported positive correlation between fish genome size and genomic abundance of TEs [3–5, 7, 15], we generated an example plot using cytological genome size estimates, i.e. C-value in picograms (pg; Fig. 1a). Species included are 29 teleosts that underwent the teleost-specific whole-genome duplication (WGD) of which five salmonid species underwent another round of WGD, the salmonid-specific one [35]. Further, we included the spotted gar (*Lepisosteus oculatus*), i.e., a deep-branching non-teleost ray-finned fish that has not undergone any further WGD after the two basal vertebrate ones but that shows the mammalian-like situation of AT/GC heterogeneity [36]. Finally, we analysed one lamprey species (*Petromyzon marinus*), one shark (*Callorhinchus milii*) and one coelacanth (*Latimeria chalumnae*). This correlation represents an important starting point for our following considerations. Detailed lists of species analysed are in Additional files 1 and 2: Tables S1 and S2.

Genome size negatively correlates with the genomic GC% in fish excluding salmonids

Data on GC_G of genome assemblies currently available in NCBI GenBank [33] and in the literature permit us to identify another crucial association – a negative correlation between fish genome size (as C-value in picograms from the Animal Genome Size Database [32]) and their genomic GC% (Fig. 1b).

To avoid any potential bias conditioned by incompleteness of currently available genome assemblies (e.g., differences in amounts of heterochromatic repeats assembled and in assembly quality sensu [37]), we compared two types of genome size datasets: one based on C-values, i.e., the non-genomics (cytological) genome size estimation (Fig. 1b) and another based on genome assembly size (Fig. 1d). Despite slight differences between these datasets, both show comparable trends, suggesting that both are usable for further analyses.

In this analysis, we excluded the eight sampled salmonid species (details in Additional file 1: Table S1) because their large genomes exhibit a salmonid-specific WGD and extremely amplified ribosomal (rRNA) genes that are exceptionally GC-rich. This feature is well known from cytogenetics [31]. Including these large and GC-enrich salmonid genomes distorts the clear correlation between GC_G and genome size in other teleost fish (cf. Additional file 3: Figure S1).

GC% of TEs positively correlates with genomic GC% in fish

Comparison of GC_{TE} with the respective GC_G uncovered a positive correlation. Firstly, we calculated the GC_{TE} out of the sum of individual consensus sequences of TEs annotated for each fish species from FishTEDB [34] (Fig. 1c) and not out of the entire mobilome reflecting the TEs' copy numbers in the respective genome. As consensus sequences are approximations of the TE copies at their time point of insertion, we consider their consensus GC_{TE} to be more appropriate here because it should not reflect the genomic location of individual TE copies. Note that FishTEDB does not include any salmonid species. For comparison, we calculated GC_{REP} of repeats including low-complexity regions and compared it with the remaining non-repetitive fraction of the relevant genomes, i.e. GC_{NONREP} (Fig. 2). For this analysis, we used masked genome assemblies from the Ensembl (Release 98, [38]) as the FishTEDB lists only consensus sequences of TEs per fish species.

The GC_{TE} is mostly higher than the overall GC_G, with two exceptions. These exceptions are cod and European

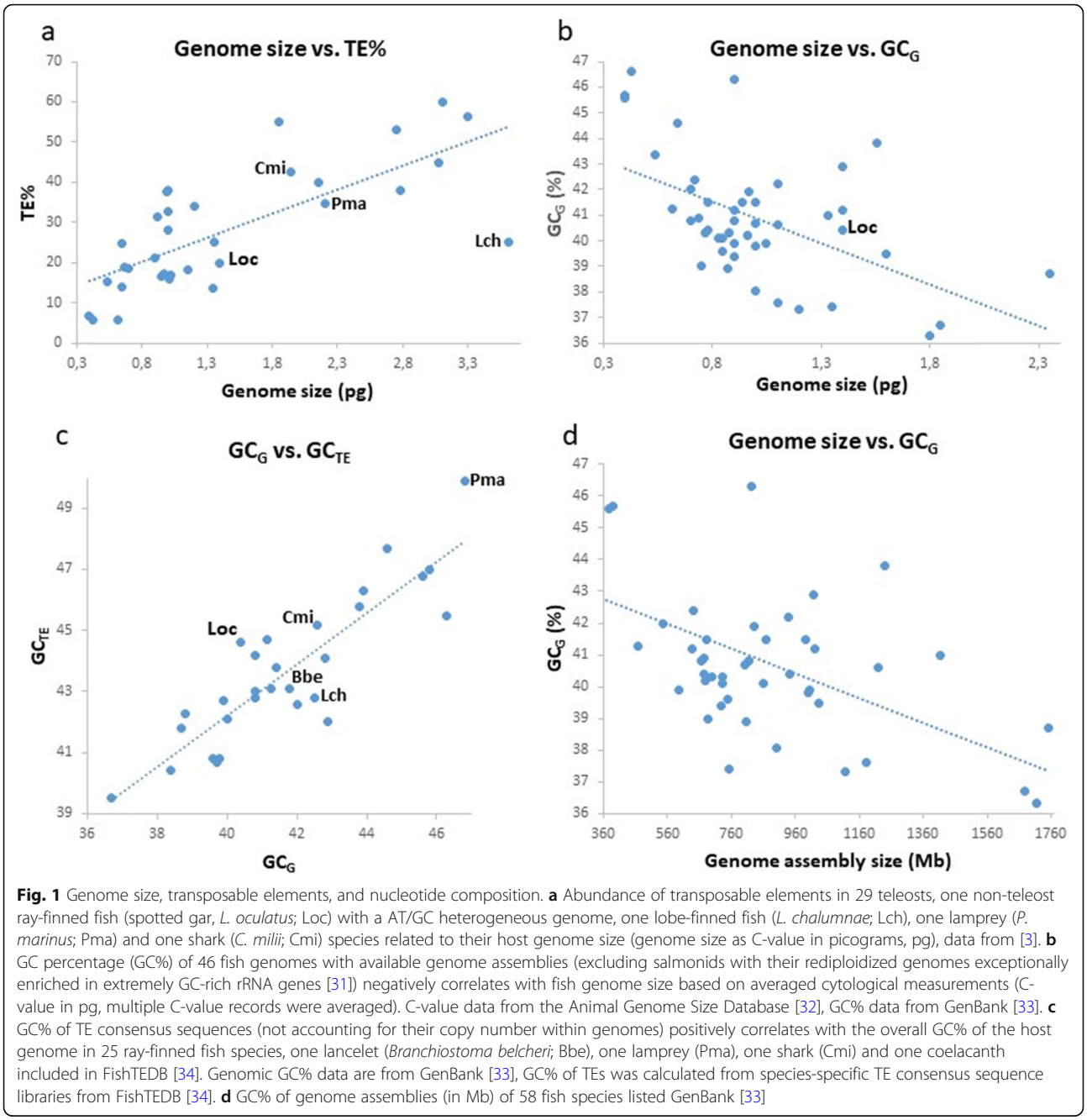


Fig. 1 Genome size, transposable elements, and nucleotide composition. **a** Abundance of transposable elements in 29 teleosts, one non-teleost ray-finned fish (spotted gar, *L. oculatus*; Loc) with a AT/GC heterogeneous genome, one lobe-finned fish (*L. chalumnae*; Lch), one lamprey (*P. marinus*; Pma) and one shark (*C. milii*; Cmi) species related to their host genome size (genome size as C-value in picograms, pg), data from [3]. **b** GC percentage (GC%) of 46 fish genomes with available genome assemblies (excluding salmonids with their rediploidized genomes exceptionally enriched in extremely GC-rich rRNA genes [31]) negatively correlates with fish genome size based on averaged cytological measurements (C-value in pg, multiple C-value records were averaged). C-value data from the Animal Genome Size Database [32], GC% data from GenBank [33]. **c** GC% of TE consensus sequences (not accounting for their copy number within genomes) positively correlates with the overall GC% of the host genome in 25 ray-finned fish species, one lancelet (*Branchiostoma belcheri*; Bbe), one lamprey (Pma), one shark (Cmi) and one coelacanth included in FishTEDB [34]. Genomic GC% data are from GenBank [33], GC% of TEs was calculated from species-specific TE consensus sequence libraries from FishTEDB [34]. **d** GC% of genome assemblies (in Mb) of 58 fish species listed GenBank [33]

eel, however, the difference is within the range of 1%, i.e., for the eel GC_G = 42.9% vs. GC_{TE} = 42.0% and for the cod GC_G = 46.3% vs. GC_{TE} = 45.5% (more details in Additional file 4: Figure S2).

GC% varies widely among particular groups of TEs in fish

Dissecting the GC anatomy of the sum of individual TE consensus sequences in fish genomes, we further disentangled GC_{TE} of the major TE groups: Class I retrotransposons are GC-richer with an averaged consensus GC_{TE} of 45.6% than Class II DNA transposons with an averaged consensus GC_{TE} of 40.1% (Fig. 3). Within Class I, LTR retrotransposons are GC-richer than LINEs. The Class I DIRS retrotransposons are the GC-richest fish TEs with GC_{TE} of 53.8%. The Class II CMC transposons are the AT-richest fish TEs with GC_{TE} of 35.8%.

Details on the variability of species-specific GC_{TE} in 19 selected species from FishTEDB are presented in Figure S3 (Additional file 5; 16 ray-finned species, one lancelet, one shark, and one lamprey species; some species displayed in FishTEDB do not contain sequences).

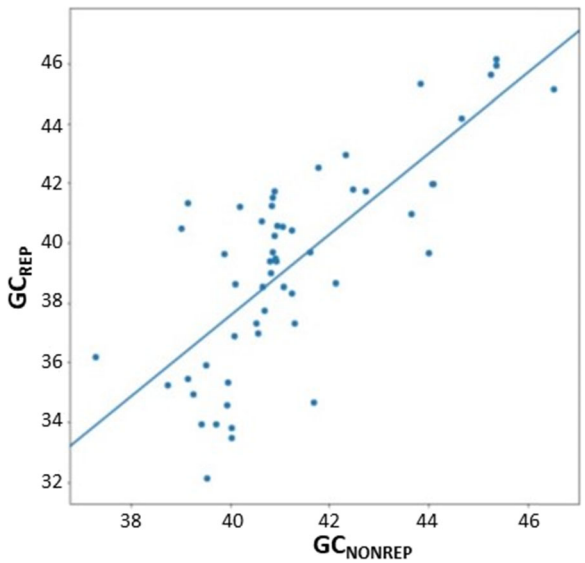


Fig. 2 Comparison of GC% of repetitive and non-repetitive genomic fractions in 54 fish species from the Ensembl database (Release 98). The Y-axis shows GC_{REP}, i.e. GC% of repeats (including low-complexity regions) masked with the RepeatMasker tool, while the X-axis shows GC_{NONREP} of the non-repetitive fraction of each assembly. Data used for this analysis are available in the Additional file 2: Table S2

GC% of Class II DNA transposons varies heavily among different fish species

The observed variation in GC_{TE} among the major TE groups listed in the FishTEDB is particularly relevant considering that fish genomes are greatly enriched in Class II DNA transposons in contrast to avian and mammalian genomes. Therefore, we calculated the GC_{TE} of all consensus sequences of DNA transposons for 17 fish species. These data provide first insights into the GC_{TE} of fish transposons. Firstly, the compact genomes of not only pufferfishes *T. flavidus* and *T. nigroviridis* but also of cod (*G. morhua*) and stickleback (*Gasterosteus aculeatus*) show GC enrichment of their TEs as well as overall GC-richer Class II DNA transposons (Fig. 4). The same is apparent also in the non-teleost spotted gar (*L. oculatus*) with its AT/GC heterogeneous genome and an unusually high GC_{TE} in comparison with teleosts. The opposite situation occurs in teleosts with larger genomes such as *D. rerio* and *Astyanax mexicanus*: DNA transposons are GC-poor(er) as well as the overall GC_G and GC_{TE} are lower.

Discussion

Recent studies on the relative contribution of TEs to genome size in fish [3, 4, 7, 39] have become an important starting point for us to understand the evolution of nucleotide composition. The above listed results raise crucial questions about the contribution of the mobilome GC% to the entire genomic GC% and to the nucleotide compositional landscape. This has been so far addressed only for the human genome [22]. Here, we show that utilizing purely genomic data for approximating genome size (assembly vs. C-value) and GC% yield reproducible and comparable data suitable for assessing nucleotide composition of host genomes and their respective TEs. The ever-increasing number of available assemblies and TE annotations for fish and other

Fig. 3 GC_{TE} in the major groups of Class I and Class II TEs, calculated as sum of GC% for all 28 fish species available in the FishTEDB database. TE consensus sequences for these calculations are from the "Browse" section of the FishTEDB database [34]

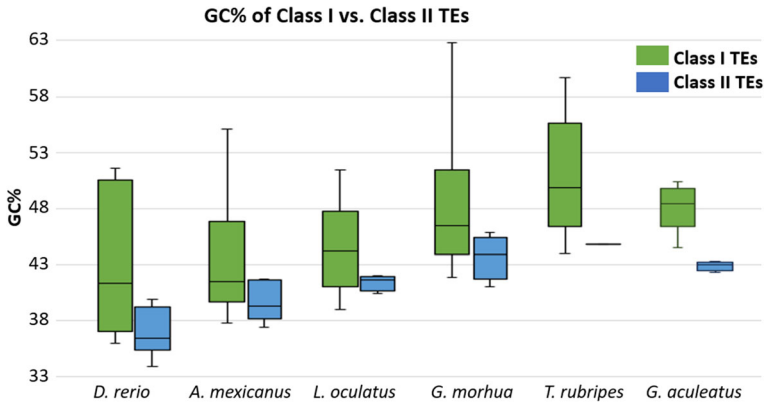


Fig. 4 Comparison GC% between TE consensus sequences from Class I (retrotransposons) and Class II (DNA transposons) in six selected fish species (highlighted in the main text) listed in the FishTEDB database [34]

vertebrates has now become sufficient to begin to address the questions raised here.

GC richness vs. AT/GC heterogeneity and TEs

It is necessary to distinguish between an overall genomic GC-richness, i.e., GC_G, and the avian or mammalian situation of AT/GC heterogeneity (recorded also in non-teleost gars [36]). This entails an alternation of GC-rich and GC-poor regions along linkage groups, thus forming banding patterns on chromosomes upon an AT/GC-specific staining (recently reviewed by [36]). In the case of AT/GC heterogeneity, the overall GC_G can be even lower than is in cases of AT/GC homogeneity typical for fish genomes as shown below. Considering that all of the currently available vertebrate genome assemblies contain gaps due to either repeat-rich or GC-rich regions [37], fish with GC-rich genomes might actually be even GC-richer than currently estimated, and potentially even more GC-rich than mammalian and avian genomes. This is indicated by the following examples: the human (GC_G = 40.9%), mouse (GC_G = 42.5%), and even chicken (GC_G = 41.9%) genomes are GC-poorer than cod (GC_G = 46.3%) and three pufferfish species (GC_G = 45.6, 45.7% and GC_G = 46.6% respectively). However, note the situation in the non-teleost spotted gar with GC_G = 40.4% and AT/GC heterogeneity. The total length of its available assembly is merely 945.878 Mb [33], which is remarkably incomplete in comparison with the cytological genome size estimate of 1.4 pg [32]. Nevertheless, the AT/GC heterogeneity evidenced cytogenetically was also confirmed using genomic data [36].

The smaller and GC-rich(er) fish genomes also contain lower TE densities (or lower densities of GC-poor TEs) and/or GC-rich (er) TEs. The fact that the averaged GC% of consensus sequences from all TE families is generally higher than the entire genomic GC% suggests that TE spread and accumulation might contribute to the overall GC_G in fish. This is further supported by our observation that genomes with a higher GC% of the repetitive genomic fraction (i.e., TEs and other repeats; GC_{REP}) have a higher GC_{NONREP}, i.e., GC% of the non-repetitive rest of the genome. However, due to the broad range of GC_{TE} of major groups of TEs in different species (Fig. 3), the activity and abundance of GC-poor(er) DNA transposons might also contribute to the AT/GC homogeneity in fish, assuming they accumulated more homogenously, compared to the AT/GC heterogeneity in avian and mammalian genomes that usually lack activity of DNA transposons.

How could TEs shape the host nucleotide compositional landscape?

Considering our findings, we anticipate at least three possible ways how TEs could influence the host nucleotide compositional landscape: 1) TEs shape it through inserting their "own" GC in a new context (i.e., increasing GC% of the region if they have high GC; lowering GC% of the region of they have low GC); 2) TEs shape nearby GC% through "spillover" of CpG methylation ('sloping shores' model of [40]), leading to CpG hypermutation and thus decrease of nearby GC%; and 3) some TEs might contain sequence motifs that increase or decrease the local recombination landscape and thus the strength of GC-biased gene conversion. There are however many more questions about GC% of TEs to be answered: Are quantitatively larger mobilomes as GC-poor as larger host genomes are overall? Why are DNA transposons GC-poor? Why are some DNA transposons GC-poorer than others and only so in some species?

Conclusion and perspectives

Here we have shown that nucleotide composition of TEs and their interplay with host genomes is an unexplored part of genome biology. The GC-poor DNA transposons

predominant in fish genomes and nearly absent in avian and mammalian genomes might have indeed contributed to shaping the nucleotide compositional landscape in vertebrates. Only the GC-heterogeneous gar and the GC-enriched pufferfishes possess GC-richer TEs and fewer DNA transposons. At the same time, among others the GC-poor genome of zebrafish possesses the GC-poorest TEs. Hence, it is possible that DNA transposon spreading and accumulation has actively contributed to the overall GC homogenization of fish genomes. On the other hand, replacement of DNA transposons by retrotransposons in avian and mammalian genomes might have contributed to their AT/GC heterogeneity through differential accumulation across chromosomes. The GC content of TEs should thus be considered as one of the factors potentially shaping the nucleotide compositional landscape in vertebrates and requires further investigations in detail. The next step envisaged is a qualitative analysis of the contribution GC% of individual TE insertions to the GC% of host genomes while accounting for TE copy number. This step can be combined with cytogenetic data to investigate the chromosomal distribution of various TEs and their potential contribution to the GC homogenization of fish genomes. With 55 fish species genome assemblies recently introduced by the 98th release of Ensembl (November 2019 [38]) and numerous others, such comprehensive analyses now appear feasible.

Methods

All species analysed in datasets produced for this study are listed in the Additional file 1: Table S1 and the datasets supporting the conclusions of this article are included in the Additional file 2: Table S2. We obtained genome size data as C-values from the www.genomesize.com database [32]. At this stage, diverse sources of datasets and databases (ref. [3], Animal Genome Size Database [32], GenBank [33], FishTEDB [34]) list different sets of fish species of which only some have been analysed for TEs. Assembly size data in Mb were obtained from the NCBI GenBank records of sequenced genomes [33]. Proportions of TEs in fish genomes were obtained from ref. [3] and compared with ref. [7]. Sequences of annotated fish TEs were obtained from Fish TE database http://www.fishtedb.org [34] and from the Repbase database at www.girinst.org [41]. Further data were extracted from literature as listed in the Additional file 2: Table S2. We used custom Python scripts to extract GC_{REP} (repeats including low-complexity regions) of fish genomes in the Ensembl database (https://www.ensembl.org/ [38]) and compared to GC% of the rest of the genome assembly (GC_{NONREP}), i.e. the non-repetitive fraction. The scripts are available at the GitHub repository https://github.com/bioinfohk/GC_TE/blob/master/GC_softmasked_genomesFISH.ipynb.

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1186/s13100-019-0195-y.

Additional file 1: Table S1. Species overview and their counts.

Additional file 2: Table S2. Datasets used for generating Figs. 1, 2, 3, 4 and Additional files 3 and 4: Figures S1-S2.

Additional file 3: Figure S1. Analysis of genome size vs. GC_G including salmonids (for comparison with Fig. 1b).

Additional file 4: Figure S2. Comparison of GC_G and GC_{TE} in 29 fish species (ray-finned fish and outgroups lancelet *Branchiostoma belcheri*, lamprey *Petromyzon marinus*, shark *Callorhinchus milii*, and coelacanth *Latimeria chalumnae*) listed in the FishTEDB [36]. In only two species analysed, GC_{TE} (orange) is lower than GC_G (blue; *A. anguilla* and *G. morhua*). Based on the dataset for Fig. 1c in Additional file 2.

Additional file 5: Figure S3. Species-specific comparisons of GC_{TE} between Class I and Class II TEs.

Abbreviations

GC%: Percentage of G + C bases, i.e., the molar ratio of guanine and cytosine in DNA; GC_G: GC% of the whole genome; GC_{NONREP}: GC% of the non-repetitive fraction of genome assemblies in Ensembl; GC_{REP}: GC% of the repetitive fraction of genome assemblies in Ensembl; GC_{TE}: GC% of TE consensus sequences; GS: Genome size; LINE: Long interspersed element; LTR: Long terminal repeat; MLE: *Mariner*-like element; SINE: Short interspersed element; TE: Transposable element; WGD: Whole genome duplication

Acknowledgements

We would like to acknowledge Carina Mugal and Cedric Feschotte for insightful discussions, and Jesper Boman and Homa Papoli Yazdi for helpful comments on an earlier version of this manuscript. We also thank two anonymous reviewers for their constructive suggestions on this manuscript. Furthermore, we would like to acknowledge Dominik Matoulek for preparation of Python scripts for GC_{REP} and GC_{NONREP} analysis and Michal Dobrovolný for his help with species-specific GC% analysis in fish from FishTEDB.

Authors' contributions

RS conceived the study, RS drafted the first version of the manuscript, RS and AS co-drafted subsequent versions of the manuscript, RS received funds for the study. Both authors read and approved the final manuscript.

Funding

The authors are grateful to the 'Excelence projekt PřF UHK 2209/2018' for the financial support.

Availability of data and materials

All data generated or analysed during this study are included in this published article and its supplementary information files.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Department of Biology, Faculty of Science, University of Hradec Králové, Hradec Králové, Czech Republic. [2]Department of Ecology and Genetics - Evolutionary Biology, Evolutionary Biology Centre (EBC), Science for Life Laboratory, Uppsala University, Uppsala, Sweden. [3]Present address: Department of Organismal Biology - Systematic Biology, Evolutionary Biology Centre (EBC), Science for Life Laboratory, Uppsala University, Uppsala, Sweden.

Received: 27 September 2019 Accepted: 5 December 2019
Published online: 12 December 2019

References

1. Li X-Q, Du D. Variation, evolution, and correlation analysis of C+G content and genome or chromosome size in different kingdoms and phyla. Zhang Z, editor. PLoS ONE. 2014;9:e88339.
2. Bernardi G. Structural and evolutionary genomics natural selection in genome evolution. Amsterdam: Elsevier; 2005. Available from: http://cmich.idm.oclc.org/login?url=http://site.ebrary.com/lib/cmich/Doc?id=10138474. [cited 2018 Nov 4]
3. Canapa A, Barucca M, Biscotti MA, Forconi M, Olmo E. Transposons, genome size, and evolutionary insights in animals. Cytogenet Genome Res. 2015;147:217–39.
4. Chalopin D, Naville M, Plard F, Galiana D, Volff J-N. Comparative analysis of transposable elements highlights mobilome diversity and evolution in vertebrates. Genome Biol Evol. 2015;7:567–80.
5. Brynildsen W. Transposable elements in teleost fish: in silico exploration of TE activity, diversity and abundance across 74 teleost fish genomes: University Oslo; 2016. Available from: http://urn.nb.no/URN:NBN:no-55565
6. Shao F, Han M, Peng Z. Evolution and diversity of transposable elements in fish genomes. Sci Rep. 2019;9 Available from: http://www.nature.com/articles/s41598-019-51888-1. [cited 2019 Nov 21].
7. Gao B, Shen D, Xue S, Chen C, Cui H, Song C. The contribution of transposable elements to size variations between four teleost genomes. Mob DNA. 2016;7 Available from: http://www.mobilednajournal.com/content/7/1/4. [cited 2018 Mar 19].
8. Braasch I, Gehrke AR, Smith JJ, Kawasaki K, Manousaki T, Pasquier J, et al. The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons. Nat Genet. 2016;48:427–37.
9. Volff J-N, Bouneau L, Ozouf-Costaz C, Fischer C. Diversity of retrotransposable elements in compact pufferfish genomes. Trends Genet. 2003;19:674–8.
10. Gao Y, Gao Q, Zhang H, Wang L, Zhang F, Yang C, et al. Draft sequencing and analysis of the genome of pufferfish Takifugu flavidus. DNA Res. 2014;21:627–37.
11. Dasilva C, Hadji H, Ozouf-Costaz C, Nicaud S, Jaillon O, Weissenbach J, et al. Remarkable compartmentalization of transposable elements and pseudogenes in the heterochromatin of the Tetraodon nigroviridis genome. Proc Natl Acad Sci. 2002;99:13636–41.
12. Neafsey DE. Genome size evolution in pufferfish: a comparative analysis of Diodontid and Tetraodontid pufferfish genomes. Genome Res. 2003;13:821–30.
13. Howe K, Clark MD, Torroja CF, Torrance J, Berthelot C, Muffato M, et al. The zebrafish reference genome sequence and its relationship to the human genome. Nature. 2013;496:498–503.
14. Tørresen OK, Star B, Jentoft S, Reinar WB, Grove H, Miller JR, et al. An improved genome assembly uncovers prolific tandem repeats in Atlantic cod. BMC Genomics. 2017;18 Available from: http://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-016-3448-x. [cited 2018 Jan 18].
15. Sotero-Caio CG, Platt RN, Suh A, Ray DA. Evolution and diversity of transposable elements in vertebrate genomes. Genome Biol Evol. 2017;9:161–77.
16. Pritham EJ. Transposable elements and factors influencing their success in eukaryotes. J Hered. 2009;100:648–55.
17. Kapusta A, Suh A, Feschotte C. Dynamics of genome size evolution in birds and mammals. Proc Natl Acad Sci. 2017;114:E1460–9.
18. Fryxell KJ, Zuckerkandl E. Cytosine deamination plays a primary role in the evolution of mammalian isochores. Mol Biol Evol. 2000;17:1371–83.
19. Mugal CF, Weber CC, Ellegren H. GC-biased gene conversion links the recombination landscape and demography to genomic base composition: GC-biased gene conversion drives genomic base composition across a wide range of species. BioEssays. 2015;37:1317–26.
20. Kent TV, Uzunović J, Wright SI. Coevolution between transposable elements and recombination. Philos Trans R Soc B Biol Sci. 2017;372:20160458.
21. Baker Z, Schumer M, Haba Y, Bashkirova L, Holland C, Rosenthal GG, et al. Repeated losses of PRDM9-directed recombination despite the conservation of PRDM9 across vertebrates. eLife. 2017;6 Available from: https://elifesciences.org/articles/24133. [cited 2018 Nov 4].
22. Duret L, Hurst LD. The elevated GC content at exonic third sites is not evidence against neutralist models of isochore evolution. Mol Biol Evol. 2001;18:757–62.
23. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. Nature. 2001;409:860–921.
24. Duret L, Mouchiroud D, Gautier C. Statistical analysis of vertebrate sequences reveals that long genes are scarce in GC-rich isochores. J Mol Evol. 1995;40:308–17.
25. Smit AF. Interspersed repeats and other mementos of transposable elements in mammalian genomes. Curr Opin Genet Dev. 1999;9:657–63.
26. Melodelima C, Gautier C. The GC-heterogeneity of teleost fishes. BMC Genomics. 2008;9:632.
27. Halaimia-Toumi N, Casse N, Demattei MV, Renault S, Pradier E, Bigot Y, et al. The GC-rich transposon bytmar1 from the deep-sea hydrothermal crab, bythograea thermydron, may encode three transposase isoforms from a single ORF. J Mol Evol. 2004;59:747–60.
28. Casse N, Bui QT, Nicolas V, Renault S, Bigot Y, Laulier M. Species sympatry and horizontal transfers of mariner transposons in marine crustacean genomes. Mol Phylogenet Evol. 2006;40:609–19.
29. Bui Q-T, Delaurière L, Casse N, Nicolas V, Laulier M, Chénais B. Molecular characterization and phylogenetic position of a new mariner-like element in the coastal crab, *Pachygrapsus marmoratus*. Gene. 2007;396:248–56.
30. Ferguson AA, Jiang N. Pack-MULEs: recycling and reshaping genes through GC-biased acquisition. Mob Genet Elem. 2011;1:135–8.
31. Dion-Côté A-M, Symonová R, Lamaze FC, Pelikánová Š, Ráb P, Bernatchez L. Standing chromosomal variation in Lake whitefish species pairs: the role of historical contingency and relevance for speciation. Mol Ecol. 2017;26:178–92.
32. Gregory TR. Animal genome size database. http://www.genomesize.com.
33. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, et al. GenBank. Nucleic Acids Res. 2013;41:D36–42.
34. Shao F, Wang J, Xu H, Peng Z. FishTEDB: a collective database of transposable elements identified in the complete genomes of fish. Database. 2018;2018. https://doi.org/10.1093/database/bax106.
35. Macqueen DJ, Johnston IA. A well-constrained estimate for the timing of the salmonid whole genome duplication reveals major decoupling from species diversification. Proc R Soc B Biol Sci. 2014;281:20132881.
36. Symonová R, Majtánová Z, Arias-Rodriguez L, Mořkovský L, Kořínková T, Cavin L, et al. Genome compositional organization in gars shows more similarities to mammals than to other ray-finned fish: cytogenomics of gars. J Exp Zool B Mol Dev Evol. 2017;328:607–19.
37. Peona V, Weissensteiner MH, Suh A. How complete are "complete" genome assemblies?-an avian perspective. Mol Ecol Resour. 2018;18:1188–95.
38. Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, et al. Ensembl 2018. Nucleic Acids Res. 2018;46:D754–61.
39. Brynildsen WR. Transposable elements in teleost fish – in silico explorations of TE activity, diversity and abundance across 74 teleost fish genomes. 2016. Available from: https://www.duo.uio.no/handle/10852/52365
40. Grandi FC, Rosser JM, Newkirk SJ, Yin J, Jiang X, Xing Z, et al. Retrotransposition creates sloping shores: a graded influence of hypomethylated CpG islands on flanking CpG sites. Genome Res. 2015;25:1135–46.
41. Bao W, Kojima KK, Kohany O. Repbase update, a database of repetitive elements in eukaryotic genomes. Mob DNA. 2015;6 Available from: http://www.mobilednajournal.com/content/6/1/11. [cited 2018 Nov 4].

Publisher's Note

genes

MDPI

Review

Integrative rDNAomics—Importance of the Oldest Repetitive Fraction of the Eukaryote Genome

Radka Symonová

Faculty of Science, Department of Biology, University of Hradec Králové, 500 03 Hradec Králové, Czech Republic; radka.symonova@gmail.com

Received: 3 February 2019; Accepted: 25 April 2019; Published: 7 May 2019

Abstract: Nuclear ribosomal RNA (rRNA) genes represent the oldest repetitive fraction universal to all eukaryotic genomes. Their deeply anchored universality and omnipresence during eukaryotic evolution reflects in multiple roles and functions reaching far beyond ribosomal synthesis. Merely the copy number of non-transcribed rRNA genes is involved in mechanisms governing e.g., maintenance of genome integrity and control of cellular aging. Their copy number can vary in response to environmental cues, in cellular stress sensing, in development of cancer and other diseases. While reaching hundreds of copies in humans, there are records of up to 20,000 copies in fish and frogs and even 400,000 copies in ciliates forming thus a literal subgenome or an rDNAome within the genome. From the compositional and evolutionary dynamics viewpoint, the precursor 45S rDNA represents universally GC-enriched, highly recombining and homogenized regions. Hence, it is not accidental that both rDNA sequence and the corresponding rRNA secondary structure belong to established phylogenetic markers broadly used to infer phylogeny on multiple taxonomical levels including species delimitation. However, these multiple roles of rDNAs have been treated and discussed as being separate and independent from each other. Here, I aim to address nuclear rDNAs in an integrative approach to better assess the complexity of rDNA importance in the evolutionary context.

Keywords: nuclear rDNA; rRNA; GC-content; secondary structure; nucleolus

1. The Eukaryotic rDNAome

RNA is essential for information flow from DNA to protein being the dominant macromolecule in protein synthesis [1]. Of the major RNA types, mRNA, tRNA, rRNA, and numerous short non-coding snRNAs, our focus here is on the nuclear rRNA encoded by ribosomal DNA (rDNA), i.e., by rRNA genes. In eukaryotic cells, up to 80% of RNA synthesis belongs to rRNA transcription indispensable to the preservation of ribosome biogenesis and protein synthesis [2]. There are about 1.5–3 million ribosomes per eukaryotic cell [3]. Hence, ribosome biogenesis consumes a tremendous amount of cellular energy and rRNA synthesis is tightly linked to cell growth and proliferation, and as such, it is responsive to general metabolism and environmental challenges [4]. In Eukaryotes, rRNA genes consist of several distinct multigene families tandemly arrayed as repeats composed of tens to hundreds or even thousands of copies. Beside two mitochondrial rRNAs, i.e., the 12S and 16S rRNA in eukaryotes, there are two fractions of nuclear rDNAs—a large, nucleolus-forming 45/47S rDNA unit and a substantially smaller extra-nucleolar 5S rDNA (Figure 1). Both the 45S and 5S rDNAs are organized into clusters of repeats often enabling their cytogenetic visualization on chromosomes [5]. The 5S rDNA can also (co)exist scattered separately within the genome e.g., in the spotted gar as shown in Figure 2a and in other organisms in S1–S4. In budding yeast, the rDNA has a very peculiar organization—the 5S rDNA unit is present in the intergenic spacers (IGSs) of the 45S rDNA and thus, alternating with 45S rDNA units [6]. The coding rDNA sequence is highly conserved among eukaryotes, while the IGSs (Figure 1) that separate the proper units of the 45S rDNA cluster can differ in length and sequence.

In budding yeast, where rDNAs are particularly well-described, IGSs contain three unique elements that are common: an origin of replication, a replication fork blocking site and a promoter that directs the synthesis of noncoding transcripts [7]. In mammals, IGSs contain regulatory regions called UCE (upstream control element), CP (core promoter) and T (termination of transcription site) [8]. Only a fraction of the numerous rDNA copies is transcribed into rRNA. The non-transcribed rDNA copies are extremely important for integrity of the entire genome [7]. In yeast, strains with artificially reduced rDNA copy numbers became sensitive to DNA damage by chemicals and ultraviolet light. This sensitivity further increased as the number of rDNA repeats decreased [7]. In rats, mice, and clawed frog *Xenopus*, the IGS contains one or more RNA polymerase I (Pol I) promoters with high homology to the core region of the main rDNA promoter [9]. Transcripts originating from spacer promoters are co-directional with pre-rRNA synthesis and enhance transcription from the main rDNA promoter, possibly by releasing Pol I [10,11]. Intergenic spacers rRNA have a crucial function in rDNA silencing. In mice, intergenic transcripts originating from a promoter located approximately 2 kb upstream from the pre-rRNA start site are processed into a heterogeneous population of 150–250 nucleotide RNAs, dubbed promoter RNA (pRNA) as their sequence matches the rDNA promoter [6,8,9,11].

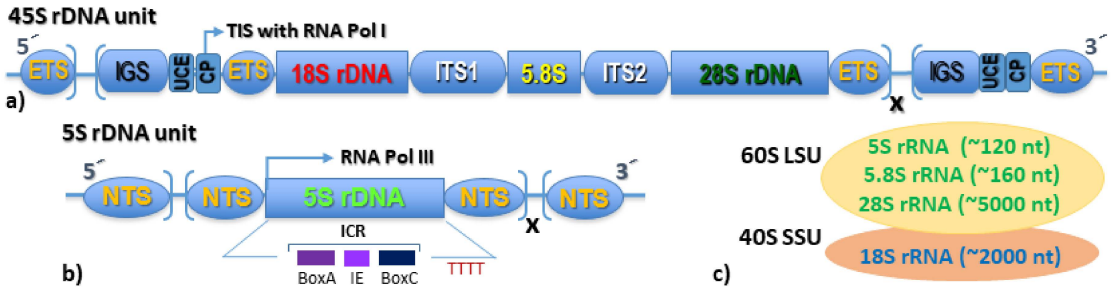


Figure 1. Brief guide to eukaryotic rDNAome—the genomic organization of the rDNA loci. (**a**) Structural organization of the 45S rDNA gene cluster (or rRNA transcription unit); the repeating or single clusters of rDNA can be found scattered throughout genome, they form the precursor pre-rRNA since ribonucleases remove spacers and release separate rRNA molecules in nucleolus—the site of ribosome biogenesis to polysome ribosome formation; (**b**) Structural organization of the 5S rDNA unit (the 5S rDNA can be also dispersed in the genome in many species); (**c**) 80S eukaryotic ribosome composed of the large subunit (LSU) and the small subunit (SSU) with outlined rRNAs. CP—core promoter, ETS—external transcribed spacer, ICR—internal control region, IE—internal element, IGS—intergenic spacer, ITS1, ITS2—internal transcribed spacer 1 and 2, RNA Pol I and III—RNA polymerase I and III, LSU—large (ribosomal) subunit, nt—nucleotides, NTS—non-transcribed spacer, SSU—small subunit, TIS—transcription initiation site, TTTT—polyT transcription termination site, UCE—upstream control element.

2. The Multifaceted Nucleolus

Multiple copies of rRNA gene clusters form nucleolar organizer regions (NORs), the NORs, around which nucleoli are built in the interphase nucleus. The nucleolar 45S rDNAs are transcribed by RNA polymerase I into rRNAs, further processed and assembled with ribosomal proteins into ribosomes [12]. Nucleoli form at the end of mitosis and persist until the onset of the next mitosis. Active nucleoli, where the pre-rRNA transcription takes place, can be visualized in nuclei by silver impregnation, the argyrophilic Ag-NOR staining [13]. From the ultrastructural viewpoint, avian and mammalian nucleoli contain three components (fibrillar centers, dense fibrillar component, and granular component) and differ from all other eukaryotes that possess bipartite nucleoli (i.e., a network of fibrillary strands embedded within granules) [14]. Interestingly, both types of nucleolar arrangement occur among living reptiles: a bicompartmentalized nucleolus in turtles and a tricompartmentalized nucleolus in lizards, crocodiles, and snakes [15]. From the functional viewpoint, the nucleolus was long regarded as a mere ribosome-producing factory. However, during recent decades numerous and crucial non-ribosomal roles were described for the nucleolus [4]. Now, there is a still growing body

of evidence that the nucleolus is central to cellular processes as varied as stress response, cell cycle regulation, RNA modification, cell metabolism, and genome stability and integrity [7]. All organisms sense and respond to stressing conditions by downregulating the transcription of rDNA to rRNA and ribosome biogenesis as these processes are extremely energy-consuming [4].

3. The Nucleolus Forming 45S rDNA

The 45S rDNA transcription unit forms a precursor pre-rRNA consisting of 18S, 5.8S, and 28S rRNAs separated by two internal transcribed spacers (ITS1, ITS2) that are removed during the rRNAs maturation process. The entire unit is further delimited by external transcribed spacers (ETS). Intergenic spacers (IGS; Figure 1a) separate each such unit with both of its sides bearing important regulatory elements [16]. This nomenclature applies to the Animal Kingdom. In plants, there is a 25S rDNA gene (instead of the 28S rDNA) within the large nucleolar rDNA multigene family 35S rDNA (instead of the 45S rDNA [17]). In unicellular organisms, where budding yeasts are the most important model system, the 35S rDNA consists of 25S, 5.8S, and 18S together with the 5S rDNA localized into the intergenic spacer within the 35S rDNA [18]. By addition of some 50–60 ribosomal proteins, the 25S/28S, 5.8S, and 5S rRNAs are fashioned into the large ribosomal subunit, 60S LSU. The 18S rRNA associates with 30–40 ribosomal proteins to form the small ribosomal subunit, 40S SSU (Figure 1c). Molecular cytogenetic localization of the rDNAs 28S rDNA fraction of the 45S rDNA unit on chromosomes is shown on Figure 2b.

The importance of nucleolus-forming rDNA and its proper functioning can be seen in the phenomenon of nucleolar dominance. Nucleolar dominance (or a nucleolus under-development due to expression of rDNA from just one parent) is a dramatic disruption in the formation of nucleoli and epigenetically controlled silencing of 45S rDNA in one of the progenitors in an interspecies hybrid. It is characteristic of some plant and animal distant hybrids and represents an example of a non-mammalian maternal imprinting of 45S rDNA [19,20]. Among animals it has been so far evidenced in details in intra-generic hybrids of *Xenopus* [20,21], in inter-generic hybrids of cyprinid fish [22,23] and in two lines of mouse-human somatic hybrids, where the human ribosomal genes were repressed, and only mouse ribosomal genes were expressed [24]. In the species *Drosophila melanogaster*, a special example of allelic inactivation resembling nucleolar dominance exists [25]. *D. melanogaster* carries its rDNA array on the X and on the Y chromosome [26], but the entire X chromosome rDNA array is normally silenced in *D. melanogaster* males, while the Y chromosome rDNA array is dominant and expressed [25].

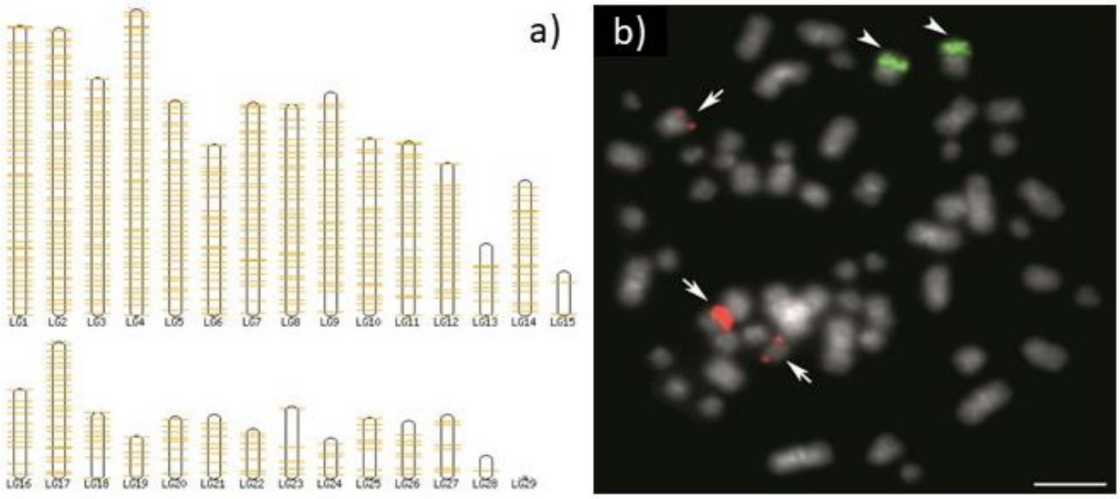


Figure 2. Comparison of two approaches of localization of rDNA on linkage groups and chromosomes in an ancient non-teleost ray-finned fish, spotted gar *(Lepisosteus oculatus*). (**a**) An in silico approach of visualization of the genomic position of rDNA loci utilizing the Ensembl genome browser tool BioMart to map 5S rDNA on linkage groups (LGs); (**b**) molecular cytogenetic localization of 5S (green, arrowheads) and 28S rDNA (red, arrows) on chromosomes by means of *fluorescence* in situ hybridization (FISH). Bar equals 5 μm (From [27], online Supplementary Material). This comparison shows the sensitivity of the in silico approach. The method enables detection of a single 5S rDNA molecule. It is possible to visualize dispersed molecules across the genome and their pseudogenes in this case. The FISH approach is limited only to huge clusters of accumulated rDNAs and has been utilized for decades particularly in cytotaxonomy in fishes, where most other cytogenetic markers work poorly. Both approaches have their own importance and justification and limits of their mutual interconnection at the current level of genomic data quality—chromosome pairs have not yet been assigned to their corresponding LGs in still too many of the sequenced species. More examples of 5S rDNA localization across LGs are available in Supplementary Materials Figures S1–S4.

4. The Extra-Nucleolar 5S rDNA

The 5S rDNA is much shorter and far less complex within its tandem array structure in comparison with the 45S rDNA. The 5S rDNA consists of a highly conserved sequence of about 120 bp coding for the 5S rRNA and including following functional elements: Box A, IE, Box C [28]. This transcribed sequence is separated at both of its ends from other transcriptional units by a highly variable non-transcribed spacer (NTS; Figure 1b). The participation of 5S rRNA on the ribosome structure is shown on Figure 1c. The 5S rRNA enhances protein synthesis by stabilizing the ribosomal structure and the peptidyl transferase activity, and potentially transmits and coordinates functional centres of the ribosome [29,30]. Two ways for visualization of 5S rDNA sites in silico on linkage groups (LGs) utilizing genomic data and employing methods of molecular cytogenetics on chromosomes are shown in Figure 2. More examples of in silico visualization of 5S rDNA on LGs are in the Supplementary Materials Figures S1–S4.

Two tissue and developmentally specific types of 5S rDNA exist in lower vertebrates, namely the somatic and the oocytes-specific ones. In fish, the oocyte type is lost during development completely (e.g., [31]). Whereas in a frog, it is lost largely [32]. The oocyte repeat comprises a 120 bp oocyte-type 5S rRNA gene placed within the few hundred bp long native AT-rich flanks, whereas the somatic repeat, i.e., a similar 120 bp somatic-type 5S rRNA gene is placed within native GC-rich flanks [33]. Instability of the oocyte 5S rRNA gene transcription complex contributes to the inactivation of the oocyte 5S rRNA gene during embryogenesis [34].

Moreover, in bony and cartilaginous fish, two types of co-occurring 5S rDNA can be distinguished. These probably paralogous type I and II were described in bony fish [35,36] and in elasmobranchs [28]. These types differ by the length and the sequence of their NTS region whereby the longer version is designated as type II [28]. Three functional variants of 5S rRNA genes exist in all life stages of common

sea urchin *Paracentrotus lividus* [37]. Systematic study of 5S rDNA sequence diversity in 97 metazoan species [38] describe several paralogous 5S rDNA sequences in 58 of the examined organisms and a flexible genome organization of 5S rDNA in animals. This study also describes three different types of termination signals and variable distances between the coding regions and the typical termination signal. Importantly, a consensus sequence and secondary structure of metazoan 5S rRNA is presented in this study [38], which can be very useful in more detailed future studies of both 5S rDNA and rRNA.

5. Copy Number Really Matters

Gene duplication is an important and frequent evolutionary process [39] and the resulting copy number variation (CNV) is the most frequent type of genetic variation per base pair in the population [40]. Although alteration of gene copy number or gene dosage has deleterious effects for a significant fraction of the genome, changes in dosage are well tolerated in many genes (reviewed by [41]). CNV of rDNA is highly studied in rDNAomics since it provides a mechanism for cellular homeostasis and for rapid and above all reversible adaptation [42–44]. Due to the tandem repetitive structure of rDNA, the repeat number can be easily reduced by homologous recombination among the repeats. However, there is a finely tuned ´gene amplification system´ compensating for these losses and another highly sophisticated system controlling the ´proper´ rDNA copy number [7]. Moreover, these systems are capable of linking external nutrients availability with rDNA copy number [45] that proves the role of rDNA in the cellular energy metabolism as described for nucleolus above. This illustrates how crucial the right copy number of rDNA is for each cell. Moreover, as uncovered at least in humans, the CNV of rDNA represents a novel and cryptic source of hypervariable genomic diversity with far-reaching global regulatory consequences [42]. However, we have accumulated only limited understanding of the immense importance of these seemingly passive and simple phenomena tightly linked with regulation of nuclear as well as mitochondrial genes expression [42] and probably with many more essential cellular mechanisms. There is an inconsistency in the quantification of rDNA copy number even in the human genome. One important study states that the rDNA copy number varies among healthy humans as a result of natural genetic diversity between 14–410 copies of the 45S rDNA unit per genome [41]. A more recent study reported that the number of rDNA repeats varies from 250 to 670 copies per diploid genome [46]. Therefore, data provided by Gibbons et al. [41,45] should be treated with caution as the low limit number of 14 copies has not been otherwise found in mammals. The genomes of higher eukaryotes harbour hundreds and thousands of copies and only prokaryotic genomes can carry fewer copies of ribosomal genes. Moreover, given that there are five pairs of clusters of ribosomal genes, located on five pairs of human acrocentric chromosomes (13, 14, 15, 21, 22), it is logically impossible that ten clusters could totally count 14 copies, i.e., just 1.4 copies per cluster. These obviously underestimated values can be explained as artifacts caused by a poor suitability of PCR-based techniques for the quantification of GC-rich moderate repeats used by Gibbons et al., 2014, 2015 [42,47] compared to more suitable nonradioactive quantitative hybridization (NQH) used by Chestkov et al., 2018 [46]. The reason is that rDNA is a specific region often forming non-canonical hairpin and loop structures and prone to oxidation in vivo and after extraction from cells [46]. Phenotypic effects of rDNA copy number were recently summarized by [48]. CNV of rDNA (loss as well as amplification) is linked to tumorigenesis [49,50].

A substantial intra-species CNV of rDNA was, among others, evidenced in a freshwater microcrustacean *Daphnia* [51]. rDNA copy number does change among tissues and during ontogenesis in multicellular organisms [31,52] and it is age-dependent in single-cell yeasts [7]. Substantial differences in rDNA CN and number of their sites on chromosomes have been repeatedly recorded in vertebrates and invertebrates on the inter-population and inter-species level of comparison [5]. Such genomic differences might also contribute to genome diversifications, reproductive barriers formation, speciation events and finally to an increased biodiversity, e.g., [53,54]. Numerous examples from fish cytogenetics show that the variation in rDNA repeats prove to be highly informative as it is subject to a more relaxed regulation than in higher vertebrates [55]. Here, traditional cytogenetics meet the currently booming

genomics to mutual usage and benefit from each other. The Animal rDNA database currently contains 539 records on fish rDNAs, namely 5S rDNA in 417 species and 45S rDNA in 479 species [5]. However, a detailed analysis of rDNA sequence organization and variation and CNV on the molecular level exists only for a handful of fish species including both 5S and 45S rDNA of zebrafish [31,51], 5S rDNA and only partial 45S rDNA of pike [56], only 5S rDNA of tilapia [57], molecular organization of the 5S rDNA type II of elasmobranchs (i.e., sharks, rays, and skates [28]) and cichlids [58]. In *Drosophila* germline stem cells, rDNA copy number decreases during aging and this age-dependent decrease in rDNA copy number is transgenerationally heritable. However, young animals are capable of recovering the normal rDNA copy number [59]. The copy number obviously plays a functional role: in *Xenopus*: the somatic 5S rDNA has about 400 copies, while the oocyte 5S rDNA has about 20,000 copies [32]. Locati et al. [31] detected about 9000 5S rRNA genes in the zebrafish genome assembly GRCz10 [31] and Symonova et al. detected about 20,000 copies of 5S rRNA genes in the Northern pike *Esox lucius* and its congener *E. cisalpinus* [56]. However, the record holders are currently protists, namely ciliates: Oligotrichia and Peritrichia [60] and representatives of the ciliate group Spirotrichea - *Oxytricha nova* with about 200,000 rDNA copies [61] and *Stylonychia lemnae* with estimated 400,000 copies of rDNA [62]. The already mentioned single-cell ciliate protozoan *Tetrahymena* amplifies its rDNA 9000-fold during development of the somatic macronucleus [63]. Whereas the copy number of 45S and 5S rDNA units is tightly coupled in mouse and human [47], such a control is apparently missing in fish [52,56,64]. This fact together with the aforementioned difference in nucleolar organization between higher and lower vertebrates and also other genomic traits (e.g., genomic GC heterogeneity) indicate that another major evolutionary transition *sensu* [65] occurred in evolution from anamniotes towards amniotes. This huge copy number variation might be linked to (or might have resulted in) the heterogeneity in rRNA genes and their variants that had been considered a peculiarity of some plants. Only recently, this heterogeneity was proved also in animal ribosomal genes, including human and mouse, where variant rRNA alleles exhibit tissue-specific expression and ribosomes bearing variant rRNA alleles are present in the actively translating ribosome pool [66].

One special topic of the rDNA CNV is based on molecular cytogenetic localization of both 5S and 45S rDNAs using FISH. FISH with rDNAs represents one of the most important chromosomal markers particularly in non-model organisms and especially in cold-blooded vertebrates, where methods like R-banding do not yield any usable and reproducible pattern. For these reasons, a heavy body of literature on molecular cytogenetics of rDNA has accumulated (for plants [17], for animals [5]). Since rDNA was omitted from many genome sequencing projects due to issues with its assembling [50,67], any precise quantification of rDNA copies is mostly still impossible. On the other hand, the still increasing availability of long-read sequencing can overcome assembling issues and provides opportunity to link the numerous results from molecular cytogenetics with genomics as was successfully demonstrated in fish cytogenomics [54,56].

A very special chapter of the rDNAomics book deals with rDNA of eukaryotic microorganisms [67]. In their 2010 review, Torres-Machorro et al. present available information on both rDNA fractions from about hundred microbial eukaryotes and show an unexpected diversity in their genomic organization [68]. Later, Drouin and Tsang [69] focus their review of 5S rDNA in protists on adaptive potential of its organization. Microbial eukaryotic rDNAs may be coded alone, in tandem repeats, linked to each other or linked to other genes. They exist in the chromosome or extrachromosomally in linear or circular units and rDNA coding regions may contain introns, sequence insertions, protein-coding genes, or additional spacers [68]. The atypical structures of rDNA have been considered as exceptions. However, it is rather likely that these organisms have preserved variations in the organization of these versatile genes that may be considered as living records of evolution [68]. A huge step in establishing the functional significance of rDNA in evolution and in ecology of organisms has been performed in protists [60].

6. Overview of Important Facts about rDNA

The most import facts about rDNA can be summarized as follows: rDNA is ubiquitous and universal across Prokaryotes, Archaea, and Eukaryotes [70]. It has a high degree of functional and sequence conservation of rDNA genes [71]. At the same time, rDNA belongs to the most copy number-hypervariable genomic segments [42] and the tandemly repeated rDNA arrays are among the most evolutionary dynamic loci of eukaryotic genomes in terms of copy number. Due to its heavy transcription, repetitive structure, and programmed replication fork pauses, the rDNA is one of the most unstable regions in the genome [7,18]. Their high genomic copy number relative to other genes appears to be much larger than required, however, unlike protein-coding genes, rDNA cannot undergo additional rounds of amplification via translation when organisms require more rRNA transcripts [72]. Copy number of 45S units is balanced with that of the 5S rDNA in mouse and human [47] but not in fish, summarized by [5]. These multiple copies of rDNA evolve in a highly coordinated manner, through unequal crossing over and/or gene conversion, two mechanisms related to homologous recombination [73]. The rRNA gene repeats use a unique gene amplification system to restore the copy number after this has been reduced due to recombination [7]. The RFB (replication fork barrier) coordinates replication and recombination, and through the latter, mediates a possible increase in the number of rDNA repeats. rDNA loci are dynamic genetic elements, their copy number changes dynamically and transgenerationally yet is maintained through a recovery mechanism in the germline (for *Drosophila* see [59]). In plants, extensive variation can exist in both rDNA copy number and rRNA expression. Among maize inbred lines, thousands of genes co-regulate with rRNA expression, including genes participating in ribosome biogenesis and other functionally relevant pathways [74]. Not only the rDNA copy number [45] but also the rRNA expression variation is a valuable source of functional diversity that affects gene expression variation and field-based phenotypic changes [74]. The intra-genomic homogenization of rDNA mostly occurs through 'concerted evolution' [75]. rDNA also shows high rates of meiotic recombination [75,76] and rDNA sites are hotspots for genome rearrangements [77]. Copy number of rDNA arrays modulates genome-wide expression of hundreds to thousands of genes and subtle changes in rDNA copy number between individuals may contribute to biologically relevant phenotypic variation also in humans [78]. rDNA contributes to global chromatin regulation and thus to a balance between heterochromatin and euchromatin in the nucleus [79]. The enormous variation in the number of rDNA copies per eukaryotic genome correlates with genome size [80] and the copy number of the 45S rDNA fraction was shown to negatively correlate with mtDNA abundance [42]. Hence, rDNA copy number variation, CNV ("rDNA dosage") is a major determinant of naturally occurring genome-wide gene expression variation in humans [42]. Ribosomal RNAs (rRNAs) account for up to 80% of all RNAs in eukaryotic cells [50]. Growth-activated rRNA synthesis may be mediated by the up-regulation of individual rDNA units, in addition to the activation of silent gene copies see e.g., Banditt et al. [81]. In mammals, 5S and 45S rDNA arrays are non-homologous, physically unlinked, transcribed by different RNA Polymerases and encode functionally interdependent RNA components of the ribosome [47]. Clusters of the 45S rDNA unit give origin to the nucleolus, the nuclear organelle that is the site of pre-45S rDNA transcription and ribosome biogenesis, see e.g., [8,82]. The rDNA contact map shows that 5S and 45S arrays each have thousands of contacts in the folded genome, with rDNA-associated regions and genes dispersed across all chromosomes [83,84]. Due to its highly repetitive nature, rDNA has been excluded from most mammalian genome-wide studies because of challenges associated with its analysis, and thus remains understudied. There is an unusual and universal GC richness of the 45S rDNA fraction in cold- as well as warm-blooded vertebrates (more details below) [72,84]. An extensive range of epigenetic modifications regulating rRNA genes transcription [67,85–87] results in only a mere subset of the multiple copies being transcribed however with far reaching implications for the entire genome (elucidation of the epigenetics of rDNA in sufficient detail would require a lot of research). On top of it, rDNA loci serve as a specialized niche for mobile elements [88].

rDNA units (so called 'rDNA-like signal') can be found scattered throughout the genome in humans [89]. These units can be described as follows: 1) highly degraded, but near full length, rDNA units, including both 45S and Intergenic Spacer (IGS), can be found at multiple sites in the human genome on chromosomes without rDNA arrays; 2) these rDNA sequences have a propensity for being centromere proximal; and 3) sequence at all human functional rDNA array ends is divergent from canonical rDNA to the point that it is pseudogenic. For this in fish, see Figure 2a and in other chordates see the Supplementary Materials Figures S1–S4.

rDNA represents a cryptic source of hypervariable genomic diversity with global regulatory consequences (ribosomal quantitative trait loci (eQTL)) in humans. The variation provides a mechanism for cellular homeostasis and for rapid and reversible adaptation [42,47].

7. GC Content of rDNA

The 45S rDNA gene clusters form the GC-richest genomic fraction particularly in Eukaryotes [90] with humans having 60%–80% GC in different parts of the rDNA [91], whereas the median genomic GC is 40.9% (NCBI, human genome assembly). This GC-richness is ascribed to the recombination rate based process known as GC-biased gene conversion [73]. On the other hand, some studies link the extremely high GC levels in rDNAs to particular requirements for stem-and-loop systems in rRNA that have an effect on the composition of the corresponding genes to thermal adaptation [92]. This line of explanations belongs to the Thermodynamic Stability Hypothesis attempting to account for the overall AT/GC heterogeneity in birds and mammals and the AT/GC homogeneity in the remaining vertebrates and the other Eukaryotes [90]. However, although Wang et al., 2006 showed support for such a thermal adaptation in Bacteria and Archaea, they did not find any for warm-blooded birds and mammals with only a slightly higher GC content of 18S (55.7%) versus cold-blooded fishes and amphibians with approximately 53.5% of GC. Their partitioning of the GC content across the 18S rRNA sequences into stem and loop regions demonstrated [93] that the differences are not concentrated in the paired stem regions as expected by Bernardi [90]. Interesting and relevant aspects of rDNA GC content exist in so-called expansion segments (ES) in 28S and 18S rRNA molecules [94,95]. Expansion of the 28S rRNA shows a clear phylogenetic increase, with a dramatic rise in mammals and especially in hominids. Here, a GC- or AU-biased expansion of rRNAs has developed in both plants and metazoans, with the GC-bias largely being preferred in extremely GC-rich ES of vertebrate 28S rRNA. This compositional bias towards GC is linked to potential roles of GC-rich rRNA during protein synthesis [96,97] and could contribute to the discussion whether the genomic GC content is driven by neutral versus selective processes. An interesting explanation of the universal GC richness of 45S rDNA comes from the GC biology—these multicopy genes should all be in a DNA region with a homogenous GC composition to allow concerted evolution and to prevent divergence through generations [98].

8. Phylogeny, Species Delimitation, and Secondary Structure of rRNAs—The Way How to Determine in Silico Whether Two Lineages Can Successfully Cross

The ITS2 sequence already belongs to the most popular and well established phylogenetic and DNA barcoding markers [99]. The rDNA sequence and its corresponding rRNA secondary structure is one of the few universal features of life without any known case of horizontal transfer and above all, identifying the organism to a unique species, making it uniquely suited to assess phylogenetic relationships [100,101]. The secondary structure of the ITS regions is well known for a wide variety of eukaryotes and have been used to aid in the alignment of these sequences for phylogenetic comparisons [101]. The RNA sequence of the ITS2 possesses another special trait so far not fully examined, namely compensatory base changes (CBCs, Figure 3). CBCs are mutations occurring simultaneously on both sides of a nucleotide pair in the ITS2 secondary structure with retention of the paired nucleotide bond, whereas hemi-CBC is a mutation of a single nucleotide of the pair still retaining the bond [102]. CBC analyses have been primarily performed in fungi and plants [92–94]. This is the reason why the majority of literature references, including methods descriptions, are on

plants (e.g., estimating structure-based phylogenetic trees from ITS2 data by [103–105]). CBCs analyses have been already successfully used to verify taxonomy of closely related species and to distinguish morphologically indistinct species in insects [102]. That shows the huge potential of mining for CBCs in ITS2 rRNA secondary structures also in the Animal Kingdom.

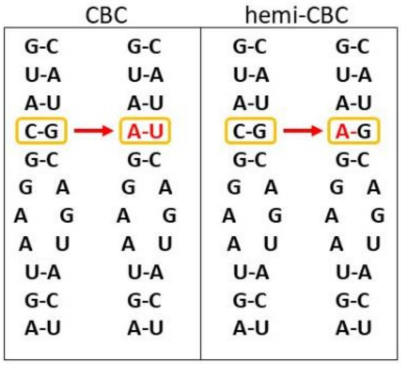


Figure 3. Visualization of compensatory base changes (CBCs) on a hypothetical internal transcribed spacer 2 (ITS2) secondary structure. The left panel depicts a conserved helix segment in which a CBC occurs, where both nucleotides of the pair underwent a mutation that resulted in retaining the paired nucleotide bond. The right panel shows a hemi-CBC, where only one nucleotide of the pair, i.e., C to A, underwent a mutation while the pair retained the nucleotide bond. Redrawn according to [106].

However, any detailed and particularly systematic survey of rRNA secondary structure in the Animal Kingdom is still in its infancy although it would be highly desirable in numerous areas of biology. Moreover, analyses of rRNA secondary structure could represent another intersection between rDNAomics based on molecular cytogenetics and on genomics since molecular cytogenetic studies frequently provide DNA sequences of rDNA fragments used in FISH experiments and further DNA sequences, in the meanwhile, became available in NCBI GenBank or could be retrieved from whole-genome datasets. Ideally, such integrative studies could contain cytogenetic results accompanied by details on rDNA/rRNA sequence and rRNA secondary structure as shown in Figure 4 to better explore any potential sequence polymorphism.

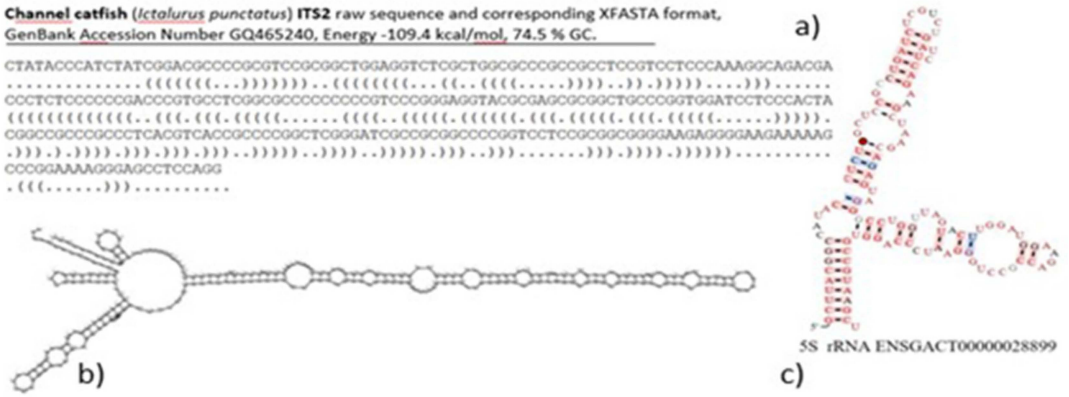


Figure 4. Text and graphic representation of prediction of rRNA secondary structure. (**a**) the Xfasta or "dot-bracket notation" way of representation of the ITS2 rRNA secondary structure of channel catfish, (**b**) visualization of the corresponding secondary structure of the sequence in (**a**) with one longest helix and four short helices, (**c**) stickleback, prediction of 5S rRNA secondary structure, Ensemble.

9. Concluding Remarks

The field of rDNAomics is extremely rapidly evolving further and has far-reaching implications for numerous areas of current biology and medicine. Medical aspects represent another crucial chapter

of rDNAomics that already exceeds the scope of this review. On the other hand, being aware of this fact might help scientists from the area of fundamental research working on non-model organisms to provide justification of their work. There are numerous diseases associated with rDNA dysfunction, particularly cancer [39,40,61,66]. Ribosomopathies are diseases caused by abnormalities in the structure or function of ribosomal component proteins or rRNA genes, or other genes whose products are involved in ribosome biogenesis [107]. Not only sequence, but also copy number of rDNAs is of particular importance in cancer—human cancer genomes show a loss of copies, accompanied by global copy number co-variation [50]. Even more relevant is the fact that rDNA repeat instability coincides with predisposition to cancer, premature aging and neurological impairment in ataxia-telangiectasia and Bloom syndrome (Warmerdam and Wolthuis, 2018). Additionally, it was shown that cancers undergo coupled 5S rDNA array expansion and 45S rDNA loss that is accompanied by increased proliferation rate and nucleolar activity. Somatic changes in rDNA copy number can exceed 10-fold the naturally occurring copy number variation across individuals [49]. Malfunction of nucleoli can be the cause of several human conditions called nucleolopathies [108]. The nucleolus is being investigated as a target for cancer chemotherapy [109,110]. Moreover, rDNA copy number may be a simple and useful indicator of whether a cancer will be sensitive to DNA damaging treatments [50]. Hence, it is desirable to understand rDNA organization, function, and its impact on the entire nucleolus and other genes' regulation (as briefly outlined here) in a broader evolutionary context.

Supplementary Materials: The following are available online at http://www.mdpi.com/2073-4425/10/5/345/s1, Figure S1: *Tetraodon nigroviridis*, karyogram with labeled 5S rDNA loci (red arrowheads). Figure S2: Zebrafish (*Danio rerio*), complete karyogram with labeled 5S rDNA loci (pink lines). Figure S3: A tunicate sea squirt (*Ciona intestinalis*), C ≈ 0.2, the thirty-five 5S rDNA site detected on four chromosomes could be assigned to a linkage group. Figure S4: Human genome illustrates here the best assembled vertebrate genome, C ≈ 3.5. Genomic localization of annotated 5S rDNA showing 5S rDNA scattered throughout fish genomes visualized on karyograms of species assembled to the chromosome level (5S rDNA sequences were filtered using the BioMart tool and the Ensemble Genes Database version 92 from Ensembl.org version 92). For comparison, two model mammalian genomes (Hsa, Mmu) and one avian genome are shown.

Funding: This study was supported by IRP PřF UHK 1903/2018.

Acknowledgments: I would like to thank Mike W. Howell for his insightful comments to this manuscript.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Bernhardt, H.S.; Tate, W.P. A Ribosome Without RNA. *Front. Ecol. Evol.* **2015**, *3*, 1–6. [CrossRef]
2. Boisvert, F.-M.; van Koningsbruggen, S.; Navascués, J.; Lamond, A.I. The multifunctional nucleolus. *Nat. Rev. Mol. Cell Biol.* **2007**, *8*, 574–585. [CrossRef] [PubMed]
3. Porokhovnik, L.; Gerton, J.L. Ribosomal DNA-connecting ribosome biogenesis and chromosome biology. *Chromosome Res. Int. J. Mol. Supramol. Evol. Asp. Chromosome Biol.* **2019**. [CrossRef] [PubMed]
4. Grummt, I. The nucleolus—guardian of cellular homeostasis and genome integrity. *Chromosoma* **2013**, *122*, 487–497. [CrossRef] [PubMed]
5. Sochorová, J.; Garcia, S.; Gálvez, F.; Symonová, R.; Kovařík, A. Evolutionary trends in animal ribosomal DNA loci: introduction to a new online database. *Chromosoma* **2018**, *127*, 141–150. [CrossRef]
6. James, S.A.; O'Kelly, M.J.T.; Carter, D.M.; Davey, R.P.; van Oudenaarden, A.; Roberts, I.N. Repetitive sequence variation and dynamics in the ribosomal DNA array of Saccharomyces cerevisiae as revealed by whole-genome resequencing. *Genome Res.* **2009**, *19*, 626–635. [CrossRef] [PubMed]
7. Kobayashi, T. Ribosomal RNA gene repeats, their stability and cellular senescence. *Proc. Jpn. Acad. Ser. B* **2014**, *90*, 119–129. [CrossRef]
8. Russell, J.; Zomerdijk, J.C.B.M. RNA-polymerase-I-directed rDNA transcription, life and works. *Trends Biochem. Sci.* **2005**, *30*, 87–96. [CrossRef] [PubMed]
9. Santoro, R.; Schmitz, K.-M.; Sandoval, J.; Grummt, I. Intergenic transcripts originating from a subclass of ribosomal DNA repeats silence ribosomal RNA genes in trans. *EMBO Rep.* **2010**, *11*, 52–58. [CrossRef] [PubMed]

10. Grimaldi, G.; Di Nocera, P.P. Multiple repeated units in Drosophila melanogaster ribosomal DNA spacer stimulate rRNA precursor transcription. *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 5502–5506. [CrossRef]
11. Mayer, C.; Neubert, M.; Grummt, I. The structure of NoRC-associated RNA is crucial for targeting the chromatin remodelling complex NoRC to the nucleolus. *EMBO Rep.* **2008**, *9*, 774–780. [CrossRef]
12. Németh, A.; Längst, G. Genome organization in and around the nucleolus. *Trends Genet. TIG* **2011**, *27*, 149–156. [CrossRef] [PubMed]
13. Howell, W.M.; Black, D.A. Controlled silver-staining of nucleolus organizer regions with a protective colloidal developer: A 1-step method. *Experientia* **1980**, *36*, 1014–1015. [CrossRef] [PubMed]
14. Thiry, M.; Lafontaine, D.L.J. Birth of a nucleolus: The evolution of nucleolar compartments. *Trends Cell Biol.* **2005**, *15*, 194–199. [CrossRef]
15. Thiry, M.; Lamaye, F.; Lafontaine, D.L.J. The nucleolus: When 2 became 3. *Nucleus* **2011**, *2*, 289–293. [CrossRef]
16. Henras, A.K.; Plisson-Chastang, C.; O'Donohue, M.-F.; Chakraborty, A.; Gleizes, P.-E. An overview of pre-ribosomal RNA processing in eukaryotes. *Wiley Interdiscip. Rev. RNA* **2015**, *6*, 225–242. [CrossRef] [PubMed]
17. Garcia, S.; Garnatje, T.; Kovařík, A. Plant rDNA database: Ribosomal DNA loci information goes online. *Chromosoma* **2012**, *121*, 389–394. [CrossRef] [PubMed]
18. Kobayashi, T.; Sasaki, M. Ribosomal DNA stability is supported by many 'buffer genes'—introduction to the Yeast rDNA Stability Database. *FEMS Yeast Res.* **2017**, *17*. [CrossRef]
19. Pikaard, C.S. Nucleolar dominance: Uniparental gene silencing on a multi-megabase scale in genetic hybrids. *Plant Mol. Biol.* **2000**, *43*, 163–177. [CrossRef] [PubMed]
20. Michalak, K.; Maciak, S.; Kim, Y.B.; Santopietro, G.; Oh, J.H.; Kang, L.; Garner, H.R.; Michalak, P. Nucleolar dominance and maternal control of 45S rDNA expression. *Proc. R. Soc. B Biol. Sci.* **2015**, *282*, 20152201. [CrossRef]
21. Maciak, S.; Michalak, K.; Kale, S.D.; Michalak, P. Nucleolar Dominance and Repression of 45S Ribosomal RNA Genes in Hybrids between ***Xenopus borealis*** and ***X. muelleri*** (2n = 36). *Cytogenet. Genome Res.* **2016**, *149*, 290–296. [CrossRef] [PubMed]
22. Xiao, J.; Hu, F.; Luo, K.; Li, W.; Liu, S. Unique nucleolar dominance patterns in distant hybrid lineage derived from Megalobrama Amblycephala × Culter Alburnus. *BMC Genet.* **2016**, *17*. [CrossRef]
23. Cao, L.; Qin, Q.; Xiao, Q.; Yin, H.; Wen, J.; Liu, Q.; Huang, X.; Huo, Y.; Tao, M.; Zhang, C.; et al. Nucleolar Dominance in a Tetraploidy Hybrid Lineage Derived From Carassius auratus red var. () × Megalobrama amblycephala (). *Front. Genet.* **2018**, *9*. [CrossRef] [PubMed]
24. Onishi, T.; Berglund, C.; Reeder, R.H. On the mechanism of nucleolar dominance in mouse-human somatic cell hybrids. *Proc. Natl. Acad. Sci. USA* **1984**, *81*, 484–487. [CrossRef] [PubMed]
25. Greil, F.; Ahmad, K. Nucleolar Dominance of the *Y* Chromosome in *Drosophila melanogaster*. *Genetics* **2012**, *191*, 1119–1128. [CrossRef] [PubMed]
26. Tautz, D.; Hancock, J.M.; Webb, D.A.; Tautz, C.; Dover, G.A. Complete sequences of the rRNA genes of Drosophila melanogaster. *Mol. Biol. Evol.* **1988**, *5*, 366–376. [PubMed]
27. Symonová, R.; Majtánová, Z.; Arias-Rodriguez, L.; Mořkovský, L.; Kořínková, T.; Cavin, L.; Pokorná, M.J.; Doležálková, M.; Flajšhans, M.; Normandeau, E.; et al. Genome Compositional Organization in Gars Shows More Similarities to Mammals than to Other Ray-Finned Fish: CYTOGENOMICS OF GARS. *J. Exp. Zoolog. B Mol. Dev. Evol.* **2017**, *328*, 607–619. [CrossRef]
28. Castro, S.I.; Hleap, J.S.; Cárdenas, H.; Blouin, C. Molecular organization of the 5S rDNA gene type II in elasmobranchs. *RNA Biol.* **2016**, *13*, 391–399. [CrossRef]
29. Szymanski, M.; Barciszewska, M.Z.; Erdmann, V.A.; Barciszewski, J. 5S Ribosomal RNA Database. *Nucleic Acids Res.* **2002**, *30*, 176–178. [CrossRef] [PubMed]
30. Dinman, J.D. 5S rRNA: Structure and Function from Head to Toe. *Int. J. Biomed. Sci. IJBS* **2005**, *1*, 2–7.
31. Locati, M.D.; Pagano, J.F.B.; Ensink, W.A.; van Olst, M.; van Leeuwen, S.; Nehrdich, U.; Zhu, K.; Spaink, H.P.; Girard, G.; Rauwerda, H.; et al. Linking maternal and somatic 5S rRNA types with different sequence-specific non-LTR retrotransposons. *RNA* **2017**, *23*, 446–456. [CrossRef]
32. Peterson, R.C.; Doering, J.L.; Brown, D.D. Characterization of two xenopus somatic 5S DNAs and one minor oocyte-specific 5S DNA. *Cell* **1980**, *20*, 131–141. [CrossRef]

33. Tomaszewski, R.; Jerzmanowski, A. The AT-rich flanks of the oocyte-type 5S RNA gene of Xenopus laevis act as a strong local signal for histone H1-mediated chromatin reorganization in vitro. *Nucleic Acids Res.* **1997**, *25*, 458–466. [CrossRef] [PubMed]
34. Wolffe, A.P.; Brown, D.D. Developmental regulation of two 5S ribosomal RNA genes. *Science* **1988**, *241*, 1626–1632. [CrossRef]
35. Martins, C.; Galetti, P.M. Two 5S rDNA arrays in neotropical fish species: is it a general rule for fishes? *Genetica* **2001**, *111*, 439–446. [CrossRef] [PubMed]
36. *Fish Cytogenetics*; Pisano, E. (Ed.) Science Publishers: Enfield, NH, USA, 2007; ISBN 978-1-57808-330-5.
37. Dimarco, E.; Cascone, E.; Bellavia, D.; Caradonna, F. Functional variants of 5S rRNA in the ribosomes of common sea urchin Paracentrotus lividus. *Gene* **2012**, *508*, 21–25. [CrossRef] [PubMed]
38. Vierna, J.; Wehner, S.; Höner zu Siederdissen, C.; Martínez-Lage, A.; Marz, M. Systematic analysis and evolution of 5S ribosomal DNA in metazoans. *Heredity* **2013**, *111*, 410–421. [CrossRef] [PubMed]
39. Conrad, B.; Antonarakis, S.E. Gene Duplication: A Drive for Phenotypic Diversity and Cause of Human Disease. *Annu. Rev. Genomics Hum. Genet.* **2007**, *8*, 17–35. [CrossRef]
40. The Wellcome Trust Case Control Consortium; Conrad, D.F.; Pinto, D.; Redon, R.; Feuk, L.; Gokcumen, O.; Zhang, Y.; Aerts, J.; Andrews, T.D.; Barnes, C.; et al. Origins and functional impact of copy number variation in the human genome. *Nature* **2010**, *464*, 704–712. [CrossRef]
41. Rice, A.M.; McLysaght, A. Dosage sensitivity is a major determinant of human copy number variant pathogenicity. *Nat. Commun.* **2017**, *8*, 14366. [CrossRef] [PubMed]
42. Gibbons, J.G.; Branco, A.T.; Yu, S.; Lemos, B. Ribosomal DNA copy number is coupled with gene expression variation and mitochondrial abundance in humans. *Nat. Commun.* **2014**, *5*. [CrossRef]
43. Long, E.O.; Dawid, I.B. Repeated genes in eukaryotes. *Annu. Rev. Biochem.* **1980**, *49*, 727–764. [CrossRef] [PubMed]
44. Oakes, M.; Siddiqi, I.; Vu, L.; Aris, J.; Nomura, M. Transcription factor UAF, expansion and contraction of ribosomal DNA (rDNA) repeats, and RNA polymerase switch in transcription of yeast rDNA. *Mol. Cell. Biol.* **1999**, *19*, 8559–8569. [CrossRef]
45. Jack, C.V.; Cruz, C.; Hull, R.M.; Keller, M.A.; Ralser, M.; Houseley, J. Regulation of ribosomal DNA amplification by the TOR pathway. *Proc. Natl. Acad. Sci.* **2015**, *112*, 9674–9679. [CrossRef] [PubMed]
46. Chestkov, I.V.; Jestkova, E.M.; Ershova, E.S.; Golimbet, V.E.; Lezheiko, T.V.; Kolesina, N.Y.; Porokhovnik, L.N.; Lyapunova, N.A.; Izhevskaya, V.L.; Kutsev, S.I.; et al. Abundance of ribosomal RNA gene copies in the genomes of schizophrenia patients. *Schizophr. Res.* **2018**, *197*, 305–314. [CrossRef] [PubMed]
47. Gibbons, J.G.; Branco, A.T.; Godinho, S.A.; Yu, S.; Lemos, B. Concerted copy number variation balances ribosomal DNA dosage in human and mouse genomes. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 2485–2490. [CrossRef]
48. Porokhovnik, L.N.; Lyapunova, N.A. Dosage effects of human ribosomal genes (rDNA) in health and disease. *Chromosome Res.* **2019**, *27*, 5–17. [CrossRef]
49. Wang, M.; Lemos, B. Ribosomal DNA copy number amplification and loss in human cancers is linked to tumor genetic context, nucleolus activity, and proliferation. *PLoS Genet.* **2017**, *13*, e1006994. [CrossRef]
50. Xu, B.; Li, H.; Perry, J.M.; Singh, V.P.; Unruh, J.; Yu, Z.; Zakari, M.; McDowell, W.; Li, L.; Gerton, J.L. Ribosomal DNA copy number loss and sequence variation in cancer. *PLoS Genet.* **2017**, *13*, e1006771. [CrossRef] [PubMed]
51. Eagle, S.H.; Crease, T.J. Copy number variation of ribosomal DNA and Pokey transposons in natural populations of Daphnia. *Mob. DNA* **2012**, *3*, 4. [CrossRef] [PubMed]
52. Locati, M.D.; Pagano, J.F.B.; Girard, G.; Ensink, W.A.; van Olst, M.; van Leeuwen, S.; Nehrdich, U.; Spaink, H.P.; Rauwerda, H.; Jonker, M.J.; et al. Expression of distinct maternal and somatic 5.8S, 18S, and 28S rRNA types during zebrafish development. *RNA* **2017**, *23*, 1188–1199. [CrossRef] [PubMed]
53. Dion-Cote, A.-M.; Symonova, R.; Rab, P.; Bernatchez, L. Reproductive isolation in a nascent species pair is associated with aneuploidy in hybrid offspring. *Proc. R. Soc. B Biol. Sci.* **2015**, *282*, 20142862. [CrossRef] [PubMed]
54. Dion-Côté, A.-M.; Symonová, R.; Lamaze, F.C.; Pelikánová, Š.; Ráb, P.; Bernatchez, L. Standing chromosomal variation in Lake Whitefish species pairs: The role of historical contingency and relevance for speciation. *Mol. Ecol.* **2017**, *26*, 178–192. [CrossRef]

55. Symonová, R.; Howell, W. Vertebrate Genome Evolution in the Light of Fish Cytogenomics and rDNAomics. *Genes* **2018**, *9*, 96. [CrossRef]
56. Symonová, R.; Ocalewicz, K.; Kirtiklis, L.; Delmastro, G.B.; Pelikánová, Š.; Garcia, S.; Kovařík, A. Higher-order organisation of extremely amplified, potentially functional and massively methylated 5S rDNA in European pikes (Esox sp.). *BMC Genomics* **2017**, *18*. [CrossRef]
57. Martins, C.; Wasko, A.P.; Oliveira, C.; Porto-Foresti, F.; Parise-Maltempi, P.P.; Wright, J.M.; Foresti, F. Dynamics of 5S rDNA in the tilapia *(Oreochromis niloticus)* genome: Repeat units, inverted sequences, pseudogenes and chromosome loci. *Cytogenet. Genome Res.* **2002**, *98*, 78–85. [CrossRef] [PubMed]
58. Nakajima, R.T.; Cabral-de-Mello, D.C.; Valente, G.T.; Venere, P.C.; Martins, C. Evolutionary dynamics of rRNA gene clusters in cichlid fish. *BMC Evol. Biol.* **2012**, *12*, 198. [CrossRef]
59. Lu, K.L.; Nelson, J.O.; Watase, G.J.; Warsinger-Pepe, N.; Yamashita, Y.M. Transgenerational dynamics of rDNA copy number in Drosophila male germline stem cells. *eLife* **2018**, *7*. [CrossRef] [PubMed]
60. Gong, J.; Dong, J.; Liu, X.; Massana, R. Extremely High Copy Numbers and Polymorphisms of the rDNA Operon Estimated from Single Cell Analysis of Oligotrich and Peritrich Ciliates. *Protist* **2013**, *164*, 369–379. [CrossRef]
61. Prescott, D.M. The DNA of ciliated protozoa. *Microbiol. Rev.* **1994**, *58*, 233–267. [CrossRef] [PubMed]
62. Heyse, G.; Jönsson, F.; Chang, W.-J.; Lipps, H.J. RNA-dependent control of gene amplification. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 22134–22139. [CrossRef]
63. Pan, W.-C.; Orias, E.; Flacks, M.; Blackburn, E.H. Allele-specific, selective amplification of a ribosomal RNA gene in tetrahymena thermophila. *Cell* **1982**, *28*, 595–604. [CrossRef]
64. Symonová, R.; Majtánová, Z.; Sember, A.; Staaks, G.B.; Bohlen, J.; Freyhof, J.; Rábová, M.; Ráb, P. Genome differentiation in a species pair of coregonine fishes: An extremely rapid speciation driven by stress-activated retrotransposons mediating extensive ribosomal DNA multiplications. *BMC Evol. Biol.* **2013**, *13*, 42. [CrossRef]
65. Szathmáry, E.; Smith, J.M. The major evolutionary transitions. *Nature* **1995**, *374*, 227–232. [CrossRef] [PubMed]
66. Parks, M.M.; Kurylo, C.M.; Dass, R.A.; Bojmar, L.; Lyden, D.; Vincent, C.T.; Blanchard, S.C. Variant ribosomal RNA alleles are conserved and exhibit tissue-specific expression. *Sci. Adv.* **2018**, *4*, eaao0665. [CrossRef]
67. Bughio, F.; Maggert, K.A. The peculiar genetics of the ribosomal DNA blurs the boundaries of transgenerational epigenetic inheritance. *Chromosome Res. Int. J. Mol. Supramol. Evol. Asp. Chromosome Biol.* **2018**. [CrossRef]
68. Torres-Machorro, A.L.; Hernández, R.; Cevallos, A.M.; López-Villaseñor, I. Ribosomal RNA genes in eukaryotic microorganisms: Witnesses of phylogeny? *FEMS Microbiol. Rev.* **2010**, *34*, 59–86. [CrossRef]
69. Drouin, G.; Tsang, C. 5S rRNA Gene Arrangements in Protists: A Case of Nonadaptive Evolution. *J. Mol. Evol.* **2012**, *74*, 342–351. [CrossRef]
70. Mallatt, J.; Chittenden, K.D. The GC content of LSU rRNA evolves across topological and functional regions of the ribosome in all three domains of life. *Mol. Phylogenet. Evol.* **2014**, *72*, 17–30. [CrossRef] [PubMed]
71. Agrawal, S.; Ganley, A.R.D. The conservation landscape of the human ribosomal RNA gene repeats. *PLOS ONE* **2018**, *13*, e0207531. [CrossRef] [PubMed]
72. Kobayashi, T. A new role of the rDNA and nucleolus in the nucleus—rDNA instability maintains genome integrity. *BioEssays* **2008**, *30*, 267–272. [CrossRef]
73. Escobar, J.S.; Glémin, S.; Galtier, N. GC-Biased Gene Conversion Impacts Ribosomal DNA Evolution in Vertebrates, Angiosperms, and Other Eukaryotes. *Mol. Biol. Evol.* **2011**, *28*, 2561–2575. [CrossRef]
74. Li, B.; Kremling, K.A.G.; Wu, P.; Bukowski, R.; Romay, M.C.; Xie, E.; Buckler, E.S.; Chen, M. Coregulation of ribosomal RNA with hundreds of genes contributes to phenotypic variation. *Genome Res.* **2018**, *28*, 1555–1565. [CrossRef]
75. Ganley, A.R.D.; Kobayashi, T. Highly efficient concerted evolution in the ribosomal DNA repeats: Total rDNA repeat variation revealed by whole-genome shotgun sequence data. *Genome Res.* **2007**, *17*, 184–191. [CrossRef]
76. Stults, D.M.; Killen, M.W.; Pierce, H.H.; Pierce, A.J. Genomic architecture and inheritance of human ribosomal RNA gene clusters. *Genome Res.* **2007**, *18*, 13–18. [CrossRef] [PubMed]

77. Stults, D.M.; Killen, M.W.; Williamson, E.P.; Hourigan, J.S.; Vargas, H.D.; Arnold, S.M.; Moscow, J.A.; Pierce, A.J. Human rRNA Gene Clusters Are Recombinational Hotspots in Cancer. *Cancer Res.* **2009**, *69*, 9096–9104. [CrossRef]
78. Paredes, S.; Branco, A.T.; Hartl, D.L.; Maggert, K.A.; Lemos, B. Ribosomal DNA deletions modulate genome-wide gene expression: "rDNA-sensitive" genes and natural variation. *PLoS Genet.* **2011**, *7*, e1001376. [CrossRef]
79. Paredes, S.; Maggert, K.A. Ribosomal DNA contributes to global chromatin regulation. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 17829–17834. [CrossRef]
80. Prokopowich, C.D.; Gregory, T.R.; Crease, T.J. The correlation between rDNA copy number and genome size in eukaryotes. *Genome* **2003**, *46*, 48–50. [CrossRef] [PubMed]
81. Banditt, M.; Koller, T.; Sogo, J.M. Transcriptional activity and chromatin structure of enhancer-deleted rRNA genes in Saccharomyces cerevisiae. *Mol. Cell. Biol.* **1999**, *19*, 4953–4960. [CrossRef] [PubMed]
82. Pederson, T. The Nucleolus. *Cold Spring Harb. Perspect. Biol.* **2011**, *3*, a000638. [CrossRef]
83. Yu, S.; Lemos, B. A Portrait of Ribosomal DNA Contacts with Hi-C Reveals 5S and 45S rDNA Anchoring Points in the Folded Human Genome. *Genome Biol. Evol.* **2016**, *8*, 3545–3558. [CrossRef] [PubMed]
84. Yu, S.; Lemos, B. The long-range interaction map of ribosomal DNA arrays. *PLoS Genet.* **2018**, *14*, e1007258. [CrossRef] [PubMed]
85. Bierhoff, H.; Postepska-Igielska, A.; Grummt, I. Noisy silence: Non-coding RNA and heterochromatin formation at repetitive elements. *Epigenetics* **2014**, *9*, 53–61. [CrossRef] [PubMed]
86. McStay, B.; Grummt, I. The Epigenetics of rRNA Genes: From Molecular to Chromosome Biology. *Annu. Rev. Cell Dev. Biol.* **2008**, *24*, 131–157. [CrossRef]
87. Schöfer, C.; Weipoltshammer, K. Nucleolus and chromatin. *Histochem. Cell Biol.* **2018**, *150*, 209–225. [CrossRef] [PubMed]
88. Eickbush, T.H.; Eickbush, D.G. Finely Orchestrated Movements: Evolution of the Ribosomal RNA Genes. *Genetics* **2007**, *175*, 477–485. [CrossRef] [PubMed]
89. Robicheau, B.M.; Susko, E.; Harrigan, A.M.; Snyder, M. Ribosomal RNA Genes Contribute to the Formation of Pseudogenes and Junk DNA in the Human Genome. *Genome Biol. Evol.* **2017**, *9*, 380–397. [CrossRef]
90. Bernardi, G. *Structural and Evolutionary Genomics: Natural Selection in Genome Evolution*; Elsevier: Amsterdam, The Netherlands, 2005; ISBN 978-0-08-046187-8.
91. Galtier, N.; Piganeau, G.; Mouchiroud, D.; Duret, L. GC-content evolution in mammalian genomes: The biased gene conversion hypothesis. *Genetics* **2001**, *159*, 907–911.
92. Varriale, A.; Torelli, G.; Bernardi, G. Compositional properties and thermal adaptation of 18S rRNA in vertebrates. *RNA* **2008**, *14*, 1492–1500. [CrossRef]
93. Wang, H.-C.; Xia, X.; Hickey, D. Thermal adaptation of the small subunit ribosomal RNA gene: A comparative study. *J. Mol. Evol.* **2006**, *63*, 120–126. [CrossRef]
94. Parker, M.S.; Balasubramaniam, A.; Sallee, F.R.; Parker, S.L. The Expansion Segments of 28S Ribosomal RNA Extensively Match Human Messenger RNAs. *Front. Genet.* **2018**, *9*. [CrossRef]
95. Parker, M.S.; Sallee, F.R.; Park, E.A.; Parker, S.L. Homoiterons and expansion in ribosomal RNAs. *FEBS Open Bio* **2015**, *5*, 864–876. [CrossRef]
96. Demeshkina, N.; Repkova, M.; Ven'Yaminova, A.; Graifer, D.; Karpova, G. Nucleotides of 18S rRNA surrounding mRNA codons at the human ribosomal A, P, and E sites: A crosslinking study with mRNA analogs carrying an aryl azide group at either the uracil or the guanine residue. *RNA* **2000**, *6*, 1727–1736. [CrossRef]
97. Barendt, P.A.; Shah, N.A.; Barendt, G.A.; Kothari, P.A.; Sarkar, C.A. Evidence for Context-Dependent Complementarity of Non-Shine-Dalgarno Ribosome Binding Sites to *Escherichia coli* rRNA. *ACS Chem. Biol.* **2013**, *8*, 958–966. [CrossRef]
98. Forsdyke, D.R. *Evolutionary Bioinformatics*, 3rd ed.; Springer: Cham, Switzerland, 2016; ISBN 978-3-319-28755-3.
99. Coleman, A.W. ITS2 is a double-edged tool for eukaryote evolutionary comparisons. *Trends Genet.* **2003**, *19*, 370–375. [CrossRef]
100. Coleman, A.W. Pan-eukaryote ITS2 homologies revealed by RNA secondary structure. *Nucleic Acids Res.* **2007**, *35*, 3322–3329. [CrossRef]

101. Coleman, A.W. Nuclear rRNA transcript processing versus internal transcribed spacer secondary structure. *Trends Genet.* **2015**, *31*, 157–163. [CrossRef]
102. Ruhl, M.W.; Wolf, M.; Jenkins, T.M. Compensatory base changes illuminate morphologically difficult taxonomy. *Mol. Phylogenet. Evol.* **2010**, *54*, 664–669. [CrossRef] [PubMed]
103. Muller, T.; Philippi, N.; Dandekar, T.; Schultz, J.; Wolf, M. Distinguishing species. *RNA* **2007**, *13*, 1469–1472. [CrossRef] [PubMed]
104. Song, J.; Shi, L.; Li, D.; Sun, Y.; Niu, Y.; Chen, Z.; Luo, H.; Pang, X.; Sun, Z.; Liu, C.; et al. Extensive Pyrosequencing Reveals Frequent Intra-Genomic Variations of Internal Transcribed Spacer Regions of Nuclear Ribosomal DNA. *PLoS ONE* **2012**, *7*, e43971. [CrossRef] [PubMed]
105. Schultz, J.; Wolf, M. ITS2 sequence–structure analysis in phylogenetics: A how-to manual for molecular systematics. *Mol. Phylogenet. Evol.* **2009**, *52*, 520–523. [CrossRef] [PubMed]
106. Ruhl, W.M. Compensatory Base Changes Illuminate Morphologically Difficult Taxonomy. Master Thesis, University of Georgia in Athens, Athens, Greece, 2009.
107. Nakhoul, H.; Ke, J.; Zhou, X.; Liao, W.; Zeng, S.X.; Lu, H. Ribosomopathies: mechanisms of disease. *Clin. Med. Insights Blood Disord.* **2014**, *7*, 7–16. [CrossRef] [PubMed]
108. Hetman, M. Role of the nucleolus in human diseases. *Biochim. Biophys. Acta BBA Mol. Basis Dis.* **2014**, *1842*, 757. [CrossRef] [PubMed]
109. Quin, J.E.; Devlin, J.R.; Cameron, D.; Hannan, K.M.; Pearson, R.B.; Hannan, R.D. Targeting the nucleolus for cancer intervention. *Biochim. Biophys. Acta BBA Mol. Basis Dis.* **2014**, *1842*, 802–816. [CrossRef] [PubMed]
110. Woods, S.J.; Hannan, K.M.; Pearson, R.B.; Hannan, R.D. The nucleolus as a fundamental regulator of the p53 response and a new target for cancer therapy. *Biochim. Biophys. Acta BBA Gene Regul. Mech.* **2015**, *1849*, 821–829. [CrossRef] [PubMed]

Chromosoma (2018) 127:141–150
https://doi.org/10.1007/s00412-017-0651-8

ORIGINAL ARTICLE

Evolutionary trends in animal ribosomal DNA loci: introduction to a new online database

Jana Sochorová[1] · Sònia Garcia[2] · Francisco Gálvez[3] · Radka Symonová[4] · Aleš Kovařík[1]

Received: 24 February 2017 / Revised: 6 November 2017 / Accepted: 7 November 2017 / Published online: 30 November 2017

Abstract

Ribosomal DNA (rDNA) loci encoding 5S and 45S (18S-5.8S-28S) rRNAs are important components of eukaryotic chromosomes. Here, we set up the animal rDNA database containing cytogenetic information about these loci in 1343 animal species (264 families) collected from 542 publications. The data are based on in situ hybridisation studies (both radioactive and fluorescent) carried out in major groups of vertebrates (fish, reptiles, amphibians, birds, and mammals) and invertebrates (mostly insects and mollusks). The database is accessible online at www.animalrdnadatabase.com. The median number of 45S and 5S sites was close to two per diploid chromosome set for both rDNAs despite large variation (1–74 for 5S and 1–54 for 45S sites). No significant correlation between the number of 5S and 45S rDNA loci was observed, suggesting that their distribution and amplification across the chromosomes follow independent evolutionary trajectories. Each group, irrespective of taxonomic classification, contained rDNA sites at any chromosome location. However, the distal and pericentromeric positions were the most prevalent (> 75% karyotypes) for 45S loci, while the position of 5S loci was more variable. We also examined potential relationships between molecular attributes of rDNA (homogenisation and expression) and cytogenetic parameters such as rDNA positions, chromosome number, and morphology.

Keywords 5S rDNA · 45S rDNA · Ribosomal RNA · Animal · Cytogenetics · Database

Introduction

Ribosomal DNA (rDNA) encodes the four essential genes needed for ribosome function: the 5S, 5.8S, 18S, and 28S rRNAs. They have been intensively studied at the cytogenetic and molecular levels. Probes derived from their conserved regions hybridise to chromosomes of diverged biological taxa, making rDNAs the first choice chromosome marker. This is probably the reason why a molecular cytogenetic approach became popular among (cyto)taxonomists in systematics studies. Development and widespread usage of fluorescence in situ hybridisation (FISH) techniques (Pinkel et al. 1986; Leitch et al. 1994) enabled to map rDNA loci on chromosomes of thousands of species over past decades until present. Employing rDNA-FISH may also provide information about the condensation status of rDNA chromatin representing, thus, a useful complement to molecular and cytogenetic (silver staining) expression studies.

rDNA evolves under the concept of concerted evolution (Zimmer et al. 1981; Dover 1982), a process maintaining its homogeneity and functionality. The process is believed to be mediated by homologous and non-homologous recombination and gene conversion. One puzzling feature is that despite overall sequence conservation (Averbeck and Eickbush 2005), rDNA tends to change the copy number (McTaggart et al. 2007; Wang et al. 2017) and position on chromosomes rapidly (Schubert and Wobus 1985; Dubcovsky and Dvorak 1995; Roy et al. 2005). Occasionally, studies have detected changes in the chromosomal location and size of specific rDNA arrays (loci). For example, the location of arrays differed between sibling

Jana Sochorová and Sònia Garcia contributed equally to this work.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00412-017-0651-8) contains supplementary material, which is available to authorized users.

✉ Aleš Kovařík
kovarik@ibp.cz

[1] Institute of Biophysics, Academy of Sciences of the Czech Republic, CZ-61265 Brno, Czech Republic

[2] Institut Botànic de Barcelona (IBB-CSIC-ICUB), Passeig del Migdia s/n, 08038 Barcelona, Catalonia, Spain

[3] Bioscripts—Centro de Investigación y Desarrollo de Recursos Científicos, 41012 Sevilla, Andalusia, Spain

[4] Faculty of Science, University of Hradec Kralove, Hradecka 1285, CZ-50003 Hradec Kralove, Czech Republic

species within the *Drosophila melanogaster* complex (Lohe and Roberts 1990). Rapid changes were also suggested in populations of the brown trout, *Salmo trutta* (Castro et al. 2001), and in the grasshoppers, *Eyprepocnemis plorans* (Cabrero et al. 2003) and *Podisma pedestris* (Veltsos et al. 2009).

The amount of literature containing cytogenetic rDNA data has been steadily increasing in the last years (Fig. 1). For illustration purposes, the searches of WOS (Web of Science) database and Google Scholar using keywords such as "*rDNA* AND *chromosome* AND *in situ hybridisation* AND *animal*" have yielded more than 500 results receiving annually more than 1100 citations. The literature is probably even more extensive since our search conditions were quite stringent and some works are published in non-indexed journals, conference proceedings, doctoral theses, and various monographs. Given the interest of such data through the number of publications in this area in recent times (approximately 50% of the publications listed in the database are from the last 6 years; more than 60 papers related with the topic have been published just in 2016), there is a need of assembling, storing, and analysing such information. Therefore, with the purpose of providing a tool allowing a better and easier use of animal rDNA cytogenetic information on the number and position of loci available, we have constructed the animal rDNA database. The resource is freely accessible at www.animalrdnadatabase.com representing a parallel to the plant rDNA database (www.plantrdnadatabase.com), created by our team previously (Garcia et al. 2012), providing the same information on plants. We have also analysed the database searching for relationships between the number of 5S and 45S (nucleolus organiser regions (NOR)) loci and for their preferential position (if any) on chromosomes.

Methods

Data assembly

The database comprises information about the number and position of rDNA in animal species collected until the end of 2016, coming from 541 publications. Papers were compiled by searching Thomson Reuters WOS, MEDLINE/PubMed, Scopus, and Google Scholar using the queries "rDNA AND chromosome," "rRNA genes AND chromosome," "rDNA AND karyotype," "rRNA AND karyotype," and "rDNA AND *in situ* hybridisation." Most journals were categorised within the areas of Genetics and Heredity, Biochemistry and Molecular Biology, Zoology, and Multidisciplinary Journals. The majority (~95%) of data are coming from fluorescent *in situ* hybridisation using 45S (18S, 28S, and internal transcribed spacer) and 5S rDNA probes. A smaller (~5%) proportion of entries was obtained from older studies based on radioactive hybridisation methods and morphological observation of secondary constrictions after the staining by classical histochemical dyes. We also aimed to include as many model representative species as possible. Together with basic information on the number, position, and linked/unlinked arrangement of rRNA genes, the resource also supplies data on chromosome number and genome size (taken from Gregory 2017) and whether the rDNA occurs on B chromosomes. The diploid locus numbers (sites) are presented as a range and mean. Three main categories of rDNA positions in chromosomes were distinguished: (i) pericentromeric = proximal sites (counting pericentromeric and centromeric positions together); (ii) distal = terminal sites (counting telomeric and subtelomeric positions together); and (iii) interstitial sites. In few cases where the hybridisation signals occupy whole chromosome arm, the database returns a "whole arm" position.

Web site construction

The tabular database structure comprising the information on number and position of rDNA loci and the source publications was created in SQL (structured query language) tables on a MySQL server. Each table had its own different field type and size. The initial spreadsheet table in which the data were compiled was exported to a CSV (comma-separated values) file. A unique ID was given for each entry together with the date and time of the export and the version of the data. Then it was imported to the SQL database (www.animalrdnadatabase.com). The website was programmed in HTML (HyperText Markup Language), CSS (Cascading Style Sheets), and JS (Javascript) for visualisation; the custom functions in PHP (PHP: Hypertext Preprocessor) were written to query the database, process, and display the data.

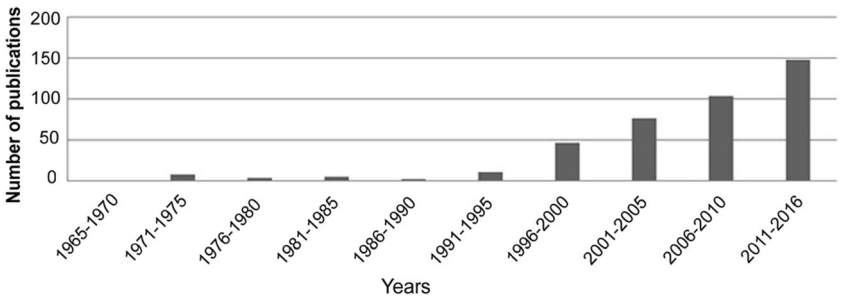


Fig. 1 Number of publications included in the database over nine successive 5-year periods and the 6-year period 2011–2016, between 1965 and 2016

Statistical analysis

Basic statistics such as mode, average, and median were obtained through MsExcel functions. Shapiro-Wilk test for normality, Pearson's correlation, and Mann-Whitney *U* test were performed in MsExcel and RStudio, v.0.98.1078, a user interface for R (www.rstudio.com). If a species had rDNA locus numbers that differed between or within populations, we treated each difference as a separate record. The assessment of rDNA numbers, rDNA positions, and chromosome morphologies comes exclusively from the literature, i.e. based on each authors' evaluation, when available. In few cases where this piece of information was not explicitly mentioned in the article, the assessment was done by evaluating of provided in situ hybridisation images.

Availability of data and materials

All data generated or analysed during this study are included in this published article, its supplementary information files, and the internet web site (http://www.animalrdnadatabase.com/).

Results and discussion

The created database (www.animalrdnadatabase.com) includes 1358 karyotypes (Table 1). The total number of species is 1343 from eight phyla, which roughly reflect animal kingdom diversity. Yet, despite this good deal of information, much of the cytogenetic data is still missing, e.g. for model species such as *Daphnia magna* (a small planktonic crustacean) or *Columba livia* (pigeon). Surprisingly, little cytogenetic information also exists for domesticated animals such as cats and reports in *Felidae* are limited to classical karyological studies in the leopard (Tanomtong et al. 2008). Furthermore, except for a few fish (Mantovani et al. 2005) and insect (Cabrero et al. 2003) genera, the interpopulation-level studies assessing cytogenetic variability have been rarely attempted. Hence, the database content could be a useful source for further research.

Number of loci per karyotype

Considering the whole database (karyotypes), the average number of 45S and 5S sites per diploid chromosome set (2C) was 3.8 and 4.5, respectively (Supplementary Table S1). The median was two sites (single locus/1C) for both 45S and 5S rDNA, respectively, indicating that most karyotypes tend to maintain locus numbers moderately low. Relatively large differences between means and medians indicated a non-Gaussian distribution of values (also revealed by significant results in the Shapiro-Wilk tests). Indeed, in each group, we identified several karyotypes largely deviating from the average (Fig. 2). The maximum numbers of 45S sites were found in the Amazonian

Table 1 Species representation of the rDNA database

Taxonomy/group		Database content			Group diversity[a]
		Karyotypes		Families	
		N	%[b]		
Vertebrates	Actinopterygian fish	539	39.8	95	30,000
	Mammals	169	12.4	33	5500
	Amphibians	44	3.2	10	6200
	Reptiles	72	5.3	26	8200
	Lampreys	2	0.1	1	38
	Cartilaginous fish	5	0.4	3	1100
	birds	15	1.1	7	10,000
Invertebrates	Arthropods	435	32.0	57	> 1,000,000
	Mollusks	54	4.0	17	81,000
	Annelids	10	0.7	5	9000
	Flatworms	8	0.6	6	25,000
	Thorny-head-worms	2	0.1	1	1500
	Echinoderms	1	0.1	1	6000
	Nematodes	1	0.1	1	2200
	Tunicates	1	0.1	1	7000
Total		1358	100.0	264	

[a] Estimated number of species in a group. Source: http://www.encyclopedia.com/

[b] Percentage of total karyotypes

fish *Schizodon fasciatus* (54/2C, de Barros et al. 2017) and the brook trout *Salvelinus fontinalis* (50/2C, Fujiwara et al. 1998). In mammals, the maximum number of 45S loci was identified in *Mus pahari* (rodent) having 42 sites/2C (Britton-Davidian et al. 2012). The maximum numbers of 5S sites were found in the neotropical lizards from the Teiidae family, *Kentropyx calcarata* (68/2C) and *K. pelviceps* (74/2C) (Carvalho et al. 2015). These karyotypes apparently account for relatively high average number of 5S loci in reptiles (Fig. 2). In mammals, the highest number of 5S loci was found in *Rhinolophus hipposideros* (bat) having 18 sites/2C (Puerma et al. 2008). About 12% species showed variation at the species level (Supplementary Table S2). The variation is explained by the presence of rDNA loci in sex chromosomes and supernumerary B chromosomes (both particularly frequent in insects), polyploidy (mainly in fish), and overall interpopulation variation. One also has to consider variation arising from differential experimental approaches used in labs.

Factors influencing rDNA loci multiplicity

About 60% karyotypes (766/1277) had a single 45S locus and 57% of karyotypes (358/628) had a single 5S locus. Karyotypes with multiple loci (both 5S and 45S) occurred almost in every group. In insects, multiple loci were found mostly in Orthoptera (e.g. grasshoppers, crickets, and locusts). These species are known to have relatively large genomes (Gregory 2017). Because of the known correlation between genome size and number of rDNA copies (Prokopowich et al. 2003), it is possible that dispersion of rDNA across chromosomes is related to their large genome sizes (~ 10 pg/2C, (Rees et al. 1978)). However, genome size cannot explain the high number of rDNA loci in actinopterygian fishes (e.g. Ráb et al. 2002; Mantovani et al. 2005; Cioffi et al. 2010; da Silva et al. 2011; Lima-Filho et al. 2014; Sember et al. 2015; Symonová et al. 2017) that generally harbour small genomes (~ 1 pg/2C). The increased number of loci could also be related to the large number of rDNA pseudogenes reported in some grasshopper genomes (Keller et al. 2006). In contrast, a whole genomic study in *Esox lucius* (Northern pike, fish) did not reveal increased pseudogenisation of highly (> 20,000 copies) amplified 5S genes (Symonová et al. 2017), suggesting that amplification does not automatically lead to pseudogenisation and that retention of pseudogenes varies between the genomes.

The amplification of rDNA has often been attributed to polyploidy (Gornung 2013). However, species with extremely large number of chromosomes (> 100/2C) do not automatically exhibit a high number of loci (Fig. 3 and Supplementary Table S3). For example, members of the arthropod genus *Austropotamobius* (2n = 176) harbour only four 45S sites (Mlinarec et al. 2016). Similarly, the fish *Acipenser baerii* and *A. transmontanus* (2n = 262) show only moderate

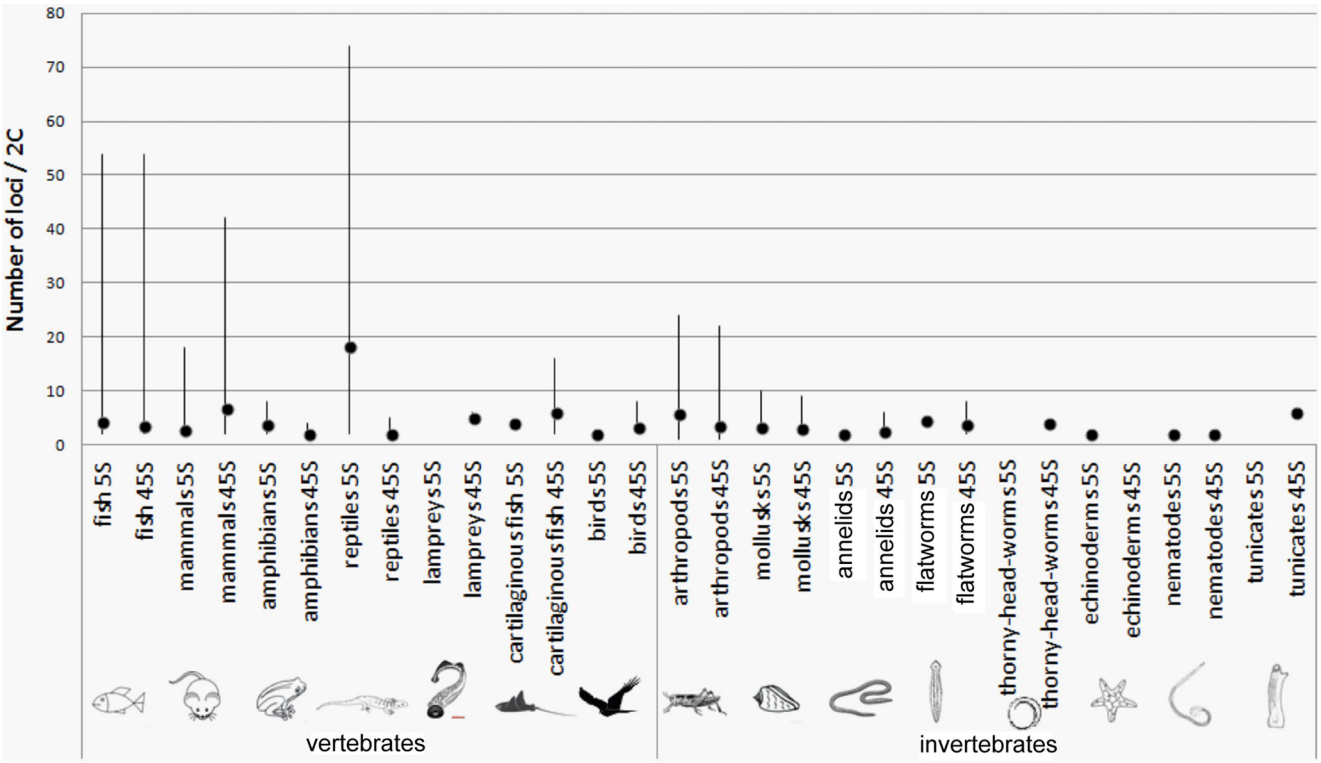


Fig. 2 Number of 5S and 45S rDNA sites in different animal taxa. Values are presented for the diploid karyotypes. Black dots indicate the average number of sites per group; lines show the range. The relatively high average number of 5S sites in reptiles is explained by an exceptionally high number of loci recorded in some members of the Teiidae family (Carvalho et al. 2015) and generally few data available for this group

numbers of 5S (four) and 45S (11) sites (Fontana et al. 2003). Reduction of rDNA in these enlarged karyotypes could be related to the "genomic shock" following polyploidy events (Semon and Wolfe 2007; Garcia et al. 2017). On the other hand, some moderate karyotypes harbour high number of rDNA loci. For example, in *Ctenogobius smaragdus* (emerald goby, 2n = 48, Lima-Filho et al. 2014), *S. fontinalis* (fish, 2n = 84, Fujiwara et al. 1998), and *M. pahari* (mouse, 2n = 48, Cazaux et al. 2011), the loci were distributed across 91, 88, and 50% of chromosomes, respectively. Certainly, polyploidy cannot explain large numbers of loci in these species and other mechanisms such as interlocus recombination (Cazaux et al. 2011), transposon activity (Symonová et al. 2013), and integration of extrachromosomally replicated rDNA (Cohen et al. 2010) should be considered.

Mutual relationships between 5S and 45S rDNA

Numerous case studies indicate likely independent amplification events of 5S and 45S rDNA in the genomes. For example, in the *Mus* genus, large (> 10-fold) variation occurs in 45S locus numbers (Britton-Davidian et al. 2012) without concomitant variation in 5S loci (Matsuda et al. 1994) (Fig. 3). Furthermore, two different cytotypes (2n = 52 and 2n = 54) of the Amazonian fish *Erythrinus erythrinus* (Cioffi et al. 2010) varied as much as 11-fold in the number of 5S loci while that of 45S loci was constant. Similarly, grasshopper genomes show extensive but independent variation in the number of 5S and 45S rDNA clusters (Cabral-de-Mello et al. 2011). Our comparative analysis of more than 500 karyotypes (Supplementary Fig. S1 and Supplementary Table S4) revealed that the numbers of both loci are not correlated (Pearson, $r = 0.047$, p value > 0.05). On the other hand, 43% karyotypes showed the same number of 45S and 5S loci, suggesting a potential relationship. However, the majority (89%) of equinumber karyotypes harboured a single locus of each, and the equinumber karyotypes with multiple loci were relatively rare (11%) which can be explained by a general tendency of genomes to keep the number of both loci low (Fig. 2).

In plants, equality of 45S and 5S loci was detected in 33% of karyotypes and a significant correlation ($p < 0.005$) between the number of 5S and 45S was observed (Garcia et al. 2017). This can be accounted to frequent whole genome duplications in plants through which both loci are equally multiplied. In animals, about 75% of karyotypes had 5S and 45S loci on different chromosomes (separate arrangement), while 25% of karyotypes

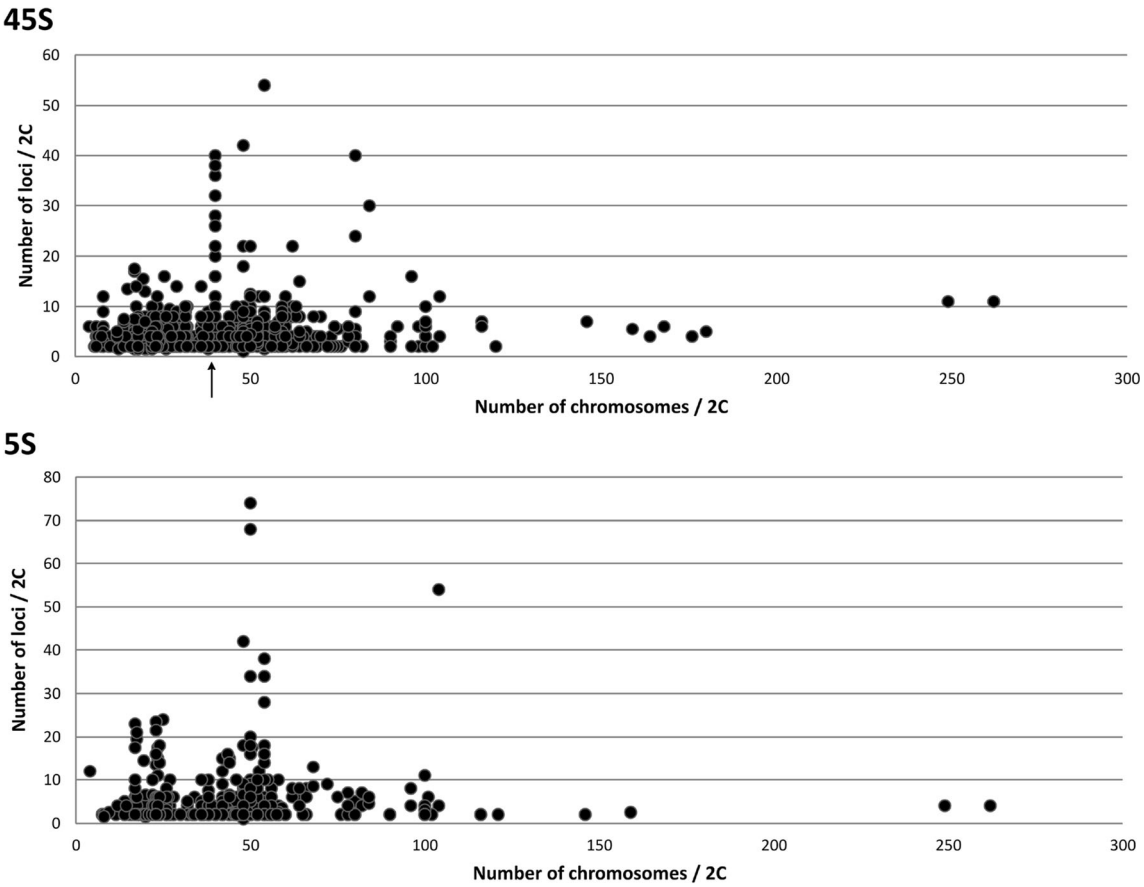


Fig. 3 Plots showing a relationship between chromosome number (*x*-axis) and rDNA sites (*y*-axis). The arrow marks a typical 2n = 40 karyotype in the *Mus* genus showing the striking variation in the number of 45S but not 5S rDNA sites

had at least one chromosome bearing both loci (colocalised). Thus, a tendency towards 5S and 45S colocalisation on the same chromosome does not appear to be as strong as in plants where colocalisation occurs in 58% of genera (Roa and Guerra 2015). Perhaps, this could be related to the increased number of loci in plants (median for 45S and 5S sites is 4/2C) (Roa and Guerra 2012; Garcia et al. 2017)) compared to the animals where medians are generally lower (2 sites/2C (Supplementary Table S1)). Colocalisation may also stimulate recombination frequency between both loci possibly leading to their physical linkage and formation of 45S-5S units. Of note, linked 45S-5S units are relatively common in plants (Garcia et al. 2009; Wicke et al. 2011; Garcia and Kovarik 2013) while in animals, they have been described in few arthropods (Drouin et al. 1992) and crustaceans (Drouin and de Sá 1995) so far. The number of 5S and 45S rRNA gene copies seems to be harmonised following concerted copy number variation in human and mouse (Gibbons et al. 2015). Thus, there may not be a simple relationship between the number of loci and the number of copies since gene richness may differ between loci. In this context, the size of nucleoli has been correlated with the number of 45S ribosomal RNA genes in amphibians (Miller and Brown 1969).

Position of rDNA on chromosomes

In the literature, there have been considerable debates over "randomness" of rDNA chromosomal positions (Hillis and Dixon 1991; Gornung 2013; Roa and Guerra 2015; Garcia et al. 2017). The information gathered in this database allowed us to address the question of preferential position (if any) of rDNA in chromosomes. We selected groups (Fig. 4) containing at least 40 species allowing robust statistical evaluation. The pie charts (Fig. 4 and Supplementary Table S5) show distribution of loci along different parts of chromosomes. Although it is clear that rDNA may occur at nearly any chromosomal position, there were significant trends in particular groups of animals. A distal location of 45S is clearly preferred in mammals, fish, and mollusks while in arthropods, its distribution is more balanced. The 5S loci were more evenly placed along the chromosomes than the 45S loci, consistent with previous observations (Baumlein and Wobus 1976; Roa and Guerra 2015; Garcia et al. 2017). In arthropods, the proximal positions of rDNA loci (both 5S and 45S) were significantly more common than in other groups (Fig. 4, for a statistical support, see Supplementary Table S6). Arthropods are the largest and most diversified phylum, representing around 70% of all animals (IUCN 2014). Since insects are highly represented within arthropods and also in our database (Table 1), we analysed 45S rDNA positions in its two largest orders, Coleoptera (beetles) and Orthoptera (mostly grasshoppers and crickets). Strikingly, significant (Supplementary Table S7) differences in 45S rDNA positions were found between both groups: Coleoptera had mostly distal distribution of 45S loci while Orthoptera had these genes preferentially located at pericentromeric positions (Fig. 5), and terminal positions were found only exceptionally (Veltsos et al. 2009).

There are several caveats in determining rDNA position of chromosomes. First, many karyotypes harbour chromosomes

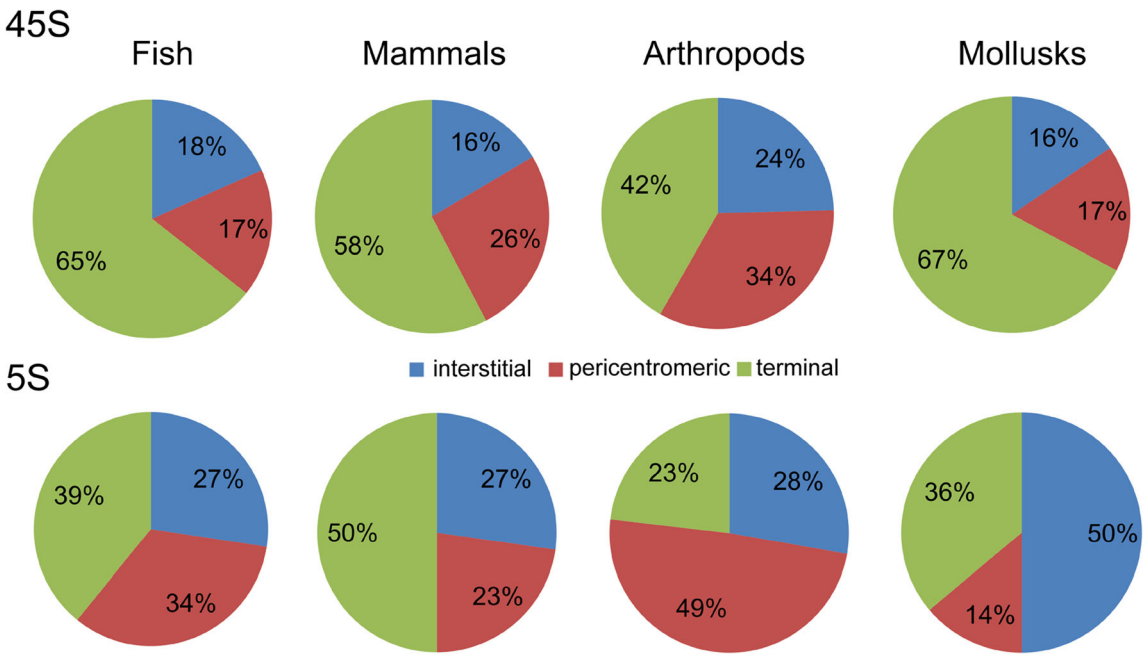


Fig. 4 Position of rDNA sites in chromosomes. The numbers of 45S and 5S sites counted in each group are as follows: fish ($N = 479$ and $N = 417$, respectively), mammals ($N = 156$ and $N = 40$, respectively), arthropods ($N = 424$ and $N = 96$, respectively), and mollusks ($N = 54$ and $N = 33$, respectively). The source data are given in Supplementary Table S5

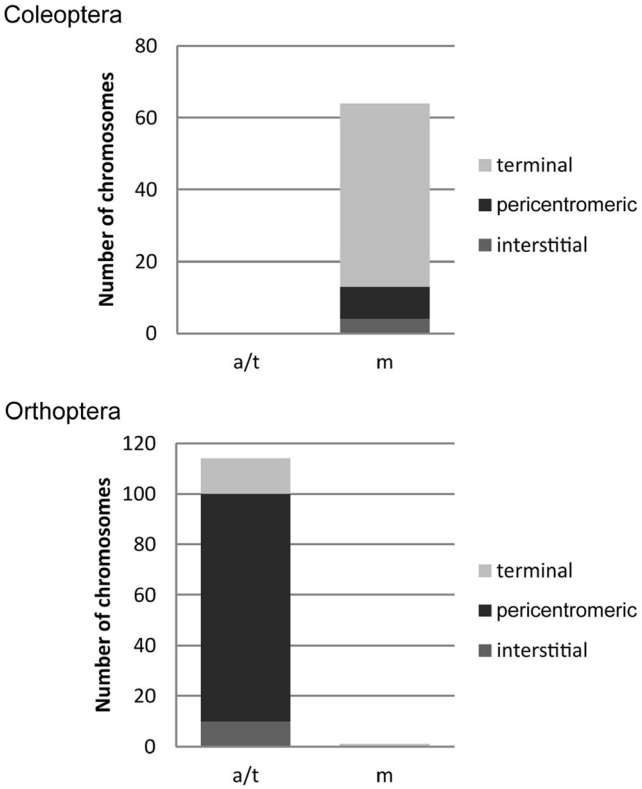


Fig. 5 Relationship between chromosome morphology (x-axis) and 45S rDNA positions (y-axis) in two of the largest orders of insects, Coleoptera ($N = 85$) and Orthoptera ($N = 141$). Only chromosomes with well-resolved morphologies were considered for the analysis. Chromosome type: m—metacentric/submetacentric; a/t—acrocentric/telocentric. The source datasets are in Supplementary Table S7

that are too small, preventing the accurate determination of loci positions. This is particularly the case of species with high number of chromosomes and relatively small genomes. Second, the resolution of FISH experiments may not permit to ascertain whether a site is located closer to the centromere or to the telomere in telocentric chromosomes or in short arms of acrocentric chromosomes. In these morphological types, the rDNA position could be considered either distal (as it appears at the end of the chromosome) or proximal (as it is located in the terminal centromere characterising these chromosomes). Hence, the "proximal-distal" location could be a more appropriate term for "pericentromeric" rDNA in acrocentric and telocentric chromosomes. For such reasons, the information about the position on chromosomes should be taken with great care since interpretation of FISH signals may vary between the researchers.

Are there functional constrains for the maintenance of distinct rDNA positions?

More than 50% of karyotypes in the database had 45S rDNA at distal (subtelomeric) positions. The number of sites located close to the chromosome ends could actually be even higher since many proximal locations can be considered as proximal-distal (78%, Supplementary Table S8). The question arises as to the functional significance (if any) of these observations, made independently in both animals and plants (Lima-de-Faria 1976; Roa and Guerra 2012; Garcia et al. 2017):

I. Position of 45S rDNA close to chromosome ends may be important for accurate positioning of 45S rDNA chromatin within and around the nucleolus (Gornung 2013). It is known that during mitosis, parts of the nucleolar proteins remain at the NORs (Schwarzacher and Wachtler 1993). Perhaps, association of partially decondensed rDNA chromatin with these proteins is better maintained at distal (or "distal-proximal") than at interstitial or centromeric positions during the transfer through mitosis. If so, distally positioned loci may better secure that rDNA transcription is rapidly resumed following mitosis early after cell division, perhaps via specific chromatin configuration. However, pericentromeric NORs were found in metacentric chromosomes of several single locus karyotypes (e.g. Barth et al. 2013; Singh and Barman 2013) suggesting that these positions, although infrequent (2% karyotypes; Supplementary Table S8), are probably compatible with expression of residing rDNA. Furthermore, secondary constrictions, thought to be remnants of activity in previous interphase, were identified at interstitial positions in some species (Fagundes et al. 2003). These studies suggest that there are no functional constrains limiting position of NOR in chromosomes with respect to the nuclear topology.
II. Non-coding functions of 45S rDNA should be considered (Kobayashi 2008). Perhaps, rDNA heterochromatin could fulfil a structural function contributing to the stabilisation of telomeric and centromeric (in distal-proximal positions) domains. In this regard, the principal determinant in rDNA silencing, the nucleolar remodelling complex (NoRC), is also important to maintain genome stability (Guetg et al. 2010) and the formation of heterochromatin (Postepska-Igielska and Grummt 2014). Cazaux et al. (2011) proposed that rDNA may predispose the chromatin to centromere formation. Indeed, pseudogenised rDNA copies that seem to regularly occur in different genomes at variable frequencies (Mentewab et al. 2011; Wang et al. 2016; Robicheau et al. 2017) may homogenise and even evolve in independent satellites (Lim et al. 2004; Ferreira et al. 2007).
III. Concerted evolution may be more efficient at chromosome termini than in other regions. It is well established that rDNA evolves via the "concerted evolution" model that maintains homogeneity of multigenic families (Zimmer et al. 1981; Dover 1982). Gene conversion and non-homologous recombination are the major

players of concerted evolution (reviewed in Nieto Feliner and Rosselló 2012). The regions near the ends of chromosomes of several organisms show higher recombination rates than more centric sequences (McKim et al. 1988; Jensen-Seaman et al. 2004). Functional rDNA copies may be located in chromosome sites with intensive recombination in subtelomeric regions and hence these positions would be favoured by natural selection. Yet, patterns of 45S rDNA unit divergence seem to be similar in species with distal (humans, Gonzalez and Sylvester 2001) and proximal locations (house mouse, Sasaki et al. 1987). However, in the *Mus* genus, 45S rDNA loci are preferentially positioned at telocentric chromosome close to the chromosome ends, which better correspond to proximal-distal location defined above. Cazaux et al. (2011) proposed that a specific configuration of these specific domains in interphase may stimulate meiotic recombination between non homologous loci.

Funding information The Czech Science Foundation (P506/16/02149J) and the Dirección General de Investigación Científica y Técnica from the Government of Spain (CGL2016-75694-P) are acknowledged for funding. JS participates on the PhD program Functional Genomics at the Masaryk University, Brno, CZ. SG benefits from a "Ramón y Cajal" contract from the Government of Spain.

Compliance with ethical standards

Competing interests The authors declare that they have no competing interests.

Consent for publication Not applicable.

Research involving human participants and/or animals Not applicable.

References

Averbeck KT, Eickbush TH (2005) Monitoring the mode and tempo of concerted evolution in the *Drosophila melanogaster* rDNA locus. Genetics 171(4):1837–1846. https://doi.org/10.1534/genetics.105.047670

Barth A, Souza VA, Sole M, Costa MA (2013) Molecular cytogenetics of nucleolar organizer regions in *Phyllomedusa* and *Phasmahyla* species (Hylidae, Phyllomedusinae): a cytotaxonomic contribution. Geneti Mol Res 12(3):2400–2408. https://doi.org/10.4238/2013.July.15.3

Baumlein H, Wobus U (1976) Chromosomal localization of ribosomal 5S RNA genes in *Chironomus thumni* by in situ hybridization of iodinated 5S RNA. Chromosoma 57(2):199–204. https://doi.org/10.1007/BF00292918

Britton-Davidian J, Cazaux B, Catalan J (2012) Chromosomal dynamics of nucleolar organizer regions (NORs) in the house mouse: micro-evolutionary insights. Heredity 108(1):68–74. https://doi.org/10.1038/hdy.2011.105

Cabral-de-Mello DC, Oliveira SG, de Moura RC, Martins C (2011) Chromosomal organization of the 18S and 5S rRNAs and histone H3 genes in *Scarabaeinae coleopterans*: insights into the evolutionary dynamics of multigene families and heterochromatin. BMC Genet 12(1):88. https://doi.org/10.1186/1471-2156-12-88

Cabrero J, Perfectti F, Gomez R, Camacho JPM, Lopez-Leon MD (2003) Population variation in the A chromosome distribution of satellite DNA and ribosomal DNA in the grasshopper *Eyprepocnemis plorans*. Chromosom Res 11:375–381, 4, DOI: https://doi.org/10.1023/A:1024127525756

Carvalho NDM, Pinheiro VSS, Carmo EJ, Goll LG, Schneider CH, Gross MC (2015) The organization of repetitive DNA in the genomes of Amazonian lizard species in the family Teiidae. Cytogenet Genome Res 147(2-3):161–168. https://doi.org/10.1159/000443714

Castro J, Rodriguez S, Pardo BG, Sanchez L, Martinez P (2001) Population analysis of an unusual NOR-site polymorphism in brown trout (*Salmo trutta* L.) Heredity (3):291–286, 302. https://doi.org/10.1046/j.1365-2540.2001.00834.x

Cazaux B, Catalan J, Veyrunes F, Douzery EJP, Britton-Davidian J (2011) Are ribosomal DNA clusters rearrangement hotspots? A case study in the genus *Mus* (Rodentia, Muridae). BMC Evol Biol 11(124):1. https://doi.org/10.1186/1471-2148-11-124

Cioffi MB, Martins C, Bertollo LAC (2010) Chromosome spreading of associated transposable elements and ribosomal DNA in the fish *Erythrinus erythrinus*. Implications for genome change and karyoevolution in fish. BMC Evol Biol 10(1):271. https://doi.org/10.1186/1471-2148-10-271

Cohen S, Agmon N, Sobol O, Segal D (2010) Extrachromosomal circles of satellite repeats and 5S ribosomal DNA in human cells. Mob DNA 8:11

da Silva M, Matoso DA, Vicari MR, de Almeida MC, Margarido VP, Artoni RF (2011) Physical mapping of 5S rDNA in two species of knifefishes: *Gymnotus pantanal* and *Gymnotus paraguensis* (Gymnotiformes). Cytogenet Genome Res 134(4):303–307. https://doi.org/10.1159/000328998

de Barros LC, Galetti PM, Feldberg E (2017) Mapping 45S and 5S ribosomal genes in chromosomes of Anostomidae fish species (Ostariophysi, Characiformes) from different Amazonian water types. Hydrobiologia 789(1):77–89. https://doi.org/10.1007/s10750-015-2583-8

Dover GA (1982) Molecular drive: a cohesive mode of species evolution. Nature 299(5879):111–117. https://doi.org/10.1038/299111a0

Drouin G, de Sá MM (1995) The concerted evolution of 5S ribosomal genes linked to the repeat units of other multigene families. Mol Biol Evol 12(3):481–493

Drouin G, Sevigny JM, McLaren IA, Hofman JD, Doolittle WF (1992) Variable arrangement of 5S ribosomal genes within the ribosomal DNA repeats of arthropods. Mol Biol Evol 9(5):826–835

Dubcovsky J, Dvorak J (1995) Ribosomal RNA multigene loci—nomads of the Triticeae genomes. Genetics 140(4):1367–1377

Fagundes V, Christoff AU, Amaro-Ghilard RC, Scheibler DR, Yonenaga-Yassuda Y (2003) Multiple interstitial ribosomal sites (NORs) in the Brazilian squirrel *Sciurus aestuans ingrami* (Rodentia, Sciuridae) with 2n = 40. An overview of *Sciurus cytogenetics*. Genet Mol Biol 26(3):253–257. https://doi.org/10.1590/S1415-47572003000300007

Ferreira IA, Bertollo LAC, Martins C (2007) Comparative chromosome mapping of 5S rDNA and 5SHindIII repetitive sequences in Erythrinidae fishes (Characiformes) with emphasis on the *Hoplias*

malabaricus 'species complex'. Cytogenet Genome Res 118(1):78–79. https://doi.org/10.1159/000106445

Fontana F, Lanfredi M, Congiu L, Leis M, Chicca M, Rossi R (2003) Chromosomal mapping of 18S-28S and 5S rRNA genes by two-colour fluorescent *in situ* hybridization in six sturgeon species. Genome 46(3):473–477. https://doi.org/10.1139/g03-007

Fujiwara A, Abe S, Yamaha E, Yamazaki F, Yoshida MC (1998) Chromosomal localization and heterochromatin association of ribosomal RNA gene loci and silver-stained nucleolar organizer regions in salmonid fishes. Chromosom Res 6(6):463–471. https://doi.org/10.1023/A:1009200428369

Garcia S, Garnatje T, Kovařík A (2012) Plant rDNA database: ribosomal DNA *loci* data including other karyological and cytogenetic information in plants. Chromosoma 121(4):389–394. https://doi.org/10.1007/s00412-012-0368-7

Garcia S, Kovarik A (2013) Dancing together and separate again: gymnosperms exhibit frequent changes of fundamental 5S and 35S rRNA genes (rDNA) organisation. Heredity 111(1):23–33. https://doi.org/10.1038/hdy.2013.11

Garcia S, Kovařík A, Leitch AR, Garnatje T (2017) Cytogenetic features of rRNA genes across land plants: analysis of the plant rDNA database. Plant J 89(5):1020–1030. https://doi.org/10.1111/tpj.13442

Garcia S, Lim KY, Chester M, Garnatje T, Pellicer J, Valles J, Leitch AR, Kovařík A (2009) Linkage of 35S and 5S rRNA genes in *Artemisia* (family Asteraceae): first evidence from angiosperms. Chromosoma 118(1):85–97. https://doi.org/10.1007/s00412-008-0179-z

Gibbons JG, Branco AT, Godinho SA, Yu S, Lemos B (2015) Concerted copy number variation balances ribosomal DNA dosage in human and mouse genomes. Proc Natl Acad Sci U S A 112(8):2485–2490. https://doi.org/10.1073/pnas.1416878112

Gonzalez IL, Sylvester JE (2001) Human rDNA: evolutionary patterns within the genes and tandem arrays derived from multiple chromosomes. Genomics 73(3):255–263. https://doi.org/10.1006/geno.2001.6540

Gornung E (2013) Twenty years of physical mapping of major ribosomal RNA genes across the teleosts: a review of research. Cytogenet Genome Res 141(2-3):90–102. https://doi.org/10.1159/000354832

Gregory TR (2017) Animal genome size database. http://www.genomesize.com/. Accessed 26 January 2017

Guetg C, Lienemann P, Sirri V, Grummt I, Hernandez-Verdun D, Hottiger MO, Fussenegger M, Santoro R (2010) The NoRC complex mediates the heterochromatin formation and stability of silent rRNA genes and centromeric repeats. EMBO J 29(13):2135–2146. https://doi.org/10.1038/emboj.2010.17

Hillis DM, Dixon MT (1991) Ribosomal DNA—molecular evolution and phylogenetic inference. Q Rev Biol 66(4):411–453. https://doi.org/10.1086/417338

IUCN (2014) Red list of threatened species. Version 2014.3. Summary statistics for globally threatened species. Table 1: numbers of threatened species by major groups of organisms (1996–2014), www.iucnredlist.org. Accessed 20 January 2017

Jensen-Seaman MI, Furey TS, Payseur BA, Lu Y, Roskin KM, Chen CF, Thomas MA, Haussler D, Jacob HJ (2004) Comparative recombination rates in the rat, mouse, and human genomes. Genome Res 14(4):528–538. https://doi.org/10.1101/gr.1970304

Keller I, Chintauan-Marquier IC, Veltsos P, Nichols RA (2006) Ribosomal DNA in the grasshopper *Podisma pedestris*: escape from concerted evolution. Genetics 174(2):863–874. https://doi.org/10.1534/genetics.106.061341

Kobayashi T (2008) A new role of the rDNA and nucleolus in the nucleus-rDNA instability maintains genome integrity. BioEssays 30(3):267–272. https://doi.org/10.1002/bies.20723

Leitch AR, Schwarzacher T, Jackson D, Leitch IJ (1994) *In situ* hybridization: a practical guide. Bios Scientific Publishers Ltd., Oxford

Lim KY, Skalická K, Koukalová B, Volkov RA, Matyasek R, Hemleben V, Leitch AR, Kovařík A (2004) Dynamic changes in the distribution of a satellite homologous to intergenic 26-18S rDNA spacer in the evolution of *Nicotiana*. Genetics 166(4):1935–1946. https://doi.org/10.1534/genetics.166.4.1935

Lima-de-Faria A (1976) The chromosome field I. Prediction of the location of ribosomal citrons. Hereditas 83:1–22

Lima-Filho PA, Bertollo LA, Cioffi MB, Costa GW, Molina WF (2014) Karyotype divergence and spreading of 5S rDNA sequences between genomes of two species: darter and emerald gobies (Ctenogobius, Gobiidae). Cytogenet Genome Res 142(3):197–203. https://doi.org/10.1159/000360492

Lohe AR, Roberts PA (1990) An unusual Y chromosome of *Drosophila simulans* carrying amplified rDNA spacer without rRNA genes. Genetics 125(2):399–406

Mantovani M, Abel LD, Moreira-Filho O (2005) Conserved 5S and variable 45S rDNA chromosomal localisation revealed by FISH in *Astyanax scabripinnis* (Pisces, Characidae). Genetica 123(3):211–216. https://doi.org/10.1007/s10709-004-2281-3

Matsuda Y, Moriwaki K, Chapman VM, Hoi-Sen Y, Akbarzadeh J, Suzuki H (1994) Chromosomal mapping of mouse 5S rRNA genes by direct R-banding fluorescence in situ hybridization. Cytogenet Cell Genet 66(4):246–249. https://doi.org/10.1159/000133704

McKim KS, Howell AM, Rose AM (1988) The effects of translocations on recombination frequency in *Caenorhabditis elegans*. Genetics 120(4):987–1001

McTaggart S, Dudycha JL, Omilian A, Crease TJ (2007) Rates of recombination in the ribosomal DNA of apomictically propagated *Daphnia obtusa* lines. Genetics 175(1):311–320. https://doi.org/10.1534/genetics.105.050229

Mentewab AB, Jacobsen MJ, Flowers RA (2011) Incomplete homogenization of 18 S ribosomal DNA coding regions in *Arabidopsis thaliana*. BMC Res Notes 4(1):93. https://doi.org/10.1186/1756-0500-4-93

Miller L, Brown DD (1969) Variation in the activity of nucleolar organizers and their ribosomal gene content. Chromosoma 28(4):430–444

Mlinarec J, Porupski I, Maguire I, Klobucar G (2016) Comparative karyotype investigations in the white-clawed crayfish *Austropotamobius pallipes* (Lereboullet, 1858) species complex and stone crayfish *A. torrentium* (Schrank, 1803) (Decapoda: Astacidae). J Crust. Biol 36:87–93

Nieto Feliner G, Rosselló JA (2012) Concerted evolution of multigene families and homeologous recombination. In: Wendel JF (ed) Plant Genome Diversity, vol 1. Springer-Verlag, Wien, pp 171–194. https://doi.org/10.1007/978-3-7091-1130-7_12

Pinkel D, Straume T, Gray JW (1986) Cytogenetic analysis using quantitative, high-sensitivity, fluorescence hybridization. Proc Natl Acad Sci U S A 83(9):2934–2938. https://doi.org/10.1073/pnas.83.9.2934

Postepska-Igielska A, Grummt I (2014) NoRC silences rRNA genes, telomeres, and centromeres. Cell Cycle 13(4):493–494. https://doi.org/10.4161/cc.27783

Prokopowich CD, Gregory TR, Crease TJ (2003) The correlation between rDNA copy number and genome size in eukaryotes. Genome 46(1):48–50. https://doi.org/10.1139/g02-103

Puerma E, Acosta MJ, Barragán MJ, Martínez S, Marchal JA, Bullejos M, Sánchez A (2008) The karyotype and 5S rRNA genes from Spanish individuals of the bat species *Rhinolophus hipposideros* (Rhinolophidae; Chiroptera). Genetica 134:287–295

Ráb P, Crossman EJ, Reed KM, Rábová M (2002) Chromosomal characteristics of ribosomal DNA in two extant species of North American mudminnows Umbra pygmaea and U. limi (Euteleostei: Umbridae). Cytogenet Genome Res 98(2-3):194–198

Rees H, Shaw DD, Wilkinson P (1978) Nuclear DNA Variation among Acridid Grasshoppers. Proc R Soc B Biol Sci 202(1149):517–525

Roa F, Guerra M (2012) Distribution of 45S rDNA sites in chromosomes of plants: Structural and evolutionary implications. BMC Evol Biol 12(1):225

Roa F, Guerra M (2015) Non-Random Distribution of 5S rDNA Sites and Its Association with 45S rDNA in Plant Chromosomes. Cytogenet Genome Res 146(3):243–249

Robicheau BM, Susko E, Harrigan AM, Snyder M (2017) Ribosomal RNA Genes Contribute to the Formation of Pseudogenes and Junk DNA in the Human Genome. Genome Biol Evol 9(2):380–397

Roy V, Monti-Dedieu L, Chaminade N, Siljak-Yakovlev S, Aulard S, Lemeunier F, Montchamp- Moreau C (2005) Evolution of the chromosomal location of rDNA genes in two Drosophila species subgroups: ananassae and melanogaster. Heredity (Edinb) 94:388–395

Schubert I, Wobus U (1985) In situ hybridization confirms jumping nucleolus organizing regions in Allium. Chromosoma 92(2):143–148

Schwarzacher HG, Wachtler F (1993) The nucleolus. Anat Embryol (Berl) 188:515–536

Sember A, Bohlen J, Slechtová V, Altmanová M, Symonová R, Rab P (2015) Karyotype differentiation in 19 species of river loach fishes (Nemacheilidae, Teleostei): extensive variability associated with rDNA and heterochromatin distribution and its phylogenetic and ecological interpretation. BMC Evol Biol 15:251

Singh M, Barman AS (2013) Chromosome breakages associated with 45S ribosomal DNA sequences in spotted snakehead fish Channa punctatus. Mol Biol Rep 40(1):723–729

Symonová R, Majtanová Z, Sember A, Staaks GBO, Bohlen J, Freyhof J, Rabová M, Rab P (2013) Genome differentiation in a species pair of coregonine fishes: an extremelyrapid speciation driven by stress-activated retrotransposons mediating extensive ribosomal DNA multiplications. BMC Evol Biol 13:42

Symonová R, Ocalewicz K, Kirtiklis L, Delmastro GB, Pelikánová Š, Garcia S, Kovařík A (2017) Higher-order organisation of extremely amplified, potentially functional and massively methylated 5S rDNA in European pikes (Esox sp.) BMC Genomics 18(1):391

Tanomtong A, Khunsook S, Keawmad P, Pintong K (2008) Cytogenetic Study of the Leopard, Panthera pardus (Carnivora, Felidae) by Conventional Staining, G-banding and High-resolution Staining Technique. Cytologia 73(1):81–90

Veltsos P, Keller I, Nichols RA (2009) Geographically localised bursts of ribosomal DNA mobility in the grasshopper Podisma pedestris. Heredity 103(1):54–61

Wang W, Lu M, Becher H, Garcia S, Kovarikova A, Leitch IJ, Leitch AR, Kovarik A (2016) Astonishing 35S rDNA diversity in the gymnosperm species Cycas revoluta Thunb. Chromosoma 125(4):683–699

Wang M, Lemos B, Eng C (2017) Ribosomal DNA copy number amplification and loss in human cancers is linked to tumor genetic context, nucleolus activity, and proliferation. PLoS Genet 13(9):e1006994

Wicke S, Costa A, Muñoz J, Quandt D (2011) Restless 5S: The re arrangement(s) and evolution of the nuclear ribosomal DNA in land plants. Mol Phylogenet Evol 61(2):321–332

Zimmer EA, Martin SL, Beverley SM, Kan YW, Wilson AC (1981) The untranslated regions of beta-globin mRNA evolve at a functional rate in higher primates. Proc Natl Acad Sci U S A 77(4):2158–2162

genes

MDPI

Review

Vertebrate Genome Evolution in the Light of Fish Cytogenomics and rDNAomics

Radka Symonová [1,*] and W. Mike Howell [2]

[1] Faculty of Science, Department of Biology, University of Hradec Králové, 500 03 Hradec Králové, Czech Republic
[2] Department of Biological and Environmental Sciences, Samford University, Birmingham, AL 35229, USA; wmhowell@samford.edu
* Correspondence: radka.symonova@gmail.com; Tel.: +420-776-121-054

Received: 30 November 2017; Accepted: 29 January 2018; Published: 14 February 2018

Abstract: To understand the cytogenomic evolution of vertebrates, we must first unravel the complex genomes of fishes, which were the first vertebrates to evolve and were ancestors to all other vertebrates. We must not forget the immense time span during which the fish genomes had to evolve. Fish cytogenomics is endowed with unique features which offer irreplaceable insights into the evolution of the vertebrate genome. Due to the general DNA base compositional homogeneity of fish genomes, fish cytogenomics is largely based on mapping DNA repeats that still represent serious obstacles in genome sequencing and assembling, even in model species. Localization of repeats on chromosomes of hundreds of fish species and populations originating from diversified environments have revealed the biological importance of this genomic fraction. Ribosomal genes (rDNA) belong to the most informative repeats and in fish, they are subject to a more relaxed regulation than in higher vertebrates. This can result in formation of a literal 'rDNAome' consisting of more than 20,000 copies with their high proportion employed in extra-coding functions. Because rDNA has high rates of transcription and recombination, it contributes to genome diversification and can form reproductive barrier. Our overall knowledge of fish cytogenomics grows rapidly by a continuously increasing number of fish genomes sequenced and by use of novel sequencing methods improving genome assembly. The recently revealed exceptional compositional heterogeneity in an ancient fish lineage (gars) sheds new light on the compositional genome evolution in vertebrates generally. We highlight the power of synergy of cytogenetics and genomics in fish cytogenomics, its potential to understand the complexity of genome evolution in vertebrates, which is also linked to clinical applications and the chromosomal backgrounds of speciation. We also summarize the current knowledge on fish cytogenomics and outline its main future avenues.

Keywords: fish cytogenomics; repetitive sequences; rDNAome; genome evolution; AT/GC compositional evolution; quantitative cytogenomics.

1. Introduction

1.1. Introduction into Cytogenomics

Cytogenomics, i.e., an integration of cytogenetic and genomic data and approaches supported by bioinformatics, is traditionally well established in clinical areas, particularly in cancer research and diagnostics [1,2]. Recently, cytogenomics has gained in importance in veterinary, e.g., [3] and in plant "-omics" research [4]. Along with progress in genomics, there was a literal call for "integrated cytogenomics" [5] in the recent endeavor towards the third-generation genome assemblies in avian genomics. In fish genome research, the first steps have already been taken [6–8]. Fish cytogenomics

has been mentioned as the future pathway along which traditional fish cytogenetics should move [9] and is the logical outcome of the current integrating of fish biology research.

What can this synergy between cytogenetics and genomics offer in groups of fishes with less explored and non-model genomes? What is the potential of cytogenomics in fish with their sometimes very large genomes consisting of numerous small chromosomes stuffed with repetitive sequences? What is the status quo and what might or what should be the future of fish cytogenomics? Do we at all need any cytogenomic research of basal vertebrates so different from mammals and we humans? Is there any potential practical use of detailed knowledge on fish genomes? We definitely sense a strong need for fish cytogenomics to develop and unfold itself further and to help us to understand the intricate fish genome evolution. Moreover, we also need fish cytogenomics to gain insights into the evolution of other vertebrate genomes, not only because all other vertebrates originate from a fish-like ancestor. Hence, fish genomes should be viewed as the first stage towards the even higher structural and functional complexity of avian and mammalian genomes. However, fish genomes are far less explored (particularly regarding their diversity), the novel methods are applied with a delay in fishes and therefore, there is lot of work to do to compensate this delay. Fish represent a diversified spectrum of ways of genome organization (Table 1) and mirror the complex genome evolution of all tetrapod vertebrates (Tetrapodomorpha) embedded within Sarcopterygian fishes, the sister branch of Actinopterygian fishes [10].

1.2. How Can This Review Been Utilized?

The goal of this review is to summarize existing data and approaches usable in downstream analyses of the genomic makeup of fishes and other vertebrates and to integrate them in contexts enabling a comprehensive understanding of the evolution of vertebrate genomes. In this way, we aim to provide a solid baseline for future fish as well as vertebrate cytogenomic evolution research showing the importance and advantages of work on fish chromosomes and genomes. This shall assist cytogeneticists to effectively utilize genomics resources. We outline the crucial and inevitable directions of fish cytogenetics and genomics as both fields of study move toward the new integrative cytogenomics with their databases growing daily. We also delineate and highlight the importance of the shift from the qualitative karyology and cytogenetics to the quantitative evolutionary-ecological and environmental cytogenomics. Hence, we have to start viewing the cytogenomic results in a broader, evolutionary and phylogenomic context and linking cytogenomic phenomena with their corresponding ecological and physiological causes and consequences. This review shall further serve as a reference to non-fish vertebrate genome researchers, where however, fishes are necessarily included as the most basal vertebrates representing the starting point of the vertebrate's genome evolution. Finally, we present datasets of sequenced fish species listing their essential cytogenomic traits and information with potential for down-stream cytogenomic research and potential pitfalls and discrepancies to be regarded.

1.3. The Complex Evolution of Vertebrates Began with Fish

Understanding at least a basic framework and timescales of fish history substantially helps global understanding genome evolution in vertebrates. Jawless fishes as the first vertebrates with true bones arose in the early Paleozoic Era some 485 million years ago (mya). This crucial stage of our evolution remained conserved in lampreys and hagfishes that both considerably differ in their cytogenomic organizations from the rest of vertebrates (Table 1), although they show remarkable traits of convergent evolution with higher vertebrates (e.g., their adaptive immunity comparable with that of ours, [11]). However, lampreys and hagfishes differ also among each other, which is easily explained by their position on the phylogenetic tree, where five large extinct (†) groups are embedded between hagfishes and lampreys (although together known as cyclostomates):

Pteraspidiformes†, Anaspida†, Thelodontiformes†, Galeaspidiformes† and Cephalaspidiformes† [21,22]. It took another 65 myr for these fish to evolve jaws (chondrichthyans, ca. 420 mya). Here, we must be aware of the large gaps on the phylogenetic tree of *living* fishes, where other extinct lineages should have been

otherwise placed (Placodermi†, Acanthodii†; [21,22]. After the jawed fishes appeared, it took another 169 myr to give rise to the ancestors of the living bowfin and gars (ca. 251 mya). However, both bowfin and gars are only the last survivors of much larger and very specious groups of once diversified and widespread lineages [23]. About 3 myr later, the ancestors of modern-day sturgeons and paddlefishes (Acipenseriformes) appeared (ca. 248 mya). The recent Acipenseriformes again represent only a small remnant group of once important radiations, whereas the order Paleonisciformes† went extinct completely without any survivors left [21,22]. The modern-day teleosts, which rule the world's waters today, did not evolve until 225 mya. However, there was another completely extinct group embedded between Ginglymodi and teleosts, namely Pholidophoriformes†, making the phylogeny around to the origin of teleosts extremely difficult and so far unresolved (the Halecostomi-Holostei problems, [24]). This large gap among the surviving lineages is beside outstanding issues in their morphology reflected also in their cytogenomic organization [7,25] and illustrates the immense complexity and the current incongruence in the border(s) among Amiiformes (bowfin with its ancestors), Lepisosteiformes (gars) and teleosts. Originated in the marine realm, fishes radiated extensively in the sea [22] and colonized freshwater environments several times [26]. Immense selective pressures of these evolving environments must have shaped the genomes of untold thousands of fishes as they adapted to every conceivable ecological niche. Only recently, we have started realizing the extent to which the environment shaped their cytogenomic features [27,28] but we still have to fully understand them. This long time that elapsed during fish evolution reflects in the broad variety of ways that the organization of the fish genome has been structured and re-structured among cyclostomates, chondrichthyans, non-teleost actinopterygians, fish-like sarcopterygians and teleosts (Table 1). In parallel, a comparable variety in genome organization occurs within teleosts, although more than 50% of their species displays $2n$ = 48–50 [29]. This has to be regarded whenever any (cyto)genomic "generalizing" analyses are planned/performed. It is extremely simplifying to include merely 2–5 teleost species and consider them sufficiently representative of approximately 26,000 living teleost species, not including other Actinopterygia e.g., [30]. Unfortunately, there are frequent reductions of "fish" to teleosts (e.g., [31], etc.) leaving the entire potential of Actinopterygia described above unutilized.

2. The Importance of Fish Cytogenomic Research Demonstrated on Their Role in Medicine

Approximately 70% of the genes associated with human diseases have their functional homologs in teleosts [32,33], although teleosts diverged from lineages leading to humans more than 400 mya and underwent a teleost specific whole-genome duplication (WGD). On the other hand, the copy number of 5S and 45S rDNA is tightly regulated in mammals [34] whereas incomparably relaxed in fishes [8,35]. Copy number variation between 5S and 45S rDNAs in humans and mice can result in cancer [36]. In fishes, variations in a much higher extent can result in a speciation event as documented e.g., in *Coregonus* (Salmoniformes, [35]), in *Erythrinus* (Characiformes, [37]) and other numerous examples. This demonstrates the importance of fish cytogenomic research in medicine and in our understanding of vertebrates' genome evolution. However, whereas the knowledge on mammalian (and yeast) rDNAs have already become established and currently receives its adequate attention with crucial findings being reported steadily. In fish, the research remains limited mostly to FISH (fluorescence in situ hybridization) detections of rDNA on chromosomes (for a survey see [38] or [39]) and to only several detailed molecular analyses of rDNAs in salmonids [40,41] and cichlids [42].

2.1. Why Is the Zebrafish the Most Important Fish Model?

Zebrafish (*Danio rerio*) have played an important role in accelerating our knowledge in embryonic development, regeneration, gene expression, transgenesis, environmental monitoring, drug discovery, cardiovascular diseases, immunology, infectious diseases, RNA splicing, stem cell biology and a host of other areas of importance to medical science [33,43–46]. At least 20% of the zebrafish's duplicated gene pairs have been retained from this WGD [47–50]. This often causes altered gene expression and protein functions, such that the complement of the expression domains of both fish paralogs are equivalent to the single orthologue in other vertebrates. Despite this WGD, zebrafish and humans

have about the same number of chromosomes and zebrafish chromosomes are mosaically orthologous to several human chromosomes [50]. Cancer researchers use the zebrafish model to study vertebrate gene function [51] since its embryos are virtually transparent. This has led to much knowledge in gene function and genetic diseases [52–54]. Zebrafish have been used to make several transgenic models of cancer (melanoma, leukemia, pancreatic cancer and hepatocellular carcinoma; [44]). Zebrafish express mutated forms of either the *BRAF* (B-Raf) or *NRAS* (Neuroblastoma RAS) oncogenes and develop melanoma when placed onto a p53 deficient background. These tumors strongly resemble human melanoma. The BRAF melanoma model was used to understand the function of genes known to be overexpressed or amplified in human melanoma [51,55]. One gene, histone methyltransferase called *SETDB1*, markedly accelerated tumor formation in the zebrafish, demonstrating its role as a new melanoma oncogene. *SETDB1* is further known to be involved in the epigenetic regulation central to tumorigenesis [51].

2.2. Other Fish Species Used as Medical Models Have Genomes Suited for Specific Human Diseases

The Japanese Medaka (*Oryzias latipes*) has a gene sharing over 95% identity with the human *HRAS* gene, which is one of the most frequently mutated genes in cancers [45]. The medaka is a complementary model to the zebrafish as it has many of its desirable traits [56]. Among Antarctic notothenioids, the Antarctic rockcod (*Notothenia corilceps*) has an extra stout mineralized skeleton. Some of these fishes demineralize their bone to increase their buoyancy to the extent that they develop osteopenia, a bone loss very similar to osteoporosis. Certain cichlids serve as models of craniofacial developmental disorders to predict and treat human craniofacial disorders [57]. They have evolved very different craniofacial morphologies dependent upon the diet to which the particular species has specialized and genes responsible for their craniofacial adaptations have been discovered. The blind cavefish (*Astyanax mexicanus*) has both surface populations living in the light and cave populations living in total or near-total darkness with retinal degeneration and albinism [45,58]. Some studies have suggested that certain eye development genes (*Paired Box 6* (*PAX6*), *sonic hedgehog*) are linked to eye degeneration [59]. Further questions are being currently solved since the full genome sequence of *Astyanax* is available [58] to determine the affected genes and elucidate their role in eye development [45]. Fair skin in fish and humans is predisposed to skin cancer. Also, the absence of melanin in the retina of cavefish may cause vision disease via often mutated gene OCA2 causing oculocutaneous albinism type II in humans [59] and is proof that the mutations in the same gene can result in the same phenotype in fish and humans [45]. Many killifishes (Cyprinodontiformes) make excellent models for aging research since they have adapted to life in the extreme conditions in the wet-dry savannahs of Africa. Eggs and embryos survive dry periods by undergoing diapause in the dry lake beds. When the rains come, they hatch quickly, and their life history is completed in the few months before the dry season arrives again. They have a very short lifespan even in aquaria under optimal conditions. Certain strains of this fish live only 10 weeks while others live 31 weeks. The fast aging in killifish show many of the same signs of older organisms, such as decreased fertility, cognitive decline, age-related molecular markers and high morbidity [60]. In the turquoise killifish, *Nothobranchius furzeri* aging is linked to an increase in cancer, infectious diseases, neurodegenerative diseases and circulation problems, hence this model is a useful model to study the aging process [60]. Toadfish (Batrachoididae) protects itself from predators by releasing urea to mask its scent and hiding spot and has a unique nitrogen excretion which makes them resistant to ammonia. Such a situation in the external or internal environment of humans would be harmful. The plain midshipman, *Porichthys notatus*, is a model for a human hepatic (portosystemic) encephalopathy, which is due to liver failure or excessive amount of nitrogen after kidney failure. The toadfish also serves as a model to study human sickle cell anemia and erythrocyte sickling under anoxic conditions. Under low oxygen levels, the toadfish's erythrocyte sickles similar to the human mutant, malfunctioning hemoglobin known as HbS [61]. Swordtails and platyfishes are one of the oldest animal models for cancer research. It was also the first to present evidence that cancer has a genetic basis. Certain

hybrids of platyfish (*Xiphophorus maculatus*) and swordtails (*X. hellerii*) develop malignant melanoma. The genetics of the tumor formation is complex and has been well documented [62]. The *Xiphophorus* model has been widely studied in an effort to better understand the mechanism of melanoma formation in humans. The *Xiphophorus* model is well established as its three genomes has been sequenced ([63] and Appendix A). Eels (Anguilliformes) are excellent models of bone demineralization and childhood kidney cancer (Wilms' tumor). Physiological stress, fasting, or extensive migrations in eels cause bone resorption. Thyroid hormone has also been shown to be in involved in demineralization of their bone. An excess of the thyroid hormone can cause osteoporosis in humans, too. Wilms' tumor occurs naturally in a high percentage of eels in nature [45] and in 1 in 10,000 children at an early age. Eels form a natural model which is not available elsewhere in order to study this childhood tumor. The bicolor damselfish *Stegastes partitus* develops multiple neurofibromas and pigment-cell tumors [64]. Rainbow Trout (*Oncorhynchus mykiss*) is one of the oldest models for studying human liver cancer (hepatoma) since it is particularly susceptible to environmental carcinogens, especially aflatoxin B1 produced by *Aspergillus*. The hepatoma in trout is strikingly similar in histopathology to that in humans. Mutations in the Ki-ras2 Kirsten rat sarcoma viral oncogene homolog (*KRAS*) oncogene often resulting in hepatoma are common in both trout and humans. The trout has since been used to identify other environmental carcinogens. Conversely, several chemoprevention studies have been done to determine dietary supplements that inhibit hepatoma in trout previously exposed to aflatoxin [65]. The mummichog (*Fundulus heteroclitus*, Cyprinodontiformes) has adapted to hypoxic water conditions, extreme high and low water temperatures, high and low salinity and has developed a tolerance to toxic pollutants produced by municipal, agricultural and industrial sources, as well as to oils and gasoline from fishing and pleasure boat traffic. Its ability to adapt to such conditions make it a model for the study of physiological resilience and adaptation [66] and to study human health concerns relative to the widely varying environmental insults that we encounter almost daily.

2.3. Amazon Molly (Poecilia formosa) as a Cytogenomic and Epigenetic Model Species in Human Health Research

Amazon molly is a clonal species [67] and females do not undergo meiosis producing diploid eggs. When a male sperm from a related species stimulates the egg, it develops parthenogenetically—all offspring are clones genetically identical to their mother. This species has experienced first the classical cytogenetic phase of research (e.g., [68]), followed by genomics and (human) cancer research (oncogenomics) [66,67]. As a human disease model, the Amazon molly has been used in research into melanoma [69], infectious diseases [70] and thyroid cancer [71]. It is of interest that [69] introduced a microchromosome into Amazon molly genome inducing susceptibility to melanomas. The next stage was using zebrafish as a single gene knockout model in epigenetic cancer research. The histone demethylase, *KDM2A*, is thought to play a role in silencing transcription in humans. Otherwise, very little is known about its role in vivo in development and disease. Scahill et al. [72] discovered that the loss of the orthologous *kdm2aa* in zebrafish is disruptive to transcriptional processes and produces a high frequency of melanomas. The discovery of the *kdm2aa* mutants represents the first single gene knockout available for the study of melanoma induction. This zebrafish model is important as the World Health Organization reports 132,000 human melanoma skin cancers occur globally each year [72].

Fishes further represent important models also for environmental genomics [73] and for aquaculture genomics, genetics and breeding [74].

3. rDNAomics—Where Fish Ecological Cytogenomics Meets Human Cancer Genomics

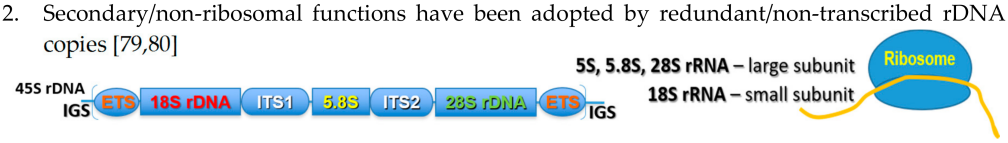
Fish cytogenetics is largely based on chromosomal mapping DNA repeats that still represent serious obstacles in genome sequencing and assembling, even in model species. This has resulted in an immense amount of cytogenetic records including a still increasing number of "bursts" or "explosive spreading" of rDNA across chromosomes mostly in freshwater fishes. Only the availability

of combined genomic data (e.g., Illumina and PacBio in the case of pikes, [8]), has enabled to quantify these extremely amplified rDNAs and to analyze them in a broader molecular context. At the moment, such data are available only for *Esox lucius* and they show that the copy number of the 5S rDNA fraction corresponds to the entire human gene number, i.e., about 20,000 5S rDNA copies. This means that in pike solely the 5S rDNA expanded to such an extent inconceivable in human or mouse genome and left the 45S rDNA fraction unamplified. Based on currently available cytogenetic data, similar situation in the 45S rDNA fraction can be expected in salmonids and in other fish groups (e.g., erythrinids, etc.) with the awaited increasing availability of hybrid sequencing. Then, we will be able to better analyze the causes and consequences of these phenomena. The results presented by the Animal rDNA Database [39] indicate that the freshwater environment might have favored these extreme amplifications of only one of the two rRNA gene fractions. This clearly rules out the primary ribosomal function of amplified rDNA molecules. Below in Figure 1 we summarize and visualize our current knowledge on coding and non-coding (non-ribosomal) functions and roles of rDNA generally and in the case of formation of reproductive barriers in fishes. The origin of these copy number bursts of rDNAs might be potentially related to the "rRNA gene amplification system," which is finely tuned to maintain or, when necessary, to recover a particular and species-specific number of rDNA copies (explained by [75], section 10.4). The exact molecular mechanism(s) are the matter of vivid discussions and speculations at the moment (e.g., [8]) and include among others unequal sister chromatid recombination or retrotransposition. However, evidence accumulates that nucleolus (i.e., sites of the active rDNA transcription) and alone the rDNA copy number are involved in human diseases, cancer predisposition and other oncogenic activities related to genomic instability [76–78].

As one of the most archaic genomic fractions, rDNA became involved in diversified levels of essential cellular functions

1. Primary/ribosomal function of transcribed rDNA (= rRNA) as backbone of ribosomes

2. Secondary/non-ribosomal functions have been adopted by redundant/non-transcribed rDNA copies [79,80]



   i. Recombination-based DNA repair
   ii. Maintenance of genome integrity mediated by multiplied rDNA copies as reaction to environmental changes
   iii. Cellular senescence
   iv. Human health and diseases [74–76]

On this level a specialized „**rDNA gene amplification system**" evolved to maintain an optimal number of rDNA repeats.

Protacathopterygian fishes represent a suitable model system of rDNA evolution due to their tendency of (extreme) multiplication of: for 5S rDNA in two European species of *Esox* [8], for 45S rDNA in the endemic *Coregonus fontanae* [33] and 45S rDNA in the species *Salvelinus fontinalis* [81]

3. Tertiary roles/evolutionary consequences as reproductive barrier and chromosomal background of speciation leading to genome differentiation and finally to speciation: e.g., statistically supported role of 45S rDNA sites in genome differentiation within the *Coregonus clupeaformis* complex in Canada [82,83] or endemic origin of *Coregonus fontanae* from the cosmopolitan *Coregonus albula* [33] accelerated via extensive multiplication of 45S rDNA.

Figure 1. Three levels of ribosomal DNAs (rDNAs) functionality, generally and specifically in fishes.

4. State-of-the-Art in Fish (Cyto)Genomics—The Starting Point

Several reviews on fish genomics have been published during 2003–2013 [77,78,84] and one recent (2016) special issue called 'Fish, Genes and Genomes: Contributions to Ecology, Evolution and Management' [85], which is mostly focused on population genetics and population and conservation genomics. In the meanwhile, numerous fish genomes have been published and several major milestones have occurred in the field of fish molecular cytogenetics: (1) Fish Karyotypes was published [86]—this immense work summarized previous knowledge and still provides a valuable reference overview of karyotypes and where available other karyological and cytogenetic traits in 3425 species/subspecies of extant jawless, cartilaginous, ray-finned and lobe-finned fishes and becomes thus an important reference tool; (2) A still continuing boom of FISH technique in fish resulting in the still increasing amount of molecular cytogenetic data; (3) Availability of more sophisticated and informative methods like GISH (genomic in situ hybridization) and CGH (comparative genomic hybridization) in fish resulting in comparative studies ([35,87,88] and more); (4) Large fish genomes with record chromosome numbers have been documented and analyzed and shifted the limits of fish genome size ([89] *Acipenser brevirostrum*, *A. mikadoi*, *Diptychus dipogon*). In parallel, the first non-human vertebrate and at the same time the first fish genome has been published—the *Takifuga rubripes* and two other pufferfish genomes thereafter (*T. flavidus* and *Tetraodon nigroviridis*) followed by newer versions of their genome assembly (details below). These fishes are of tremendous interest to fish cytogenomicists and generally to vertebrate genomics since they involve several crucial phenomena of the fish genome evolution that still represent outstanding questions. They have the smallest known of vertebrate genomes (350–500 Mb, [90]; Appendix A), which was the impetus for their sequencing. This among others enabled the first accurate prediction of the number of human protein coding genes [91]. However, their compact genomes retain similar chromosome numbers as teleosts with doubled genome size (i.e., 2n = 42/44), ([90]; Appendix A). Another notable feature of the tetraodontid genome its increased GC content—their GC-rich regions are gene-rich as in mammals [90], although the typical mammalian type of chromosome banding does not occur in pufferfishes [7]. On the other hand, there are other teleost species with comparably high genomic GC content (e.g., *Clupea, Gasterosteus, Gadus,* for details see Appendix A), however, without the extreme genome size reduction and genome compactness as in tetraodontids. A situation so far unprecedented among all fishes has been documented in extant gar genera (*Atractosteus* and *Lepisosteus*, Lepisosteiformes). Although their genomic GC content is not so increased as in pufferfishes, or as in other aforementioned species with unreduced genome size, their AT/GC compositional heterogeneity is unparalleled among cold-blooded vertebrates and is cytogenetically detectable in the same way as in mammals i.e., G-banding and AT/GC banding functionality [7]. This briefly illustrates the extent and the complexity of genome organization within fishes and the need to exploit the available resources in integrative cytogenomic approaches to at least partly clarify factors involved in the evolution of functional genome organization in vertebrates.

Table 1. Main cytogenomic traits in fish-like chordates.

Group	2*n* Chromosome Counts (Basic Features)	Micro- Chromosomes	CG Heterogeneity	WGD after the First Two Basal Vertebrates´ WGDs	C-Value/Haploid DNA Content (pg) [12]	Specific Features in the Genome History and Chromosomal Evolution
Myxiniformes (hagfishes)	14–48	NO	unknown	not observed	Myxinidae 2.5–4.59	chromatin diminution, programmed genome rearrengement [13,14]
Petromyzontiformes (lampreys)	76–178	NO	GC-rich DNA repeats	not observed	Petromyzontidae 1.29–2.5	programmed genome rearrengement [14]
Chondrichthyes (cartilaginous fishes)	54–102	YES	observed, presumably satellite DNA	not observed	Chimeriformes 1.5-2 Selachimorpha ~3–17 Rajimorphii 2.7–17	AT/GC heterogeneity positively correlated with genome size [15]
Ceratodontiformes (lungfishes)	34–68	Only *N. forsteri* [15] otherwise not	unknown	not observed	*Neoceradotus* 52.75–74.86 *Lepidosiren* 80.55–123.9 *Protopterus* 40–132.8	"genomic obesity" without WGD documented [16,17]
Coelacanthiformes (lobe-finned fishes)	48	YES	unknown	not observed	*Latimeria* 2.8–6.6	chromosomes similar to ancient frogs [18]
Acipenseriformes (sturgeons, paddlefish)	~120–240–360	YES ~50%	NORs and GC-rich microchromosomes [7] and G-banding [19,20]	multiple in sturgeons, one in paddlefish	*Acipenser* 1.8–9.3 *Polyodon* 1.6–2.4	multiple WGD, ploidy diversity
Lepisosteiformes (gars)	56–58	small sized chromosomes	in both genera	not observed	*Atractosteus* 1.2 *Lepisosteus* 1.4	regionally high recombination rate
Amiiformes (bowfin)	46	NO	only NORs	not observed	*Amia calva* 1.2	convergent evolution with teleosts?
Polypteriformes (bichirs)	36–38 biarmed, extremelly large	NO	only NORs	not observed	*Erpetoichthys* 4.5 *Polypterus* 3.6-7.2	not investigated
Teleostei	~50 (exceptions up to 100–150 or more)	Micro B-chromosomes	only NORs	TGD and lineage specific WGDs	mostly 0.4- ~1.0	from genome compaction to lineage specific WGD

WGD: Whole-genome duplication; NOR: Nucleolar organizer region; pg: picograms.

5. Fish Cytogenomics—Practical Application of Integrated Cytogenetics and Genomics

5.1. Assignment of Linkage Groups to Specific Chromosomes Using FISH with Sequenced BACs

This approach has been so far applied only in several model fish species: zebrafish [92], rainbow trout [93], Atlantic salmon [94], Nile tilapia [95]. However, due to the workload required not all species have been processed in this way and even in species with genomes assembled to the chromosome level their linkage groups (LGs) still have not been assigned to chromosomes. Hence, there still exists need to continue in this effort in model as well as non-model species e.g., spotted gar [7,82], Lake Whitefish [83,96], Northern pike [97], etc. that will finally allow for down-stream analyses and combinations of data obtained by molecular cytogenetics and genome sequencing. This is crucial e.g., to explore the (1) genomic context of the rDNA sites on specific chromosomes; (2) the DNA sequence of centromeric and pericentromeric regions that frequently harbor diverse repetitive elements and show either AT- or GC-richness; finally, (3) identification of residual tetrasomic sites in paleo-tetraploids (salmonids) which is currently impossible by means of genomic and bioinformatics tools [98]. However, using FISH mapping of BACs, we have identified already two of eight predicted residually tetrasomic sites (i.e., sites preserving the ancestral tetraploid condition within the otherwise secondary diploidized genome) in the Lake Whitefish (*C. clupeaformis*, [99]).

5.2. Quick Qualitative Analysis and Visualization of AT- vs. GC-Rich Chromosomal Regions

Quick qualitative analysis and visualization of AT vs. GC rich chromosomal regions utilizing fluorescent staining specific for AT (DAPI) and for GC (Chromomycin A_3 (CMA_3), 7-amino-actinomycin D (7-AAD), propidium iodide). This is a particularly useful approach in combination with bioinformatics analysis producing AT/GC profile across chromosomes/LGs hence quantifying cytogenetic data. Basically, there are two possible approaches producing AT/GC profiles across linkage group/chromosome: (1) chromoplot [7] calculates and plots the absolute GC percentage (Figure 2) or (2) the tool isoSegmenter [100] segmenting genome into pre-defined and broadly used concept of "isochores" (i.e., large genomic regions homogeneous in their GC content, *sensu* [101]). This DNA compositional cytogenomics gained in importance due to the recently uncovered AT/GC compositional heterogeneity in the ancient gars [7]. This finding means that it is crucial (a) to change our attitude to generally accepted compositionally homogenous fish genomes, which is not true anymore; (b) to revisit the so far obtained results and to employ e.g., the simultaneous DAPI/CMA_3 staining together with attempts of G-banding in fishes. Some authors tend to present DAPI stained metaphases separately from the CMA_3 stained ones or not to show one of them. This makes the situation complicated for any serious reason and does not allow for proper exploitation of the data (i.e., a part of information would be missing despite its actual availability); (c) this sheds new light on the vertebrate genomic DNA composition generally because so far the broadly accepted concept has considered fish and amphibians AT/GC homogenous whereas only birds and mammals were AT/GC heterogeneous, with transient situations in reptiles [101]. However, the bromodeoxyuridine (BrdU)/pulse replication labeling does produce reproducible bands in both compositionally heterogeneous higher vertebrates and homogenous fish and amphibians (e.g., for fish [102]). Therefore, in the light of findings in gars and the bowfin, it would be highly desirable to perform a large-scale comparative cytogenomic study across fishes to be able to exactly analyze the potential banding pattern in fishes so far considered compositionally homogenous. Namely, the BrdU banding pattern should provide a "scaffold" of expected bands, which should be then analyzed with DAPI/CMA_3 and in parallel on the LG profiles. In this way, we should be able to quantify the cytogenetic thresholds of chromosomal band visualization and make the approach of chromosomal banding more sensitive to the putative less heterogeneous pattern in fish genomes. This step will be essential to properly understand the issue of compositional organization of fish and amphibian versus avian and mammalian genomes and finally the vertebrates' genomes generally. The quantification on AT/GC profiles will require higher versions of better genome assemblies assembled to the chromosome level and with already filled gaps (or filled as much as

possible). This means, that the so far available versions of genomes were not yet suitable for these analyses and hence, fish researchers have not yet missed anything.

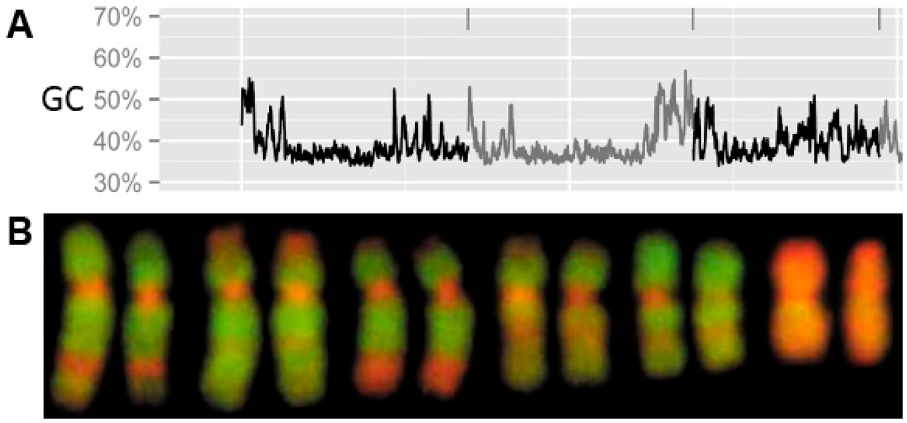


Figure 2. (**A**) GC-profiles across three linkage groups (LGs) so far unassigned to their corresponding chromosome pairs showing fluctuations in GC-percentage produced using the chromoplot tool. These three linkage LGs are arranged according to their numbers in the Ensembl [103] and separated by vertical lines above profiles and by the alternation of gray and black colors (*x*-axis—genome position, *y*-axis—GC%); (**B**) Partial karyotype stained with 4',6-diamidino-2-phenylindole/ Chromomycin A_3 (DAPI/CMA_3) showing six pairs of larger chromosomes of the spotted gar with altering GC-rich regions in red and AT-rich regions in green. Reproduced with permission from [7].

5.3. Cytogenetic Mapping of Repetitive Sequences on Conspecific Populations

Cytogenetic mapping of repetitive sequences on conspecific populations exploring their participation in evolutionary diversifications of different vertebrate species. In studies by [83,96], *population* molecular cytogenetics was shown to be an irreplaceable tool in exploring the population dynamics of rDNA sites. On the example of the Lake Whitefish (*Coregonus clupeaformis*, Salmonidae), we have documented that the differential dynamics of rDNAs across chromosomes participate in the evolutionary diversification of fish genomes. Similar studies performed in Erythrinidae (Characiformes) e.g., by [37,104] (this issue and citations therein) show comparably high intra-species evolutionary dynamics in rDNA and also other repetitive DNA in the genera *Erythrinus* and *Hoplias*. These and other similar finding originating from *freshwater* fish groups show the relevance of the FISH-based rDNA mapping (1) for understanding evolutionary mechanisms underlying ecological speciation [35] and (2) as an important tool how to tackle incipient and established biodiversity [105].

5.4. Cytogenomics of Duplicated Genomes—Understanding Mechanisms of Genome Evolution in Vertebrates Which Have Undergone Whole-Genome Duplication(s)

Whereas there are no further WGD events in higher (warm-blooded) vertebrates, in lower vertebrates, particularly in fishes (recent review by [106]) but also in amphibians and to some extent in reptiles [107], there are numerous examples of WGDs. Hence, fish genomes offer irreplaceable insights into the diversity of evolutionary patterns in Sarcopterygia and Actinopterygia (Table 1). However, to sequence and above all to assemble such genomes remains challenging (bichirs, lungfishes, sturgeons, salamanders). Hence, the "classical" cytogenetics still represents an important tool to analyze these genomes and the post-WGD evolution [20,89,108]. Not only in these cases, cytogenetics can largely benefit from genomics and bioinformatics and compensate for the obstacles during sequencing and genome assembling ([19], this special issue).

6. Fish Cytogenetics and Cytogenomics in the Era of Digitalized Biology—The Boom of Databases

6.1. Databases for Fish Biology

Currently, fish genetics is beginning to benefit from increasing diversified efforts to build databases to compile and curate the increasing amount of molecular data on fish genomes. Hence, despite the devotion to fish genomics of the Indian subcontinent, we can explore and utilize in evolutionary studies e.g.: "FBIS: A regional DNA barcode archival and analysis system for Indian fishes" introduced by [109], "FMiR: A Curated Resource of Mitochondrial DNA Information for Fish" introduced by [110], "FishMicrosat: a microsatellite database of commercially important fishes and shellfishes of the Indian subcontinent" by [111]. Microsatellites of 31 fish species not confined to India are available in the microsatellite database MSDB [112]. The particularly important tool "Fish Karyome" is now available in its upgraded version [113]. Beside these four tools originating from India, the Animal rDNA database [39] including in its current version 546 fish and fish-like species is a highly relevant and informative tool usable in evolutionary-ecological fish cytogenomics or rDNAomics. Genome size of more than 2000 fish and fish-like species is currently available at the online database [12]. General data about fish biology, ecology and biogeography are already traditionally provided by Fishbase [114] and represent the first-choice tool for any evolutionary studies, including cytogenomics, in the ecological context. An online database specialized on B chromosomes and involving fishes is described in the Section 8.2 dealing specifically with B chromosomes.

6.2. History of the Zebrafish Reference Genome

History of the zebrafish reference genome started in 2001, when the Wellcome Trust Sanger Institute initiated the zebrafish genome-sequencing project and selected the Tübingen zebrafish reference strain as it had been widely used to identify mutations affecting embryogenesis. The Zv8 assembly was a hybrid of high-quality clone sequence (83%) and whole-genome shotgun (WGS) sequence (17%), with a total size of 1.412 gigabases (Gb). The clone and WGS sequence is tied to a high-resolution, high-density meiotic map called the Sanger AB Tübingen map (SATmap). This full genome sequence was made available to the public at the NCBI Zebrafish Genome Page and is maintained by the Genome Reference Consortium (FishMap Zv8 [115]). In 2009, the Institute of Genomics and Integrative Biology in India sequenced the genome of a wild zebrafish strain. The genome contained about 1.7 Gb and when compared to the Zv8 variations were found in over 5 million nucleotides and over 1.6 million indels. Later, the zebrafish reference genome was published, consisting of 1.4 Gb and over 26,000 protein-coding genes [32]. After the release of the Zv8 project, they joined the Genome Reference Consortium (GRC) for further improvement and maintenance. The GRC has now released a new reference assembly, GRCz11. Sanger's GRC partners at the ZebraFish Information Network (ZFIN) continue in updates and maintenance. In parallel, Amores & Postlethwait [116] recognized a need for molecular cytogenetics in zebrafish and performed an exhausting study to facilitate the unambiguous cytogenetic identification of each individual chromosome. Several further molecular cytogenetic studies localized repetitive sequences in zebrafish [117–119].

6.3. Fish Genomes Available

Using the NCBI Genome Browser [120] and filtering for the Kingdom "Eukaryota," Group "Animals" and Subgroup "Fishes" we find further about 90 fish species with genomes sequenced and assembled to the diverse levels. Additional fish species with a sequenced genome available can be found in a literature search resulting at the moment with about 95 species. These currently available species are listed in Appendix A, an online continuously updated version of this list will be available on the web page [121]. This list in alphabetical order includes: 2 lancelets, 1 hagfish, 2 lampreys, 1 shark, 1 ray, 1 gar representing the non-telesost Actinopterygia, more than 80 teleosts and 1 coelacanth

of the group Sarcopterygia. In this dataset, we integrate basic genomic with basic cytogenomic traits—genome size in Mb based on sequencing, genome size as C-value originating from [12]; linkage groups (haploid) and diploid chromosome numbers; level and number of assemblies (i.e., draft, contig, scaffold, chromosome level), genomic GC percentage and sequencing platform applied emphasizing the PacBio technology with its particular relevance for chromosomal and repetitive DNA studies. The phylogenetic "coverage" of sequenced fish and fish-like species is provided in Appendix A and will be also online and continuously updated. Some of the model fish species have been (re)sequenced and (re)assembled up to 4, 5 times (see column #Assembly in Appendix A; details e.g., at [122] or at the Genome Assembly Database of the European Nucleotide Archive (ENA, [123]). Importantly, each new version means filling more gaps, improving scaffolds towards the chromosome level, improving of genome size estimations and proportion of repetitive sequences e.g., [124].

Sometimes substantially different results are found than in the previous genome versions (e.g., in Atlantic cod, [124]). Moreover, there are strong indications that numerous non-model genomes contain widespread and predictable assembly and annotation errors [30], which means that using their improved versions will be critical. Other de novo draft genome assemblies of 66 teleost species are available through [125] (these species will be included in further online version of Appendix A). Further, the Genome 10K Project aims to sequence and analyze 10,000 vertebrate species including fish and hence other fish genomes are expected to be available in the future [126]. Beside the aforementioned genomic sequences, 24 fish transcriptomes are available through the PhyloFish database [127]. On the top of that the Fish T1K platform announced transcriptomes of 124 actinopterygian fish species (covering 46 orders and 99 families) to have been already sequenced and further 59 species were in progress [128]. Since this web-based information has been updated only in 30 April 2015, the results might have been already substantially more progressed. Genomes assembled to the chromosomal level are particularly important for cytogenomics—currently there are 18 species (see Appendix A). Once the available linkage groups (LGs) are assigned to their chromosomes, the cytogenetic results will be directly applicable in genomic data and vice versa. This is however, still in the future, hopefully not too far away.

6.4. Future of Fish Cytogenomics = Phylogenomics

The abovementioned tools are together with the most recent fish phylogenetics published by [22] and [10], an excellent starting point for the application of modern phylogenetic comparative [129] and other quantitative methods exploring cytogenomic phenomena in the broader eco-evo context.

7. Quantitative Eco-Evo Cytogenomics

Fishes are the only group of vertebrates, where relationships between genome size and essential cellular parameters, sometimes called cytogenomic ratio, remain uncertain [130]. In fish, similarly as in other vertebrates and eukaryotes generally, two opposite sets of theories attempt to explain the mechanisms behind the large variation in the amount of non-coding DNA—adaptive and non-adaptive theories [131–133]. Genome size is negatively correlated with GC percentage in several fishes, like *Tetraodon*, *Takifugu*, *Gasterosteus*, etc. (Appendix A). On the other hand, also larger fish genomes show GC increment (*Clupea*, *Gadus*) or even the mammalian like GC heterogeneity (gars, [7]). There is also a clear effect of environment and life-style on the GC content in teleosts [134]. Further, genome size obviously correlates with chromosome numbers, although considerable modifications in genome size can occur largely independent of changes in chromosome counts [29]. Here, the role of environment has been already repeatedly evidenced: occurrence of larger genomes in freshwater versus marine fishes [28,135,136], so far not recorded polyploidy in marine actinopterygians versus frequent incidence of polyploidy among chondrichthyans [12,15] and freshwater actinopterygians [28,29] and higher cytogenomic ratios in chondrosteans than in actinopterygians as in cold-water fishes relative to their warm-water counterparts [130]. These few examples show the complex network of

numerous interactions on diverse levels ranging from molecular, over developmental, physiological and life-history traits to the environmental.

However, we still need to better explore and understand all these levels and the interactions, since there are several major topics and related outstanding issues in fish quantitative eco-evo cytogenomics, where the genomics resources, e.g., data on exon/intron sizes and counts will be highly desirable to complement the existing robust datasets obtained from cytogenetics and flow cytometry: (1) the conundrum of remarkably constant diploid chromosome number (48–50) even among species that differ significantly in DNA content [28,135]—are there any phylogenetic constrains and/or selective pressures directing chromosomal evolution towards $2n$ = 48–50? In other words, do we have to look for potential nucleotypic limitations of genome size on cellular and organismal phenotypes? i.e., any links between rate of development, life-span length, metabolic rate, body size and cytogenomic parameters?; (2) the role of water environment (freshwater x marine) in the dynamics of evolution of genome size, chromosome number, fundamental number etc.—occurrence of larger genomes in freshwater versus marine fishes [28,135,136]; (3) generally the incidence of polyploidy only in the freshwater environment with the exception of chondrichthyans; (4) origin of the genomic gigantism (obesity) in lungfishes and bichirs [12,17,137].

Utilizing increasingly available karyological and cytogenetic data, several attempts have been performed to quantify these purely qualitative data and assess them in ecological, life history traits and physiological context. In this way, chromosome and chromosome arms (FN) numbers have been compared as a measure of chromosomal dynamics several times [138]. These studies indicate a chromosomal stability and conservatism in marine fishes although limited to restricted regions and groups [138]. All these studies represent important first steps although they are largely limited to a narrow subset of fish lineages, limited geographically and certainly not exploiting the available data and advanced cytogenomic methods. However, there is an essential shift from the previous mostly descriptive approach towards understanding the phenomena observed at the functional stage.

8. Sex and B Chromosomes, Nuclear Architecture and Genome Evolution Research

Here we summarize three further aspects of genome evolution, where rDNA does play an important role and where the integrative cytogenomic already proved its potential. B chromosomes were among others proved to originate from sex chromosomes [139] or to interact with sex determination [140].

8.1. Sex Chromosomes in Fish

Sex chromosomes in fish are exhaustively discussed in another paper of this special issue and illustrated on the well explored example of Neotropical fishes [104]. We would like to highlight the contribution of rDNA regions in sex chromosomes evolution [104,141] and the cytogenomic approach that proved successful in combining both cytogenetic and genomic data e.g., [141] in the two most important model species. The mechanism of sex-determination in zebrafish is of importance to understand if it is to serve as a vertebrate model system to study human development and disease. However, past researchers have failed to find either an XY, ZW, multiple sex-determining mechanism, or environmental determination. Recently, using cytogenomics, this question has been mostly answered. A novel genetic map of single-nucleotide polymorphism (SNP) was used in a genome-wide linkage study of sex-determination in zebrafish [142]. Loci were identified on zebrafish chromosomes 5 and 16. Chromosome 5 locus contains *dmrt1*, a gene found in sex determination from fruit flies to humans. A mutation in the orthologue of this gene in humans results in complete sex reversal of XY individuals. Chromosome 16 contains *cyp21a2*. Mutation of the human orthologue of this gene is a common cause of pseudohermaphroditism. Recently, zebrafish chromosome 4 has been identified as a sex chromosome along with the sex-linked genes on chromosomes 5 and 6 discussed above [143]. In zebrafish, there is a combination of effects on the genome, germ cells and the environment with influences from epigenetic factors. However, the primary factors in

sex-determination in zebrafish remain controversial [143]. The Japanese Medaka has an XX-XY chromosome-based sex determination similar to mammals with the male determining master regulator genes on the Y chromosome. Interestingly, this mode is not conserved even within the genus *Oryzias* [144].

8.2. B Chromosomes (Not Only) in Fish

B chromosomes (not only) in fish is another area of biology, where cytogenetics successfully and productively meets genomics and rDNAomics [145,146]. B chromosomes are known to contain rDNA frequently e.g., [147], summarized by [79,139]. B chromosomes have been identified e.g., in seven South American and in fourteen African cichlid species [143], in the genus *Poecilia* as we already mentioned by [148], in the bleak (*Alburnus alburnus,* Cyprinidae) [149] and in three species of thorny catfishes [150]. The complete list of B chromosomes identified in fishes is available in a specialized database 'B chrom' [151] by [152]. There are 278 entries are listed comprising approx. 120 species (depending on the species status and the level of species identification; accessed on 24 January 2018).

8.3. Nuclear Architecture

Nuclear architecture—nucleolus and rDNA emerged as important components of the nuclear architecture [78] and as indispensable components of mechanisms maintaining genomic integrity [81,153,154]. Moreover, the repetitive nature of rDNA and other repetitive genes (e.g., histones) results in a high evolutionary dynamics, known also as evolutionary hotspots [80,155]. Therefore, to understand the complex evolutionary structural as well as functional, mechanisms in vertebrate genomes, it is crucial to view the current cytogenomic knowledge also in the context of nuclear architecture, regulation of gene expression, role of transposons and epigenetics. At this stage, any attempts to explore structural aspects of interphase nuclei in fishes generally are missing. The single study on basic organization of two cold-blooded vertebrate genomes by [156] demonstrated that gene-rich regions in one amphibian (*Rana esculenta*) and one reptile (*Podarcis sicula*) occupy the more internal part of the nuclei, whereas the gene-poor regions occupy the periphery. This finding is similar to that previously reported in warm-blooded vertebrates, despite the lower GC levels of the gene-rich regions of cold-blooded vertebrates [156] and citations therein. In Atlantic cod, Kirubakaran et al. showed an example of putative directional selection for retaining two adjacent inversions on LG1 [157]. These inversions repress meiotic recombination in crosses. Moreover, the chromosomal block with these inversions harbors 763 genes, including candidates regulating swim bladder pressure, heme synthesis and skeletal muscle organization conferring adaptation to long-distance migrations and vertical movements down to large depths. Despite interbreeding between forms with (migratory ecotype) and without (stationary ecotype), the inversions are maintaining genetic differentiation [157].

9. Roots of (Population) rDNAomics in Fish Cytogenetics—rDNAomics as Another Dimension of Environmental Genomics

There is a long tradition in chromosomal mapping of rDNA sites in the fish cytogenetics. This descriptive and qualitative work resulted in astonishing findings of extremely multiplied rDNAs, both 45S and 5S rDNA [8,35,37,158], for numerous other examples, see the animal rDNA database [39]. Mapping of rDNAs (but also of other highly repetitive DNA fractions as e.g., histones and transposons ([159,160]) across populations proved to be an important tool to explore the (sub)chromosomal background of populations' diversification, incipient speciation and finally completed speciation events. These repetitive sequences appear to evolve at a higher rate and their mapping hence enables to catch various stages on the gradient of genome diversification, sometimes even stages that are not distinguishable on the morphological or genetic level ([161] vs. [35]). However, only in combination with other molecular methods and genomic data, a precise quantification and detailed insight became possible [8]. In this way, we documented a peculiar higher-order organization of the extremely amplified, potentially functional and massively methylated 5S rDNA in two species of European pikes,

whereas the 45S rDNA fraction was ascertained in both of the pike species not to have undergone any amplification (Figure 3).

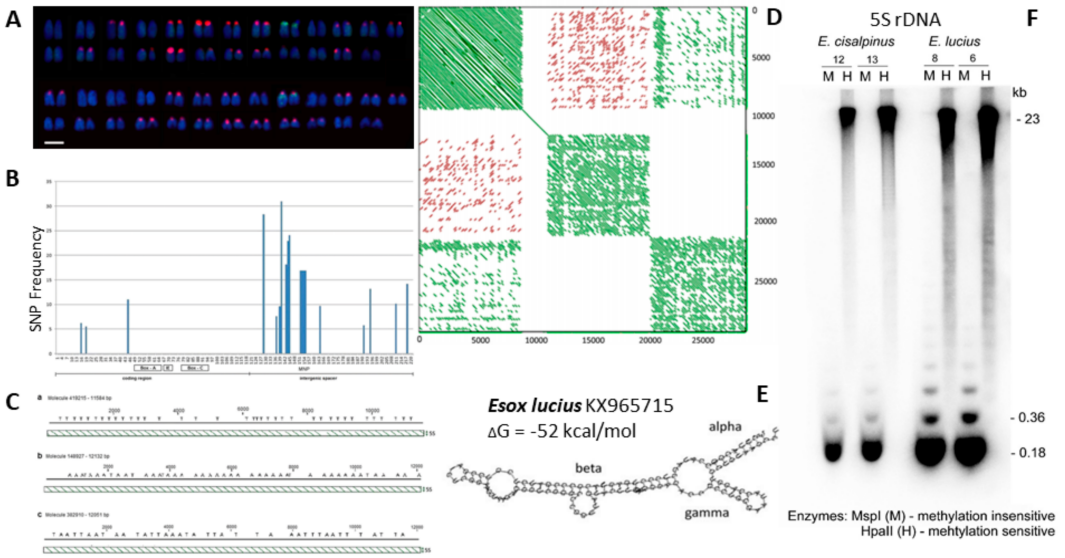


Figure 3. Summary of an integrative cytogenomic study on the rDNAome in European pikes (*Esox lucius* and *E. cisalpinus*). (**A**) Fluorescence *in situ* hybridization (FISH) with 5S rDNA (red) and 45S rDNA (green); (**B**) Distribution of single-nucleotide polymorphism (SNPs) along the *E. lucius* 5S rDNA unit obtained from Illumina reads showing absence of SNPs in the internal controlling region composed of Box-A, IE and Box-C elements; (**C**) Distribution of variants in intergenic spacer regions (IGS) in three PacBio reads (a–c). Slanted lines indicate tandemly arranged units visualized through the alignment of reads (*x*-axis) with a 5S gene (*y*-axis); (**D**) Higher-order organization of 5S rDNA arrays in *E. lucius*. Self-to-self comparison of long PacBio molecules representing three groups; (**E**) 5S rRNA domain reconstruction of *E. lucius* indicating its potential functionality; (**F**) Methylation analysis of 5S rDNA by the methylation-sensitive HpaII (H) restriction enzyme and its methylation-insensitive MspI (M) isoschizomere. Reproduced from [8].

Interestingly, Salmoniformes, the sister lineage of Esociformes, where pikes belong to, tend to amplify 45S rDNA with mostly stable and low 5S rDNA copy numbers [35]. Hence, this whole group of Protacanthopterygia (i.e., Salmoniformes and Esociformes, [20]) provides a suitable model system for further exploring and above all understanding the evolutionary dynamics of their rDNAome. Already the available findings in fish rDNAs are crucial since they show the immense differences between the genome and rDNAome complexity in lower and higher vertebrates. Namely, the discrepancy between copy number of 5S and 45S rDNA (e.g., [76]), which is being tolerated by fish and contributes to genomes diversification, speciation and finally to increase in biodiversity [8,35,83,96]. In mammals, differences in copy number between 5S and 45 rDNAs has been proved to be involved in pathological conditions.

10. Conclusions

Nucleolus and rDNAs are the hub integrating environmental and intracellular signals [81] and the cellular stress sensor [153]. Moreover, rDNA copy number has been shown to play a crucial role in maintenance of the genome integrity and onset of diseases and senescence [162]. Therefore, the integrative cytogenomic analysis of not only fish rDNAomes represents another, so far unexploited, dimension of the genomics in fish *sensu* [73] and allows alternative insights into the complex interactions between cells and organisms and their environment. In future, more systematic studies on molecular cytogenetic detection of 5S and 45S rDNA in different populations and/or species

will enable us to assess the role of rDNA spreading across chromosomes in genome differentiation at different environmental conditions. As we showed, the vertebrates' rDNAome still represents a largely underestimated and unexploited genomic fraction with the huge potential to elucidate and proper understand crucial genomic functions and above all genome's interactions with the environment.

Acknowledgments: R.S. was supported by the Faculty of Science of the University of Hradec Králové. The costs of open access were covered by funds of the University of Hradec Králové provided to R.S. This study was supported by Specific Research Project from University of Hradec Kralove Nr. 2105/2017.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. List of Fish Species with Sequenced Genomes Currently Available

Table A1. Detailed overview of fish species with a sequenced genome. Diverse levels of genome assemblies (draft, contig, scaffold, fully assembled genomes to the chromosome level) and numbers of assembly versions are listed together with basic cytogenetic traits (2*n*, C-value, GC%).

Species	Order [a]	Family [a]	Size (Mb)	C-val [b]	*n*/2*n* [c]	Assembly/Type [d]	GC%	Notes [e]
Acanthochromis polyacanthus	Ovalentaria	Pomacentridae	991.585	0.94	—	1/S	41.5	
Amphilophus citrinellus	Cichliformes	Cichlasomat., Heroini	844.903	—	—/48	1/S	41.4	
Anguilla anguilla	Anguilliformes	Anguillidae	1018.7	1.11–1.67	—/38	1/S	42.9	
Anguilla japonica	Anguilliformes	Anguillidae	1288.6	1.09	—/38	1/S	38.8	
Anguilla rostrata	Anguilliformes	Anguillidae	1413.03	1.01–1.66	—/38	1/S	41.0	
Anoplopoma fimbria	Perciform × Scorpaeniform	Anoplomatidae	699.326	0.71–0.84	—	1/C	40.3	
Astyanax mexicanus	Characiformes	Characidae	1335.24	—	25/50	1/Ch	38.4	
Austrofundulus limnaeus	Cyprinodontiformes	Rivulidae	866.963	—	—	1/S	41.1	
Boleophthalmus pectinirostris	Gobiiformes	Gobiid., Oxudercinae	955.752	—	—/46	1/S	40.1	
Branchiostoma belcheri	Cephalochordata	Branchiostomidae	426.124	—	—/36	2/S	40.15	
Branchiostoma floridae	Cephalochordata	Branchiostomidae	521.895	—	—	1/S	41.8	
Callorhinchus milii	Chimaeriformes	Callorhinchidae	974.499	1.94	—	1/S	42.6	
Channa argus	Perciformes	Channidae	615.3	0.63–0.77	—/48	1/D	—	
Clupea harengus	Clupeiformes	Clupeidae	807.712	—	—/50-52	1/S	44.5	
Cottus rhenanus	Perciformes	Cottidae	563.609	—	—	1/S	36.8	
Cynoglossus semilaevis	Pleuronectiformes	Cynoglossidae	470.199	0.62 *C. bilineatus*	22	1/Ch	41.27	
Cyprinodon nevadensis	Cyprinodontiformes	Cyprinodontidae	1011.85	—	—/48	1/S	39.0	
Cyprinodon variegatus	Cyprinodontiformes	Cyprinodontidae	1035.18	1.6	—/48	1/S	39.5	
Cyprinus carpio	Cypriniformes	Cyprinidae	1713.66	1.6–2.0	50/100	2/S, Ch	med 37.1	
Danio rerio	Cypriniformes	Cyprinidae	1679.2	1.6–2.3	25/48	4/3S, 1 Chr	med 36.7	
Dicentrarchus labrax	Moroniformes	Moronidae	675.917	0.78	—	2/S	40.4	
Eptatretus burgeri	Myxiniformes	Myxinidae/Eptatretinae	2608.38	2.98	/36	1/S	—	
Esox lucius	Esociformes	Esocidae	904.497	0.8–1.4	25/50	3/Ch	42.2	PB
Fundulus heteroclitus	Cyprinodontiformes	Fundulidae	1021.9	1.3–1.5	-/48	1/S	41.2	
Gadus morhua	Gadiformes	Gadidae	824.311	0.4–0.9	-/46	2/S	46.3	454, I, PB [124]
Gasterosteus aculeatus	Perciformes	Cottiodei, Gasterosteid.	446.611	0.6–0.7	—/42	2/S, Ch	44.6	LM [163]
Haplochromis burtoni	Perciformes × Cichliformes	Cichlidae (Afr.)	831.412	0.97	—/40	1/S	41.9	
Hippocampus comes	Syngnathiformes	Syngnathidae	493.776	—	—	1/S	43.7	
Ictalurus punctatus	Siluriformes	Ictaluridae	783.275	1.0	29/58	1/Ch	39.8	PB, I [164,165]
Kryptolebias marmoratus	Cyprinodontiformes	Rivulidae	680.367	—	—/48	2/S	med 39.5	
Labeotropheus fuelleborni	Perciformes × Cichliformes	Cichlidae (Afr.)	70.8584	—	—/44	1/S	42.1	
Labrus bergylta	Labriformes	Labridae	805.481	—	—	1/S	40.9	
Larimichthys crocea	Acanthuriformes	Sciaenidae	678.938	—	—	2/S	med 41.4	
Lates calcarifer	Perciformes	Centropomidae	668.481	0.7	—/48	2/S	med 40.6	
Latimeria chalumnae	Ceolacanthiformes	Coelacanthidae	2860.59	2.8–6.6	—/48	2/S	med 42.5	
Lepisosteus oculatus	Lepisosteiformes	Lepisosteidae	945.878	1.4	29/58	1/Ch	40.4	
Lethenteron camtschaticum	Petromyzontiformes	Petromyzontidae	1030.66	~1.4	—/144-162	1/S	48.1	
Leuciscus waleckii	Cypriniformes	Cyprinidae	752.539	congeners 1.2–1.5	—/50	1/S	37.4	
Leucoraja erinacea	Rajiformes	Rajidae	1555.46	3.5–4.6	—	1/C	40.3	
Maccullochella peelii	Perciformes ×	Percichthyidae	633.241	0.83	—	1/S	—	
Maylandia zebra	Perciformes × Cichliformes	Cichlidae (Afr.)	859.842	—	—	2/S	41.4	
Mchenga conophoros	Perciformes × Cichliformes	Cichlidae (Afr.)	73.4256	—	—	1/S	41.8	
Megalobrama amblycephala	Cypriniformes	Cyprinidae, Cultrinae	1116.0	1.12–1.35	—	1/D	37.3	[166]
Melanochromis auratus	Perciformes × Cichliformes	Cichlidae (Afr.)	68.2386	—	—/46	1/S	41.5	
Miichthys miiuy	Acanthuriformes	Sciaenidae	619.301	—	—	1/S	39.3	
Mola mola	Tetraodontiformes	Molidae	639.452	0.8–0.9	—/46	1/S	41.2	[167]

Table A1. *Cont.*

Species	Order [a]	Family [a]	Size (Mb)	C-val [b]	*n*/2*n* [c]	Assembly/Type [d]	GC%	Notes [e]
Monopterus albus	Synbranchiformes	Synbranchidae	684.144	0.6–0.9	—/24	1/S	41.5	
Morone saxatilis	Moroniformes	Moronidae	585.167	0.9	—/48	1/S	40.0	
Neolamprologus brichardi	Perciformes × Cichliformes	Cichlidae (Afr.)	847.91	—	-	1/S	42	
Nothobranchius furzeri	Cyprinodontiformes	Notobranchiidae	1132.74	1.56	19/38	4/S, Chr	43, 8	
Nothobranchius kuhntae	Cyprinodontiformes	Notobranchiidae	5.23461	—	—/38	1/S	44.8	
Notothenia coriiceps	Perciformes	Nototheniidae	636.614	—	—	1/S	40.8	
Oncorhynchus kisutch	Salmoniformes	Salmonidae	2369.93	2.6–3.0	30/60	1/Ch	43.6	
Oncorhynchus mykiss	Salmoniformes	Salmonidae	2179	1.9–2.9	29/60	2/Ch	med 43.7	
Oreochromis niloticus	Perciformes × Cichliformes	Cichlidae (Afr.)	1009.86	0.9–1.2	23/44	2/Ch	med 39.9	PB, I [168,169]
Oryzias latipes	Beloniformes	Adrianichthyidae	869.818	0.9–1.1	24/48	5/Ch	med 40.8	
Pampus argenteus	Scombriformes	Stromateidae	350.449	—	—	1/S	38.2	
Paralichthys olivaceus	Pleuronectiformes	Paralichthyidae	643.911	0.7	24/46–48	2/S, Ch	med 42.4	
Periophthalmodon schlosseri	Gobiiformes	Gobiid., Oxudercinae	679.761	0.96	—	1/S	40.2	
Periophthalmus magnuspinnatus	Gobiiformes	Gobiid., Oxudercinae	701.697	—	—	1/S	40	
Petromyzon marinus	Petromyzontiformes	Petromyzontidae	885.535	1.6–2.4	—/168	1/S	46.8	
Pimephales promelas	Cypriniformes	Cyprinidae	1219.33	1.1	—/50	2/S	med 40.6	
Poecilia formosa	Cyprinodontiformes	Poeciliidae	748.923	0.75–0.97	—/46	1/S	39.6	
Poecilia latipinna	Cyprinodontiformes	Poeciliidae	815.145	0.9–1.0	—/46	1/S	40.8	
Poecilia mexicana	Cyprinodontiformes	Poeciliidae	801.711	0.7–1.38	—/46	1/S	40.7	
Poecilia reticulata	Cyprinodontiformes	Poeciliidae	731.622	0.77–1.0	23/46	1/Ch	40.3	
Protosalanx hyalocranius	Osmeriformes	Salangidae	525	—	—	1/D	—	
Pseudopleuronectes yokohamae	Pleuronectiformes	Pleuronectidae	547.831	0.67	—/48	1/C	42	
Pundamilia nyererei	Perciformes × Cichliformes	Cichlidae (Afr.)	830.133	—	—	1/S	41.9	
Pygocentrus nattereri	Characiformes	Serrasalmidae	1285.35	—	—	1/S	40.6	
Rhamphochromis esox	Perciformes × Cichliformes	Cichlidae (Afr.)	71.2951	—	—	1/S	42.3	
Rhincodon typus	Orectolobiformes	Rhincodontidae	2931.6	—	—	1/S	41.8	
Salmo salar	Salmoniformes	Salmonidae	2966.89	3.0–3.3	29/60	1/Ch	43.9	PB, I, S [170]
Scartelaos histophorus	Gobiiformes	Gobiid., Oxudercinae	695.009	—	—	1/S	39.1	
Scleropages formosus	Osteoglossiformes	Osteoglossidae	777.359	—	—	4/S, Chr	med 43.9	
Sebastes aleutianus	Scorpaeniformes	Sebastidae	899.65	—	—	1/S	40.9	
Sebastes minor	Scorpaeniformes	Sebastidae	681.653	—	—	1/S	40.8	
Sebastes nigrocinctus	Scorpaeniformes	Sebastidae	746.045	—	—	1/S	40.8	
Sebastes rubrivinctus	Scorpaeniformes	Sebastidae	756.297	—	—	1/S	40.7	
Sebastes steindachneri	Scorpaeniformes	Sebastidae	648.011	—	—	1/S	40.7	
Seriola dumerili	Carangiformes	Carangidae	677.67	0.74	—	1/S	40.9	
Seriola lalandi	Carangiformes	Carangidae	685	—	—	1/S	-	PB, I
Seriola quinqueradiata	Carangiformes	Carangidae	med 750.45	0.83	—	2/S	40.25	
Sinocyclocheilus anshuiensis	Cypriniformes	Cyprinidae	1632.72	—	—	1/S	38	
Sinocyclocheilus grahami	Cypriniformes	Cyprinidae	1750.29	2.35	—/96	1/S	38.7	
Sinocyclocheilus rhinocerous	Cypriniformes	Cyprinidae	1655.79	—	—	1/S	38.1	
Squalius pyrenaicus	Cypriniformes	Cyprinidae	48.1393	2.4	—/50	1/C	51.8	
Stegastes partitus	Ovalentaria	Pomacentridae	800.492	—	—	1/S	42.1	
Takifugu flavidus	Tetraodontiformes	Tetraodontidae	378.032	—	—	1/S	45.6	
Takifugu rubripes	Tetraodontiformes	Tetraodontidae	391.485	0.4	22/44	1/Ch	45.8	
Tetraodon nigroviridis	Tetraodontiformes	Tetraodontidae	342.403	0.35–0.51	21/42	7/S, Ch	46.6	Genoscope
Thunnus orientalis	Scombriformes	Scombridae	684.497	—	—/48	1/C	39.7	
Xiphophorus couchianus	Cyprinodontiformes	Poeciliidae	708.396	0.75	—	1/S	40.9	
Xiphophorus hellerii	Cyprinodontiformes	Poeciliidae	733.802	0.7–1.0	—/48	1/S	41.2	
Xiphophorus maculatus	Cyprinodontiformes	Poeciliidae	729.664	0.8–1.0	—/48	1/S	39.8	

[a] orders and families mostly based on [22]; [b] c-value, based on [12] database; [c] *n*: based on genomic/sequence data, originates from NCBI, 2*n* based on cytogenetic data [86]; [d] number of assemblies currently released and level of assembly (C = contig, D = draft, S = scaffold, Ch = chromosomal level); [e] Sequencing methods (I = Illumina, PB = PacBio, S = Sanger), linkage map (LM) available.

References

1. Bernheim, A. Cytogenomics of cancers: From chromosome to sequence. *Mol. Oncol.* **2010**, *4*, 309–322. [CrossRef] [PubMed]
2. Xiang, B.; Leon, A.; Li, M.M.; Iqbal, A.M.; Li, P.; Li, S.; Papenhausen, P.R.; Schwartz, S.; Zhang, X.-X.; Geiersbach, K.B.; et al. Atlas of Cytogenomics in Oncology and Hematology: A Platform-Neutral Clinical Cancer Genomics Database. *Cancer Genet.* **2012**, *205*, 420. [CrossRef]
3. McPherson, M.C.; Robinson, C.M.; Gehlen, L.P.; Delany, M.E. Comparative cytogenomics of poultry: Mapping of single gene and repeat loci in the Japanese quail (*Coturnix japonica*). *Chromosome Res.* **2014**, *22*, 71–83. [CrossRef] [PubMed]
4. Barh, D.; Khan, M.S.; Davies, E. *PlantOmics: The Omics of Plant Science*; Springer: New Delhi, India, 2015; ISBN 978-81-322-2172-2.
5. Kapusta, A.; Suh, A. Evolution of bird genomes-a transposon's-eye view. *Ann. N. Y. Acad. Sci.* **2017**, *1389*, 164–185. [CrossRef] [PubMed]
6. Nakajima, R.T.; Cabral-de-Mello, D.C.; Valente, G.T.; Venere, P.C.; Martins, C. Evolutionary dynamics of rRNA gene clusters in cichlid fish. *BMC Evol. Biol.* **2012**, *12*, 198. [CrossRef] [PubMed]
7. Symonová, R.; Majtánová, Z.; Arias-Rodriguez, L.; Mořkovský, L.; Kořínková, T.; Cavin, L.; Pokorná, M.J.; Doležálková, M.; Flajšhans, M.; Normandeau, E.; et al. Genome Compositional Organization in Gars Shows More Similarities to Mammals than to Other Ray-Finned Fish: Cytogenomics of Gars. *J. Exp. Zool. B Mol. Dev. Evol.* **2017**, *328*, 607–619. [CrossRef] [PubMed]
8. Symonová, R.; Ocalewicz, K.; Kirtiklis, L.; Delmastro, G.B.; Pelikánová, Š.; Garcia, S.; Kovařík, A. Higher-order organisation of extremely amplified, potentially functional and massively methylated 5S rDNA in European pikes (*Esox* sp.). *BMC Genom.* **2017**, *18*, 391. [CrossRef] [PubMed]
9. Cioffi, M.B.; Bertollo, L.A.C. Chromosomal distribution and evolution of repetitive DNAs in fish. *Genome Dyn.* **2012**, *7*, 197–221. [CrossRef] [PubMed]
10. Betancur-R, R.; Wiley, E.O.; Arratia, G.; Acero, A.; Bailly, N.; Miya, M.; Lecointre, G.; Ortí, G. Phylogenetic classification of bony fishes. *BMC Evol. Biol.* **2017**, *17*, 162. [CrossRef] [PubMed]
11. Saha, N.R.; Smith, J.; Amemiya, C.T. Evolution of adaptive immune recognition in jawless vertebrates. *Semin. Immunol.* **2010**, *22*, 25–33. [CrossRef] [PubMed]
12. Gregory, T.R. Animal Genome Size Database. Available online: http://genomesize.com (accessed on 9 November 2017).
13. Caputo Barucchi, V.; Giovannotti, M.; Nisi Cerioni, P.; Splendiani, A. Genome Duplication in Early Vertebrates: Insights from Agnathan Cytogenetics. *Cytogenet. Genome Res.* **2013**, *141*, 80–89. [CrossRef] [PubMed]
14. Smith, J.J.; Saha, N.R.; Amemiya, C.T. Genome biology of the cyclostomes and insights into the evolutionary biology of vertebrate genomes. *Integr. Comp. Biol.* **2010**, *50*, 130–137. [CrossRef] [PubMed]
15. Stingo, V.; Rocco, L. Selachian cytogenetics: A review. *Genetica* **2001**, *111*, 329–347. [CrossRef] [PubMed]
16. Rock, J.; Eldridge, M.; Champion, A.; Johnston, P.; Joss, J. Karyotype and nuclear DNA content of the Australian lungfish, *Neoceratodus forsteri* (Ceratodidae: Dipnoi). *Cytogenet. Cell Genet.* **1996**, *73*, 187–189. [CrossRef] [PubMed]
17. Koch, J.; Lüdemann, J.; Spies, R.; Last, M.; Amemiya, C.T.; Burmester, T. Unusual Diversity of Myoglobin Genes in the Lungfish. *Mol. Biol. Evol.* **2016**, *33*, 3033–3041. [CrossRef] [PubMed]
18. Bogart, J.P.; Balon, E.K.; Bruton, M.N. The chromosomes of the living coelacanth and their remarkable similarity to those of one of the most ancient frogs. *J. Hered.* **1994**, *85*, 322–325. [CrossRef] [PubMed]
19. Andreyushkova, D.A.; Makunin, A.I.; Beklemisheva, V.R.; Romanenko, S.A.; Druzhkova, A.S.; Biltueva, L.B.; Serdyukova, N.A.; Graphodatsky, A.S.; Trifonov, V.A. Next Generation Sequencing of Chromosome-Specific Libraries Sheds Light on Genome Evolution in Paleotetraploid Sterlet (*Acipenser ruthenus*). *Genes* **2017**, *8*, 318. [CrossRef] [PubMed]
20. Romanenko, S.A.; Biltueva, L.S.; Serdyukova, N.A.; Kulemzina, A.I.; Beklemisheva, V.R.; Gladkikh, O.L.; Lemskaya, N.A.; Interesova, E.A.; Korentovich, M.A.; Vorobieva, N.V.; et al. Segmental paleotetraploidy revealed in sterlet (*Acipenser ruthenus*) genome by chromosome painting. *Mol. Cytogenet.* **2015**, *8*, 90. [CrossRef] [PubMed]

21. Helfman, G.S.; Collette, B.B.; Facey, D.E.; Bowen, B.W. *The Diversity of Fishes: Biology, Evolution, and Ecology*, 2nd ed.; Wiley-Blackwell: Oxford, UK, 2009; ISBN 978-1-4051-2494-2.
22. Nelson, J.S.; Grande, T.; Wilson, M.V.H. *Fishes of the World*, 5th ed.; John Wiley & Sons: Hoboken, NJ, USA, 2016; ISBN 978-1-118-34233-6.
23. Cavin, L. *Freshwater Fishes: 250 Million Years of Evolutionary History*; ISTE Press/Elsevier: London, UK, 2017.
24. Sallan, L.C. Major issues in the origins of ray-finned fish (*Actinopterygii*) biodiversity. *Biol. Rev. Camb. Philos. Soc.* **2014**, *89*, 950–971. [CrossRef] [PubMed]
25. Majtánová, Z.; Symonová, R.; Arias-Rodriguez, L.; Sallan, L.; Ráb, P. "Holostei versus Halecostomi" Problem: Insight from Cytogenetics of Ancient Nonteleost Actinopterygian Fish, Bowfin *Amia calva*. *J. Exp. Zool. B Mol. Dev. Evol.* **2017**, *328*, 620–628. [CrossRef] [PubMed]
26. Vega, C.G.; Wiens, J.J. Why are there so few fish in the sea? *Proc. R. Soc. B Biol. Sci.* **2012**, *279*, 2323–2329. [CrossRef] [PubMed]
27. Comber, S.C.L.; Smith, C. Polyploidy in fishes: Patterns and processes: Polyploidy in fishes. *Biol. J. Linn. Soc.* **2004**, *82*, 431–442. [CrossRef]
28. Smith, E.M.; Gregory, T.R. Patterns of genome size diversity in the ray-finned fishes. *Hydrobiologia* **2009**, *625*, 1–25. [CrossRef]
29. Mank, J.E.; Avise, J.C. Phylogenetic conservation of chromosome numbers in Actinopterygiian fishes. *Genetica* **2006**, *127*, 321–327. [CrossRef] [PubMed]
30. Francis, W.R.; Wörheide, G. Similar Ratios of Introns to Intergenic Sequence across Animal Genomes. *Genome Biol. Evol.* **2017**, *9*, 1582–1598. [CrossRef] [PubMed]
31. Sarropoulou, E.; Fernandes, J.M.O. Comparative genomics in teleost species: Knowledge transfer by linking the genomes of model and non-model fish species. *Comp. Biochem. Physiol. Part D Genom. Proteom.* **2011**, *6*, 92–102. [CrossRef] [PubMed]
32. Howe, K.; Clark, M.D.; Torroja, C.F.; Torrance, J.; Berthelot, C.; Muffato, M.; Collins, J.E.; Humphray, S.; McLaren, K.; Matthews, L.; et al. The zebrafish reference genome sequence and its relationship to the human genome. *Nature* **2013**, *496*, 498–503. [CrossRef] [PubMed]
33. Santoriello, C.; Zon, L.I. Hooked! Modeling human disease in zebrafish. *J. Clin. Investig.* **2012**, *122*, 2337–2343. [CrossRef] [PubMed]
34. Gibbons, J.G.; Branco, A.T.; Godinho, S.A.; Yu, S.; Lemos, B. Concerted copy number variation balances ribosomal DNA dosage in human and mouse genomes. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 2485–2490. [CrossRef] [PubMed]
35. Symonová, R.; Majtánová, Z.; Sember, A.; Staaks, G.B.; Bohlen, J.; Freyhof, J.; Rábová, M.; Ráb, P. Genome differentiation in a species pair of coregonine fishes: An extremely rapid speciation driven by stress-activated retrotransposons mediating extensive ribosomal DNA multiplications. *BMC Evol. Biol.* **2013**, *13*, 42. [CrossRef] [PubMed]
36. Xu, B.; Li, H.; Perry, J.M.; Singh, V.P.; Unruh, J.; Yu, Z.; Zakari, M.; McDowell, W.; Li, L.; Gerton, J.L. Ribosomal DNA copy number loss and sequence variation in cancer. *PLOS Genet.* **2017**, *13*, e1006771. [CrossRef] [PubMed]
37. Cioffi, M.B.; Martins, C.; Bertollo, L.A.C. Chromosome spreading of associated transposable elements and ribosomal DNA in the fish *Erythrinus erythrinus*. Implications for genome change and karyoevolution in fish. *BMC Evol. Biol.* **2010**, *10*, 271. [CrossRef] [PubMed]
38. Animal rDNA Database. Available online: www.animalrdnadatabase.com (accessed on 9 November 2017).
39. Sochorová, J.; Garcia, S.; Gálvez, F.; Symonová, R.; Kovařík, A. Evolutionary trends in animal ribosomal DNA loci: Introduction to a new online database. *Chromosoma* **2017**. [CrossRef] [PubMed]
40. Reed, K.M.; Phillips, R.B. Structure and organization of the rDNA intergenic spacer in lake trout (*Salvelinus namaycush*). *Chromosome Res. Int. J. Mol. Supramol. Evol. Asp. Chromosome Biol.* **2000**, *8*, 5–16. [CrossRef]
41. Reed, K.M.; Hackett, J.D.; Phillips, R.B. Comparative analysis of intra-individual and inter-species DNA sequence variation in salmonid ribosomal DNA cistrons. *Gene* **2000**, *249*, 115–125. [CrossRef]
42. Mazzuchelli, J.; Kocher, T.D.; Yang, F.; Martins, C. Integrating cytogenetics and genomics in comparative evolutionary studies of cichlid fish. *BMC Genom.* **2012**, *13*, 463. [CrossRef] [PubMed]
43. Chitramuthu, B. Modeling Human Disease and Development in Zebrafish. *Hum. Genet. Embryol.* **2013**, *3*, e108. [CrossRef]

44. Liu, S.; Leach, S.D. Zebrafish Models for Cancer. *Annu. Rev. Pathol. Mech. Dis.* **2011**, *6*, 71–93. [CrossRef] [PubMed]
45. Schartl, M. Beyond the zebrafish: Diverse fish species for modeling human disease. *Dis. Model. Mech.* **2014**, *7*, 181–192. [CrossRef] [PubMed]
46. Kari, G.; Rodeck, U.; Dicker, A.P. Zebrafish: An emerging model system for human disease and drug discovery. *Clin. Pharmacol. Ther.* **2007**, *82*, 70–80. [CrossRef] [PubMed]
47. Amores, A.; Force, A.; Yan, Y.L.; Joly, L.; Amemiya, C.; Fritz, A.; Ho, R.K.; Langeland, J.; Prince, V.; Wang, Y.L.; et al. Zebrafish *HOX* clusters and vertebrate genome evolution. *Science* **1998**, *282*, 1711–1714. [CrossRef] [PubMed]
48. Braasch, I.; Postlethwait, J.H. Polyploidy in Fish and the Teleost Genome Duplication. In *Polyploidy and Genome Evolution*; Soltis, P.S., Soltis, D.E., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; pp. 341–383. ISBN 978-3-642-31441-4.
49. Meyer, A.; Schartl, M. Gene and genome duplications in vertebrates: The one-to-four (-to-eight in fish) rule and the evolution of novel gene functions. *Curr. Opin. Cell Biol.* **1999**, *11*, 699–704. [CrossRef]
50. Postlethwait, J.H.; Woods, I.G.; Ngo-Hazelett, P.; Yan, Y.L.; Kelly, P.D.; Chu, F.; Huang, H.; Hill-Force, A.; Talbot, W.S. Zebrafish comparative genomics and the origins of vertebrate chromosomes. *Genome Res.* **2000**, *10*, 1890–1902. [CrossRef] [PubMed]
51. Ceol, C.J.; Houvras, Y.; Jane-Valbuena, J.; Bilodeau, S.; Orlando, D.A.; Battisti, V.; Fritsch, L.; Lin, W.M.; Hollmann, T.J.; Ferré, F.; et al. The histone methyltransferase *SETDB1* is recurrently amplified in melanoma and accelerates its onset. *Nature* **2011**, *471*, 513–517. [CrossRef] [PubMed]
52. Golzio, C.; Willer, J.; Talkowski, M.E.; Oh, E.C.; Taniguchi, Y.; Jacquemont, S.; Reymond, A.; Sun, M.; Sawa, A.; Gusella, J.F.; et al. *KCTD13* is a major driver of mirrored neuroanatomical phenotypes of the 16p11.2 copy number variant. *Nature* **2012**, *485*, 363–367. [CrossRef] [PubMed]
53. Panizzi, J.R.; Becker-Heck, A.; Castleman, V.H.; Al-Mutairi, D.A.; Liu, Y.; Loges, N.T.; Pathak, N.; Austin-Tse, C.; Sheridan, E.; Schmidts, M.; et al. *CCDC103* mutations cause primary ciliary dyskinesia by disrupting assembly of ciliary dynein arms. *Nat. Genet.* **2012**, *44*, 714–719. [CrossRef] [PubMed]
54. Roscioli, T.; Kamsteeg, E.-J.; Buysse, K.; Maystadt, I.; van Reeuwijk, J.; van den Elzen, C.; van Beusekom, E.; Riemersma, M.; Pfundt, R.; Vissers, L.E.L.M.; et al. Mutations in *ISPD* cause Walker-Warburg syndrome and defective glycosylation of α-dystroglycan. *Nat. Genet.* **2012**, *44*, 581–585. [CrossRef] [PubMed]
55. Patton, E.E.; Widlund, H.R.; Kutok, J.L.; Kopani, K.R.; Amatruda, J.F.; Murphey, R.D.; Berghmans, S.; Mayhall, E.A.; Traver, D.; Fletcher, C.D.M.; et al. *BRAF* mutations are sufficient to promote nevi formation and cooperate with p53 in the genesis of melanoma. *Curr. Biol.* **2005**, *15*, 249–254. [CrossRef] [PubMed]
56. Wittbrodt, J.; Shima, A.; Schartl, M. Medaka—A model organism from the far East. *Nat. Rev. Genet.* **2002**, *3*, 53–64. [CrossRef] [PubMed]
57. Albertson, R.C.; Cresko, W.; Detrich, H.W.; Postlethwait, J.H. Evolutionary mutant models for human disease. *Trends Genet.* **2009**, *25*, 74–81. [CrossRef] [PubMed]
58. McGaugh, S.E.; Gross, J.B.; Aken, B.; Blin, M.; Borowsky, R.; Chalopin, D.; Hinaux, H.; Jeffery, W.R.; Keene, A.; Ma, L.; et al. The cavefish genome reveals candidate genes for eye loss. *Nat. Commun.* **2014**, *5*, 5307. [CrossRef] [PubMed]
59. Protas, M.E.; Hersey, C.; Kochanek, D.; Zhou, Y.; Wilkens, H.; Jeffery, W.R.; Zon, L.I.; Borowsky, R.; Tabin, C.J. Genetic analysis of cavefish reveals molecular convergence in the evolution of albinism. *Nat. Genet.* **2006**, *38*, 107–111. [CrossRef] [PubMed]
60. Terzibasi, E.; Valenzano, D.R.; Cellerino, A. The short-lived fish *Nothobranchius furzeri* as a new model system for aging studies. *Exp. Gerontol.* **2007**, *42*, 81–89. [CrossRef] [PubMed]
61. Hárosi, F.I.; von Herbing, I.H.; Van Keuren, J.R. Sickling of anoxic red blood cells in fish. *Biol. Bull.* **1998**, *195*, 5–11. [CrossRef] [PubMed]
62. Meierjohann, S.; Schartl, M. From Mendelian to molecular genetics: The *Xiphophorus* melanoma model. *Trends Genet.* **2006**, *22*, 654–661. [CrossRef] [PubMed]
63. Schartl, M.; Walter, R.B.; Shen, Y.; Garcia, T.; Catchen, J.; Amores, A.; Braasch, I.; Chalopin, D.; Volff, J.-N.; Lesch, K.-P.; et al. The genome of the platyfish, *Xiphophorus maculatus*, provides insights into evolutionary adaptation and several complex traits. *Nat. Genet.* **2013**, *45*, 567–572. [CrossRef] [PubMed]

64. Schmale, M.C.; Hensley, G.T.; Udey, L.R. Neurofibromatosis in the bicolor damselfish (*Pomacentrus partitus*) as a model of von Recklinghausen neurofibromatosis. *Ann. N. Y. Acad. Sci.* **1986**, *486*, 386–402. [CrossRef] [PubMed]
65. Williams, D.E. The rainbow trout liver cancer model: Response to environmental chemicals and studies on promotion and chemoprevention. *Comp. Biochem. Physiol. Part C Toxicol. Pharmacol.* **2012**, *155*, 121–127. [CrossRef] [PubMed]
66. Burnett, K.G.; Bain, L.J.; Baldwin, W.S.; Callard, G.V.; Cohen, S.; Di Giulio, R.T.; Evans, D.H.; Gómez-Chiarri, M.; Hahn, M.E.; Hoover, C.A.; et al. *Fundulus* as the premier teleost model in environmental biology: Opportunities for new insights using genomics. *Comp. Biochem. Physiol. Part D Genom. Proteom.* **2007**, *2*, 257–286. [CrossRef] [PubMed]
67. Lampert, K.; Schartl, M. The origin and evolution of a unisexual hybrid: *Poecilia formosa*. *Philos. Trans. R. Soc. B Biol. Sci.* **2008**, *363*, 2901–2909. [CrossRef] [PubMed]
68. Schartl, M.; Nanda, I.; Schlupp, I.; Wilde, B.; Epplen, J.T.; Schmid, M.; Parzefall, J. Incorporation of subgenomic amounts of DNA as compensation for mutational load in a gynogenetic fish. *Nature* **1995**, *373*, 68–71. [CrossRef]
69. Schartl, A.; Hornung, U.; Nanda, I.; Wacker, R.; Müller-Hermelink, H.K.; Schlupp, I.; Parzefall, J.; Schmid, M.; Schartl, M. Susceptibility to the development of pigment cell tumors in a clone of the Amazon molly, *Poecilia formosa*, introduced through a microchromosome. *Cancer Res.* **1997**, *57*, 2993–3000. [PubMed]
70. Tobler, M.; Schlupp, I. Parasites in sexual and asexual mollies (*Poecilia*, Poeciliidae, Teleostei): A case for the Red Queen? *Biol. Lett.* **2005**, *1*, 166–168. [CrossRef] [PubMed]
71. Woodhead, A.D.; Setlow, R.B.; Pond, V. The Amazon molly, *Poecilia formosa*, as a test animal in carcinogenicity studies: Chronic exposures to physical agents. *Natl. Cancer Inst. Monogr.* **1984**, *65*, 45–52. [PubMed]
72. Scahill, C.M.; Digby, Z.; Sealy, I.M.; Wojciechowska, S.; White, R.J.; Collins, J.E.; Stemple, D.L.; Bartke, T.; Mathers, M.E.; Patton, E.E.; et al. Loss of the chromatin modifier *Kdm2aa* causes BrafV600E-independent spontaneous melanoma in zebrafish. *PLoS Genet.* **2017**, *13*, e1006959. [CrossRef] [PubMed]
73. Cossins, A.R.; Crawford, D.L. Fish as models for environmental genomics. *Nat. Rev. Genet.* **2005**, *6*, 324–333. [CrossRef] [PubMed]
74. Aquaculture Genomics, Genetics and Breeding Workshop; Abdelrahman, H.; ElHady, M.; Alcivar-Warren, A.; Allen, S.; Al-Tobasei, R.; Bao, L.; Beck, B.; Blackburn, H.; Bosworth, B.; et al. Aquaculture genomics, genetics and breeding in the United States: Current status, challenges and priorities for future research. *BMC Genom.* **2017**, *18*, 191. [CrossRef]
75. Kobayashi, T. Genome Instability of Repetitive Sequence: Lesson from the Ribosomal RNA Gene Repeat. In *DNA Replication, Recombination and Repair*; Hanaoka, F., Sugasawa, K., Eds.; Springer: Tokyo, Japan, 2016; pp. 235–247. ISBN 978-4-431-55871-2.
76. Wang, M.; Lemos, B. Ribosomal DNA copy number amplification and loss in human cancers is linked to tumor genetic context, nucleolus activity and proliferation. *PLoS Genet.* **2017**, *13*, e1006994. [CrossRef] [PubMed]
77. Boisvert, F.-M.; van Koningsbruggen, S.; Navascués, J.; Lamond, A.I. The multifunctional nucleolus. *Nat. Rev. Mol. Cell Biol.* **2007**, *8*, 574–585. [CrossRef] [PubMed]
78. Guetg, C.; Santoro, R. Formation of nuclear heterochromatin: The nucleolar point of view. *Epigenetics* **2012**, *7*, 811–814. [CrossRef] [PubMed]
79. Makunin, A.I.; Dementyeva, P.V.; Graphodatsky, A.S.; Volobouev, V.T.; Kukekova, A.V.; Trifonov, V.A. Genes on B chromosomes of vertebrates. *Mol. Cytogenet.* **2014**, *7*, 99. [CrossRef] [PubMed]
80. Terencio, M.L.; Schneider, C.H.; Gross, M.C.; do Carmo, E.J.; Nogaroto, V.; de Almeida, M.C.; Artoni, R.F.; Vicari, M.R.; Feldberg, E. Repetitive sequences: The hidden diversity of heterochromatin in prochilodontid fish. *Comp. Cytogenet.* **2015**, *9*, 465–481. [CrossRef] [PubMed]
81. Tsekrekou, M.; Stratigi, K.; Chatzinikolaou, G. The Nucleolus: In Genome Maintenance and Repair. *Int. J. Mol. Sci.* **2017**, *18*, 1411. [CrossRef] [PubMed]
82. Braasch, I.; Gehrke, A.R.; Smith, J.J.; Kawasaki, K.; Manousaki, T.; Pasquier, J.; Amores, A.; Desvignes, T.; Batzel, P.; Catchen, J.; et al. The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons. *Nat. Genet.* **2016**, *48*, 427–437. [CrossRef] [PubMed]

83. Dion- Côté, A.-M.; Symonová, R.; Ráb, P.; Bernatchez, L. Reproductive isolation in a nascent species pair is associated with aneuploidy in hybrid offspring. *Proc. R. Soc. B Biol. Sci.* **2015**, *282*, 20142862. [CrossRef] [PubMed]
84. Clark, M.S. Genomics and Mapping of Teleostei (Bony Fish). *Comp. Funct. Genom.* **2003**, *4*, 182–193. [CrossRef] [PubMed]
85. Carvalho, G.R.; Hauser, L.; Martinsohn, J.; Naish, K. Fish, genes and genomes: Contributions to ecology, evolution and management. *J. Fish Biol.* **2016**, *89*, 2471–2478. [CrossRef] [PubMed]
86. Arai, R. *Fish Karyotypes*; Springer: Tokyo, Japan, 2011; ISBN 978-4-431-53876-9.
87. Majtánová, Z.; Choleva, L.; Symonová, R.; Ráb, P.; Kotusz, J.; Pekárik, L.; Janko, K. Asexual Reproduction Does Not Apparently Increase the Rate of Chromosomal Evolution: Karyotype Stability in Diploid and Triploid Clonal Hybrid Fish (Cobitis, Cypriniformes, Teleostei). *PLoS ONE* **2016**, *11*, e0146872. [CrossRef] [PubMed]
88. Rampin, M.; Bi, K.; Bogart, J.P.; Collares-Pereira, M.J. Identifying parental chromosomes and genomic rearrangements in animal hybrid complexes of species with small genome size using Genomic In Situ Hybridization (GISH). *Comp. Cytogenet.* **2012**, *6*, 287–300. [CrossRef] [PubMed]
89. Havelka, M.; Bytyutskyy, D.; Symonová, R.; Ráb, P.; Flajšhans, M. The second highest chromosome count among vertebrates is observed in cultured sturgeon and is associated with genome plasticity. *Genet. Sel. Evol.* **2016**, *48*, 12. [CrossRef] [PubMed]
90. Jaillon, O.; Aury, J.-M.; Brunet, F.; Petit, J.-L.; Stange-Thomann, N.; Mauceli, E.; Bouneau, L.; Fischer, C.; Ozouf-Costaz, C.; Bernot, A.; et al. Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature* **2004**, *431*, 946–957. [CrossRef] [PubMed]
91. Roest Crollius, H.; Weissenbach, J. Fish genomics and biology. *Genome Res.* **2005**, *15*, 1675–1682. [CrossRef] [PubMed]
92. Phillips, R.B.; Amores, A.; Morasch, M.R.; Wilson, C.; Postlethwait, J.H. Assignment of zebrafish genetic linkage groups to chromosomes. *Cytogenet. Genome Res.* **2006**, *114*, 155–162. [CrossRef] [PubMed]
93. Phillips, R.B.; Nichols, K.M.; DeKoning, J.J.; Morasch, M.R.; Keatley, K.A.; Rexroad, C.; Gahr, S.A.; Danzmann, R.G.; Drew, R.E.; Thorgaard, G.H. Assignment of rainbow trout linkage groups to specific chromosomes. *Genetics* **2006**, *174*, 1661–1670. [CrossRef] [PubMed]
94. Phillips, R.B.; Keatley, K.A.; Morasch, M.R.; Ventura, A.B.; Lubieniecki, K.P.; Koop, B.F.; Danzmann, R.G.; Davidson, W.S. Assignment of Atlantic salmon (*Salmo salar*) linkage groups to specific chromosomes: Conservation of large syntenic blocks corresponding to whole chromosome arms in rainbow trout (*Oncorhynchus mykiss*). *BMC Genet.* **2009**, *10*, 46. [CrossRef] [PubMed]
95. Guyon, R.; Rakotomanga, M.; Azzouzi, N.; Coutanceau, J.P.; Bonillo, C.; D'Cotta, H.; Pepey, E.; Soler, L.; Rodier-Goud, M.; D'Hont, A.; et al. A high-resolution map of the Nile Tilapia genome: A resource for studying cichlids and other percomorphs. *BMC Genom.* **2012**, *13*, 222. [CrossRef] [PubMed]
96. Dion-Côté, A.-M.; Symonová, R.; Lamaze, F.C.; Pelikánová, Š.; Ráb, P.; Bernatchez, L. Standing chromosomal variation in Lake Whitefish species pairs: The role of historical contingency and relevance for speciation. *Mol. Ecol.* **2017**, *26*, 178–192. [CrossRef] [PubMed]
97. Rondeau, E.B.; Minkley, D.R.; Leong, J.S.; Messmer, A.M.; Jantzen, J.R.; von Schalburg, K.R.; Lemon, C.; Bird, N.H.; Koop, B.F. The genome and linkage map of the northern pike (*Esox lucius*): Conserved synteny revealed between the salmonid sister group and the Neoteleostei. *PLoS ONE* **2014**, *9*, e102089. [CrossRef] [PubMed]
98. Sutherland, B.J.G.; Gosselin, T.; Normandeau, E.; Lamothe, M.; Isabel, N.; Audet, C.; Bernatchez, L. Salmonid Chromosome Evolution as Revealed by a Novel Method for Comparing RADseq Linkage Maps. *Genome Biol. Evol.* **2016**, *8*, 3600–3617. [CrossRef] [PubMed]
99. Symonová, R.; Sutherland, B.J.G.; Bernatchez, L. Residually tetrasomic sites in Coregonus clupeaformis. Under preparation.
100. Cozzi, P.; Milanesi, L.; Bernardi, G. Segmenting the Human Genome into Isochores. *Evol. Bioinform.* **2015**, *11*, 253–261. [CrossRef] [PubMed]
101. Bernardi, G. *Structural and Evolutionary Genomics: Natural Selection in Genome Evolution*; New Comprehensive Biochemistry; Elsevier: Amsterdam, The Netherlands, 2004; ISBN 978-0-444-51255-0.

102. Daniel-Silva, M.F.Z.; Almeida-Toledo, L.F. Chromosome evolution in fish: BrdU replication patterns demonstrate chromosome homeologies in two species of the genus Astyanax. *Cytogenet. Genome Res.* **2005**, *109*, 497–501. [CrossRef] [PubMed]
103. Zerbino, D.R.; Achuthan, P.; Akanni, W.; Amode, M.R.; Barrell, D.; Bhai, J.; Billis, K.; Cummins, C.; Gall, A.; Girón, C.G.; et al. Ensembl 2018. *Nucleic Acids Res.* **2018**, *46*, D754–D761. [CrossRef] [PubMed]
104. Cioffi, M.B.; Yano, C.F.; Sember, A.; Bertollo, L.A.C. Chromosomal Evolution in Lower Vertebrates: Sex Chromosomes in Neotropical Fishes. *Genes* **2017**, *8*, 258. [CrossRef] [PubMed]
105. Cioffi, M.B.; Franco, W.; Ferreira, R.; Carlos Bertollo, L.A. Chromosomes as Tools for Discovering Biodiversity—The Case of Erythrinidae Fish Family. In *Recent Trends in Cytogenetic Studies—Methodologies and Applications*; Tirunilai, P., Ed.; InTech: Rijeka, Croatia, 2012; ISBN 978-953-51-0178-9.
106. Glasauer, S.M.K.; Neuhauss, S.C.F. Whole-genome duplication in teleost fishes and its evolutionary consequences. *Mol. Genet. Genom.* **2014**, *289*, 1045–1060. [CrossRef] [PubMed]
107. Mable, B.; Alexandrou, M.; Taylor, M. Genome duplication in amphibians and fish: An extended synthesis. *J. Zool.* **2011**, *284*, 151–182. [CrossRef]
108. Symonová, R.; Havelka, M.; Amemiya, C.T.; Howell, W.M.; Kořínková, T.; Flajšhans, M.; Gela, D.; Ráb, P. Molecular cytogenetic differentiation of paralogs of *HOX* paralogs in duplicated and re-diploidized genome of the North American paddlefish (*Polyodon spathula*). *BMC Genet.* **2017**, *18*, 19. [CrossRef] [PubMed]
109. Nagpure, N.S.; Rashid, I.; Pathak, A.K.; Singh, M.; Singh, S.P.; Sarkar, U.K. FBIS: A regional DNA barcode archival and analysis system for Indian fishes. *Bioinformation* **2012**, *8*, 483–488. [CrossRef] [PubMed]
110. Nagpure, N.S.; Rashid, I.; Pathak, A.K.; Singh, M.; Pati, R.; Singh, S.P.; Sarkar, U.K. FMiR: A Curated Resource of Mitochondrial DNA Information for Fish. *PLoS ONE* **2015**, *10*, e0136711. [CrossRef] [PubMed]
111. Nagpure, N.S.; Rashid, I.; Pati, R.; Pathak, A.K.; Singh, M.; Singh, S.P.; Sarkar, U.K. FishMicrosat: A microsatellite database of commercially important fishes and shellfishes of the Indian subcontinent. *BMC Genom.* **2013**, *14*, 630. [CrossRef] [PubMed]
112. Avvaru, A.K.; Saxena, S.; Sowpati, D.T.; Mishra, R.K. MSDB: A Comprehensive Database of Simple Sequence Repeats. *Genome Biol. Evol.* **2017**, *9*, 1797–1802. [CrossRef] [PubMed]
113. Nagpure, N.S.; Pathak, A.K.; Pati, R.; Rashid, I.; Sharma, J.; Singh, S.P.; Singh, M.; Sarkar, U.K.; Kushwaha, B.; Kumar, R.; et al. Fish Karyome version 2.1: A chromosome database of fishes and other aquatic organisms. *Database* **2016**, *2016*, baw012. [CrossRef] [PubMed]
114. Froese, R.; Pauly, D. FishBase. World Wide Web Electronic Publication. 2017. Available online: www.fishbase.org (accessed on 9 November 2017).
115. Bhartiya, D.; Maini, J.; Sharma, M.; Joshi, P.; Laddha, S.V.; Jalali, S.; Patowary, A.; Purkanti, R.; Lalwani, M.; Singh, A.R.; et al. FishMap Zv8 update—A genomic regulatory map of zebrafish. *Zebrafish* **2010**, *7*, 179–180. [CrossRef] [PubMed]
116. Amores, A.; Postlethwait, J.H. Banded Chromosomes and the Zebrafish Karyotype. In *Methods in Cell Biology*; Elsevier: Amsterdam, The Netherlands, 1998; Volume 60, pp. 323–338. ISBN 978-0-12-544162-9.
117. Gornung, E.; De Innocentiis, S.; Annesi, F.; Sola, L. Zebrafish 5S rRNA genes map to the long arms of chromosome 3. *Chromosome Res. Int. J. Mol. Supramol. Evol. Asp. Chromosome Biol.* **2000**, *8*, 362. [CrossRef]
118. Phillips, R.B.; Reed, K.M. Localization of repetitive DNAs to zebrafish (*Danio rerio*) chromosomes by fluorescence in situ hybridization (FISH). *Chromosome Res. Int. J. Mol. Supramol. Evol. Asp. Chromosome Biol.* **2000**, *8*, 27–35. [CrossRef]
119. Sola, L.; Gornung, E. Classical and molecular cytogenetics of the zebrafish, *Danio rerio* (Cyprinidae, Cypriniformes): An overview. *Genetica* **2001**, *111*, 397–412. [CrossRef] [PubMed]
120. National Center for Biotechnology Information (NCBI). Genome Browser. Available online: www.ncbi.nlm.nih.gov/genome/browse (accessed on 9 November 2017).
121. Personal Webpage of Radka Symonová. Available online: http://lide.uhk.cz/Symonra1 (accessed on 9 November 2017).
122. Kitts, P.A.; Church, D.M.; Thibaud-Nissen, F.; Choi, J.; Hem, V.; Sapojnikov, V.; Smith, R.G.; Tatusova, T.; Xiang, C.; Zherikov, A.; et al. Assembly: A resource for assembled genomes at NCBI. *Nucleic Acids Res.* **2016**, *44*, D73–D80. [CrossRef] [PubMed]
123. European Nucleotide Archive (ENA). Genome Assembly Database. Available online: www.ebi.ac.uk/ena/browse/genome-assembly-database (accessed on 9 November 2017).

124. Tørresen, O.K.; Star, B.; Jentoft, S.; Reinar, W.B.; Grove, H.; Miller, J.R.; Walenz, B.P.; Knight, J.; Ekholm, J.M.; Peluso, P.; et al. An improved genome assembly uncovers prolific tandem repeats in *Atlantic cod*. *BMC Genom.* **2017**, *18*, 95. [CrossRef] [PubMed]
125. Malmstrøm, M.; Matschiner, M.; Tørresen, O.K.; Jakobsen, K.S.; Jentoft, S. Whole genome sequencing data and de novo draft assemblies for 66 teleost species. *Sci. Data* **2017**, *4*, 160132. [CrossRef] [PubMed]
126. Koepfli, K.-P.; Paten, B. Genome 10K Community of Scientists; O'Brien, S.J. The Genome 10K Project: A way forward. *Annu. Rev. Anim. Biosci.* **2015**, *3*, 57–111. [CrossRef] [PubMed]
127. Pasquier, J.; Cabau, C.; Nguyen, T.; Jouanno, E.; Severac, D.; Braasch, I.; Journot, L.; Pontarotti, P.; Klopp, C.; Postlethwait, J.H.; et al. Gene evolution and gene expression after whole genome duplication in fish: The PhyloFish database. *BMC Genom.* **2016**, *17*, 368. [CrossRef] [PubMed]
128. China National Genebank. *FishT1K*. Available online: https://db.cngb.org/fisht1k/status (accessed on 9 November 2017).
129. Garamszegi, L. Z. *Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology: Concepts and Practice*; Springer: Berlin, Germany, 2014; ISBN 978-3-662-43550-2.
130. Hardie, D.C.; Hebert, P.D.N. The nucleotypic effects of cellular DNA content in cartilaginous and ray-finned fishes. *Genome* **2003**, *46*, 683–706. [CrossRef] [PubMed]
131. Lefébure, T.; Morvan, C.; Malard, F.; François, C.; Konecny-Dupré, L.; Guéguen, L.; Weiss-Gayet, M.; Seguin-Orlando, A.; Ermini, L.; Sarkissian, C.D.; et al. Less effective selection leads to larger genomes. *Genome Res.* **2017**, *27*, 1016–1028. [CrossRef] [PubMed]
132. Gregory, T.R.; Witt, J.D.S. Population size and genome size in fishes: A closer look. *Genome* **2008**, *51*, 309–313. [CrossRef] [PubMed]
133. Elliott, T.A.; Gregory, T.R. Do larger genomes contain more diverse transposable elements? *BMC Evol. Biol.* **2015**, *15*, 69. [CrossRef] [PubMed]
134. Tarallo, A.; Angelini, C.; Sanges, R.; Yagi, M.; Agnisola, C.; D'Onofrio, G. On the genome base composition of teleosts: The effect of environment and lifestyle. *BMC Genom.* **2016**, *17*, 173. [CrossRef] [PubMed]
135. Hardie, D.C.; Hebert, P.D. Genome-size evolution in fishes. *Can. J. Fish. Aquat. Sci.* **2004**, *61*, 1636–1646. [CrossRef]
136. Yi, S.; Streelman, J.T. Genome size is negatively correlated with effective population size in ray-finned fish. *Trends Genet.* **2005**, *21*, 643–646. [CrossRef] [PubMed]
137. Vervoort, A. Tetraploidy in *Protopterus* (Dipnoi). *Experientia* **1980**, *36*, 294–296. [CrossRef]
138. Paim, F.G.; da Hora Almeida, L.A.; de Mell Affonso, P.R.A.; Sobrinho-Scudeler, P.E.; Oliveira, C.; Diniz, D. Chromosomal stasis in distinct families of marine Percomorpharia from South Atlantic. *Comp. Cytogenet.* **2017**, *11*, 299–307. [CrossRef] [PubMed]
139. Camacho, J.P.M.; Sharbel, T.F.; Beukeboom, L.W. B-chromosome evolution. *Philos. Trans. R. Soc. B Biol. Sci.* **2000**, *355*, 163–178. [CrossRef] [PubMed]
140. Valente, G.T.; Nakajima, R.T.; Fantinatti, B.E.A.; Marques, D.F.; Almeida, R.O.; Simões, R.P.; Martins, C. B chromosomes: From cytogenetics to systems biology. *Chromosoma* **2017**, *126*, 73–81. [CrossRef] [PubMed]
141. Anderson, J.L.; Rodríguez Marí, A.; Braasch, I.; Amores, A.; Hohenlohe, P.; Batzel, P.; Postlethwait, J.H. Multiple sex-associated regions and a putative sex chromosome in zebrafish revealed by RAD mapping and population genomics. *PLoS ONE* **2012**, *7*, e40701. [CrossRef] [PubMed]
142. Bradley, K.M.; Breyer, J.P.; Melville, D.B.; Broman, K.W.; Knapik, E.W.; Smith, J.R. An SNP-Based Linkage Map for Zebrafish Reveals Sex Determination Loci. *G3* **2011**, *1*, 3–9. [CrossRef] [PubMed]
143. Nagabhushana, A.; Mishra, R.K. Finding clues to the riddle of sex determination in zebrafish. *J. Biosci.* **2016**, *41*, 145–155. [CrossRef] [PubMed]
144. Matsuda, M.; Nagahama, Y.; Shinomiya, A.; Sato, T.; Matsuda, C.; Kobayashi, T.; Morrey, C.E.; Shibata, N.; Asakawa, S.; Shimizu, N.; et al. DMY is a Y-specific DM-domain gene required for male development in the Medaka fish. *Nature* **2002**, *417*, 559–563. [CrossRef] [PubMed]
145. De Andrade Silva, D.M.Z.; Utsunomia, R.; Ruiz-Ruano, F.J.; Daniel, S.N.; Porto-Foresti, F.; Hashimoto, D.T.; Oliveira, C.; Camacho, J.P.M.; Foresti, F. High-throughput analysis unveils a highly shared satellite DNA library among three species of fish genus *Astyanax*. *Sci. Rep.* **2017**, *7*, 12726. [CrossRef] [PubMed]
146. Ruiz-Estévez, M.; López-León, M.D.; Cabrero, J.; Camacho, J.P.M. B-chromosome ribosomal DNA is functional in the grasshopper *Eyprepocnemis plorans*. *PLoS ONE* **2012**, *7*, e36600. [CrossRef] [PubMed]
147. Utsunomia, R.; de Andrade Silva, D.M.Z.; Ruiz-Ruano, F.J.; Araya-Jaime, C.; Pansonato-Alves, J.C.; Scacchetti, P.C.; Hashimoto, D.T.; Oliveira, C.; Trifonov, V.A.; Porto-Foresti, F.; et al. Uncovering the

Ancestry of B Chromosomes in *Moenkhausia sanctaefilomenae* (Teleostei, Characidae). *PLoS ONE* **2016**, *11*, e0150573. [CrossRef] [PubMed]
148. Lamatsch, D.K.; Trifonov, V.; Schories, S.; Epplen, J. T.; Schmid, M.; Schartl, M. Isolation of a cancer-associated microchromosome in the sperm-dependent parthenogen *Poecilia formosa*. *Cytogenet. Genome Res.* **2011**, *135*, 135–142. [CrossRef] [PubMed]
149. Schmid, M.; Ziegler, C.G.; Steinlein, C.; Nanda, I.; Schartl, M. Cytogenetics of the bleak (*Alburnus alburnus*), with special emphasis on the B chromosomes. *Chromosome Res. Int. J. Mol. Supramol. Evol. Asp. Chromosome Biol.* **2006**, *14*, 231–242. [CrossRef] [PubMed]
150. Takagui, F.H.; Dias, A.L.; Birindelli, J.L.O.; Swarça, A.C.; da Rosa, R.; Lui, R.L.; Fenocchio, A.S.; Giuliano-Caetano, L. First report of B chromosomes in three neotropical thorny catfishes (*Siluriformes*, *Doradidae*). *Comp. Cytogenet.* **2017**, *11*, 55–64. [CrossRef] [PubMed]
151. Jones, R.N.; Diez, M. The B chromosome database. *Cytog. Gen. Res.* **2004**, *106*, 149–150. [CrossRef] [PubMed]
152. D'Ambrosio, U.; Alonso-Lifante, M.P.; Barros, K.; Kovařík, A.; Mas de Xaxars, G.; Garcia, S. B-chrom: A database on B-chromosomes of plants, animals and fungi. *New Phytol.* **2017**, *216*, 635–642. [CrossRef] [PubMed]
153. Grummt, I. The nucleolus—Guardian of cellular homeostasis and genome integrity. *Chromosoma* **2013**, *122*, 487–497. [CrossRef] [PubMed]
154. Ide, S.; Miyazaki, T.; Maki, H.; Kobayashi, T. Abundance of ribosomal RNA gene copies maintains genome integrity. *Science* **2010**, *327*, 693–696. [CrossRef] [PubMed]
155. Charlesworth, B.; Sniegowski, P.; Stephan, W. The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* **1994**, *371*, 215–220. [CrossRef] [PubMed]
156. Federico, C.; Scavo, C.; Cantarella, C.D.; Motta, S.; Saccone, S.; Bernardi, G. Gene-rich and gene-poor chromosomal regions have different locations in the interphase nuclei of cold-blooded vertebrates. *Chromosoma* **2006**, *115*, 123–128. [CrossRef] [PubMed]
157. Kirubakaran, T.G.; Grove, H.; Kent, M.P.; Sandve, S.R.; Baranski, M.; Nome, T.; De Rosa, M.C.; Righino, B.; Johansen, T.; Otterå, H.; et al. Two adjacent inversions maintain genomic differentiation between migratory and stationary ecotypes of Atlantic cod. *Mol. Ecol.* **2016**, *25*, 2130–2143. [CrossRef] [PubMed]
158. Fujiwara, A.; Abe, S.; Yamaha, E.; Yamazaki, F.; Yoshida, M.C. Chromosomal localization and heterochromatin association of ribosomal RNA gene loci and silver-stained nucleolar organizer regions in salmonid fishes. *Chromosome Res. Int. J. Mol. Supramol. Evol. Asp. Chromosome Biol.* **1998**, *6*, 463–471. [CrossRef]
159. Costa, G.W.W.F.; Cioffi, M.B.; Bertollo, L.A.C.; Molina, W.F. Unusual dispersion of histone repeats on the whole chromosomal complement and their colocalization with ribosomal genes in *Rachycentron canadum* (Rachycentridae, Perciformes). *Cytogenet. Genome Res.* **2014**, *144*, 62–67. [CrossRef] [PubMed]
160. Costa, G.W.W.F.; Cioffi, M.B.; Bertollo, L.A.C.; Molina, W.F. The Evolutionary Dynamics of Ribosomal Genes, Histone H3 and Transposable Rex Elements in the Genome of Atlantic Snappers. *J. Hered.* **2016**, *107*, 173–180. [CrossRef] [PubMed]
161. Mehner, T.; Pohlmann, K.; Elkin, C.; Monaghan, M.T.; Nitz, B.; Freyhof, J. Genetic population structure of sympatric and allopatric populations of Baltic ciscoes (*Coregonus albula* complex, Teleostei, Coregonidae). *BMC Evol. Biol.* **2010**, *10*, 85. [CrossRef] [PubMed]
162. Kobayashi, T. Ribosomal RNA gene repeats, their stability and cellular senescence. *Proc. Jpn. Acad. Ser. B Phys. Biol. Sci.* **2014**, *90*, 119–129. [CrossRef] [PubMed]
163. Glazer, A.M.; Killingbeck, E.E.; Mitros, T.; Rokhsar, D.S.; Miller, C.T. Genome Assembly Improvement and Mapping Convergently Evolved Skeletal Traits in Sticklebacks with Genotyping-by-Sequencing. *G3* **2015**, *5*, 1463–1472. [CrossRef] [PubMed]
164. Chen, X.; Zhong, L.; Bian, C.; Xu, P.; Qiu, Y.; You, X.; Zhang, S.; Huang, Y.; Li, J.; Wang, M.; et al. High-quality genome assembly of channel catfish, *Ictalurus punctatus*. *GigaScience* **2016**, *5*. [CrossRef] [PubMed]
165. Liu, Z.; Liu, S.; Yao, J.; Bao, L.; Zhang, J.; Li, Y.; Jiang, C.; Sun, L.; Wang, R.; Zhang, Y.; et al. The channel catfish genome sequence provides insights into the evolution of scale formation in teleosts. *Nat. Commun.* **2016**, *7*, 11757. [CrossRef] [PubMed]
166. Liu, H.; Chen, C.; Gao, Z.; Min, J.; Gu, Y.; Jian, J.; Jiang, X.; Cai, H.; Ebersberger, I.; Xu, M.; et al. The draft genome of blunt snout bream (*Megalobrama amblycephala*) reveals the development of intermuscular bone and adaptation to herbivorous diet. *GigaScience* **2017**, *6*, 1–13. [CrossRef] [PubMed]

167. Pan, H.; Yu, H.; Ravi, V.; Li, C.; Lee, A. P.; Lian, M.M.; Tay, B.-H.; Brenner, S.; Wang, J.; Yang, H.; et al. The genome of the largest bony fish, ocean sunfish (*Mola mola*), provides insights into its fast growth rate. *GigaScience* **2016**, *5*. [CrossRef] [PubMed]
168. Brawand, D.; Wagner, C.E.; Li, Y.I.; Malinsky, M.; Keller, I.; Fan, S.; Simakov, O.; Ng, A.Y.; Lim, Z.W.; Bezault, E.; et al. The genomic substrate for adaptive radiation in African cichlid fish. *Nature* **2014**, *513*, 375–381. [CrossRef] [PubMed]
169. Conte, M.A.; Gammerdinger, W.J.; Bartie, K.L.; Penman, D.J.; Kocher, T.D. A high quality assembly of the Nile Tilapia (*Oreochromis niloticus*) genome reveals the structure of two sex determination regions. *BMC Genom.* **2017**, *18*. [CrossRef] [PubMed]
170. Lien, S.; Koop, B.F.; Sandve, S.R.; Miller, J.R.; Kent, M.P.; Nome, T.; Hvidsten, T.R.; Leong, J.S.; Minkley, D.R.; Zimin, A.; et al. The Atlantic salmon genome provides insights into rediploidization. *Nature* **2016**, *533*, 200–205. [CrossRef] [PubMed]

RESEARCH ARTICLE

Genome Compositional Organization in Gars Shows More Similarities to Mammals than to Other Ray-Finned Fish

RADKA SYMONOVÁ[1,2,3*], ZUZANA MAJTÁNOVÁ[1,2], LENIN ARIAS-RODRIGUEZ[4], LIBOR MOŘKOVSKÝ[2], TEREZA KOŘÍNKOVÁ[1], LIONEL CAVIN[5], MARTINA JOHNSON POKORNÁ[1,6], MARIE DOLEŽÁLKOVÁ[1,2], MARTIN FLAJŠHANS[7], ERIC NORMANDEAU[8], PETR RÁB[1], AXEL MEYER[9], AND LOUIS BERNATCHEZ[8]

[1]Laboratory of Fish Genetics, Institute of Animal Physiology and Genetics, The Czech Academy of Sciences, Liběchov, Czech Republic
[2]Department of Zoology, Faculty of Science, Charles University, Prague 2, Czech Republic
[3]Research Institute for Limnology, University of Innsbruck, Mondsee, Austria
[4]División Académica de Ciencias Biológicas, Universidad Juárez Autónoma de Tabasco (UJAT), Villahermosa, Tabasco, México
[5]Muséum d'Histoire Naturelle, Geneva 6, Switzerland
[6]Department of Ecology, Faculty of Science, Charles University, Prague 2, Czech Republic
[7]Faculty of Fisheries and Protection of Waters, South Bohemian Research Centre of Aquaculture and Biodiversity of HydrocenosesUniversity of South Bohemia in České Budějovice, Vodňany, Czech Republic
[8]IBIS, Department of Biology, University Laval, Pavillon Charles-Eugène-Marchand, Avenue de la Médecine Quebec City, Canada
[9]Chair in Zoology and Evolutionary Biology, Department of Biology, University of Konstanz, Konstanz, Germany

Data archival location: https://github.com/libor-m/vertebrate-GC
Conflicts of interest: None
Grant Sponsor: This study was supported by projects P506/11/P596 and 14-02940S of the Czech Science Foundation and by the project between the Academy of Sciences of the Czech Republic and the National Council for Science and Technology of Mexico 167686/204264, the project 43-251468 of the Charles University Grant Agency and Research Support Grant SVV 260087/2014, supported by the Ministry of Education, Youth and Sports of the Czech Republic projects "CENAKVA" No. CZ.1.05/2.1.00/01.0024, "CENAKVA II" No. LO1205 under the NPU I program. LC was supported partly by the Swiss National Research Fund (200021-140827).
Additional Supporting Information may be found in the online version of this article.
*Correspondence to: Radka Symonová; Research Institute for Limnology, University of Innsbruck, Mondseestraße 9, A-5310 Mondsee, Austria.
E-mail: Radka.Symonova@uibk.ac.at
Received 17 May 2016; Revised 13 November 2016; Accepted 22 November 2016
DOI: 10.1002/jez.b.22719
Published online in Wiley Online Library (wileyonlinelibrary.com).

ABSTRACT

Genomic GC content can vary locally, and GC-rich regions are usually associated with increased DNA thermostability in thermophilic prokaryotes and warm-blooded eukaryotes. Among vertebrates, fish and amphibians appeared to possess a distinctly less heterogeneous AT/GC organization in their genomes, whereas cytogenetically detectable GC heterogeneity has so far only been documented in mammals and birds. The subject of our study is the gar, an ancient "living fossil" of a basal ray-finned fish lineage, known from the Cretaceous period. We carried out cytogenomic analysis in two gar genera (*Atractosteus* and *Lepisosteus*) uncovering a GC chromosomal pattern uncharacteristic for fish. Bioinformatic analysis of the spotted gar (*Lepisosteus oculatus*) confirmed a GC compartmentalization on GC profiles of linkage groups. This indicates a rather mammalian mode of compositional organization on gar chromosomes. Gars are thus the only analyzed extant ray-finned fishes with a GC compartmentalized genome. Since gars are cold-blooded anamniotes, our results contradict the generally accepted hypothesis that the phylogenomic onset of GC compartmentalization occurred near the origin of amniotes. Ecophysiological findings of other authors indicate a metabolic similarity of gars with mammals. We hypothesize that gars might have undergone convergent evolution with the tetrapod lineages leading to mammals on both metabolic and genomic levels. Their metabolic adaptations might have left footprints in their compositional genome evolution, as proposed by the metabolic rate hypothesis. The genome organization described here in gars sheds new light on the compositional genome evolution in vertebrates generally and contributes to better understanding of the complexities of the mechanisms involved in this process. *J. Exp. Zool. (Mol. Dev. Evol.) 00B:1–13, 2016.* © 2016 Wiley Periodicals, Inc.

J. Exp. Zool. (Mol. Dev. Evol.) 00B:1–13, 2016

How to cite this article: Symonová R, Majtánová Z, Arias-Rodriguez L, Mořkovský L, Kořínková T, Cavin L, Pokorná MJ, Doležálková M, Flajšhans M, Normandeau E, Ráb P, Meyer A, Bernatchez L. 2016. Genome compositional organization in gars shows more similarities to mammals than to other ray-finned fish. J. Exp. Zool. (Mol. Dev. Evol.) XXB: 1–13.

INTRODUCTION

Genomic compositional architecture is a complex and nonrandom organization of DNA bases (AT/GC). Genomes of higher vertebrates can be described as mosaics of compositionally homogeneous DNA domains (sometimes called isochores) with a distinct GC content (Bernardi, 2005). So called "light" domains with lower GC proportions represent the late-replicating gene-poor genome "desert" and "heavy" domains with higher GC proportions represent the early-replicating gene-rich genome "core" (Bernardi, 2005). Ray-finned fish genomic architecture with a narrower range in GC proportions differs considerably from mammalian architecture where a broad range of GC proportions occurs (Costantini et al., 2007). Genomes of hundreds of vertebrates, including about 300 fish species, were investigated for genomic base composition (Bernardi and Bernardi, '90). Since gars were never covered in this work, we have analyzed the genome and chromosomes of the spotted gar (*Lepisosteus oculatus*) and the chromosomes of the tropical gar (*Atractosteus tropicus*).

Extant gars (Lepisosteiformes) are represented by only seven species divided into two genera– *Atractosteus* and *Lepisosteus* (Nelson, 2006; Fig. 1) and are the only survivors of an early radiation of ray-finned fishes highly diversified and widely distributed in the Mesozoic (Cavin, 2010; López-Arbarello, 2012). Gars are believed to be sister to another ancient lineage, Amiiformes, represented by a single extant species, the bowfin (*Amia calva*) (Grande, 2010). Together with other "ancient fishes" sensu Inoue et al. (2003), gars are sometimes referred to as "living fossils" since they do not appear to have undergone many apparent morphological changes during the last 100 million years (Wright et al., 2012; Sallan, personal communication). Gars and bowfin did not go through any further whole genome duplications (WGD) after the two basal rounds of vertebrate genome duplications (VGD1+2/R1+2) that followed the origin of vertebrates (Fig. 1; Meyer and Malaga-Trillo '99; Vandepoele et al., 2004; Dehal and Boore, 2005). They thus exemplify one of the basal "nonteleost" actinopterygian lineages that branched off the stem lineage before the teleost-specific WGD (TGD; Amores et al., '98; Taylor et al., 2003; Crow et al., 2006) and can therefore serve as an outgroup for exploring evolutionary mechanisms associated with TGD (Amores et al., 2011; Braasch et al., 2016). Ecophysiologically, gars embody a derived group highly specialized to extreme conditions (low oxygen and high-temperature environments) with a metabolic organization that is unique among

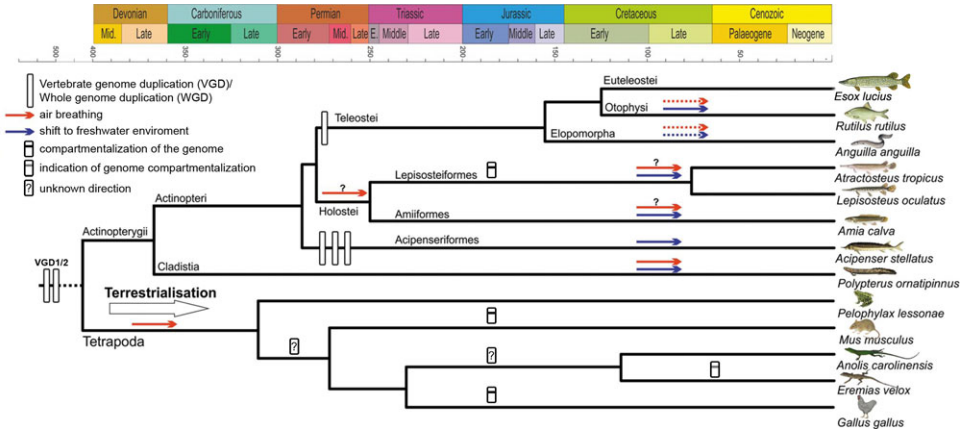


Figure 1. Time-scaled consensus tree of osteichthyes focused on the nonteleost actinopterygians and tetrapod clades in sarcopterygians. Relevant genomic and ecophysiological adaptations are plotted on the tree: vertebrate genome duplications (VGD1/2), following WGD, genome base compartmentalization, air breathing, and terrestrialization. The genomic base compartmentalization is shown to be isolated in gars within the group Actinopterygii. In this phylogeny, we favored a double appearance of heterogeneity within the amniotes (i.e., warm-blooded mammals and birds), although we cannot exclude the possibility that heterogeneity occurred at the amniotes node, then disappeared in several "reptilian" lineages.

fishes in their heightened capacity to synthesize and oxidize glucose, heightened capacity for anaerobic glycolysis, and the low number of oxidative enzymes in gar tissues (Frick et al., 2007). The metabolic organization of their air-breathing organ resembles that of lungs more than that of the swim bladder of most teleosts (Frick et al., 2007). The spotted gar gene expression pattern exhibits many similarities with those of tetrapod genomes (Braasch et al., 2014) and has been utilized to link teleost models influenced by the TGD with the human genome in disease research (Amores et al., 2011; Braasch et al., 2014, 2016).

There has been only limited investigation into gar chromosomes so far, with predominantly just basic karyotypic characteristics and GC-rich ribosomal DNA (rDNA) sites being identified (Ráb et al., '99), along with the very recent research of Braasch et al. (2016) which uncovered the striking conservation between chromosomes of the spotted gar and the chicken. Braasch et al. (2016) found that the karyotypes of these two species differed by merely 17 large fissions, fusions, or translocations.

Independently, rDNA sites in Eukaryotes are generally known as regions of substantial GC enrichment in association with GC-biased gene conversion (Escobar et al., 2011). The genome GC content was initially investigated using analytical centrifugation (Thiery et al., '76). From these studies, later supplemented with genomics, a generally accepted concept arose about the genomic GC compositional heterogeneity in birds and mammals and a substantially less heterogeneous GC organization in fishes and amphibians, with transitional states in reptiles (reviewed in Bernardi, 2005). A "thermodynamic stability hypothesis" proposed to explain this phenomenon stated that the GC compositional heterogeneity was an adaptation to homeothermy since higher GC levels stabilize coding DNAs along with their RNAs and proteins (Costantini et al., 2009).

We have combined cytogenetics and bioinformatics to obtain cytogenomic insights into the organization of the gar genome and analyzed it further by comparing gars with other vertebrates. In both gar genera, we found GC genome heterogeneity that is uncharacteristic for cold-blooded vertebrates. This was detected bioinformatically and cytogenetically in *L. oculatus* and cytogenetically in *A. tropicus*. This led us to reassess the applicability of the conventional cytogenetic method of G-banding in fishes, a technique routinely working so far only in mammals and birds. G-banding visualizes the alternation of gene-rich, early-replicating and gene-poor, late-replicating chromosome regions sensu Bernardi (2005). A comparable banding pattern can be obtained with AT/GC-specific fluorochromes and replication banding, both of which are less prone to interference by artifacts (Sumner, '90). The unavailability of G-banding in lower vertebrates represents a serious obstacle to the study of their cytogenetics, as it makes it impossible to properly identify homologous chromosomes, which is one of the basic prerequisites for all downstream cytogenetic analyses. There are reports showing "G-banding" in fishes but these concluded that the banding pattern is irreproducible and incomparable with patterns routinely produced in mammals (e.g., Medrano et al., '88; Ueda & Naoi, 1999; Romanenko et al., 2015; this study). Genuine G-banding

has been reported to work in eels (Coluccia et al., 2010), cyprinids (Luo, '98), and in tonguefish (Zhuang et al., 2006). However, these results have never been confirmed with any AT/GC-specific fluorochrome staining. Since G-banding provides a reproducible pattern in both gar genera, and genomic data of several fish species are available, we could resolve this issue. We analyzed GC profiles in following teleosts; zebrafish, representing a GC homogenized teleost genome (Costantini et al., 2007), stickleback and pufferfish, both representing more GC heterogeneous genomes as reported by Costantini et al., 2007. Analysis of GC profiles among linkage groups elucidated the relationship between the GC distribution at the sequence and the chromosomal level.

A link between environment, metabolic rate, and genomic GC content has recently been established (e.g., Chaurasia et al., 2011, 2014; Tarallo et al., 2016). Gars, with their physiological adaptations to extreme temperatures and low oxygen conditions (Frick et al., 2007) combined with their peculiar genome organization reported here, represent an excellent model system to bring new insights into this long-lasting and multilevel effort. Namely, the major evolutionary forces driving the GC content variation within and among genomes and the compositional differences between cold- and warm-blooded vertebrates still remained unclear. The GC heterogeneity across the genome results from complex interactions and can be investigated through various techniques involving cytogenetics, ecophysiology, molecular evolution, (phylo)genomics, bioinformatics, regulation of gene expression, DNA methylation, and spatial distribution of DNA in interphase nuclei (Varriale 2014; Mugal et al., 2015a).

MATERIALS AND METHODS

Blood Culturing and Chromosome Preparation

Details on specimen acquisition are listed in the Supporting Information.

Leucocytes from blood samples of *A. tropicus*, *L. oculatus*, *A. calva, Polypterus ornatipinnis,* and *Polypterus lapradei* were cultivated, and chromosome spreads were prepared according to the protocol of Fujiwara et al. (2001) with some modifications. Briefly, 0.2–0.5 mL of blood was collected from an anesthetized fish by puncturing the *vena cava caudalis* using a heparinized syringe. The leucocyte-rich plasma was used to set up primary cultures in 5 mL Medium 199 (Sigma, St. Louis, MO, USA) supplemented with 10% fetal calf serum (FBS Superior, Biochrom, Berlin, Germany),1% antibiotic antimycotic solution (Sigma), Lipopolysaccharides (LPS) from *E. coli* (0.5 mg per 5 mL medium), Phytohemagglutinin (PHA-P) (90 μg per 5 mL medium, Remel, Lenexa, KS, USA), kanamycin (0.3 mg per 5 mL medium, Sigma), and 0.175 μL 10% mercaptoethanol (Sigma). After 120 hr of incubation at 20°C, 5 mL cultures were harvested using standard colchicine (two drops of 0.1% colchicine per 5 mL media) and 0.075 M KCl hypotonic (8 min.) treatments followed by fixation in a freshly prepared fixative (methanol: acetic acid 3:1, v/v) three times. Cell suspensions of *Anolis carolinensis*, *Eremias velox*, and *Gallus gallus* were obtained by leucocyte cultivation sensu Sohn and Ryu ('99) and Pokorná et al. (2010) with slight modifications. Chromosome preparations of *M. musculus* were obtained from bone marrow sensu Ford and Hamerton ('56).

All national guidelines for the care and use of animals were followed. This study was covered by the "Valid Animal Use Protocols" Nr. CZ00221 of the Laboratory of Fish Genetics, IAPG, issued by the Czech Ministry of Agriculture; by the valid Mexican permission SAGARPA/CONAPESCA, DGOPA 09004.04111.3088 and by the Animal facilities accreditation of the Faculty of Science, Charles University in Prague, Czech Republic (24773/2008-10001).

CMA_3/DAPI (CDD) Staining. Chromomycin A_3 (CMA_3, DNA dye–specific for GC-rich regions) and DAPI (AT-specific) fluorescent staining was performed as described by Sola et al. ('92) on metaphases of the following species: *A. tropicus, L. oculatus, A. calva, A. stellatus, R. rutilus, A. anguilla, P. ornatipinnis, P. lapradei, P. lessonae, A. carolinensis, E. velox, G. gallus*, and *M. musculus*.

G- and C-Banding. G-banding was performed sensu Seabright ('71, '72) on the same chromosome preparations treated with CDD staining to directly compare banding patterns. During G-banding, we tested 15, 20, and 30 sec of trypsin treatment. G-banding was tested on chromosomes of *Amia calva*, *Acipenser baeri, A. gueldenstaedtii, A. ruthenus*, and *A. stellatus*.

C-banding was performed sensu Sumner ('90) with the slight modifications described in Pokorná et al. (2014) to determine whether heterochromatin accumulations resemble the G-banding pattern as described previously (Schmid and Guttenbach, '88; Graphodatsky, '89; Holmquist and Ashley, 2006). Chromosomes were counterstained by DAPI to enhance contrast. Microphotographs were taken in the fluorescent regime and inverted. Chromosomes were classified sensu Levan et al. ('64). Further details on the molecular cytogenetic and image analyses are listed in the Supporting Information.

Bioinformatic Analysis

We calculated GC profiles for each linkage group (LG, are expected to correspond to chromosomes, Braasch et al., 2016) in *L. oculatus* and compared them with GC profiles of *Danio rerio*, *Tetraodon nigroviridis,* and *Gasterosteus aculeatus*. For comparison with representative mammalian genomes, we calculated GC profiles of the house mouse and human. The profiles were calculated with a sliding window of 10 kbp, striding the chromosome in 10 kbp steps. Profile plots were made after combining the original data to 100 kbp windows. Genome assemblies of analyzed species were downloaded from www.ensembl.org (version published in Flicek et al., 2014).

In addition, for the spotted gar genome, an analysis of GC content in genes versus intergenic regions was conducted. This analysis was performed twice, once with all sequence data and once with transposons and low-complexity regions excluded from the intergenic regions. To identify the genomic regions, transposons, and low-complexity regions, we relied on annotations provided by ENSEMBL.

Genome assemblies were preprocessed using Python, in that A/C/G/T bases were counted for each genome. These data sets were then subjected to a GC content analysis performed for each LG separately using R v. 3.0, packages ggplot, dplyr, tidyr (R Development Core Team, 2012). All custom-generated Python and R scripts are available on the codes repository GitHub (https://github.com/libor-m/vertebrate-GC).

RESULTS

Alternation of AT- and GC-Rich Regions on Gar Chromosomes

In *A. tropicus*, the diploid chromosome number was $2n = 56$, the chromosomal arms number NF = 90 in both females and males (Fig. 2A). The karyotype was composed of 18 pairs of macrochromosomes and 10 pairs of very small chromosomes (Figs. 2A, C, E, and G; marked as st/a*). The chromosomes consisted of five pairs of meta- (m), 11 pairs of submeta- (sm), and 12 pairs of subtelo- (st) to acrocentric (a) chromosomes (Figs. 2A, C, E, and G). In metaphases with less condensed chromosomes after CDD staining, two pairs of biarmed and eight pairs of uniarmed very small chromosomes could be distinguished among the st/a category. There were no detectable differences between males and females.

In *L. oculatus*, $2n = 58$ in unsexed animals, NF = 92, the karyotype composed of 19 pairs of macrochromosomes and 10 pairs of very small chromosomes (Figs. 2B, D, F, and H; marked as st/a*). Chromosomes consisted of 4 pairs of metacentrics, 10 pairs of submetacentrics, and 14 pairs of subtelo- to acrocentric chromosomes (Figs. 2B, D, F, and H). In less condensed metaphases after CDD staining, two pairs of bi- and eight pairs of uniarmed very small chromosomes could be distinguished among the st/a category.

The base-specific CDD (DAPI/CMA_3) staining of chromosomes of *A. tropicus* and *L. oculatus* demonstrated a clear CMA_3^+-banding pattern (GC-rich regions) along with a $DAPI^+$ pattern (AT-rich regions). This alternation of AT/GC bands enabled the unambiguously distinguishing between homologous and homeologous chromosomes, excluding the very small (the ten smallest) chromosomes (Figs. 2A and B). Giemsa-stained karyotypes of *A. tropicus* and *L. oculatus* are shown in Figures 2G and H. This banding pattern enabled us to generate ideograms for both species (Fig. S1a and b in the Supporting Information). Fluorescence in situ hybridization (FISH) with rDNA yielded two (*A. tropicus*) or three (*L. oculatus*) signals for 28S rDNA and two signals for 5S rDNA were detected in both species (Fig. S2 in the Supporting Information). Contrary to gars, all other ray-finned fishes included in this analysis showed a "teleost pattern" with homogeneously stained chromosomes, that is a balanced proportion of AT/GC and GC-rich rDNA sites (Figs. 3C–G). The same pattern was observed in amphibians represented by *P. lessonae* (Fig. 3H). In reptiles, we identified a homogeneous pattern in *Anolis carolinensis* (Fig. 3J) and indications of compositional heterogeneity in *Eremias velox* (Fig. 3I). In the chicken, CDD staining yielded mostly homogeneously stained macrochromosomes (some with GC^+ regions terminally) and GC-rich microchromosomes (Fig. 3K). In the mouse, the pattern of alternating AT^+/GC^+ bands occurred on all chromosomes (Fig. 3L; details are given in the Supporting Information).

Heterochromatic C-Bands Do Not Mimic G-Bands

G-banding in gars produced a pattern enabling the karyotyping of chromosomes (Figs. 2C and D). The G-banding pattern corresponds to the CDD pattern: The positive G-bands (dark; sensu Sumner, '90) match up with the AT-rich bands, and the negative G-bands (pale; sensu Sumner '90) match up with the GC-rich bands. No clear G-banding pattern was observed in sturgeons and bowfin (not shown). Further details are given in the Supporting Information.

C-banding, which is used to visualize constitutive heterochromatin, produced bright signals in the GC-rich centromeres and weaker signals interstitially in *L. oculatus* and *A. tropicus* (Fig. 2E and F).

Based on our results, we can summarize the relationship between the constitutive heterochromatin (the C-bands), G-bands, and AT-/GC-rich regions in gars as follows: (i) The centromeric constitutive heterochromatin overlaps with the GC-rich centromeric regions but does not overlap with G-bands that are mostly absent in centromeres; (ii) the usually terminally located GC-rich regions (Fig. 2; e.g., the 1. and 2. metacentric, the 3. and 4. submetacentric chromosomes) are clearly C-negative as well as G-negative; (iii) the interstitial GC-rich bands appear to be present in the vicinity of the C-bands; however, this finding needs to be verified by specialized analysis; (iv) The G-bands can be found in clearly C-negative, C-positive, and slightly C-positive regions; (v) The very GC-rich, and also to a large extent heterochromatinized, chromosomal pair is G-positive on the q arm. Details are described in the Supporting Information. This indicates the euchromatic genome occupation of a significant proportion of G-bands in gars similar to the situation in warm-blooded vertebrates (sensu Bickmore and Craig, '96). By comparing the results with the CDD patterns, we can rule out that the C-bands mimic the G-banding pattern in gars.

GC Profiles in Genomes of Spotted Gar and Other Relevant Species

GC profiles of LGs of the spotted gar (*L. oculatus)* genome appear to correspond to our cytogenetic results and demonstrate the AT/GC heterogeneity observed along gar chromosomes at

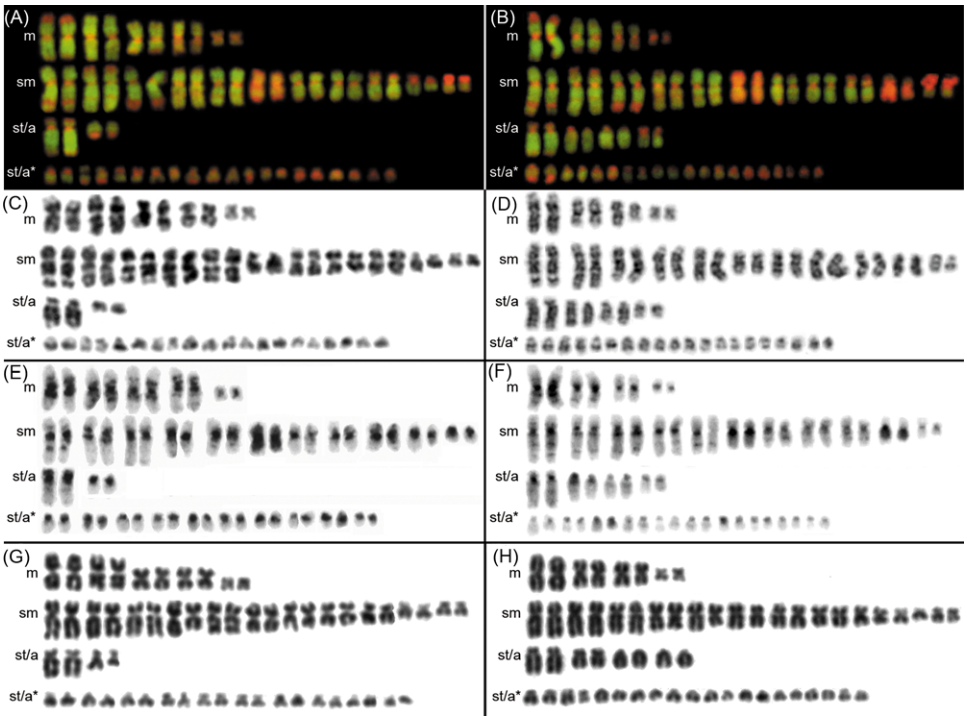


Figure 2. Karyotypes of tropical gar (A) CDD-stained; (C) G-banded; (E) C-banded; (G) Giemsa stained and of spotted gar (B) CDD-stained; (D) G-banded; (F) C-banded; (H) Giemsa stained. The CMA_3 signal was inserted from green into red and the DAPI signal from blue into the green channel. The C-banding pattern was counterstained with DAPI and inverted. m, metacentric; sm, submetacentric; st/a, subtelocentric-acrocentric chromosomes; st/a*, subtelocentric-acrocentric very small chromosomes. Images were adjusted separately before being incorporated into the plate.

the sequence level (Fig. 4A). The GC profiles also enabled the quantification of the qualitative cytogenetic data showing an alternation of GC-rich and GC-poor regions along chromosomes. Bernardi (2005) defines "heavy" GC-rich DNA domains as having >45% GC and "light" domains with 35–45% GC (Fig. 4A). The GC-rich regions in *L. oculatus* frequently exceeded 50% GC over DNA regions of several Mbp, for example, in LG1-2, LG4-6, LG8-12, and most of the small chromosomes. The GC-rich and AT-rich regions form discernible compositionally homogeneous domains. The AT-rich regions form flattened stretches. The GC-rich regions occur as GC-rich peaks interrupted by sharp decreases in GC content. The GC-rich regions are mostly telomerically but also interstitially situated and correspond to the GC-rich bands after CDD staining. However, it was impossible at this stage to identify exactly which linkage group of the *L. oculatus* genome represents which pair of chromosomes, since there were no identification markers present to link both types of data sets. GC profiles of LGs in the pufferfish with a compact teleost genome and a shift toward a higher GC content show a pattern of very densely distributed fluctuations between 40 and 50% GC with terminal (telomeric) sharp narrow increases up to 55–60% of GC (Fig. 4B). GC profiles of the stickleback also show a slight shift toward a higher GC content, compositional fluctuations between 40 and 50% GC, no prominent GC peaks and several regions of profile flattened around 45% GC (Fig. 4C). GC profiles of the zebrafish (*Danio rerio*), representing homogenized teleost genomes, are distinctly flattened between 35 and 40% GC. Only very narrow peaks reaching up to 50% GC occur in the terminal locations in the majority of LGs (Fig. 4D). In the house mouse and human, the same procedure shows fluctuations in the range of 35–55% GC, with the exception of the Y chromosome with a narrow compositional range of around 40% GC (Figs. 4E and F; Figs. S5a and b (in the Supporting Information)). Comparable patterns were obtained with window lengths of 10, 250, 500, and 1 000 kbp (not shown).

The analysis of GC content in genes versus intergenic regions of the spotted gar and other tested actinopterygian genomes shows that gene-rich regions are more GC-rich in the spotted gar as well as in pufferfish and stickleback, but not in zebrafish (Figs. S4a–d in the Supporting Information). In the spotted gar, the gene-rich regions contain 25–75% GC, whereas the intergenic regions 0–55% GC. Both the genic and intergenic

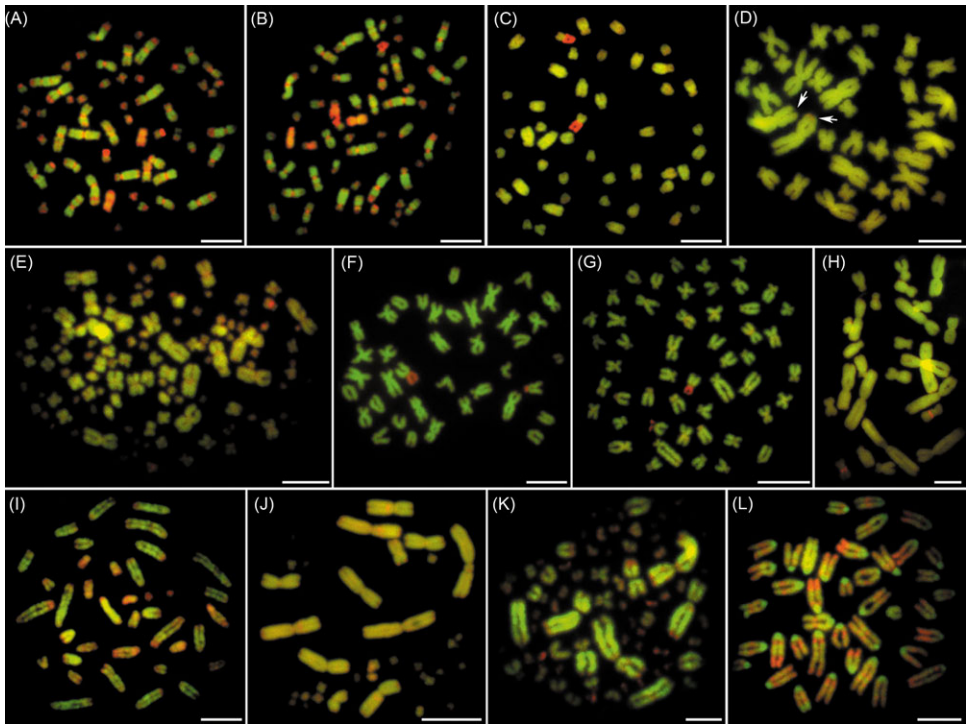


Figure 3. CDD-stained chromosomes of different vertebrate representatives: (A) tropical gar *Atractosteus tropicus;* (B) spotted gar *Lepisosteus oculatus;* (C) bowfin *Amia calva;* (D) ornate bichir *Polypterus ornatipinnis;* (E) starry sturgeon *Acipenser stellatus;* (F) European eel *Anguilla anguilla;* (G) roach *Rutilus rutilus;* (H) pool frog *Pelophylax lessonae;* (I) rapid racerunner *Eremias velox;* (J) green anole *Anolis carolinensis;* (K) chicken *Gallus gallus* and (L) house mouse *Mus musculus.* The CMA_3 signal was inserted from green into red and the DAPI signal from blue into the green channel to enhance the contrast between these signals. Images were adjusted separately before being incorporated into the plate. Bars equal 5 μm.

regions show maximal density at approximately 38% GC in the spotted gar (Fig. S4a in the Supporting Information). The GC-richness of gene-rich regions is even more pronounced in stickleback and pufferfish (Figs. S4b and c in the Supporting Information). In both of these species, the maximal density in genic regions shows a shift toward higher GC content when compared with the intergenic regions (Figs. S4b and c in the Supporting Information).

DISCUSSION

Cytogenetically Detectable AT/GC Compartmentalization in Gars

CDD staining is a versatile and well-understood tool visualizing the global AT/GC distribution in plant and animal chromosomes (Sumner, '90). This is not the case for G-banding where uncertainties about exact interactions between the dye and DNA persist (Sumner, '90). In gars, we have combined both these stainings and excluded that constitutive heterochromatin mimics G-bands by C-banding in both gar species (Figs. 2E and F). This was further supported by bioinformatic analyses showing that GC-rich regions are more gene-rich (Fig. S4 in the Supporting Information). So far, eels were the only exception among fishes with repeatedly reported functioning G-bands (Wiberg, '83; Sola et al., '84). However, our CDD staining in the eel yielded the "typical teleost pattern," that is without any GC heterogeneity along chromosomes up to two GC^+ rDNA sites (Fig. 3F). Equilibrium centrifugation in eels also did not show any pattern similar to the GC-heavy isochores (Bernardi, 2005) characteristic for birds and mammals. *A. anguilla* exhibits an intermolecular compositional heterogeneity (Medrano et al., '88; Bernardi and Bernardi, '90) similar to what was found in *A. rostrata* and explained by the presence of GC-rich satellites (Hudson et al., '80). Medrano et al. ('88) admitted that the G-bands they had produced in eels were not of the same quality as the results produced by the same technique in mammals. Moreover, CDD staining in several anguilliform fishes demonstrated that rDNA sites are the only GC-rich regions on otherwise homogeneously stained chromosomes (Salvadori et al., 2009; Coluccia et al., 2010). Therefore,

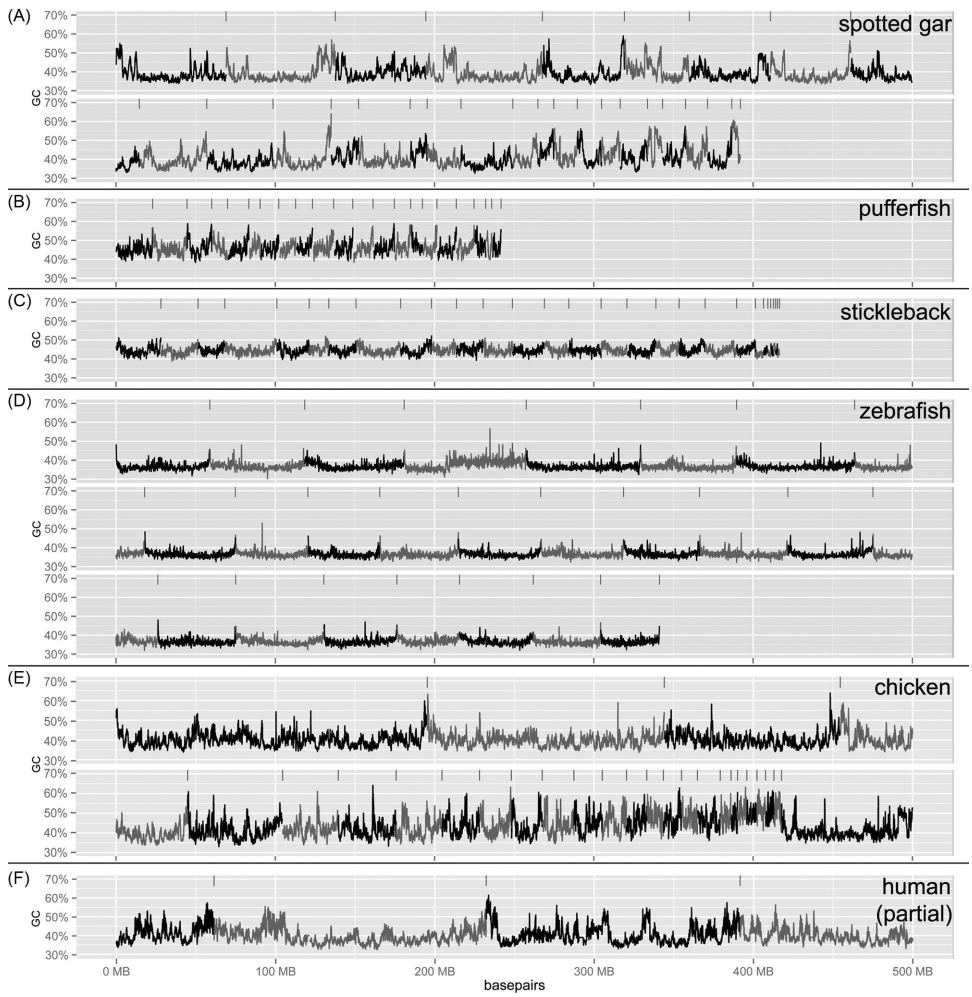


Figure 4. GC profiles along linkage groups (LGs) in (A) spotted gar, (B) pufferfish, (C) stickleback and (D) zebrafish. For comparison, chicken (E) and a partial human (F) GC profile are provided (for full human GC profile, see Fig. S3a in the Supporting Information). GC profiling was performed with uniform scaling, and differing genome sizes are reflected in sizes of profiles. LGs are arranged according to their numbers in the Ensembl database and separated by vertical lines above profiles and by the alternation of gray and black colors (*x* axis – genome position, *y* axis – percentage of GC).

no reliable report has yet been published on the successful application of G-banding in fishes as also noted by Sharma et al. (2002).

For anurans, Bernardi (2005) summarized equilibrium centrifugation results in *Xenopus, Leptodactylus, Bufo*, and *Rana* and one representative of the Caudata (*Pleurodeles waltlii*) and concluded that the genome compositional properties of Amphibians are similar to those described for fishes (Bernardi, 2005). Gazoni et al. (2012) performed CDD staining in eight species of the *Leptodactylus* frog and showed a pattern similar to the ones we observed in *P. lessonae* (Fig. S1 in the Supporting Information) and teleosts in this study. Reproducible G-banding was never successful in anurans, whereas replication banding was routinely possible (Schmid and Steinlein 2015).

There is a comparable pattern of compositional heterogeneity in gars and the house mouse with similar counts of discernible AT/GC bands per chromosome differing in AT-rich centromeres in the mouse and GC-rich in gars (Fig. 4L).

The G-bands previously reported in teleosts should be considered "G-like" bands unrelated to the genome GC compartmentalization observed in homeotherm vertebrates. We ascribe these G-like bands to a specific phase of DNA spiralization (less condensed), when some chromosomal regions are more susceptible to Giemsa staining and/or trypsin digestion.

Nevertheless, the alternation of early and late replicating DNA regions exists also in cold-blooded vertebrates as demonstrated by replication banding in fish and amphibians (Vinas et al., '96; Jankun et al., '98; Boroń, 2003; Schmid and Steinlein, 2015). The structural role of gene-rich and gene-poor regions sensu Cremer and Cremer (2001) appears to be universal across all vertebrates (Federico et al., 2006). Hence, the genome organization in cold-blooded vertebrates is not accompanied with such prominent GC segregation as in warm-blooded genomes. Even in species with more GC heterogeneous genomes (pufferfish, stickleback), the base distribution appears more shuffled and without accumulation in any distinctly GC homogeneous (GC-rich or GC-poor) domains. This may be the reason why it is impossible to visualize any bands in fish chromosomes (with the exception of gars). Therefore, it is necessary to distinguish between the ranges of GC values represented in fish genomes globally and between the AT/GC distributions along chromosomes. In anamniotes, another important factor can be the fact that noncoding sequences with various GC content may blur the higher GC content of coding sequences (Fortes et al., 2007). This is particularly apparent in Figs. S4 a–d in the Supporting Information, where stickleback and pufferfish represent genomes with reduce genome size (for comparison of genome sizes Figs. 4A–D) and a clear shift toward GC-richness. This shift is however not accompanied by such pronounced GC heterogeneity as in gars and amniotes.

How Did the GC Heterogeneity in Gar Arise?

DNA nucleotides are not randomly distributed within genomes (Straalen and Roelofs, 2012). Currently, six mutually nonexclusive hypotheses, with different weight, have been proposed to explain the variation in AT/GC organization between cold- and warm-blooded vertebrates: (1) the thermodynamic stability hypothesis, (2) selection for GC content, (3) AT-biased mutations, (4) GC-biased gene conversion (gBGC), (5) DNA methylation (Mugal et al., 2015a), and (6) the footprint of metabolism (Berná et al., 2012). With our cytogenomic data and literature sources, we can assess three of these hypotheses: (1), (4), and (6).

Thermostability Hypothesis

The "thermostability hypothesis" (Bernardi, 2005 for review) explains the higher GC heterogeneity of warm-blooded genomes as an adaptation to increased body temperature, where higher GC levels contribute to the stabilization of coding DNA fractions, RNAs, and proteins (Costantini et al., 2009). Specifically, the stabilization of GC-rich and gene-rich open chromatin structures with their high transcriptional activity associated with an increased DNA bendability and decreased nucleosome formation potential (Bernardi, 2005; Vinogradov, 2005). It has not been explicitly tested whether gars are cold- or warm-blooded (Miller et al., 2006; Nelson, 2006). However, since they are actinopterygians (Cavin, 2010; Near et al., 2012), they are definitely anamniotes which rules out the possibility that GC heterogenization occurred at the anamniotes – amniotes transition. This opposes the "thermostability hypothesis" linking GC heterogenization with the homeothermy of amniotes (Costantini et al., 2009; Uliano et al., 2010; Tarallo et al., 2016). Based on the report of Farrell (2007), we assume that gars are cold-blooded. The tropical gar in Mexico is able to tolerate the extremely high temperatures and very low oxygen conditions of tropical swamps (Burleson et al.,'98; Miller et al., 2006; Arias-Rodriguez own observations). Whereas, gars from the temperate regions of North America, including the spotted gar analyzed in this study, thrive also in low temperatures. Further investigations should be carried out to study their thermoregulation adaptations, as for example, in elasmobranchs and tuna fish (Block and Finnerty, '94).

GC-Biased Gene Conversion

GC-biased gene conversion (gBGC) belongs to widely evidenced hypotheses and states that regions with higher recombination rates contain higher local GC content (Lartillot 2013; Mugal et al., 2013, 2015b). In gars, the centromeres, the majority of very small chromosomes, and some short chromosomal arms (all anticipated high-recombining regions), together with a number of other regions, are distinctly GC-rich (Fig. 2A and B; Figs. S3a and b in the Supporting Information). This supports gBGC in terms introduced, for example, by Pessia et al., 2012, who successfully used chromosome size as a proxy for recombination rate. High GC levels also occur in high-recombining 45S rDNA regions in gars, which is consistent with the gBGC model demonstrated by Escobar et al. (2011). Consequently, low recombination rates are associated with decreased GC proportions, for example, in the short-tail opossum where the low number of large chromosomes was proven to be linked with the low recombination rate in autosomes and an overall low GC content (Mikkelsen et al., 2007). We show similar situations in the pool frog and the anole, where the large chromosomes were homogeneously CDD-stained (Fig. 3). The same correlation between chromosome length and GC distribution has been reported in birds and mammals (Romiguier et al., 2010; Weber et al., 2014).

Gars' Unique Metabolism

A growing number of studies, some also in teleosts (Uliano et al., 2010; Chaurasia et al., 2011, 2014; Berná et al., 2012; Tarallo et al., 2016) introduced a novel hypothesis that environment temperature, oxygen content, and lifestyle may be factors correlating with the genome GC composition. Specifically, an increment in metabolic rate, known to be higher in mammals and birds, affects transcription activity and chromatin structure (Berná et al., 2012).

This hypothesis may be of potential importance in the future since in the Florida gar, the following metabolic adaptations were identified: (i) a heightened capacity to synthesize and oxidize glucose; (ii) a heightened capacity for anaerobic

glycolysis; (iii) a low number of oxidative enzymes in gar tissues, and (iv) the evolution of their modified swim bladder resembling the lungs of tetrapods in form and function (Frick et al., 2007; Echelle and Grande, 2014 and references therein). Gars are bimodal in respiration—they use their gills for water breathing and their lung(s) for air breathing. This gives them a greater respiratory efficiency in exhaustive situations, allowing them to continue normal activity under conditions that would incapacitate virtually every other fish (Echelle and Grande, 2014).

Chaurasia et al. (2014) proposed a correlation between *temperature-corrected* metabolic rate (MR), entire genomic GC content, and the length and GC content of introns in teleosts. This will also have to be tested in gars. Hence, the overall proportion of introns and their GC content may indirectly influence the GC content of the entire genome (e.g., in compact genomes). In this regard, gars may differ from other tropical fish in their modified metabolism and/or increased MR (Uliano et al., 2010). Therefore, the temperature component might be involved in the compositional genome evolution as suggested by Bernardi (2005). However, the final modulation might instead be influenced by metabolism-driven traits (Uliano et al., 2010; Chaurasia et al., 2011). These correlations and potential causes are yet to be properly analyzed, particularly in gars and bowfin. Differences in physiology between gars and the bowfin indicate specific metabolic traits in gars resembling those of mammals (Frick et al., 2007). This might help to explain why gars possess cytogenetically and bioinformatically detectable GC heterogeneity whereas their putative sister lineage represented by the bowfin exhibit the typical "teleost pattern" on its chromosomes.

Evolutionary Rate in Gars' Evolution

Gars are considered to have evolved slowly and to show low rates of speciation (Rabosky et al., 2013; Braasch et al., 2016). However, these analyses are based only on the living subset of vertebrate lineages, overlooking the great diversity seen in the fossil record, more so than that found in Polypteriformes, Chondrostea, and Amiidae (Cavin, 2010; López-Arbarello 2012; Sallan, personal communication). The palaeobiology of gars and the bowfin based on solid fossil evidence (Sallan, personal communication) indicates a 50 million year long period of accelerated evolution that occurred in gars but not in amiids with an otherwise very similar evolutionary-ecological history (Grande, 2010). This period of accelerated evolution might have also been accompanied by dynamic changes in gars, which may have left a footprint in their genome in the form of GC compartmentalization. (More details are presented in the Supporting Information). Moreover, based on the fossil evidence, gars and early tetrapods could have faced similar selective pressures and converged on similar solutions (Sallan, personal communication). On the other hand, at this stage, we also have to take into account that gars could have retained the ancestral genome structure of Gnathostomata.

The compositional genome heterogeneity among vertebrates cannot be attributed merely to the transition from anamniotes to amniotes as generally accepted before this study. The genomic DNA base architecture in vertebrates appears to rather result from a interplay of multiple opposing forces, for example, AT-biased mutations and GC-biased gene conversions (e.g., Weller et al., 2014), genome and chromosome sizes (e.g., Romiguier et al., 2010), selection for GC and drift, and by diverse cellular, metabolic (Chaurasia et al., 2014) and environmental (e.g., Uliano et al., 2010; Chaurasia et al., 2011) factors influenced by physical and chemical properties of AT/GC and their role in the regulation of gene expression (Vinogradov, 2005). In gars, such interplay should be explored with regard to their metabolic adaptations (sensu Uliano et al., 2010; Chaurasia et al., 2011, 2014). The situation presented here in gars, in the light of the known genomic traits of their closest extant relatives, implies important consequences for the phylogenomic evolution not only of basal vertebrates but also for amniotes.

ACKNOWLEDGMENTS

We thank J. Zima for comments about the manuscript, Š. Pelikánová, J. Čechová, and P. Šejnohová of the LFG for assistance in the laboratory, L. Choleva for *P. lessonae* cell suspensions, M. Altmanová and L. Kratochvíl for help during chromosome preparations of reptiles, R. Černý for providing *P. lapradei* blood samples, T. Kučera for blood samples of *P. ornatipinnis,* K. Trejbalová for the introduction to ultracentrifugation, J. R. Indy for help with *A. tropicus*, and L. Sallan for the elucidation of paleobiological aspects of the evolution of gars. All colleagues who are acknowledged agreed to the final version of the text.

All custom generated Python and R scripts produced for this study are available at the codes repository GitHub: https://github.com/libor-m/vertebrate-GC

AUTHOR CONTRIBUTIONS

R.S. designed the study. R.S. and Z.M. performed cytogenetic analyses and codrafted the manuscript together with P.R., A.M., E.N., and L.B. L.M. performed bioinformatic analyses. L.C. provided the paleobiological context and codrafted the manuscript together with M.J.P. and Z.M. T.K. and L.A.R. participated on cytogenetic analyses. L.A.R., M.J.P., M.F., and M.D. provided chromosome preparations. All coauthors revised the text and agreed to the final version.

LITERATURE CITED

Amores A, Catchen J, Ferrara A, Fontenot Q, Postlethwait JH. 2011. Genome evolution and meiotic maps by massively parallel DNA sequencing: spotted gar, an outgroup for the teleost genome duplication. Genetics 188:799–808.

Amores A, Force A, Yan YL, Joly L, Amemiya C, Fritz A, Ho RK, Langeland J, Prince V, Wang YL, Westerfield M, Ekker M, Postlethwait

JH. 1998. Zebrafish hox clusters and vertebrate genome evolution. Science 282:1711-1714.

Berná L, Chaurasia A, Angelini C, Federico C, Saccone S, D'Onofrio G. 2012. The footprint of metabolism in the organization of mammalian genomes. BMC Genomics 13:174.

Bernardi G. 2005. Structural and evolutionary genomics: natural selection in genome evolution. Amsterdam: Elsevier.

Bernardi G, Bernardi G. 1990. Compositional patterns in the nuclear genome of cold-blooded vertebrates. J Mol Evol 31:265–281.

Bickmore WA, Craig J. 1996. Chromosome bands: Patterns in the genome, 1st ed. New York: Springer.

Block BA, Finnerty JR. 1994. Endothermy in fishes: a phylogenetic analysis of constraints, predispositions, and selection pressures. Environ Biol Fishes 40:283–302.

Boroń A. 2003. Replication banding patterns in the spined loach, *Cobitis taenia* L. (Pisces, Cobitidae). Genetica 119:51–55.

Braasch I, Gehrke AR, Smith JJ, Kawasaki K, Manousaki T, Pasquier J, Amores A, Desvignes T, Batzel P, Catchen J, Berlin AM, Campbell MS, Barrell D, Martin KJ, Mulley JF, Ravi V, Lee AP, Nakamura T, Chalopin D, Fan S, Wcisel D, Cañestro C, Sydes J, Beaudry FEG, Sun Y, Hertel J, Beam MJ, Fasold M, Ishiyama M, Johnson J, Kehr S, Lara M, Letaw JH, Litman GW, Litman RT, Mikami M, Ota T, Saha NR, Williams L, Stadler PF, Wang H, Taylor JS, Fontenot Q, Ferrara A, Searle SMJ, Aken B, Yandell M, Schneider I, Yoder JA, Volff J-N, Meyer A, Amemiya CT, Venkatesh B, Holland PWH, Guiguen Y, Bobe J, Shubin NH, Di Palma F, Alföldi J, Lindblad-Toh K, Postlethwait JH. 2016. The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons. Nat Genet 48:427–437.

Braasch I, Guiguen Y, Loker R, Letaw JH, Ferrara A, Bobe J, Postlethwait JH. 2014. Connectivity of vertebrate genomes: paired-related homeobox (Prrx) genes in spotted gar, basal teleosts, and tetrapods. Comp Biochem Physiol C Toxicol Pharmacol 163:24–36.

Burleson M, Shipman B, Smatresk N. 1998. Ventilation and acid-base recovery following exhausting activity in an air-breathing fish. J Exp Biol 201:1359–1368.

Cavin L. 2010. Diversity of Mesozoic semionotiform fishes and the origin of gars (Lepisosteidae). Naturwissenschaften 97:1035–1040.

Chaurasia A, Tarallo A, Bernà L, Yagi M, Agnisola C, D'Onofrio G. 2014. Length and GC content variability of introns among teleostean genomes in the light of the metabolic rate hypothesis. PloS One 9:e103889.

Chaurasia A, Uliano E, Berná L, Agnisola C, D'Onofrio G. 2011. Does habitat affect the genomic GC content? A lesson from teleostean fish: a mini review. In: Fish ecology. Hauppauge, NY: Nova Science Publishers. p 61–80.

Cohen N, Dagan T, Stone L, Graur D. 2005. GC composition of the human genome: in search of isochores. Mol Biol Evol 22:1260–1272.

Coluccia E, Deiana AM, Libertini A, Salvadori S. 2010. Cytogenetic characterization of the moray eel Gymnothorax tile and chromosomal banding comparison in Muraenidae (Anguilliformes). Mar Biol Res 6:106–111.

Costantini M, Auletta F, Bernardi G. 2007. Isochore patterns and gene distributions in fish genomes. Genomics 90:364–371.

Costantini M, Cammarano R, Bernardi G. 2009. The evolution of isochore patterns in vertebrate genomes. BMC Genomics 10:146.

Cremer T, Cremer C. 2001. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. Nat Rev Genet 2:292–301.

Crow KD, Smith CD, Cheng J-F, Wagner GP, Amemiya CT. 2012. An independent genome duplication inferred from Hox paralogs in the American paddlefish–a representative basal ray-finned fish and important comparative reference. Genome Biol Evol 4:937–953.

Crow KD, Stadler PF, Lynch VJ, Amemiya C, Wagner GP. 2006. The "fish-specific" Hox cluster duplication is coincident with the origin of teleosts. Mol Biol Evol 23:121–136.

Dehal P, Boore JL. 2005. Two rounds of whole genome duplication in the ancestral vertebrate. PLoS Biol 3:e314.

Duret L, Mouchiroud D, Gautier C. 1995. Statistical analysis of vertebrate sequences reveals that long genes are scarce in GC-rich isochores. J Mol Evol 40:308–317.

Echelle AA, Grande L. 2014. Lepisosteidae: Gars. In: Warren ML, Burr BM, editors. Freshwater fishes of North America: Volume 1: Petromyzontidae to Catostomidae. Baltimore: JHU Press. p 243–278.

Escobar JS, Glémin S, Galtier N. 2011. GC-biased gene conversion impacts ribosomal DNA evolution in vertebrates, angiosperms, and other eukaryotes. Mol Biol Evol 28:2561–2575.

Farrell AP. 2007. Cardiovascular systems in primitive fishes. In: McKenzie DJ, Farrel AP and Branner CJ, editors. Fish physiology, Vol. 26: Primitive fishes. San Diego, CA: Academic Press. p 53–120.

Federico C, Scavo C, Cantarella CD, Motta S, Saccone S, Bernardi G. 2006. Gene-rich and gene-poor chromosomal regions have different locations in the interphase nuclei of cold-blooded vertebrates. Chromosoma 115:123–128.

Flicek P, Amode MR, Barrell D, Beal K, Billis K, Brent S, Carvalho-Silva D, Clapham P, Coates G, Fitzgerald S, Gil L, Girón CG, Gordon L, Hourlier T, Hunt S, Johnson N, Juettemann T, Kähäri AK, Keenan S, Kulesha E, Martin FJ, Maurel T, McLaren WM, Murphy DN, Nag R, Overduin B, Pignatelli M, Pritchard B, Pritchard E, Riat HS, Ruffier M, Sheppard D, Taylor K, Thormann A, Trevanion SJ, Vullo A, Wilder SP, Wilson M, Zadissa A, Aken BL, Birney E, Cunningham F, Harrow J, Herrero J, Hubbard TJP, Kinsella R, Muffato M, Parker A, Spudich G, Yates A, Zerbino DR, Searle SMJ. 2014. Ensembl 2014. Nucleic Acids Res 42:D749-755.

Ford CE, Hamerton JL. 1956. A colchicine, hypotonic citrate, squash sequence for mammalian chromosomes. Stain Technol 31:247–251.

Fortes GG, Bouza C, Martínez P, Sánchez L. 2007. Diversity in isochore structure among cold-blooded vertebrates based on GC content of coding and non-coding sequences. Genetica 129:281–289.

Frick NT, Bystriansky JS, Ballantyne JS. 2007. The metabolic organization of a primitive air-breathing fish, the Florida gar (Lepisosteus platyrhincus). J Exp Zool A Ecol Genet Physiol 307:7–17.

Fujita MK, Edwards SV, Ponting CP. 2011. The Anolis lizard genome: an amniote genome without isochores. Genome Biol Evol 3:974–984.

Fujiwara A, Nishida-Umehara C, Sakamoto T, Okamoto N, Nakayama I, Abe S. 2001. Improved fish lymphocyte culture for chromosome preparation. Genetica 111:77–89.

Gazoni T, Gruber SL, Silva APZ, Araújo OGS, Narimatsu H, Strüssmann C, Haddad CFB, Kasahara S. 2012. Cytogenetic analyses of eight species in the genus *Leptodactylus* Fitzinger, 1843 (Amphibia, Anura, Leptodactylidae), including a new diploid number and a karyotype with multiple translocations. BMC Genet 13:109.

Grande L. 2010. An empirical synthetic pattern study of gars (Lepisosteiformes) and closely related species, based mostly on skeletal anatomy the resurrection of Holostei. Lawrence, KS: American Society of Ichthyologists and Herpetologists.

Graphodatsky AS. 1989. Conserved and variable elements of mammalian chromosomes. In: Halnan CRE, editor. Cytogenetics of animals. Oxon, UK: CAB International Press. p 95–124.

Holmquist GP. 1989. Evolution of chromosome bands: molecular ecology of noncoding DNA. J Mol Evol 28:469–486.

Holmquist GP, Ashley T. 2006. Chromosome organization and chromatin modification: influence on genome function and evolution. Cytogenet Genome Res 114:96–125.

Hudson AP, Cuny G, Cortadas J, Haschemeyer AE, Bernardi G. 1980. An analysis of fish genomes by density gradient centrifugation. Eur J Biochem 112:203–210.

Inoue JG, Miya M, Tsukamoto K, Nishida M. 2003. Basal actinopterygian relationships: a mitogenomic perspective on the phylogeny of the "ancient fish." Mol Phylogenet Evol 26:110–120.

Jankun M, Ocalewicz K, Woznicki P. 1998. Replication, C- and fluorescent chromosome banding patterns in european whitefish, *Coregonus lavavetus* L. Hereditas 128:195–199.

Lartillot N. 2013. Phylogenetic patterns of GC-biased gene conversion in placental mammals, and the evolutionary dynamics of recombination landscapes. Mol Biol Evol 30:489–502.

Levan A, Fredga K, Sandberg AA. 1964. Nomenclature for centromeric position on chromosomes. Hereditas 52:201–220.

López-Arbarello A. 2012. Phylogenetic interrelationships of ginglymodian fishes (Actinopterygii: Neopterygii). PLoS One 7:e39370.

Ludwig A, Belfiore NM, Pitra C, Svirsky V, Jenneckens I. 2001. Genome duplication events and functional reduction of ploidy levels in sturgeon (Acipenser, Huso and Scaphirhynchus). Genetics 158:1203–1215.

Luo C. 1998. Multiple chromosomal banding in grass carp, *Ctenopharyngodon idellus*. Heredity 81:481–485.

Medrano L, Bernardi G, Couturier J, Dutrillaux B, Bernardi G. 1988. Chromosome banding and genome compartmentalization in fishes. Chromosoma 96:178–183.

Meyer A, Málaga-Trillo E. 1999. Vertebrate genomics: More fishy tales about Hox genes. Curr Biol 9:R210–213.

Mikkelsen TS, Wakefield MJ, Aken B, Amemiya CT, Chang JL, Duke S, Garber M, Gentles AJ, Goodstadt L, Heger A, Jurka J, Kamal M, Mauceli E, Searle SMJ, Sharpe T, Baker ML, Batzer MA, Benos PV, Belov K, Clamp M, Cook A, Cuff J, Das R, Davidow L, Deakin JE, Fazzari MJ, Glass JL, Grabherr M, Greally JM, Gu W, Hore TA, Huttley GA, Kleber M, Jirtle RL, Koina E, Lee JT, Mahony S, Marra MA, Miller RD, Nicholls RD, Oda M, Papenfuss AT, Parra ZE, Pollock DD, Ray DA, Schein JE, Speed TP, Thompson K, VandeBerg JL, Wade CM, Walker JA, Waters PD, Webber C, Weidman JR, Xie X, Zody MC, Broad Institute Genome Sequencing Platform, Broad Institute Whole Genome Assembly Team, Graves JAM, Ponting CP, Breen M, Samollow PB, Lander ES, Lindblad-Toh K. 2007. Genome of the marsupial *Monodelphis domestica* reveals innovation in non-coding sequences. Nature 447:167–177.

Miller RR, Minckley WL, Norris SM. 2006. Freshwater fishes of México. Chicago, IL: University of Chicago Press.

Mugal CF, Arndt PF, Ellegren H. 2013. Twisted signatures of GC-biased gene conversion embedded in an evolutionary stable karyotype. Mol Biol Evol 30:1700–1712.

Mugal CF, Arndt PF, Holm L, Ellegren H. 2015a. Evolutionary consequences of DNA methylation on the GC content in vertebrate genomes. G3 (Bethesda) 5:441–447.

Mugal CF, Weber CC, Ellegren H. 2015b. GC-biased gene conversion links the recombination landscape and demography to genomic base composition. Bioessays 37:1317–1326

Near TJ, Eytan RI, Dornburg A, Kuhn KL, Moore JA, Davis MP, Wainwright PC, Friedman M, Smith WL. 2012. Resolution of ray-finned fish phylogeny and timing of diversification. Proc Natl Acad Sci U S A 109:13698–13703.

Nelson JS. 2006. Fishes of the world, 4th ed. Hoboken, NJ: Wiley.

Neusser M. 2004. Karyotypevolution, Genomorganisation und Zellkernarchitektur der Neuweltaffen. Dissertation, LMU München: Fakultät für Biologie.

Pessia E, Popa A, Mousset S, Rezvoy C, Duret L, Marais GA. 2012. Evidence for widespread GC-biased gene conversion in eukaryotes. Genome Biol Evol 4:675-682

Pokorná M, Rábová M, Ráb P, Ferguson-Smith MA, Rens W, Kratochvíl L. 2010. Differentiation of sex chromosomes and karyotypic evolution in the eye-lid geckos (Squamata: Gekkota: Eublepharidae), a group with different modes of sex determination. Chromosome Res 18:809–820.

Pokorná M, Rens W, Rovatsos M, Kratochvíl L. 2014. A ZZ/ZW sex chromosome system in the thick-tailed Gecko (Underwoodisaurus milii; Squamata: Gekkota: Carphodactylidae), a member of the ancient gecko lineage. Cytogenet Genome Res 142:190–196.

Ráb P, Rábová M, Reed KM, Phillips RB. 1999. Chromosomal characteristics of ribosomal DNA in the primitive semionotiform fish, longnose gar *Lepisosteus osseus*. Chromosome Res Int J Mol Supramol Evol Asp Chromosome Biol 7:475–480.

Rabosky DL, Santini F, Eastman J, Smith SA, Sidlauskas B, Chang J, Alfaro ME. 2013. Rates of speciation and morphological evolution

are correlated across the largest vertebrate radiation. Nat Commun 4:1958.

R Development Core Team. 2012. R: a language and environment for statistical computing. Vienna, Austria. Available at http://www.R-project.org/

Romanenko SA, Biltueva LS, Serdyukova NA, Kulemzina AI, Beklemisheva VR, Gladkikh OL, Lemskaya NA, Interesova EA, Korentovich MA, Vorobieva NV, Graphodatsky AS, Trifonov VA. 2015. Segmental paleotetraploidy revealed in sterlet (*Acipenser ruthenus*) genome by chromosome painting. Mol Cytogenet 8:90.

Romiguier J, Ranwez V, Douzery EJP, Galtier N. 2010. Contrasting GC-content dynamics across 33 mammalian genomes: relationship with life-history traits and chromosome sizes. Genome Res 20:1001–1009.

Sallan LC. 2014. Major issues in the origins of ray-finned fish (Actinopterygii) biodiversity. Biol Rev 89:950–971.

Salvadori S, Coluccia E, Cannas R, Cau A, Deiana AM. 2009. A ZZ-ZW sex chromosome system in the finless eel *Dalophis imberbis* (Anguilliformes, Ophichtidae). Genetica 135:283–288.

Schmid M, Guttenbach M. 1988. Evolutionary diversity of reverse (R) fluorescent chromosome bands in vertebrates. Chromosoma 97:101–114.

Schmid M, Steinlein C. 2015. Chromosome banding in Amphibia. XXXII. The genus *Xenopus* (Anura, Pipidae). Cytogenet Genome Res 145:201–217.

Seabright M. 1971. A rapid banding technique for human chromosomes. Lancet 2:971–972.

Seabright M. 1972. The use of proteolytic enzymes for the mapping of structural rearrangements in the chromosomes of man. Chromosoma 36:204–210.

Sharma OP, Tripathi NK, Sharma KK. 2002. A review of chromosome banding in fishes. In: Sobit RC, Obe G, Athwal RS, editors. Some aspects of chromosome structure and functions. Dordrecht, the Netherlands: Springer. p 109–122.

Sohn SH, Ryu EK. 1999. Chromosome preparation from chick leucocyte culture using ficoll treatment. J AgricTechnol Res 12:131–140.

Sola L, Camerini B, Cataudella S. 1984. Cytogenetics of Atlantic eels: C- and G-banding, nucleolus organizer regions, and DNA content. Cytogenet Genome Res 38:206–210.

Sola L, Rossi AR, Iaselli V, Rasch EM, Monaco PJ. 1992. Cytogenetics of bisexual/unisexual species of Poecilia. II. Analysis of heterochromatin and nucleolar organizer regions in *Poecilia mexicana mexicana* by C-banding and DAPI, quinacrine, chromomycin A3, and silver staining. Cytogenet Cell Genet 60:229–235.

Straalen NM van, Roelofs D. 2012. An introduction to ecological genomics. Oxford, UK: Oxford University Press.

Sumner AT. 1990. Chromosome banding. London: Unwin Hyman.

Tarallo A, Angelini C, Sanges R, Yagi M, Agnisola C, D'Onofrio G. 2016. On the genome base composition of teleosts: the effect of environment and lifestyle. BMC Genomics. 17:173.

Taylor JS, Braasch I, Frickey T, Meyer A, Van de Peer Y. 2003. Genome duplication, a trait shared by 22000 species of ray-finned fish. Genome Res 13:382-390.

Thiery J, Macaya G, Bernardi G. 1976. Analysis of eukaryotic genomes by density gradient centrifugation. J Mol Biol 108:219–235.

Ueda T, Naoi H. 2011. BrdU-4Na-EDTA-Giemsa band karyotypes of 3 small freshwater fish, *Danio rerio*, *Oryzias latipes*, and *Rhodeus ocellatus*. Genome 3:531–535.

Uliano E, Chaurasia A, Bernà L, Agnisola C, D'Onofrio G. 2010. Metabolic rate and genomic GC. What we can learn from teleost fish. Mar Genomics 3:29–34.

Vandepoele K, De Vos W, Taylor JS, Meyer A, Van de Peer Y. 2004. Major events in the genome evolution of vertebrates: paranome age and size differ considerably between ray-finned fishes and land vertebrates. Proc Natl Acad Sci U S A 101:1638–1643.

Varriale A. 2014. DNA methylation, epigenetics, and evolution in vertebrates: facts and challenges. Int J Evol Biol 2014:e475981.

Vinas A, Gómez C, Martínez P, Sánchez L. 1996. Replication banding in the chromosomes of the European eel (*Anguilla anguilla*). Genetica 98:107–110.

Vinogradov AE. 2005. Noncoding DNA, isochores and gene expression: nucleosome formation potential. Nucleic Acids Res 33:559–563.

Weber CC, Boussau B, Romiguier J, Jarvis ED, Ellegren H. 2014. Evidence for GG-biased gene conversion as a driver of between-lineage differences in avian base composition. BMC Genome Biol 15:549.

Weller AM, Rödelsperger C, Eberhardt G, Molnar RI, Sommer RJ. 2014. Opposing forces of A/T-biased mutations and G/C-biased gene conversions shape the genome of the nematode *Pristionchus pacificus*. Genetics 196:1145–1152.

Wiberg UH. 1983. Sex determination in the European eel (*Anguilla anguilla*, L.). A hypothesis based on cytogenetic results, correlated with the findings of skewed sex ratios in eel culture ponds. Cytogenet Cell Genet 36:589–598.

Wright JJ, David SR, Near TJ. 2012. Gene trees, species trees, and morphology converge on a similar phylogeny of living gars (Actinopterygii: Holostei: Lepisosteidae), an ancient clade of ray-finned fishes. Mol Phylogenet Evol 63:848–856.

Zhuang ZM, Wu D, Zhang SC, Pang QX, Wang CL, Wan RJ. 2006. G-banding patterns of the chromosomes of tonguefish *Cynoglossus semilaevis* Günther, 1873. J Appl Ichthyol 22:437–440.

RESEARCH ARTICLE

"Holostei versus Halecostomi" Problem: Insight from Cytogenetics of Ancient Nonteleost Actinopterygian Fish, Bowfin *Amia Calva*

ZUZANA MAJTÁNOVÁ[1,2,†], RADKA SYMONOVÁ[1,3*,†], LENIN ARIAS-RODRIGUEZ[4], LAUREN SALLAN[5], AND PETR RÁB[1]

[1]Laboratory of Fish Genetics, Institute of Animal Physiology and Genetics, Academy of Sciences of the Czech Republic, Liběchov, Czech Republic
[2]Department of Zoology, Faculty of Science, Charles University in Prague, Prague 2, Czech Republic
[3]Research Institute for Limnology, University of Innsbruck, Mondsee, Austria
[4]División Académica de Ciencias Biológicas, Universidad Juárez Autónoma de Tabasco (UJAT), Villahermosa, Tabasco, México
[5]Department of Earth and Environmental Science, University of Pennsylvania, Philadelphia, Pennsylvania

ABSTRACT

Bowfin belongs to an ancient lineage of nonteleost ray-finned fishes (actinopterygians) and is the only extant survivor of a once diverged group, the Halecomorphi or Amiiformes. Owing to the scarcity of extant nonteleost ray-finned lineages, also referred as "living fossils," their phylogenetic interrelationships have been the target of multiple hypotheses concerning their sister group relationships. Molecular and morphological data sets have produced controversial results; bowfin is considered as either the sister group to genome-duplicated teleosts (together forming the group of Halecostomi) or to gars (Lepisosteiformes; together forming the group of Holostei). However, any detailed cytogenetic analysis of bowfin chromosomes has never been performed to address this issue. Here we examined bowfin chromosomes by conventional (Giemsa-staining, C-banding, base-specific fluorescence and silver staining) and molecular (FISH with rDNA probes) cytogenetic protocols. We identified diploid chromosome number $2n = 46$ with a middle-sized submetacentric chromosome pair as the major ribosomal DNA-bearing (45S rDNA), GC-positive and silver-positive element. The minor rDNA (5S rDNA) sites were localized in the pericentromeric region of one middle-sized acrocentric chromosome pair. Comparison with available cytogenetic data of other nonteleost actinopterygians (bichirs, sturgeons, gars) and

[†]These authors contributed equally.

Conflicts of interest: None
Author contributions detailed in the Acknowledgements.
Grant sponsor: This study was supported by project of Czech Science Foundation (www.gacr.cz; project number 14-02940S PR, RS and ZM), the National Council for Science and Technology of Mexico (CONACYT) No. 167686/204264, and the University of Pennsylvania.

*Correspondence to: Radka Symonová, Research Institute for Limnology, University of Innsbruck, Mondseestraße 9, A-5310 Mondsee, Austria.
E-mail: Radka.Symonova@uibk.ac.at

Received 27 June 2016; Revised 12 September 2016; Accepted 22 November 2016
DOI: 10.1002/jez.b.22720
Published online in Wiley Online Library (wileyonlinelibrary.com).

teleost species including representative of basally branching lineages showed bowfin chromosomal characteristics more similar to the teleost type than to any other nonteleosts. Particularly striking differences were identified between bowfin and gars, the latter of which were found to mimic mammalian AT/GC genomic organisation. Such conclusion however contradicts the most recent phylogenomic results and raises the question what states are ancestral and what are derived. *J. Exp. Zool. (Mol. Dev. Evol.) 00:1–9, 2017.* © 2017 Wiley Periodicals, Inc.

J. Exp. Zool. (Mol. Dev. Evol.) 00:1–9, 2017

How to cite this article: MAJTÁNOVÁ Z, SYMONOVÁ R, ARIAS-RODRIGUEZ L, SALLAN L, RÁB P. 2017. "Holostei versus Halecostomi" Problem: Insight from Cytogenetics of Ancient Nonteleost Actinopterygian Fish, Bowfin *Amia Calva*. J. Exp. Zool. (Mol. Dev. Evol.) XXX:1–9

INTRODUCTION

The bowfin *Amia calva*, a voracious freshwater predator occurring in eastern North America, is the only extant survivor of a once taxonomically diverse and widespread halecomorph (or amiiform) fish fauna from the Mesozoic and Paleogene periods (Grande and Bemis, '98; Nelson, 2006). As a single extant halecomorph species belonging to a genus, which first appeared over 100 million years ago (Grande and Bemis '98) and representing one of only five living nonteleost actinopterygian lineages, bowfin represent exceptional example of species with extremely long-retained ecomorphologies. The bowfin has also been of interest to biologists due to its longevity, aestivation, and air breathing (see Grande and Bemis, '98 and reference therein). More recently, bowfin became an indispensable component of molecular and morphological phylogenetic analyses of actinopterygians (in-depth review in Sallan, 2014). Multiple studies have sought to determine the interrelationships of nonteleost and teleost ray-finned fishes to reconstruct deep branching events within the clade. These have produced two contradictory hypotheses: (i) molecular phylogenies overwhelmingly place bowfin as the sister group to gars, Lepisosteiformes or Ginglymodi within the monophyletic Holostei lineage (Inoue et al., 2003; Near et al., 2012; Betancur-R et al., 2013; Broughton et al., 2013; Faircloth et al., 2013; Braasch et al., 2016; Pasquier et al., 2016); and (ii) the majority of morphological studies favor a sister group relationship of bowfin with teleosts within the monophyletic Halecostomi (in details reviewed by Sallan, 2014; see Grande, 2010, for another view). This has produced the famous "gar–bowfin–teleost" or "Holostei versus Halecostomi" problem (e.g., Grande, 2010; López-Arbarello, 2012; Sallan, 2014) and obscured the circumstances of origins of teleosts and their innovations, such as genome duplication and morphological novelties. To address this phylogenetic debate, numerous osteological, paleontological, as well as molecular characteristics and combinations thereof have been investigated (Sallan, 2014), including, most recently, ultraconserved elements of the genome (Faircloth et al., 2013) or analyses of the spotted gar genome and bowfin transcriptome (Braasch et al., 2016). Available cytogenetic data, that is, chromosome number, karyotype, and other (sub)chromosomal characteristics, have not been well investigated so far though they might have shed light to the problem from another point of view. Previously, the bowfin chromosome number and karyotype were reported merely in two studies, by Ohno et al. ('69) and Suzuki and Hirata ('91), both based on conventionally Giemsa-stained chromosomes.

In this study, we report on bowfin chromosomal characteristics, as investigated by means of conventional (Giemsa staining, C-banding, Chromomycin A_3/DAPI staining, Ag-staining) and molecular (fluorescence *in situ* hybridization with rDNA probes) cytogenetic protocols. We address the "Holostei versus Halecostomi" problem from the cytogenetic point of view based on available data of nonteleost actinopterygians, namely Polypteriformes, Acipenseridae, Polyodontidae, Lepisosteiformes, and teleost representatives. We show that bowfin share more karyological and cytogenetic similarities with teleosts than with gars or other nonteleost actinopterygians.

MATERIALS AND METHODS

Material/Specimen Acquisition

Blood samples of two unsexed juvenile individuals of the bowfin *Amia calva* were collected from the broodstock belonging to the DACBiol, UJAT, Villahermosa, Tabasco, Mexico. All institutional and national guidelines was covered by the "Valid Animal Use Protocols" Nr. CZ 00386 of the Laboratory of Fish genetics, the Institute of Animal Physiology and Genetics issued by the Czech Ministry of Agriculture on September 25, 2014 and by the ethical and research regulations at 2016 from the Universidad Juárez Autónoma de Tabasco, Mexico.

Blood Culturing and Chromosome Preparation

Leucocytes from blood samples of *A. calva* were cultivated, and chromosome spreads were prepared according to the protocol of Fujiwara et al. (2001) with some modifications. Briefly, 0.2–0.5 mL of blood was collected from an anesthetized fish by puncture of the *vena cava caudalis* using a heparinized sterile syringe. The leucocyte-rich plasma was used to set up primary cultures in 5 mL of the Medium 199 (Sigma, St. Louis, MO, USA) supplemented with 10% fetal calf serum (FBS Superior, Biochrom, Berlin, Germany), 1% antibiotic antimycotic solution (Sigma),

Lipopolysaccharide from *E. coli* (0.5 mg per 5 mL of medium), Phytohaemagglutinin-P (90 μg per 5 mL of medium, Remel, Lenexa, KS, USA), kanamycin (0.3 mg per 5 mL of medium (Sigma), and 0.175 μL 10% mercaptoethanol (Sigma). After 120 hr of incubation at 20°C, 5 mL of culture was harvested using standard colchicine (two drops of 0.1% colchicine per 5 mL of media) and 0.075 M KCl hypotonic (8 min) treatments followed by fixation in a freshly prepared fixative (methanol: acetic acid 3:1, v/v) three times.

Karyological Analysis

Chromosomes were initially stained in buffered Giemsa (5%, 10 min) to visualize their morphology. In total, 20 metaphases of each individual have been inspected to confirm the chromosome morphology and composition of the karyotype. For localization of nucleolar organizer regions (NORs; that is transcriptionally active major rDNA unit), we have performed $AgNO_3$ impregnation (Ag-staining) as described by Howell and Black ('80). Chromomycin A_3 (CMA_3, DNA dye specific for GC-rich regions) and DAPI (AT-specific) fluorescent staining was performed to reveal GC genome composition as described by Sola et al. ('92). A C-banding technique visualizing blocks of constitutive heterochromatin was performed according to Sumner ('72) with slight modifications described in Pokorná et al. (2014). The chromosomes after C-banding were counterstained by DAPI to enhance the contrast, and the microphotographs were taken in the fluorescent regime (not in the light microscopy regime as in the Giemsa-stained C-banding) and inverted.

Probes Preparation, FISH

Probes for FISH experiments were produced by PCR with the primers pairs and thermal cycling conditions according Komiya and Takemura ('79) for 5S rDNA and White et al. ('90) for 18S rDNA (primers NS1 and NS4). The PCR reactions were carried out in a final volume of 25 μL consisting of 100 ng genomic DNA, 12.5 μL PPP master mix, 0.01 mM of each primer and PCR water to complete the volume (all reagents from TopBio, Prague, Czech Republic). Cycling conditions were as follows: (a) 18S: 2 min at 95°C; 35 cycles of 1 min at 95°C, 40 sec at 55°C and 2 min at 72°C; 5 min at 72°C; (b) 5S rDNA: 5 min at 94°C; two cycles of 1 min at 95°C, 30 sec at 61°C, and 45 s at 72°C; two cycles of 1 min at 95°C, 30 sec at 59°C and 45 sec at 72°C; two cycles of 1 min at 95°C, 30 sec at 57°C and 45 sec at 72°C; 25 cycles of 1 min at 95°C, 30 sec at 61°C and 45 sec at 72°C; 7 min at 72°C. The amplified fragments were sequenced at the ABI 3700 sequencer prior FISH experiments. Probes were indirectly labeled with biotin-16-dUTP (Roche, Mannheim, Germany) and digoxigenin-11-dUTP (Roche) through PCR reamplification of previously sequenced PCR products. Reamplification was carried out under the same condition as previous PCR reaction. Labeled PCR products were precipitated. A hybridization mixture was made consisting of hybridization buffer (Symonová et al., 2015), sonicated salmon sperm blocking DNA (15 μg/slide; Sigma-Aldrich, St. Louis, MO, USA) and differently labeled PCR products of both genes.

The hybridization and detection procedure were carried out under conditions described by Symonová et al. (2015). The biotin-dUTP-labeled probes were detected by either the Invitrogen Cy™3-Streptavidin (Invitrogen, San Diego, CA, USA; cat. no. 43-4315) or by the FITC-Streptavidin (cat. no. 43-4311). The digoxigenin-dUTP-labeled probes were detected either by the Roche Anti-Digoxogenin-Fluorescein (cat. no. 11207741910) or by the Anti-Digoxigenin-Rhodamin (cat. no. 11207750910). The chromosomes were counterstained with Vectashield/DAPI (1.5 mg/ mL) (Vector, Burlingame, CA, USA). Hybridization lasted for 1 day in a dark humid chamber at 37°C.

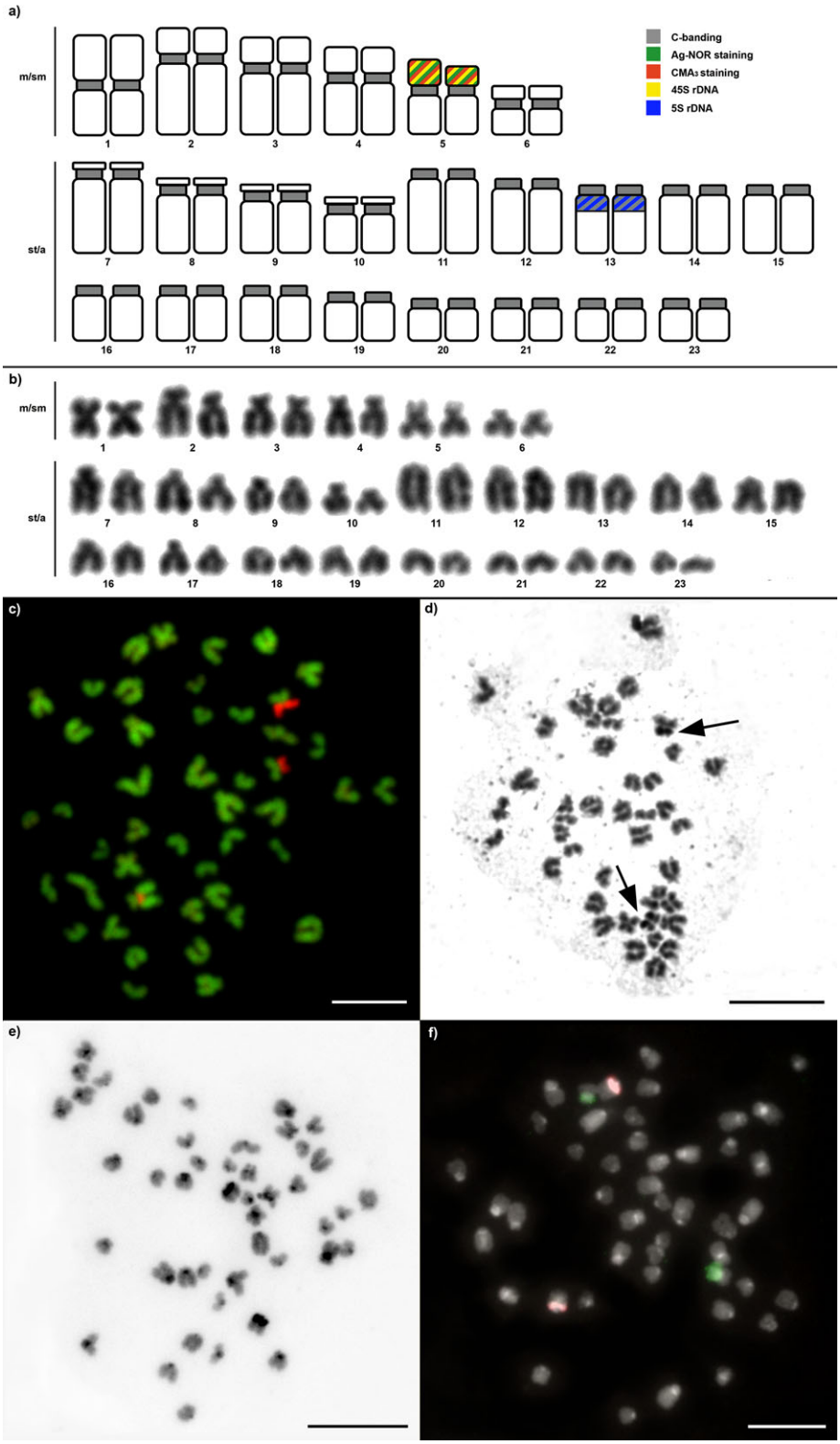
Image Analysis

Chromosomal preparations were examined by an Olympus Provis AX 70 epifluorescence microscope. Images of metaphase chromosomes were recorded with a cooled Olympus DP30BW CCD camera. The IKAROS and ISIS imaging programs (Metasystems, Altlussheim, Germany) were used to analyze grey-scale images. The captured digital images from FISH experiments were pseudocolored (red for Anti-Digoxigenin-Rhodamin, green for Invitrogen FITC-Streptavidin) and superimposed using Microimage and Adobe Photoshop software, version CS5, respectively. In the case of CMA_3/DAPI staining, the CMA_3 signal was inserted into the red and the DAPI signal into the green channel to enhance the contrast between these two types of signals. The chromosomal categories were classified according to Levan et al. ('64). Chromosomes were ordered within each group in decreasing size order.

RESULTS

Karyological Analysis, FISH

The diploid chromosome number in *A. calva* was $2n = 46$, the fundamental number (i.e., number of chromosomal arms determined by scoring metacentrics-submetacentrics as biarmed and subtelocentrics-acrocentrics as uniarmed) was NF = 58. Results of karyological analyses together with results of FISH experiments performed on *A. calva* chromosomes are summarized in the ideogram (Fig. 1a). Karyotype was composed of one pair of meta-, five pairs of submeta-, and four pairs of subtelo-, and 13 pairs of acrocentric chromosomes (Figs. 1a and b). The submetacentric pair no. 5 (Figs. 1a and b) possessed a polymorphic secondary constriction in the terminal short *p* arm causing differences in the size of homologs. The CMA_3/DAPI fluorescence demonstrated homogeneously stained chromosomes, that is a balanced proportion of AT/GC up to the extremely GC-rich that is CMA_3 positive NOR sites on the chromosome pair no. 5 (Figs. 1a and c). This position was also an Ag-NOR site (Figs. 1a and d). C-banded chromosomes with visualized


a)
m/sm
st/a
C-banding
Ag-NOR staining
CMA3 staining
45S rDNA
5S rDNA
1 2 3 4 5 6
7 8 9 10 11 12 13 14 15
16 17 18 19 20 21 22 23
b)
m/sm
st/a
1 2 3 4 5 6
7 8 9 10 11 12 13 14 15
16 17 18 19 20 21 22 23
c)
d)
e)
f)

constitutive heterochromatin showed signals in centromeres of all chromosomes; in addition, strong signals, corresponding to 5S rDNA sites, have been detected on chromosome pair no. 13 (Figs. 1a and e).

The FISH experiments with 5S and 45S rDNA probes showed a nonsyntenic pattern, that is signals on distinct chromosomal pairs have been detected in both individuals (Figs. 1a and f). Although the 45S rDNA site (28S rDNA, 5.8S rDNA, and 18S rDNA) was located at the telomeric region on short (*p*) arms of the submetacentric pair no. 5, the 5S rDNA sites were situated at the centromeric regions of the acrocentric pair no. 13 (Figs. 1a and f).

DISCUSSION

Bowfin Karyotype

Our results confirmed the $2n = 46$ in bowfin, as reported by previous studies (Ohno et al., '69; Suzuki and Hirata, '91). Ohno et al. ('69) described its karyotype with 10 pairs of meta- and 13 pairs of acrocentric chromosomes. These authors used squash tissue protocol for chromosome preparations, and our inspection of published karyotype documents theirs low quality. Suzuki and Hirata ('91), using direct kidney preparation protocol, reported its karyotype with one pair of meta-, 12 pairs of submeta-subtelocentric, and 10 pairs of acrocentric chromosomes. Both studies examined bowfin karyotype using Giemsa-stained chromosomes. However, Suzuki and Hirata ('91) improved their study with C-banding technique to document the distribution of constitutive heterochromatin. Their study revealed a middle-sized acrocentric chromosome pair bearing secondary constriction in *p* arms and also hypothesized that only metacentric pair in bowfin karyotype originated via centric fusion decreasing thus 2n from 48 to 46.

Our interpretation of the bowfin karyotype here based on leucocyte culture confirmed the presence of one distinct pair of metacentric chromosomes and distinguished in more detail five pairs of submeta-, four pairs of subtelo-, and 14 pairs of subtelo- to acrocentric chromosomes with some chromosomes, however, at borderline between chromosome categories. We further identified a pair of middle-sized acrocentric chromosome pair with secondary constriction in *p* arms as the major rDNA-bearing (45S rDNA), CMA_3-positive and Ag-positive element. Since bowfin is the only extant survivor of one of the major actinopterygian lineages, it is impossible to compare cytogenetic results with other species of amiiform fishes. Instead, we can discuss possible cytotaxonomic interrelationships of bowfin with other major actinopterygian groups based on available data.

Cytotaxonomic Comparison of Bowfin with Major Lineages of Actinopterygii

Actinopterygii includes five major lineages, four of which exhibit long-retained ecomorphologies alongside low species counts and are thus referred to as "living fossils" (Amiiformes, Polypteriformes, Acipenseriformes, Lepisosteiformes) (Sallan 2014). The fifth lineage, the teleosts, represents the vast majority of extant ray-finned fishes and about half of all living vertebrates (Nelson, 2006; Sallan, 2014). Summary of genome and chromosome organization characteristics of major clades of Actinopterygii is given in Table 1.

In comparison with other actinopterygian lineages, Polypteriformes and Acipenseriformes significantly differ in genomes, karyotypes, and other chromosomal characteristics. These differences are ascribed to separate evolutionary histories of these lineages (Sallan, 2014). Polypterids are among the clades of ray-finned fishes that exhibit a significantly low rate of lineage diversification (Near et al., 2014). Recent phylogenies shows they diverged as the first branch of the actinopterygian stem at least 390 MYA and hence are placed as sister group to all other lineages (Near et al., 2012; Betancur-R et al., 2013; Broughton et al., 2013). Polypterids exhibits the lowest chromosomal counts among other nonteleost actinopterygians and at the same time huge chromosome sizes. All analyzed bichirs and ropefish (Morescalchi et al., 2011 and references therein) possess $2n = 36$, except *P. weeksii* where $2n = 38$ was likely caused by a centric fission of one metacentric chromosome and *P. dolloi* with $2n = 68$, an apparent case of autopolyploidy without noticeable morphological effects (Vervoort, '80). Except the last mentioned, their karyotypes composed exclusively by biarmed and extremely large chromosomes. The C-value of the Polypteriformes (up to 7.25 pg; Table 1.; Gregory, 2015).

Figure 1. Cytogenetics of *Amia calva*. (a) Schematic representations of chromosomes after cytogenetic experiments: C-banded chromosomal regions are marked by grey color; active NORs visualized by Ag-NOR staining are marked by green color; CMA3-stained regions are marked by red color; rDNA signals are marked by yellow (45S rDNA) and blue (5S rDNA). Legend: m/sm, metacentric-submetacentric; st/a, subtelocentric-acrocentric chromosomes. (b) Representative karyotype arranged from Giemsa-stained chromosomes. (c) CMA_3/DAPI-stained chromosomes. The CMA_3 signal was inserted from green into red and the DAPI signal from blue into the green channel to enhance the contrast between these signals. (d) Ag-NOR-stained chromosomes; (e) C-banded chromosomes. The C-banding pattern was counterstained with DAPI and inverted; (f) Metaphase chromosomes after FISH with 45S rDNA (green signals) and 5S rDNA (red signals) probes. The DAPI-counterstaining signal was reverted from blue to grey to enhance the contrast. Particular images were adjusted separately before assembling the plate. Bars equal 5 μm. [Color figure can be viewed at wileyonlinelibrary.com]

Table 1. Summary of traits and factors involved in genome and chromosomes organisation of Actinopterygia

Group	2*n* chromosome counts (basic features)	Micro-chromosomes	CG hetero-geneity	Whole genome duplication after R1 and R2	C-value/haploid DNA content (pg) www.genomesize.com	Specific features in the genome history and chromosomal evolution
Acipenseriformes (sturgeons, paddlefish)	~ 120–240–360	YES ~ 50%	Only NOR	Multiple in sturgeons, one in paddlefish	Acipenser 1.8–9.3 Polyodon 1.6–2.4	Multiple WGD, ploidy diversity
Lepisosteiformes (gars)	56–58	Small sized chromosomes	In both extant genera	Not observed	Atractosteus 1.2 Lepisosteus 1.4	Regionally high recombination rate (see Pessia et al., 2012)
Amiiformes (bowfin)	46	No	Only NOR	Not observed	Amia calva 1.2	Convergent genome evolution with teleosts?
Polypteriformes (bichirs)	36–38 biarmed, extremely large	NO	only NOR	not observed	Erpetoichthis 4.5 Polypterus 3.6–7.2	Not investigated
Teleostei	~ 50 (exceptionally up to 100–150)	Micro B-chromosomes	Only NOR	TGD and lineage-specific WGDs	0.4 to ~ 1.0 for exceptions see*	From genome compaction to lineage-specific WGD

*Lineages with undergone WGD: Salmoniformes, Cypriniformes, Synbranchiformes, Siluriformes (Soltis and Soltis, 2012)

Acipenseriformes represent one of the oldest crown group lineages within ray-finned fishes, dating back to at least the early Jurassic (some 200–175 MYA; Sallan, 2014). This group, together with gars and bowfin, diverged from teleosts before the teleost-specific whole genome duplication (TGD) (Volff, 2005; Froschauer et al., 2006; Amores et al., 2011). Molecular phylogenies of Acipenseriformes (Ludwig et al., 2001; Peng et al., 2007; Fontana, 2016; Havelka et al., 2016), and recently also genomic studies (Crow et al., 2012), have documented several independent whole genome duplication (WGD) events. Unlike all other actinopterygian genomes, acipenseriforms have karyotypes composed of macrochromosomes and numerous dot-like microchromosomes of gradually decreasing size (Table 1). Acipenseriformes represent lineage with large genome sizes which vary according to ploidy level from 1.9 to 9.3 pg in *Acipenser* and from 1.6 to 2,4 pg in *Polyodon* (Gregory, 2015).

The evolutionary history of the remaining ray-finned lineages, that is bowfin, gars, and teleosts, is more complicated, causing the "gar–bowfin–teleost" problem. It is important to note that extant species numbers of these lineages do not reflect their past diversity, and all molecular phylogenetic studies are based only on the subset of living species. All three lineages diverged in the Triassic and were widely distributed during the Mesozoic era (Cavin, 2010; Grande, 2010; López-Arbarello, 2012), whereas teleosts underwent their WGD during the first 100 million years of the Mesozoic era and have subsequently speciated at great rates; the diversity of Amiids and Lespisosteids has been lost over time (Grande, 2010). Nowadays, teleosts represent vast majority of living ray-finned fishes with more than 30,000 extant species, that is 96% of all extant fishes (Nelson, 2006), with many new species being described each year (Eschmeyer and Fricke, 2015). However, extant gars are represented by two genera, *Atractosteus* and *Lepisosteus*, and seven species (Nelson, 2006), and bowfin is the only survivors of Amiids (Cavin, 2010; Grande, 2010). Hence, a detailed cytogenetic analysis of gars–bowfin–teleost lineages is a relevant and desirable approach to address the Holostei–Halecostomi issue.

The bowfin $2n = 46$ is comparable to the hypothesized ancestral teleost protokaryotype after the TGD ($2n = 48$) (Jaillon et al., 2004; Kohn et al., 2006; Kasahara et al., 2007; but see also Nakatani et al., 2007) which is retained in approximately 50% of examined teleost taxa (Mank and Avise, 2006; Arai, 2011). However, the karyotype is known in less than 12% of existing teleost diversity (Arai, 2011) and the available data document enormous genome diversity with chromosome numbers spanning from $2n = 12$ in deepwater Spark anglemouth *Stigmops* (*Gonostoma*) *bathyphylum* (Gonostomatidae, Stomiiformes) (Post, '74) to $2n = \sim 446$ in schizothoracine species *Ptychobarbus dipogon,*

(Cyprinidae, Cypriniformes) (Yu and Yu, '90). Lepisosteiformes represent higher chromosomal counts than bowfin and teleosts. The tropical gar ($2n = 56$) and the spotted gar ($2n = 58$) demonstrates difference in $2n$ by a fusion/fission event of one chromosome pair Symonová et al. (in this issue). Since gars did not undergo the TGD (Amores et al., 2011), their higher chromosomal counts related to teleost were explained by chromosome fusions in the lineage leading to teleosts after divergence from gar but before the TGD (Braasch et al., 2016). Although genome size of bowfin, gars, and teleosts are similar (Table 1), it is not a suitable marker to resolve their phylogenetic relations because it can be affected, for example, by the presence of repetitive elements or regional duplications. Comparable C-value would suggest a possible independent WGD in Amiids, but this option has been already excluded, based on gene content analyses (e.g., Crow et al., 2006; Mulley et al., 2006; Hurley et al., 2007; Braasch et al., 2016; Pasquier et al., 2016). Another possibility is that chromosome fusions occurred in the lineage leading to bowfin, but additional analyses, for example, bowfin chromosome-level genome assembly, will be necessary to test this hypothesis.

Regarding morphology of chromosomes, bowfin and teleosts share the same type of karyotypes whereas the gar karyotype consists of macrochromosomes together with very small chromosomes. Considering the presence of very small chromosomes and conserved synteny in chromosome content between gar and chicken as an ancestral vertebrate condition, their absence in bowfin suggests that its karyotype is more derived compared to the gar lineage (Braasch et al., 2016). Moreover, Symonová et al. (in this issue) demonstrated that gars and bowfin differ in intrachromosomal organization. The application of cytogenetic approaches and bioinformatics analyses revealed genome AT/GC heterogeneity in spotted gar genome (Symonová et al., in this issue). On the other hand, both bowfin and teleosts demonstrate an overall balanced proportion of AT/GC with the exception of extremely GC-rich NOR sites.

The teleost-resembling pattern of cytogenetic characteristics in bowfin may be explained by an example of homoplasy, that is convergence between teleosts and *Amia*. Other option, that is a sister group relationship between *Amia* and teleosts, that is the monophyletic lineage Halecostomi as proposed from some morphological analyses and potentially obscured by TGD in teleosts is in strong contrast with the molecular consensus now coalescing around the nested Holostei–Teleostei topology (Sallan, 2014; Braasch et al., 2016).

CONCLUSION

In conclusion, our study partly confirmed cytogenetic results of previous analyses and extended the description of bowfin chromosomes using available cytogenetic protocols. Our cytogenetic data clearly document that chromosomal evolution of basal Actinopterygii led to different genome organization, genome size, and chromosome number among extant lineages. Despite the fact that recent phylogenomic analyses support the monophyletic relationship of the spotted gar and bowfin, cytogenetically the bowfin strongly resembles the teleost pattern of genome organization, a phenomenon also observed for several morphological characteristics.

ACKNOWLEDGMENTS

The authors would like to express their gratitude to Šárka Pelikánová, Petra Šejnohová, and Jana Čechová of the Laboratory of Fish Genetics for their assistance in the laboratory.

AUTHOR CONTRIBUTIONS

P.R. designed the study and co-drafted manuscript. R.S. and Z.M. designed the study, performed cytogenetic analyses, and co-drafted the manuscript. L.S. provided the phylogenetic and paleobiological context and codrafted the manuscript together with L.A.R. L.A.R. participated on cytogenetic analyses and provided chromosome preparations. All coauthors revised the text and agreed the final version.

LITERATURE CITED

Amores A, Catchen J, Ferrara A, Fontenot Q, Postlethwait JH. 2011. Genome evolution and meiotic maps by massively parallel DNA sequencing: spotted gar, an outgroup for the teleost genome duplication. Genetics 188:799–808.

Arai R. 2011. Fish karyotypes: a check list. Tokyo, Japan: Springer.

Betancur-R R, Broughton RE, Wiley EO, Carpenter K, López JA, Li C, Holcroft NI, Arcila D, Sanciangco M, Cureton Ii JC, Zhang F, Buser T, Campbell MA, Ballesteros JA, Roa-Varon A, Willis S, Borden WC, Rowley T, Reneau PC, Hough DJ, Lu G, Grande T, Arratia G, Ortí G. 2013. The tree of life and a new classification of bony fishes. PLoS Curr 5. ecurrents.tol.53ba26640df0ccaee75bb165c8c26288.

Braasch I, Gehrke AR, Smith JJ, Kawasaki K, Manousaki T, Pasquier J, Amores A, Desvignes T, Batzel P, Catchen J, Berlin AM, Campbell MS, Barrell D, Martin KJ, Mulley JF, Ravi V, Lee AP, Nakamura T, Chalopin D, Fan S, Wcisel D, Cañestro C, Sydes J, Beaudry FEG, Sun Y, Hertel J, Beam MJ, Fasold M, Ishiyama M, Johnson J, Kehr S, Lara M, Letaw JH, Litman GW, Litman RT, Mikami M, Ota T, Saha NR, Williams L, Stadler PF, Wang H, Taylor JS, Fontenot Q, Ferrara A, Searle SMJ, Aken B, Yandell M, Schneider I, Yoder JA, Volff J-N, Meyer A, Amemiya CT, Venkatesh B, Holland PWH, Guiguen Y, Bobe J, Shubin NH, Di Palma F, Alföldi J, Lindblad-Toh K, Postlethwait JH. 2016. The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons. Nat Genet 48:427–437.

Broughton RE, Betancur-R R, Li C, Arratia G, Ortí G. 2013. Multi-locus phylogenetic analysis reveals the pattern and tempo of bony fish evolution. PLoS Curr 5. doi:10.1371/currents.tol.2ca8041495ffafd0c92756e75247483e.

Cavin L. 2010. Diversity of Mesozoic semionotiform fishes and the origin of gars (Lepisosteidae). Naturwissenschaften 97:1035–1040.

Crow KD, Stadler PF, Lynch VJ, Amemiya C, Wagner GP. 2006. The "fish-specific" Hox cluster duplication is coincident with the origin of teleosts. Mol Biol Evol 23:121–136.

Crow KD, Smith CD, Cheng J-F, Wagner GP, Amemiya CT. 2012. An independent genome duplication inferred from Hox paralogs in the American paddlefish - a representative basal ray-finned fish and important comparative reference. Genome Biol Evol 4:937–953.

Eschmeyer WN, Fricke R. 2015. Catalog of fishes: Genera, species, references. Zootaxa 3882:1.

Faircloth BC, Sorenson L, Santini F, Alfaro ME. 2013. A phylogenomic perspective on the radiation of ray-finned fishes based upon targeted sequencing of ultraconserved elements (UCEs). PLoS One 8:e65923.

Fontana F. 2016. Geneweb. Available from http://sveb.unife.it/it/ricerca-1/laboratori/geneweb

Froschauer A, Braasch I, Volff J-N. 2006. Fish genomes, comparative genomics and vertebrate evolution. Curr Genomics 7:43–57.

Fujiwara A, Nishida-Umehara C, Sakamoto T, Okamoto N, Nakayama I, Abe S. 2001. Improved fish lymphocyte culture for chromosome preparation. Genetica 111:77–89.

Grande L. 2010. An empirical synthetic pattern study of gars (Lepisosteiformes) and closely related species, based mostly on skeletal anatomy the resurrection of Holostei. Lawrence, KS: American Society of Ichthyologists and Herpetologists.

Grande L, Bemis WE. 1998. A comprehensive phylogenetic study of Amiid fishes (Amiidae) based on comparative skeletal anatomy. an empirical search for interconnected patterns of natural history. J Vertebr Paleontol 18:1–696.

Gregory TR. 2015. Animal genome size database. Available from http://www.genomesize.com

Havelka M, Bytyutskyy D, Symonová R, Ráb P, Flajšhans M. 2016. The second highest chromosome count among vertebrates is observed in cultured sturgeon and is associated with genome plasticity. Genet Sel Evol 48:12.

Howell WM, Black DA. 1980. Controlled silver-staining of nucleolus organizer regions with a protective colloidal developer: a 1-step method. Experientia 36:1014–1015.

Hurley IA, Mueller RL, Dunn KA, Schmidt EJ, Friedman M, Ho RK, Prince VE, Yang Z, Thomas MG, Coates MI. 2007. A new time-scale for ray-finned fish evolution. Proc R Soc Lond B Biol Sci 274:489–498.

Inoue JG, Miya M, Tsukamoto K, Nishida M. 2003. Basal actinopterygian relationships: a mitogenomic perspective on the phylogeny of the "ancient fish." Mol Phylogenet Evol 26:110–120.

Jaillon O, Aury J-M, Brunet F, Petit J-L, Stange-Thomann N, Mauceli E, Bouneau L, Fischer C, Ozouf-Costaz C, Bernot A, Nicaud S, Jaffe D, Fisher S, Lutfalla G, Dossat C, Segurens B, Dasilva C, Salanoubat M, Levy M, Boudet N, Castellano S, Anthouard V, Jubin C, Castelli V, Katinka M, Vacherie B, Biémont C, Skalli Z, Cattolico L, Poulain J, de Berardinis V, Cruaud C, Duprat S, Brottier P, Coutanceau J-P, Gouzy J, Parra G, Lardier G, Chapple C, McKernan KJ, McEwan P, Bosak S, Kellis M, Volff J-N, Guigó R, Zody MC, Mesirov J, Lindblad-Toh K, Birren B, Nusbaum C, Kahn D, Robinson-Rechavi M, Laudet V, Schachter V, Quétier F, Saurin W, Scarpelli C, Wincker P, Lander ES, Weissenbach J, Roest Crollius H. 2004. Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. Nature 431:946–957.

Kasahara M, Naruse K, Sasaki S, Nakatani Y, Qu W, Ahsan B, Yamada T, Nagayasu Y, Doi K, Kasai Y, Jindo T, Kobayashi D, Shimada A, Toyoda A, Kuroki Y, Fujiyama A, Sasaki T, Shimizu A, Asakawa S, Shimizu N, Hashimoto S-I, Yang J, Lee Y, Matsushima K, Sugano S, Sakaizumi M, Narita T, Ohishi K, Haga S, Ohta F, Nomoto H, Nogata K, Morishita T, Endo T, Shin-I T, Takeda H, Morishita S, Kohara Y. 2007. The medaka draft genome and insights into vertebrate genome evolution. Nature 447:714–719.

Kohn M, Högel J, Vogel W, Minich P, Kehrer-Sawatzki H, Graves JAM, Hameister H. 2006. Reconstruction of a 450-My-old ancestral vertebrate protokaryotype. Trends Genet TIG 22:203–210.

Komiya H, Takemura S. 1979. Nucleotide sequence of 5S ribosomal RNA from rainbow trout (*Salmo gairdnerii*) liver. J Biochem (Tokyo) 86:1067–1080.

Levan A, Fredga K, Sandberg AA. 1964. Nomenclature for centromeric position on chromosomes. Hereditas 52:201–220.

López-Arbarello A. 2012. Phylogenetic interrelationships of ginglymodian fishes (Actinopterygii: Neopterygii). PLoS One 7:e39370.

Ludwig A, Belfiore NM, Pitra C, Svirsky V, Jenneckens I. 2001. Genome duplication events and functional reduction of ploidy levels in sturgeon (*Acipenser*, *Huso* and *Scaphirhynchus*). Genetics 158:1203–1215.

Mank JE, Avise JC. 2006. Phylogenetic conservation of chromosome numbers in Actinopterygiian fishes. Genetica 127:321–327.

Morescalchi MA, Stingo V, Capriglione T. 2011. Cytogenetic analysis in *Polypterus ornatipinnis* (Actinopterygii, Cladistia, Polypteridae) and 5S rDNA. Mar Genomics 4:25–31.

Mulley JF, Chiu C, Holland PWH. 2006. Breakup of a homeobox cluster after genome duplication in teleosts. Proc Natl Acad Sci 103:10369–10372.

Nakatani Y, Takeda H, Kohara Y, Morishita S. 2007. Reconstruction of the vertebrate ancestral genome reveals dynamic genome reorganization in early vertebrates. Genome Res 17:1254–1265.

Near TJ, Eytan RI, Dornburg A, Kuhn KL, Moore JA, Davis MP, Wainwright PC, Friedman M, Smith WL. 2012. Resolution of ray-finned fish phylogeny and timing of diversification. Proc Natl Acad Sci USA 109:13698–13703.

Near TJ, Dornburg A, Tokita M, Suzuki D, Brandley MC, Friedman M. 2014. Boom and bust: ancient and recent diversification in bichirs (Polypteridae: Actinopterygii), a relictual lineage of ray-finned fishes. Evolution 68:1014–1026.

Nelson JS. 2006. Fishes of the world, 4th edition. John Wiley & Sons, Hoboken, New Jersey, 601 p.

Ohno S, Muramoto J, Stenius C, Christian L, Kittrell WA, Atkin NB. 1969. Microchromosomes in holocephalian, chondrostean and holostean fishes. Chromosoma 26:35–40.

Pasquier J, Cabau C, Nguyen T, Jouanno E, Severac D, Braasch I, Journot L, Pontarotti P, Klopp C, Postlethwait JH, Guiguen Y, Bobe J. 2016. Gene evolution and gene expression after whole genome duplication in fish: the PhyloFish database. BMC Genomics 17:368.

Peng Z, Ludwig A, Wang D, Diogo R, Wei Q, He S. 2007. Age and biogeography of major clades in sturgeons and paddlefishes (Pisces: Acipenseriformes). Mol Phylogenet Evol 42:854–862.

Pessia E, Popa A, Mousset S, Rezvoy C, Duret L, Marais GAB. 2012. Evidence for Widespread GC-biased Gene Conversion in Eukaryotes. Genome Biol Evol 4:787–794.

Pokorná M, Rens W, Rovatsos M, Kratochvíl L. 2014. A ZZ/ZW sex chromosome system in the thick-tailed Gecko (*Underwoodisaurus milii*; Squamata: Gekkota: Carphodactylidae), a member of the ancient gecko lineage. Cytogenet Genome Res 142:190–196.

Post A. 1974. Chromosomes of tree species of family Gonostomatidae (Osteichthyes, Stomiatoidei). In: Results of research cruises of FRV "Walther Hertwig" to South-America. XXXIV. vol. 25. p 51–55.

Sallan LC. 2014. Major issues in the origins of ray-finned fish (Actinopterygii) biodiversity. Biol Rev 89:950–971.

Šlechtová V, Bohlen J, Tan HH. 2007. Families of Cobitoidea (Teleostei; Cypriniformes) as revealed from nuclear genetic data and the position of the mysterious genera *Barbucca*, *Psilorhynchus*, *Serpenticobitis* and *Vaillantella*. Mol Phylogenet Evol 44:1358–1365.

Sola L, Rossi AR, Iaselli V, Rasch EM, Monaco PJ. 1992. Cytogenetics of bisexual/unisexual species of Poecilia. II. Analysis of heterochromatin and nucleolar organizer regions in *Poecilia mexicana mexicana* by C-banding and DAPI, quinacrine, chromomycin A3, and silver staining. Cytogenet Cell Genet 60:229–235.

Soltis P, Soltis DE. 2012. Polyploidy and genome evolution. Heidelberg, Germany: Springer Science & Business Media.

Sumner AT. 1972. A simple technique for demonstrating centromeric heterochromatin. Exp Cell Res 75:304–306.

Suzuki A, Hirata J. 1991. Chromosomes and DNA contents of *Amia calva*. Chromosome Inf Serv 50:34–37.

Symonová R, Sember A, Majtánová Z, Ráb P. 2015. Characterization of fish genomes by GISH and CGH. In: Ozouf-Costaz C, Pisano E, Foresti F, de Almeida L, editors. Fish cytogenetic techniques. Boca Raton, FL: CRC Press. p 118–131. Available from http://www.crcnetbase.com/doi/10.1201/b18534-17

Vervoort A. 1980. Karyotypes and nuclear-DNA contents of Polypteridae (Osteichthyes). Experientia 36:646–647.

Volff JN. 2005. Genome evolution and biodiversity in teleost fish. Heredity 94:280–294.

White TJ, Bruns T, Lee S, Taylor J. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis M, Gelfand J, Sninsky J, White T, editors. PCR protocols: A guide to methods and applications. Orlando, FL: Academic Press. p 315–322.

Yu X, Yu X. 1990. A schizothoracine fish species *Diptychus dipogon* with a very high number of chromosomes. Chromosome Inf Serv 48:17–18.

Symonová et al. BMC Genomics (2017) 18:391
DOI 10.1186/s12864-017-3774-7

BMC Genomics

RESEARCH ARTICLE Open Access

Higher-order organisation of extremely amplified, potentially functional and massively methylated 5S rDNA in European pikes (*Esox* sp.)

Radka Symonová[1*], Konrad Ocalewicz[2,3], Lech Kirtiklis[4], Giovanni Battista Delmastro[5], Šárka Pelikánová[6], Sonia Garcia[7] and Aleš Kovařík[8*]

Abstract

Background: Pikes represent an important genus (*Esox*) harbouring a pre-duplication karyotype (2n = 2x = 50) of economically important salmonid pseudopolyploids. Here, we have characterized the 5S ribosomal RNA genes (rDNA) in *Esox lucius* and its closely related *E. cisalpinus* using cytogenetic, molecular and genomic approaches. Intragenomic homogeneity and copy number estimation was carried out using Illumina reads. The higher-order structure of rDNA arrays was investigated by the analysis of long PacBio reads. Position of loci on chromosomes was determined by FISH. DNA methylation was analysed by methylation-sensitive restriction enzymes.

Results: The 5S rDNA loci occupy exclusively (peri)centromeric regions on 30–38 acrocentric chromosomes in both *E. lucius* and *E. cisalpinus*. The large number of loci is accompanied by extreme amplification of genes (>20,000 copies), which is to the best of our knowledge one of the highest copy number of rRNA genes in animals ever reported. Conserved secondary structures of predicted 5S rRNAs indicate that most of the amplified genes are potentially functional. Only few SNPs were found in genic regions indicating their high homogeneity while intergenic spacers were more heterogeneous and several families were identified. Analysis of 10–30 kb-long molecules sequenced by the PacBio technology (containing about 40% of total 5S rDNA) revealed that the vast majority (96%) of genes are organised in large several kilobase-long blocks. Dispersed genes or short tandems were less common (4%). The adjacent 5S blocks were directly linked, separated by intervening DNA and even inverted. The 5S units differing in the intergenic spacers formed both homogeneous and heterogeneous (mixed) blocks indicating variable degree of homogenisation between the loci. Both *E. lucius* and *E. cisalpinus* 5S rDNA was heavily methylated at CG dinucleotides.

Conclusions: Extreme amplification of 5S rRNA genes in the *Esox* genome occurred in the absence of significant pseudogenisation suggesting its recent origin and/or intensive homogenisation processes. The dense methylation of units indicates that powerful epigenetic mechanisms have evolved in this group of fish to silence amplified genes. We discuss how the higher-order repeat structures impact on homogenisation of 5S rDNA in the genome.

Keywords: rDNA, Evolution, Chromosome, Fish, *Esox*, Single cell PacBio sequencing

* Correspondence: Radka.Symonova@uibk.ac.at; kovarik@ibp.cz
[1]Research Institute for Limnology, University of Innsbruck, Mondseestrasse 9, A-5310 Mondsee, Austria
[8]Laboratory of Molecular Epigenetics, Institute of Biophysics, Czech Academy of Science, Kralovopolska 135, 61265 Brno, Czech Republic
Full list of author information is available at the end of the article

Background

Esox, the only genus in the family Esocidae (Esociformes) contains seven pike species in two monophyletic subgenera (*Esox* and *Kenoza*) with a circumpolar distribution [1]. *Esox americanus*, *E. masquinongy* and *E. niger* live naturally in North America, *E. reicherti* is the only Euroasian esocid endemic to the Amur River basin (Russia and China) while the *E. cisalpinus* (*E. flaviae*) and *E. aquitanicus* are native to Europe [2]. The Northern pike (*E. lucius*) occurs in North America and Eurasia. Its wide distribution and easy access make the Northern pike the most studied esocid species in terms of behaviour, ecology, genetics and evolution. The Northern pike inhabits lakes, rivers and brackish waters. It is an important commercial and recreational species. In the recent years, overexploitation of the natural stocks and climate change has resulted in the dramatic decline of some pike populations [3].

In the Northern pike, nuclear and mitochondrial DNAs exhibit relatively low genetic variability while there is a considerable genetic differentiation among populations [4, 5]. This may be explained by its ecological status–the top predator population size is related to the prey density and/or the bottlenecks that accompanied post-glacial expansion of the pike. Phylogenetic studies confirmed esocids as the closest relatives of the autotetraploid ancestor of salmonid fishes (trout, chars, salmons, ciscoes and grayling) [1, 6]. Both genome size and the number of chromosome arms are about doubled in salmonids when compared to the pike [7, 8]. The Northern pike genome sequence has been recently published [9] and its linkage groups were successfully mapped on the salmonid reference genomes revealing the importance of *Esox* as the pre-duplication outgroup of salmonids [10]. The genus *Esox* possesses 50 acrocentric chromosomes and the number of chromosomal arms (FN) equalled also 50 [11]. The *Esox* karyotype is thus similar to the presumed karyotype of the diploid common ancestor of salmonids. Hypothetically, the salmonid ancestral karyotype after the salmonid-specific whole-genome duplication (WGD) was composed of 100 uni-armed chromosomes (FN = 100). Subsequent diploidization process included vast chromosome rearrangements that have resulted in formation of different karyotypes in the extant salmonids [7, 12]. Centric fusions retaining the number of chromosome arms and decreasing number of chromosome resulted in the formation of Coregoninae and Salmoninae karyotypes. The Thymallinae karyotypes experienced mostly inversions that increased the number of chromosome arms and retained chromosome number close to the presumed ancestral tetraploid karyotype [12]. Therefore, a deep knowledge concerning the *Esox* karyotype and genome organisation is of much importance to reconstruct and understand the complex evolution of salmonid lineages. However, although salmonids are one of the best cytogenetically and molecularly studied fish lineages [13–15], published data on *Esox* sp. chromosomes are limited to basic information concerning chromosome number, their morphology and location of the Nucleolus Organizer Regions (NORs) in *E. lucius*, *E. niger*, *E. masquinongy* and *E. americanus* [7, 11].

The 5S rDNA unit consists of a conserved 120 bp genic region and a more variable intergenic spacer (IGS), also known as Non-Transcribed Spacer (NTS). The genic region contains a tripartite RNA Polymerase III promoter composed of three motifs, Box-A, Internal Element (IE) and Box-C [16], which appear to be conserved throughout the tree of life [17, 18]. Thus the 5S rRNA gene appears to be under extreme selection constrains that maintain structural rRNA features essential for ribosome function and conserved transcription binding sites. The presence of RNA Polymerase III promoter within the short genic region makes the 5S gene a relatively autonomous element prone to change its location on chromosomes. Based on the available data, fish harbour an extraordinary capacity to amplify and spread their rDNA sites across chromosomes - according to the currently assembled animal rDNA database (comprising >900 species of vertebrate, arthropods and molluscs), there are eight fish genera among the top ten species with the largest number of 5S rDNA loci (http://www.animalrdnadatabase.com). In fish, up to 30–40 sites were evidenced in diverged lineages – Cypriniformes, Siluriformes, Characiformes, Salmoniformes, and Esociformes [14, 19–22].

Chromosomal mapping and sequence analysis of 5S rDNA give a powerful tool in genome evolution studies. Two basic models have been proposed for the evolution of rRNA genes (reviewed in [23, 24]). The first model, "concerted evolution", is based on the observation that rDNA units are uniform within the genome while they differ across the genomes [25, 26]. Under this model, a mutation is rapidly spread across the arrays (genome) or is lost and as a result intragenomic homogeneity of genic and non-coding regions is similar. An alternative, the "birth and death" model has been proposed for multigenic families [27]. According to this model, new genes originate by successive duplications, and are either maintained for a long time or are lost, or else degenerate into pseudogenes. Sequence diversity of coding and spacer regions have been taken as criteria to distinguish between both models and a mixed model of these two has also been proposed [28]. Intragenomic heterogeneity of rDNA paralogs has been correlated with the number of loci in the genome [29]. The known high rDNA mobility across chromosomes [30] has been usually ascribed to various translocation [31] and transposition events [19, 32] as well as to polyploidy and interspecies hybridisation [33]. The latter seems to be the most plausible explanation of the heterogeneity because parental genomes may donate divergent paralogs to their

hybrid nucleus. In fish hybrids, the intra- and intergenomic variation was exploited in numerous phylogenetic studies [28, 34].

Despite the wealth of knowledge about 5S rDNA sequence and its chromosome position little is known about the higher-order repeat organisation, in which a block of multiple basic repeat units forms a larger repeat unit of repeats. Here, we exploited PacBio genomic resources now available [9] to determine the higher-order organisation of 5S rRNA genes that to our best knowledge has not been examined to date in any system. We addressed the question of structure, sequence homogeneity and evolution of 5S rRNA genes in two related species *E. lucius* and *E. cisalpinus*. We applied an integrative approach involving classical cytogenetics, molecular biology and *in silico* genomics methods.

Methods

Species and sample collections

Following specimens of the Northern pike (*Esox lucius*) were cytogenetically studied: ten young unsexed specimens purchased in a fish farm in Libechov, Czech Republic and 22 8-month old specimens (12 males, nine females and one unsexed) from a fish farm in Olsztyn, Poland. Of the only recently described Southern Pike (*Esox cisalpinus*) we analysed cytogenetically: fourteen young unsexed specimens from the Provincial Fish hatchery of Carmagnola, Italy (young specimens obtained from wild pike parents collected in the Turin Province, determined by G. B. Delmastro and E. Sala), a single specimen from the Po river, Carmagnola, locality Gerbasso, determined by G. B. Delmastro and finally, two specimens determined by G. B. Delmastro that are deposited in the ichthyologic collection of the Muséum National d'Histoire Naturelle (Paris, France).

Molecular cytogenetics

Molecular cytogenetic methods were carried out independently in laboratories in Poland and Czech Republic. The karyotypes were assessed following standard procedures [14, 35]. Briefly, fish were injected with 0.3% colchicine (Sigma-Aldrich, St. Louis, MO, USA) solution (0.25 ml/100 g body weight) 60 min. before being sacrificed. Portions of cephalic kidneys were removed, placed in 5 ml of a hypotonic solution (0.075 M KCl) and homogenised using glass homogenizers. Cell suspensions were then transferred to the 10 ml glass tubes, hypotonised for 40 min at RT and fixed with freshly prepared ice cold fixative (methanol: acetic acid, 3: 1, v/v). The fixative was changed three times before splashing on microscopic slides. Somatic metaphase plates were prepared by conventional air-drying technique with some modifications [36]. For visualization and description of chromosome morphology, metaphase spreads were stained with 4′, 6-diamidino-2-phenylindole DAPI (Vector, Burlingame, CA, USA).

Based on the number and the quality of the metaphase spreads, we selected two (*E. cisalpinus*) and five (*E. lucius*) individuals for the Fluorescence in situ Hybridisation (FISH) carried out according to [15, 35]. The 5S rDNA probe was obtained by amplification of *Esox* genomic DNA using the forward primer 5S-1: 5′-TACGCCCG ATCT CGTCCGATC-3′ and the reverse primer 5S-2: 5′-CAGGCTGGTATGGCCGTAAGC-3′ [37]. For the 28S rDNA probe, following primers were used: 5′-AAAC TCTGGTGGAGGTCCGT-3′ and 5′-CTTACCAAAAGT GGCCCACTA-3′. PCR products were purified using the GeneElute PCR Clean-Up Kit (Sigma, USA) and labeled by incorporation of Biotin-16-dUTP (28S) and Digoxigenin-11-dUTP (5S) by nick-translation method (Roche, Switzerland). FISH with 150 ng of rDNA probe per slide was performed with RNase-pretreated and formamide-denaturated chromosome slides. Post-hybridisation wash was performed at 37 °C for 20 min. Detection of FISH signals was performed using Avidin-FITC and anti-Digoxigenin-Rhodamine, Fab fragments, respectively (Roche, Switzerland). Only high quality metaphase cells were examined under Zeiss Axio Imager A1 (Zeiss, Germany) and Nikon 90i (Nikon, Japan) microscopes equipped with epi-fluorescence and digital (Applied Spectral Imaging, Galilee, Israel) and monochromatic ProgRes MFcool (Jenoptic, Germany) cameras, respectively. Images were captured and the electronic processing of the images was performed using the Band View/FISH View software (Applied Spectral Imaging) and Lucia software ver. 2.0 (Laboratory Imaging, Czech Republic). Post-processing elaboration of the pictures was made using CorelDRAW Graphics Suite 11 (Corel Corporation, Canada).

Cloning and Sanger sequencing

DNA was isolated from blood cells, muscles and fins of adult individuals from several *E. lucius* and *E. cisalpinus* individuals using the classical phenol-chloroform extraction method. The crude DNA fraction was re-purified by DNeasy Blood and Tissue kit (Qiagen, Germany). PCR was used to amplify 5S units from genomic DNA of *E. lucius* and *E. cisalpinus* using PCR as described above. After agarose gel electrophoresis, two fragments of about 220 and 450 bp were visualised in each species. Both fragments were purified and cloned into the pGEM-T vector (Promega, USA). Two clones from each species were sequenced: the short inserts contained monomeric units carrying the 120 bp genic region and an intergenic spacer; the long inserts contained dimmers with two copies of the genic region and a spacer. In addition, one 5S-carrying clone was obtained from the *E. lucius* genomic library prepared by digestion of DNA

with the *Ase*I restriction enzyme. The sequences were submitted to the GenBank under the accession numbers (KX950799, KX965715-KX965718).

Southern and slot blot hybridisation

The procedure followed the protocol described by [38]. The 5S rDNA probe was a 243 bp insert of the 5S_Eci_a clone (GenBank KX965716) from *E. cisalpinus.* The plasmid insert was amplified and labelled with the 32P-dCTP (DekaPrime kit, Fermentas, Lithuania). The probe was hybridised at high stringency conditions (washing 2x SSC, 0.1% SDS followed by 0.1xSSC, 0.1% SDS at 65 °C). The hybridisation signals were visualised by Phosphor imaging (Typhoon 9410, GE Healthcare, PA, USA) and signals were quantified using ImageQuant software (GE Healthcare, PA, USA). The copy number of 5S genes was estimated using slot blot hybridisation. Briefly, the DNA concentration was estimated spectrophotometrically at OD_{260nm} using Nanodrop 3300 Fluorospectrometer (Thermo Fisher Scientific, USA). Concentrations were verified by the electrophoresis in agarose gels using dilutions of lambda DNA as standards. The three dilutions of genomic DNA (100, 50 and 25 ng), together with a serial dilutions of unlabelled plasmid inserts corresponding to the 5S monomers (GenBank KX965715-6), were denatured in 0.4 M NaOH and blotted onto a positively charged Nylon membrane (Hybond XC, GE Healthcare, USA) using a vacuum slot blotter (Schleicher-Schuell, Germany). The probe and the hybridisation conditions and visualisation of signals were the same as described above.

Intragenomic variation and rDNA copy number determined from NGS reads

The source data was the SRR1197512 archive containing Illumina HiSeq 2000 reads from the whole genome sequencing project of *Esox lucius* (SRX494131, University of Victoria, isolate CL-BC-CA-002) [9]. Sequence downloads and basic read manipulations of the genomic reads were done with the aid of the Galaxy Server [39]. The starting read pool of NGS reads consisted of more than 900 million unpaired reads. Before mapping all reads with Ns, reads less than 90 nt in length or reads below quality Phred scores 30 were removed using 'QC and Manipulations' tools at the Galaxy server. The data in FASTQ formats were imported into the CLC Genomics Workbench 6.5.1 (Qiagen, Germany) (CLC). The number of reads was then reduced (Table 1) to decrease computing time using a Sample Reads" command. The high quality reads were mapped to the reference sequences: the 220 bp fragment of 5S rDNA from *E. lucius* (GenBank KX965716) and the 1581 bp fragment of 18S sequence from the pike icefish (*Champsocephalus esox*) (AF518187) using the 'Map Read Reference' tool (CLC). The parameter settings were as follows: mismatch cost value: 2, insertion cost value: 3, deletion cost value: 3, with both the length fraction value and the similarity fraction value set at 0.5 and 0.8, respectively. Files with the mapped reads were saved and used for the downstream copy number and SNP analyses. Transcriptomic reads obtained from SRR1228710-12, SRR1228725 and SRR1228729 archives were mapped to 5S genic region (GenBank KX965716) as described above. Relatively few cDNA reads were mapped probably due to the fact that RNA was polyA-filtrated prior to library construction.

The genome proportion (GP, in percentages) of rDNA was calculated as the number of mapped reads versus total number of reads. The genome space (GS, in megabases) was calculated from the formula: genome size x GP of rDNA units. The copy number was calculated from GS values (in bp) divided by the size of the 5S monomer (220 bp) or part (1581 bp) of the 18S gene. The mapped rDNA units were relatively evenly covered by NGS (each 5S gene nucleotide was covered >3000 x) allowing reliable calculation of genome proportions (not shown).

Variants were called via the 'Probabilistic Variant Detection' function tool in CLC using default settings. SNPs were filtered as follows: minimum read coverage–300, count (the number of countable reads supporting

Table 1 Copy number of rRNA genes in *E. lucius* determined from NGS reads and Southern blot hybridisation

rDNA	Platform[a]/method	Read Archive accession	Total reads	Mapped reads[b]/BLAST hits	GP[c] (%)	GS[d] rDNA (Mb)	Copies[e] (1C)
5S	*IL*	SRR1197513	166,395,976	366,676	0.22	2.42	~20,200
	PB	SRR1930096	555,762	29,151	0.16	1.81	~18,000
	S.blot[f]				5	53	~200,000
18S	*IL*	SRR1197513	166,395,976	199,694	0.12	1.32	~820
	PB	SRR1930096	555,762	1,436	0.08	0.91	~570

[a]Sequencing platform: *IL* – Illumina; *PB* – PacBio
[b]The number refers to the number of Illumina reads mapped to 5S reference or 5S hits with the PacBio data base
[c]*GP* – genome proportion
[d]GS – 5S genome space was calculated from GP: genome size (MB) * GP (%), considering 1100 Mb/1C the *E. lucius* genome size
[e]The copies were calculated as follows: GS (bp) divided by size of the 5S unit monomer (220 bp)
[f]Experimental evaluation of copies by the slot blot hybridisation (Additional file 2: Figure S2)

the allele)–50, frequency (the ratio of "the number of 'countable' reads supporting the allele" to "the number of 'countable' reads covering the position of the variant"): ≥5% (high frequency SNPs).

Analysis of long 5S rDNA arrays within the PacBio reads

We created a BLAST library from the sequence archive SRR1930096 comprising >500,000 PacBio reads. The library was BLASTed against the NGS consensus (the 5S genic region plus the spacer) built from the *E. lucius* Illumina mapped reads. These PacBio reads (2640) were then extracted (primary 5S archive). Because the higher-order organization of units was the primary goal we filtered the primary archive to obtain longer reads of ≥10 kb. This step yielded 286 reads with Phred quality scores Q = 10–11. The average size of read was 12.5 kb; the longest read was 29,914 bp. To determine the number of 5S rRNA genes in each read we used MultiBlast search (e = 1.0) and queried the 286 sequences against the 120 bp 5S NGS consensus (genic region). MultiBlast outputs were exported in the csv format to MS Excel and further processed. Pairwise comparisons of "self to self" read or "read to the 5S rDNA unit NGS consensus" was carried out for each read using the YASS genomic similarity search tool [40]. The alignment parameters were as follows: Scoring matrix (match, transversion, transition, other): +5, −4, −3, −4; gap costs (opening, extension): −16, −4; E-value threshold 0.001. X-drop threshold: 30. In order to reveal degenerate genes, the E-value was increased to 1.0 in some cases. This less stringency search usually resulted in about 15% increase in the number of hits. Outputs were presented as dot-plots.

Spacer variants in long PacBio reads were analysed using the 'Search Motif Tool' function in CLC. The search query involved a 10 nucleotide intergenic spacer sequence containing four highly polymorphic positions in the middle. Stringent conditions were applied scoring alignments with more than 90% matches along the 10 nucleotide query. Detected motifs were annotated and counted.

Phylogenetic and secondary 5S structure analysis

Alignments were built using the 120-bp long genic sequences originating from: (i) clones isolated in this work, (ii) clones from the GenBank, (iii) sequenced rRNA and (iv) cDNA consensus sequences prepared from the mapped transcriptomic reads. ClustalW alignment was implemented within the BioEdit program [41]. A phylogeny NJ tree was constructed using the Seaview program [42]. All calculations were run at default conditions using the Jukes-Kantor model and 1000 replicates.

Phylogenetic NJ trees were constructed from Illumina reads using short 70 bp subregions derived from the 5S genic region (Elu_b clone, position 201–270, GenBank KX965717) and the intergenic spacer (the same clone, position 108–177). The stand-alone BLAST library of SRR1197512 was searched using these subregions as queries. Hit reads were extracted, trimmed for unique lengths (70 bp), sampled to 500 and aligned ('Multiple Alignment' tool function of the CLC). Unrooted NJ trees were constructed from aligned reads employing the Jukes-Cantor model and visualised in radial projections. Haplotypic diversity was calculated from aligned reads using the DNASp4 program [43].

Secondary structure modelling was carried out using an online tool at the RNAfold web server ([44], http://rna.tbi.univie.ac.at/). The secondary structures were based on minimum free energy (MFE) calculations using the Turner 2004 model. The program setting was as follows: isolated nucleotides were avoided; vote for dangling energies on both sides of a helix in any case.

DNA methylation analysis

The purified genomic DNA samples from *E. lucius* (2 individuals) and *E. cisalpinus* (3 individuals) were digested with methylation-sensitive *Hpa*II (sensitive to CG methylation) and its methylation-insensitive *Msp*I isoschizomere (both enzymes are cutting at CCGG). The restriction fragments were hybridised on blots with the alpha[P32]dCTP-labelled 5S probe (Eci_a clone, GenBank KX965716). Control of digestion efficiency was carried out by spiking the *Esox* genomic DNA with a non-methylated plasmid DNA (pBluescript, Stratagen) and subsequent hybridisation with a plasmid probe. Both *Msp*I and *Hpa*II enzymes yielded expected restriction fragments (not shown).

Results

Localisation of 5S and 45S rDNA loci and heterochromatin on *Esox* chromosomes

Both *E. lucius* and *E. cisalpinus* showed the same number of chromosomes (2n = 50) exhibiting strict acrocentric morphology (FN = 50). We used FISH to determine the number and position of rDNA loci in *Esox* chromosomes using the 5S and 28S rDNA probes (Fig. 1a and b). In *E. cisalpinus*, the 5S rDNA probe hybridised to 30–34 sites (Fig. 1a, quantitative data are summarized in Additional file 1: Figure S1), all in the (peri)centromeric regions. The 28S probe hybridised to (peri)centromeric sites corresponding to NORs on two homologs (single locus) that lacked the 5S rDNA signals (Fig. 1a). In *E. lucius*, the 5S rDNA probe hybridised to 30–38 sites in (peri)centromeric regions (Fig. 1b, quantitative data are summarised in Additional file 1: Figure S1). The 28S rDNA probe hybridised to two NORs that co-localised but not overlapped with 5S signals. The 28S rDNA probe hybridised more distally compared to the 5S rDNA probe. These different patterns of the chromosomal distribution of both major and minor rDNA

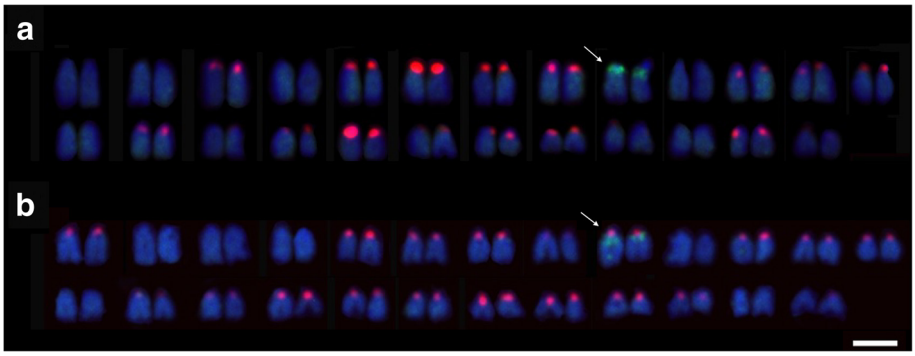


Fig. 1 Molecular cytogenetic FISH analysis with rDNA probes to *Esox cisalpinus* (**a**) and *E. lucius* (**b**) chromosomes shown on representative karyotypes. The probes were 5S rDNA in *red* (31 signals in *E. cisalpinus* and 37 signals in *E. lucius*) and 18S rDNA in *green* (two signals in both species marked by *arrows*). Chromosomes are arranged in pairs approximately in decreasing size. More quantitative data on counts of 5S rDNA signals are provided in Additional file 1: Figure S1. Scale equals to 5 μm

sequences observed in *E. lucius* and *E. cisalpinus* enabled cytogenetic identification of these two species.

Massive amplification of 5S gene copies

FISH showed an extraordinary high number of strong 5S probe signals on numerous *Esox* chromosomes indicating a high copy repeat. This observation provoked a question about the 5S gene copy number in both *Esox* genomes. To determine the 5S copy number we first applied computation approach based on the proportion of mapped reads relative to the total reads (Table 1). The 5S rDNA copy number calculated from both data sets (coming from Illumina and PacBio platforms) were in agreement. The 18S rDNA copy number was at least 20 fold lower than that of the 5S rDNA. This is in line with differences in sites number (single 18S site and about thirty six 5S sites, Fig. 1).

We also determined the 5S rDNA copy number using the classical slot blot hybridisation (Additional file 2: Figure S2). Both *E. lucius* DNA isolates yielded more than 200 thousand copies confirming extreme gene amplification. However, the copy number determined by slot blot hybridisation was >10 fold higher than that one calculated from NGS reads.

Cloning, sequencing and the phylogenetic analysis

A phylogenetic tree (Fig. 2a) was constructed from an alignment comprising several genomic clones from the GenBank, sequenced 5S rRNA genes [37] and a cDNA consensus sequence built from *E. lucius* transcriptomic reads. In addition, a 5S-derived satellite from *Hoplias malabaricus* was included. The sequences grouped into two main clades: (i) The upper clade contained the 5S-derived satellite clones from *H. malabaricus*. Relatively long branches indicate a considerable sequence divergence. (ii) The second clade was formed by a group of genomic clones from different species including those of *Esox*, sequenced 5S rRNA from rainbow trout (*Salmo*) and the cDNA consensus from *E. lucius*. The sequences of clones from *E. lucius* and *E. cisalpinus* were similar both within and across species. Consequently, clones from *E. lucius* and *E. cisalpinus* could not be separated and formed a common branch with the 5S cDNA consensus. Within the functional genes, the *H. malabaricus* sequence (accession AY624053) was relatively well separated from other sequences isolated from *Lepomis*, *Salmo*, *Cyprinus* and *Esox*. This is understandable since *Hoplias* represents the most divergent genus out of the five fish genera analysed [45] indicating that the 5S tree roughly reflected the phylogeny.

Conserved secondary structures of *Esox* 5S rRNA

Thermodynamic stability of 5S rRNA secondary structure has been considered as an important criterion for gene functionality [46]. In general, the stable three domain structure is an attribute of functional genes. To investigate whether the amplified *Esox* 5S rDNA code for any functional molecules, we predicted the rRNA structures of several *Esox* molecules by computer modelling and results were compared with those of other genera. Indeed, the *E. lucius* clones produced comparable molecule shapes as the 5S rRNA of *Salmo* (rainbow trout) (Fig. 2b) and other potentially functional fish 5S rRNA genes (Additional file 3: Figure S3). The thermodynamic stability was comparable (high) between the species. The 5S-derived satellite from *Hoplias malabaricus* had significantly lower ($\Delta G \sim$ 28 kcal/mol, in average) thermodynamic stability than the functional genes (typically $\Delta G \sim 50$ kcal/mol) explaining why the satellite structures were vastly different from functional genes of both *Salmo* and *E. lucius*. The typical Y-shape of the fish 5S rRNA was not so pronounced as in other species (Additional file 3: Figure S3) [17, 23]. The structural asymmetry of fish 5S rRNA molecules was apparently caused by a longer beta and shorter gamma domain, respectively.

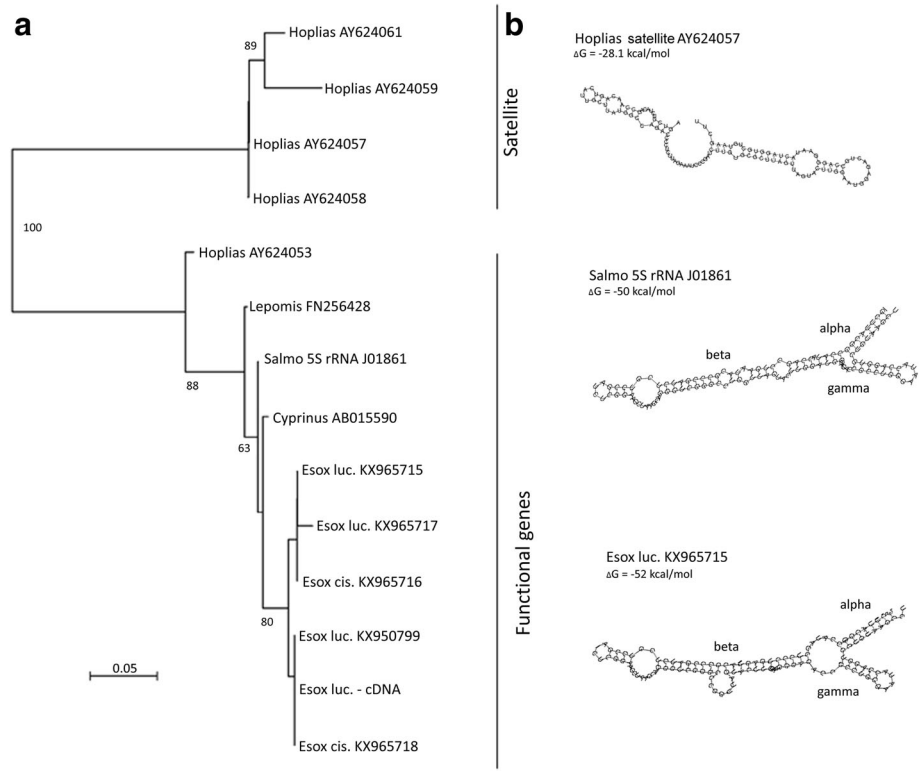


Fig. 2 a The phylogram constructed from 5S rDNA clones, 5S-related satellites, 5S rRNA and 5S cDNA consensus from the *E. lucius* transcriptome. Bootstrap support of branching (>60%) are indicated. Note, clustering of *Esox* genomic and cDNA clones. Examples of presumed rRNA secondary structures are shown in (**b**) and Additional file 3: Figure S3. Domains of 5S rRNA follow the nomenclature of [17]

Low intragenomic heterogeneity of the 5S genic region contrasts with higher diversity in the intergenic spacers

In order to determine the intragenomic homogeneity of 5S rDNA we explored the sequenced genome of *E. lucius*. The Illumina reads were mapped to the 5S rDNA reference clone 'a' from *E. lucius* (GenBank KX965715) and subjected to the analysis of variance. We considered only high confident SNPs occurring at ≥ 5% frequency, i.e. at least 1000 genes that carry such a variant (considering there may be ~20,000 copies of 5S genes in the genome). Quantification of four kinds of sequence variation (indels and substitutions) along the genic region is shown in Fig. 3A and Table S1 in Additional file 4. Indels were rare and most polymorphisms could be attributed to single nucleotide substitutions. In the 120 bp genic region, only three SNPs, all substitutions, were found. The SNPs located outsides of the regulatory motifs (Boxes A and C and IE) and did not significantly affect secondary structure (not shown). The site containing a T > C mutation at position +45 was also polymorphic among the cDNA reads (Additional file 4: Table S2) suggesting that both variants are expressed. The intragenomic SNPs were approximately five fold more abundant (12.8 SNPs per 100 bp) in the spacer than in the genic region (2.5 SNPs per 100 bp). The 28S rRNA gene was slightly more homogeneous having only 1.2 SNPs per 100 bp of sequence.

In order to determine phylogenetic relationships between the 5S families we constructed haplotypic networks from 500 randomly selected Illumina reads mapping parts of the genic and IGS regions, respectively (Additional file 5: Figure S4A, B). It is evident that the NJ tree constructed from IGS was highly bifurcated compared to the one constructed from genic sequences. The overall diversity expressed as the number of substitutions/100 bp (Pi) was higher (about three-fold) in the spacer region (Pi = 0.0508) than in the genic region (Pi = 0.0145).

Higher-order organisation of 5S arrays

To date, the regularity of 5S tandems and distribution of variants of arrays in the genome have not been assessed due to technical difficulties related with sequencing of long and repetitive molecules. Only now, the single molecule sequencing technologies, such as the PacBio, appear to be suitable for such studies since they generate longer reads than other sequencing platforms. We took advantage of the availability of the PacBio-sequenced *E. lucius* genome (ENA archive SRR1930096) and addressed the question of distribution of 5S intergenic spacer variants (i) and higher-order organisation of 5S arrays (ii):

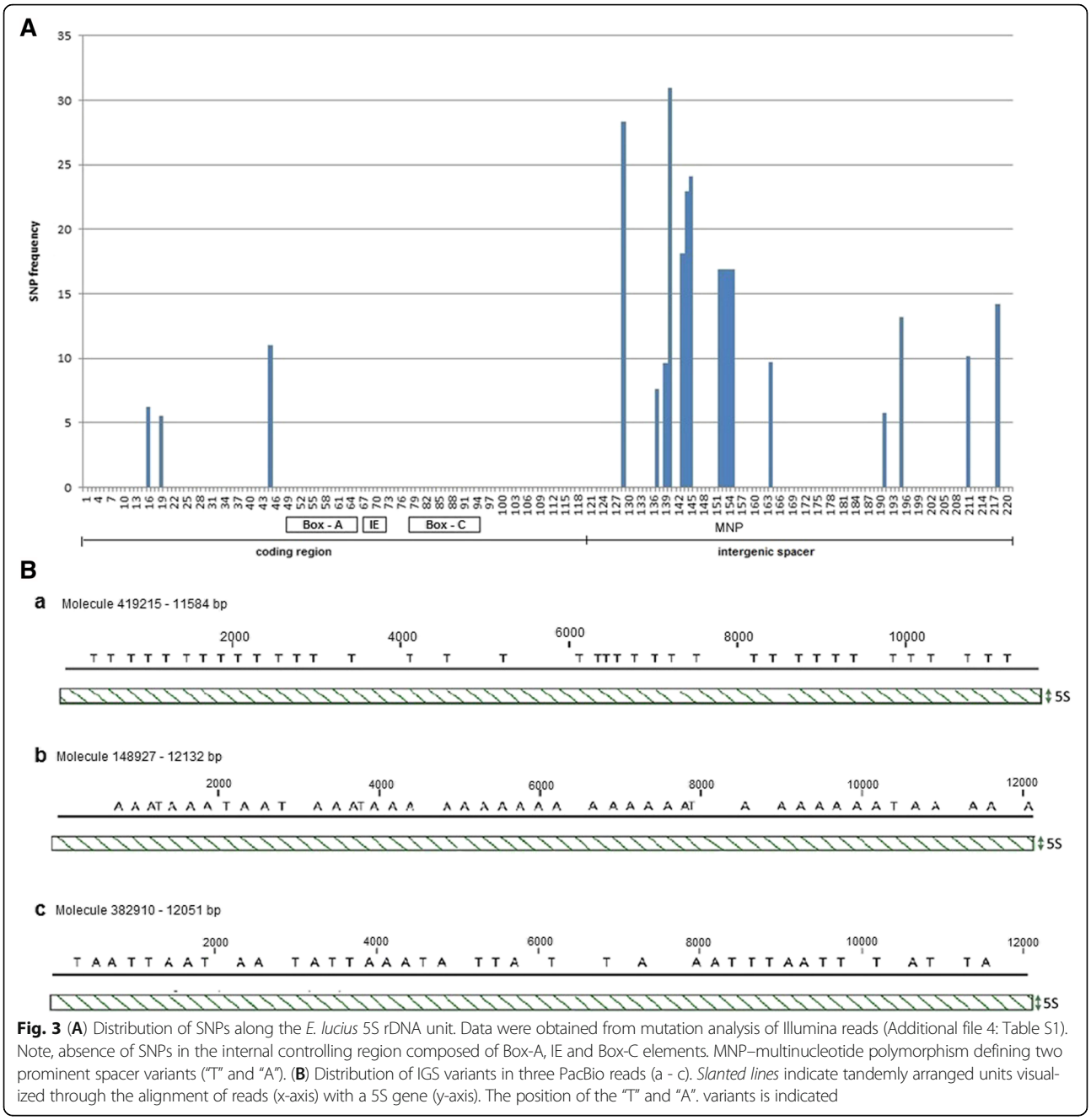


Fig. 3 (**A**) Distribution of SNPs along the *E. lucius* 5S rDNA unit. Data were obtained from mutation analysis of Illumina reads (Additional file 4: Table S1). Note, absence of SNPs in the internal controlling region composed of Box-A, IE and Box-C elements. MNP–multinucleotide polymorphism defining two prominent spacer variants ("T" and "A"). (**B**) Distribution of IGS variants in three PacBio reads (a - c). *Slanted lines* indicate tandemly arranged units visualized through the alignment of reads (x-axis) with a 5S gene (y-axis). The position of the "T" and "A". variants is indicated

i. Previous SNP analysis revealed variation in the intergenic spacer region (Fig. 3A). At the position 152–155, the tetranucleotide TCCT > AGGA variation corresponded to a more abundant (83%) "T" (TCCT motif) and the less (17%) abundant "A" (AGGA motif) variants (Additional file 4: Table S1). We determined the distribution of these variants in three randomly selected PacBio molecules that showed relatively high quality scores ($Q = 11$) and hence may be used for analysis of variants (Fig. 3B). All three molecules (a, b, c) had a comparable number of tandemly arranged genes ranging 50 to 55. The molecule in (a) contained 50 complete units out of which 37 had the "T" spacer variant. Thirteen units had neither of the two variants probably due to mutations or sequencing errors. The molecule in (b) had 51 complete units. Out of these, 33 units had the "A", three units had the "T" and 15 had other variants. The molecule in (c) was the most heterogeneous comprising 55 units out of which 20 can be assigned to the "T" and 19 to the "A" variants. Thus, variant composition and homogeneity differ from array to array.

ii. The higher-order organisation of 5S repeats was investigated in 286 PacBio molecules extracted from the 5S rDNA BLAST search dataset and size-filtrated for 10–30 kb. This subset representing about 43% of total *E. lucius* 5S rDNA was analysed for gene richness (Fig. 4). The 5S genes number varied markedly (1–110) between individual reads indicating differences in genomic organisation of repeats. To analyse the higher-order repeat organisation, each sequence was subjected to self to self comparison (Additional file 6: Figure S5). Based on the resulting dot-plot profiles, four types of higher-order organisation can be distinguished (Table 2 and Additional file 7: Tables S3): (i) Group I, molecules containing continuous blocks of 5S tandem repeats spanning the entire read length (Fig. 5a). (ii) Group II, molecules containing one or several blocks of 5S head to tail tandems plus variable portions of unrelated mostly unique sequences (Fig. 5b). In one read, the 5S block was attached to another block of 5S-unrelated tandems (Additional file 7: Table S3). (iii) Group III, molecules harbouring invertedly repeated blocks of 5S tandems that may or may not be separated by spacers (Fig. 5c and Additional file 8: Figure S6). (iv) Group IV, containing no longer blocks (Fig. 5d), but rather dispersed 5S copies. Despite relative abundance (37% reads) the 5S gene richness was low (~4%) in this group (Table 2). In contrast, gene representation in reads bearing long blocks (Groups I–III) was high (96%). Pairwise comparisons of reads with the 5S reference (the genic region consensus built from Illumina reads) allowed us to determine variation in length of intergenic spacers. The three major spacer length variants were identified: (i) Short 95–116 bp variant occurring in 7035 (94%) units; (ii) Long 321–340 bp variant found in 357 (5%) units; (iii) Ultralong 1153–1209 bp variant found in 65 (1%) units (Additional file 7: Table S3). The sequences of short and long spacer variants were unrelated. Except for one read, spacer variants formed independent arrays.

The majority of 5S rRNA genes are highly methylated

The methylation status of 5S rDNA genes in *E. lucius* and *E. cisalpinus* was determined by enzymatic digestion of genomic DNA with methylation-sensitive *Hpa*II and insensitive *Msp*I. In all samples, the probe hybridised to high molecular weight bands after the digestion of DNA with methylation-sensitive *Hpa*II (Fig. 6). In contrast, a major low molecular weight band of about 220 bp was visible after the digestion with *Msp*I. The slow-migrating oligomeric *Msp*I bands were faint confirming a high homogeneity of arrays. The near complete resistance of 5S rDNA to *Hpa*II digestion indicated high level of methylation of most of the units. There were no differences in methylation profiles between the blood and fin

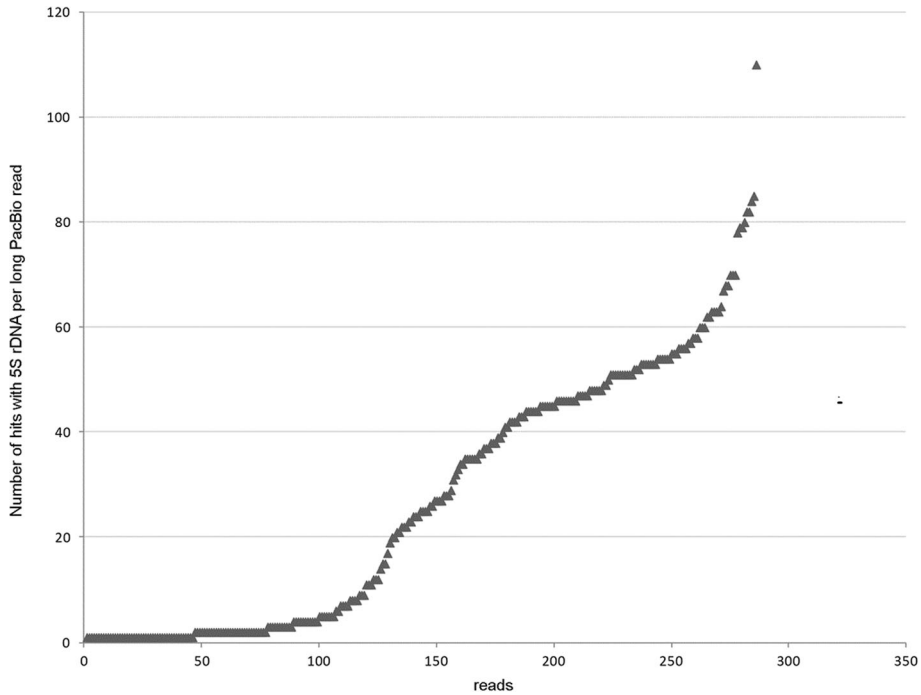


Fig. 4 5S rDNA content in long >10 kb PacBio reads of *E. lucius*. The total number of reads was 286. The sequences were queried with the 5S consensus (genic part). The MultiBlast search yielded 7768 alignments (about 43% of total 18,000 hits, Table 1) here considered as gene copies

Table 2 Classification of PacBio reads according to 5S rDNA higher-order organisation

Group	5S arrangement in a read[a]	Number of reads (percentage)	Number of 5S genes (percentage) [b]
I	Tandem repeats, no unique DNA	95 (33)	4539 (58)
II	Tandem repeats + unique DNA	72 (25)	2352 (30)
III	Blocks of inverted repeats	11 (4)	639 (8)
IV	Dispersed or short (<5 units) tandems	108 (38)	274 (4)
Total		286 (100)	7769 (100)

[a]The data sets and read annotations are in Additional file 7: Table S3
[b]The number of genes in reads was determined based on MultiBlast search using the 5S genic region (NGS consensus) as a query

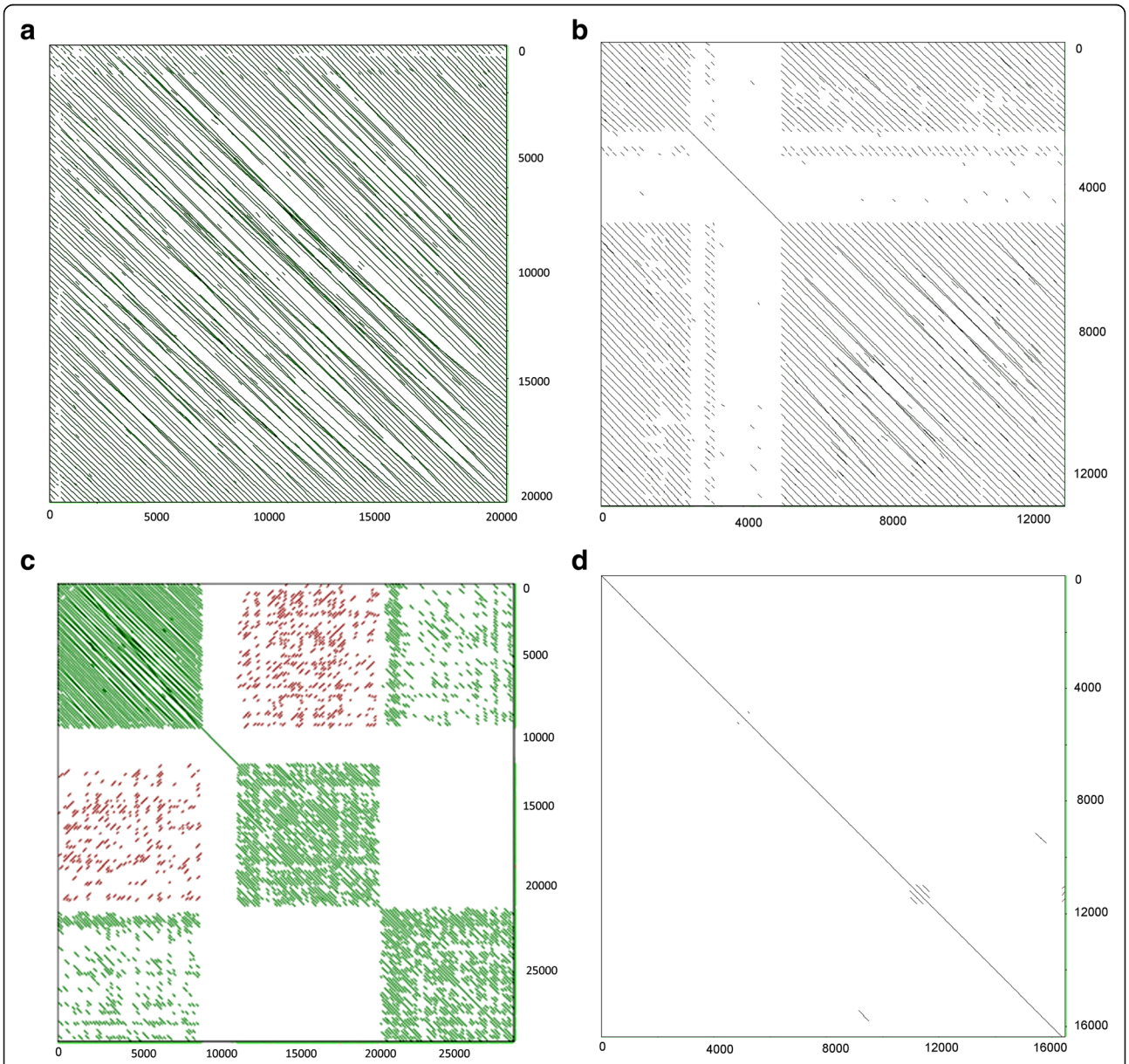


Fig. 5 Higher-order organization of 5S rDNA arrays in *E. lucius*. Self-to-self comparison of long PacBio molecules representing four groups: **a** Group I molecule #531194 (19,172 bp) containing the longest uninterrupted block of 5S genes (87). **b** Group II molecule #421211 (13,043 bp) containing two blocks of tandem repeats (52 genes) separated by a ~2 kb spacer. **c** Group III molecule #496426 (29,914 bp). It represents the longest PacBio read available harboring 110 5S copies; **d** Group IV molecule # 49823 (16,372 bp) containing large part of unique sequence plus five copies (three in tandem) of 5S rDNA. Coordinates are in base pairs

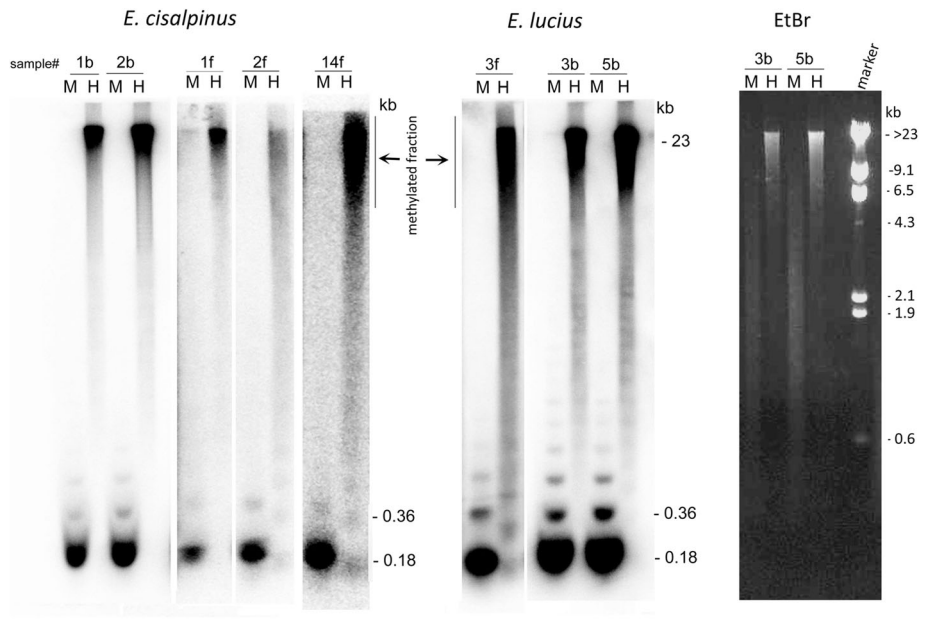


Fig. 6 Methylation analysis of 5S rRNA genes by the methylation-sensitive *Hpa*II (H) restriction enzyme and its methylation-insensitive *Msp*I (M) isoschizomere. The *left two panels* show Southern blot hybridisation with the 5S rDNA probe; the *right panel* (EtBr) shows ethidium bromide-stained DNA fragments before the hybridisation. The probe hybridised with a ~200 bp *Msp*I fragment corresponding to a monomer. Hybridisation signals with the *Hpa*II fragments were of high molecular weight. Samples "f" and "b" were fin and blood isolates, respectively. The samples from fin tissue were slightly degraded which probably accounted for smear signals

tissues. The globally high methylation level is evident from ethidium-bromide stained DNA fragments. Most *Hpa*II-fragments migrated as high molecular weight relic while the DNA was relatively efficiently digested into shorter fragments with methylation-insensitive *Msp*I.

Discussion

Massive amplification of 5S rRNA genes in *Esox lucius*

In animals, the number of rRNA genes typically reaches up to hundreds of copies [46–48]. It is therefore striking that these *Esox* species harbour tens of thousands of 5S rDNA copies. The actual copy number could be even higher since the experimental copy number estimates exceed 200 thousand copies forming about 5% of the *Esox* genomes. The apparent discrepancy (more than 10 fold) between NGS and slot blot hybridisation could be attributed to the fact that tandem repeats are typically underrepresented in the DNA sequencing libraries [49, 50]. The second explanation could be potential variation in 5S copy number in different populations of *E. lucius*. In our study, the copy number was calculated from the whole genome sequencing dataset of an American population while the experimental copy number estimation was performed in different individuals of European origin. Thus, we cannot exclude inter-population differences in copy numbers, already reported in mammals [51]. In any case, the number of 5S rRNA genes exceeding tens of thousands of copies is far more than seen in most metazoans (http://www.animalrDNAdatabase.com).

The 5S rDNA copy number, sequence and position on chromosomes are similar in *E. lucius* and *E. cisalpinus*. The genomic spreading of 5S rDNA was not accompanied by any concomitant expansion of 45S rDNA whose copy number was limited to about 800 (Table 1). Independent amplification of both types of rDNA has also been reported in other fish genera [19, 21] suggesting that concerted amplification of 45S and 5S rDNA [51] may not be operating in fish or it is limited to particular groups. Related salmonid fishes harbour far more 45S than 5S loci [13–15, 36]. Thus, it is reasonable to suggest that amplification of 5S must have occurred after the divergence of a common *Esox* ancestor from the rest of salmonids.

Possible epigenetic regulation of amplified 5S rRNA genes by DNA methylation

The extremely high copy number of 5S genes in *Esox* resembles that of some amphibians [52]. However, in amphibians amplified repeats are attributed to be pseudogenes while most repeats we analysed here are probably capable to encode functional 5S rRNA. This was evidenced by their high intragenomic homogeneity, good matches between genomic and transcriptomic reads and conserved thermodynamically stable secondary structures. This raises an important question about

their transcription regulation. In mammals, only about one hundred of 45S rRNA genes are estimated to be transcribed at any one time (for review see [53]). In *Danio rerio* (zebrafish) only twelve 5S copies appear to be active in oocytes while a large number of genes in another locus are silenced and activated only at later developmental stages [54]. Perhaps, transcription activity of amplified genes in *Esox* might be regulated via epigenetic mechanisms as well. Indeed, 5S genes in both *E. lucius* and *E. cisalpinus* were heavily methylated at CG motifs suggesting that a powerful epigenetic system has evolved in this genus. There were no apparent differences in methylation levels between different tissues reported in other systems [55] suggesting that most genes are evenly methylated. However, our restriction enzyme-based methylation assay reveals global level of 5S rDNA methylation while its resolution is unable to detect changes at the single gene level. It is worth noting that sequence polymorphisms are located within the first half of the 5S gene both in *E. lucius* (Fig. 3) and *D. rerio* [54]. Therefore, developmental regulation of 5S rDNA expression cannot be excluded in *Esox*.

The 5S genes are organised in large blocks of variable sequence homogeneity

Tandem arrays of repeated sequences are generally considered as problematic regions often refractory to in depth genomic analysis. Hence, the organisation of repeat variants in the genome has been intensively discussed [47, 56, 57]. Based on conventional Southern hybridisation methods, repeat variants are believed to form separate arrays [58]. However, at the genomic scale evidence is missing due to technical difficulties related to sequencing of long and repetitive molecules. In *E. lucius*, variation in the 5S intergenic spacer was about 5-fold higher than in the genic region indicating relaxed selection constraints on intergenic spacers. We took the single cell PacBio sequencing approach to study distribution of spacer variants in tandem arrays in this species. Being aware that the randomly occurring sequencing errors are relatively high (~13% in DNA polymerase single pass) using this technology [59] making sequence polymorphisms difficult to interpret. However, sequencing errors cannot account for all the variation observed in our data sets. This is because SNPs (spacer variants) residing in the array were regularly phased (Fig. 3B) while sequencing errors are distributed randomly [59]. Secondly, the type and position of multinucleotide variation detected in long PacBio molecules were similar to those detected in high quality Illumina reads. Thus, the analysis of PacBio reads seems to be an adequate way to address the question of distribution of major variants. We found that 5S spacer variants form both separate and mixed arrays. Strict tandem arrangement of 5S units in the heterogeneous arrays suggests that single or multiple nucleotide polymorphisms do not impair array regularity. In contrast, major spacer length variants tend to form independent arrays (Additional file 7: Table S3). Interestingly, in *Arabidopsis* where 5S loci are located on three chromosomes, a similar block-like structures composed of homo- and heterogenous arrays were detected in BAC libraries [60] and recent PacBio sequencing revealed substantial spacer polymorphisms in the 45S rDNA intergenic spacers located on two chromosomes [61]. These results suggest that the higher-order repeat structure may be evolutionary conserved. Despite the variation, evolution of 5S rRNA genes in *Esox* almost certainly fits the concept of concerted evolution considering that tens of thousands of copies are present in the genome and the SNPs being relatively infrequent. Relative heterogeneity of some arrays can be explained by reduced efficiency of interlocus recombination and/or less stringent selection constrains imposed on the spacers.

Mechanisms of arrays spreading

Accumulation of repeats at similar chromosomal positions raises questions on the mechanisms mediating spreading and homogenisation of 5S rDNA units. Several hypotheses can be drawn:

i. Chromosome location may affect recombination rate and homogenisation of 5S arrays. In mouse, a huge copy number variability in 45S rRNA genes has been associated with their purely centromeric location in mostly acrocentric/telocentric chromosomes [62]. Thus, it has been proposed that DNA breaks may appear quite frequently in such located rDNA sequences leading to translocations of rDNA to other chromosomes. In *Esox*, 5S loci are uniformly located on short arms of nearly all acrocentric chromosomes. By analogy, in *Esox*, interlocus recombination of 5S genes could be driven by their (peri)centromeric position in acrocentric chromosomes.
ii. 5S rRNA genes are able to multiply and integrate into other areas of the genome using a mechanism similar to retrotransposition among others. The 5S rRNA genes and SINEs harbor the same type of internal RNA polymerase III promoter [63]. On the top of that, some authors [64, 65] found a unique class of SINEs that have been formed by fusion of a 5S rRNA gene and a LINE, showing that 5S and retroelements may interact. In support, SINE elements with 5S features were reported in several fish [54, 64] and a retroelement was co-localised with 5S loci on *Erythrinus erythrinus* (red wolf fish) chromosomes [19] and recently among members of another fish genus, *Gymnotus* [66]. In the future it will be interesting to analyse inter-block spacers in

Group II reads (those containing one or few blocks of 5S genes head to tail, plus unrelated sequences) for the presence of transposable elements which may support their potential role in 5S rDNA mobilisation.

iii. 5S genes spread through extrachromosomal replication and reintegration in new locations. Covalent extrachromosomal circles of rDNA have been identified in diverged biological taxa [67, 68]. In *Xenopus laevis* rDNA is known to replicate extrachromosomally during development [69]. In our datasets some loci (4% PacBio reads) contained invertedly repeated large blocks of 5S tandems. There is experimental evidence that extrachromosomal DNA can be generated by replication errors at the inverted repeats [70]. Therefore, one can hypothesize that the homogenisation of 5S loci across *Esox* chromosomes is mediated by the initial replication block at the inverted repeat, excision, extrachromosomal replication and reintegration by homologous or non-homologous recombination into a new genomic location. Large palindromes may thus transpose and seed 5S blocks to distal locations. Supporting this, a recent study of a human centromeric satellite (using long read PacBio sequencing) showed increased frequencies of inversions in acrocentric chromosomes compared to other chromosomes [71]. Perhaps, acrocentromeric positions of rDNA could be particularly vulnerable to such rearrangements and/or inversions. This model may also explain why interlocus homogenisation of rDNA often occurs without extensive chromosome rearrangements [30, 72].

Long term scenario of arrays evolution

i. It can be envisaged that as long as 5S rRNA genes undergo interlocus homogenisation the number of loci would remain relatively constant. This is likely happening in both closely related species *E. lucius* and *E. cisalpinus* that show similar number of genes, sequence of the loci (including intergenic spacers) and their chromosome positions. Yet, at the cytogenetic level, small differences were noted: in contrast to *E. lucius*, *E. cisalpinus* lacked the co-localisation of 5S and 45S rDNA. Given that the 45S loci occur on homeologous chromosomes it follows that there was either 5S locus gain (*E. lucius*) or locus loss (*E. cisalpinus*) after the separation from the common ancestor. Perhaps, adjacent blocks to 45S and 5S could be represented by an unstable chromatin configuration leading to breaks and chromosomal translocation.

ii. As methylated cytosines are more susceptible to mutations [73] we may expect gradual accumulation of C > T and G > A mutations in the *Esox* 5S genes and their subsequent pseudogenisation. Indeed, in some plant species most rRNA genes were converted into pseudogenes [74]. However, we have no evidence for such significant erosion in *Esox* 5S rDNA despite its high methylation (methylation-induced mutations were not significantly enriched in *Esox* 5S rDNA, $p > 0.05$). Moreover, two major variants of a genic region seem to be expressed suggesting that not all mutations automatically lead to pseudogenisation. It is likely that *Esox* genomes seem to be in a dynamic phase of evolution, where most pseudogenes are being removed by genetic recombination.

iii. Amplified 5S genes acquire centromeric function. Retrotransposons and RNA-polymerase-III-transcribed genes, including tRNA and 5S rRNA (the so called Pol III genes) have been found to be associated with centromeres in fission yeasts. Furthermore, results of some studies suggest a functional link between the centromeric localisation of the Pol III genes and chromosome condensation resulting in the proper assembly of mitotic chromosomes [75]. In the fish *H. malabaricus*, 5S genes gave rise to an independent satellite with apparently centromeric function [76]. Similar conversion might have happened in other species [77, 78] as well. Thus, if 5S rDNA is supportive to the centromeric function and plays a role in the assembly of mitotic chromosomes then its "invasions" to centromeric positions might be favored in evolution. Natural selection would select those variants that bind centromeric histones and adopt centromere-specific chromatin configuration. Relative heterogeneity of some blocks bearing degenerated units (Fig. 5c and Additional file 6: Figure S5) suggests that the process of satellite formation could have already been started. It is also possible that satellite repeats may arise from orphanised degenerated 5S insertions accounting for about 4% of *E. lucius* rDNA (Table 2).

Conclusions

In two European *Esox* species we have witnessed the extreme amplification of 5S rRNA genes reaching up to tens of thousands of copies and their distribution across more than half of the chromosomes. Such a high number is exceptional in animals, generally thought to contain moderate number of these genes. Most of the amplified genes appear to be functional and heavily epigenetically modified. Detailed analysis of long PacBio reads suggests a considerable variation in the phasing of unit variants and in the arrangement of large blocks of 5S tandem repeats. These higher-order structure polymorphisms may potentially influence the expression and homogenisation of these genes.

Additional files

Additional file 1: Figure S1. Summary on counts of FISH signals of 5S rDNA on chromosomes of two individuals of *E. cisalpinus* (Eci1 and Eci2) and five individuals of *E. lucius* two of which originated from the Czech Republic (EluCz3, EluCz5) and three from Poland (EluP8, EluP9, EluP15). (PDF 309 kb)

Additional file 2: Figure S2. Estimation of 5S rDNA copy number by slot blot hybridisation. The results are shown for two independent genomic DNA isolates and two 5S insert standards (*E. lucius* and *E. cisalpinus*). The 5S rDNAs (genic + spacer regions) account for about 5% of *E. lucius* genome equaling to about 250,000 copies (the data collected from two blot replicates and averaged). (PDF 267 kb)

Additional file 3: Figure S3. Secondary structure models of 5S rDNA molecule for fish (A) and non-fish species (B). (PDF 186 kb)

Additional file 4: Tables S1 and S2. Type, position, frequency and coverage of 5S rDNA polymorphisms. Table S1 – Analysis of genomic Illumina reads. Table S2 – Analysis of transcriptomic Illumina reads. SNPs along whole 5S units (genic and the intergenic spacer) were analyzed using genomic reads; for the transcriptome, only the genic region was considered. Note, high number of SNPs in spacer region compared to the genic region despite the shorter length (Table S1). SNP – single nucleotide polymorphism; MNP – multiple nucleotide polymorphism. (PDF 870 kb)

Additional file 5: Figure S4. Phylogenetic NJ trees constructed from 500 aligned Illumina reads derived from the 5S genic part (A) and an intergenic spacer (B). Well supported branches (bootstrap >60%) are indicated by blue arrows. (C) A sequence of the 5S clone b from *E. lucius* (GenBank KX965717.1) with highlighted subregions used in the phylogenetic analysis. (PDF 856 kb)

Additional file 6: Figure S5. Dot plot diagrams resulting from self to self comparison of long PaBio reads. (PDF 1371 kb)

Additional file 7: Table S3. Analysis of higher-order repeat structure of 5S rDNA using long (≥10 kb) PacBio reads. The selected sequences are ordered according to lengths (descending). Number of gene copies in reads was determined by MultiBlast. Arrangement was assessed by visual inspection of dot plot matrices (Additional file 6: Figure S5). Grouping followed the nomenclature in Table 2. Intergenic spacer variants: S–short (95–116 bp); L–long (321–340 bp) and UL–ultralong (1153–1209 bp). (PDF 471 kb)

Additional file 8: Figure S6. A group III molecule #499258 (19,900 bp) organised in two large immediately linked inverted blocks of tandem repeats. Green and red slanted lines indicate direct and inverted orientation of units, respectively. (B) The junction region alignment to 5S. Note, absence of any 5S-unrelated sequence between the inverted repeats. Note, a partial deletion of IGS in the third copy. (C) Nucleotide sequence of the junction region with annotated genic sequence (brown) and IGS (green). (PDF 291 kb)

Abbreviations

ΔG: Change in Gibbs free energy; FISH: Fluorescence In situ Hybridisation; FN: Fundamental number, number of chromosomal arms; GB: Gigabase; GP: Genome proportion; GS: Genome space; IE: Internal Element; IGS: Intergenic spacer; IL: Ilumina sequencing; MB: Megabase; MNP: Multiple nucleotide polymorphism; NGS: Next Generation Sequencing; NJ: Neighbour Joining; NOR: Nucleolus Organizer Region; NTS: Non-Transcribed Spacer; PB: PacBio sequencing; rDNA: Ribosomal RNA genes; SNP: Single nucleotide polymorphism; WGS: Whole Genome Duplication

Acknowledgements

We thank to O. Cocco, G. Rocco and E. Sala (Associazione Pescatori Carmagnolesi) and to P. Lo Conte (Città Metropolitana di Torino, Servizio Caccia e Pesca) who provided young pike specimens from the Provincial Fish hatchery of Carmagnola.

Funding

This study was supported by a young researchers fellowship (NWF15/BIO-7) of the University of Innsbruck, Austria to RS (design of the study, sample collection, cytogenetic analysis, writing the manuscript); the Czech Science Foundation projects P501/12/G090 to AK (design of the study, epigenetic analysis, bioinformatics procedures, writing the manuscript) and 14-02940S to RS and ŠP (sample collection, DNA isolation); a project financed by University of Warmia and Mazury in Olsztyn, Poland (No. 18.610.003-300) to KO (writing the manuscript, cytogenetic analysis) and a project from the government of Spain (CGL2016-75694-P) to SG (genomic data analysis, manuscript editing).

Availability of data and materials

All data generated or analysed during this study are included in this published article and its supplementary information files. Newly generated sequences have been submitted to GenBank under the following accession numbers: KX950799, KX965715-KX965718. Datasets used in the current study are available from the corresponding authors upon request. Chromosomal preparations and preparations after FISH experiments are deposited in laboratories of KO and ŠP.

Authors' contributions

RS and AK designed the study and performed laboratory experiments. KO and LK performed FISH with 5S rDNA and data analysis, GD provided *E. cisalpinus* samples, ŠP and RS collected biological material and prepared microscopic samples. AK, SG and RS carried out bioinformatic analysis. RS, AK, KO, GD
co-drafted the manuscript, all authors read the manuscript. All authors approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Ethics approval

In Poland, this study was carried out in strict accordance with the recommendations in the Polish ACT of 21 January 2005 of Animal Experiments (Dz. U. z 2005 r. Nr 33, poz. 289). The protocol was approved by the Local Ethical Committee for the Experiments on Animals of the University of Warmia and Mazury in Olsztyn, Poland (Permit Number: 11/2012). In the Czech Republic, all national guidelines for the care and use of animals were followed. The experimental procedures involving fish were approved by the Institutional Animal Care and Use Committee of the Institute of Animal Physiology and Genetics, CAS, v.v.i, according with directives from State Veterinary Administration of the Czech Republic, permit number 124/2009, and by the permit number CZ 00221 issued by Ministry of Agriculture of the Czech Republic.

Publisher's Note

Author details

[1]Research Institute for Limnology, University of Innsbruck, Mondseestrasse 9, A-5310 Mondsee, Austria. [2]Department of Marine Biology and Ecology, Faculty of Oceanography and Geography, University of Gdansk, Av. Marszałka Piłsudskiego 46, Gdynia 81-378, Poland. [3]Department of Ichthyology, Faculty of Environmental Sciences, University of Warmia and Mazury, M. Oczapowskiego Str. 5, 10-718 Olsztyn, Poland. [4]Department of Zoology, Faculty of Biology and Biotechnology, University of Warmia and Mazury, M. Oczapowskiego Str. 5, 10-718 Olsztyn, Poland. [5]Museum of Natural History, Cascina Vigna, Via S. Francesco di Sales, 188, Carmagnola, Italy. [6]Laboratory of Fish Genetics, Institute of Animal Physiology and Genetics, Czech Academy of Sciences, Rumburská 89, Liběchov 277 21, Czech Republic. [7]Institut Botànic de Barcelona (IBB-CSIC-ICUB), Passeig del Migdia s/n, 08038 Barcelona, Catalonia, Spain. [8]Laboratory of Molecular Epigenetics, Institute of Biophysics, Czech Academy of Science, Kralovopolska 135, 61265 Brno, Czech Republic.

Received: 4 November 2016 Accepted: 9 May 2017
Published online: 18 May 2017

References

1. Nelson JS, Grande TC, Wilson MVH. Fishes of the World, 5th Ed. Hoboken, New Jersey, USA, Wiley; 2016.

2. FishBase, a global information on fishes. http://www.fishbase.org. Accessed 31 Mar 2017.
3. Lucentini L, Palomba A, Gigliarelli L, Sgaravizzi G, Lancioni H, Lanfaloni L, et al. Temporal changes and effective population size of an Italian isolated and supportive-breeding managed northern pike (*Esox lucius*) population. Fish Res. 2009;96:139–47.
4. Maes GE, VAN Houdt JKJ, De Charleroy D, Volckaert FAM. Indications for a recent holarctic expansion of pike based on a preliminary study of mtDNA variation. J Fish Biol. 2003;63:254–9.
5. Jacobsen BH, Hansen MM, Loeschcke V. Microsatellite DNA analysis of northern pike (*Esox lucius* L.) populations: insights into the genetic structure and demographic history of a genetically depauperate species. Biol J Linn Soc. 2005;84:91–101.
6. Ishiguro NB, Miya M, Nishida M. Basal euteleostean relationships: a mitogenomic perspective on the phylogenetic reality of the "Protacanthopterygii". Mol Phylogenet Evol. 2003;27:476–88.
7. Rab P, Flajshans M, Ludwig A, Lieckfeldt D, Ene C, Rabova M, et al. The second highest chromosome count among vertebrates is associated with extreme ploidy diversity in hybrid sturgeons. Cytogenet Genome Res. 2004;106:24.
8. Animal genome size database. http://www.genomesize.com. Accessed 4 Jan 2017.
9. Rondeau EB, Minkley DR, Leong JS, Messmer AM, Jantzen JR, von Schalburg KR, et al. The genome and linkage map of the northern pike (*Esox lucius*): conserved synteny revealed between the salmonid sister group and the Neoteleostei. PLoS One. 2014;9:e102089.
10. Sutherland B, Gosselin T, Normandeau E, Lamothe M, Isabel N, Audet C, et al. Novel method for comparing RADseq linkage maps reveals chromosome evolution in Salmonids. Genome Biol Evol. 2016;8:3600–17.
11. Rab P, Crossman EJ. Chromosomal NOR phenotypes in north-American pikes and pickerels, genus *Esox*, with notes on the Umbridae (Euteleostei, Esocae). Can J Zool. 1994;72:1951–6.
12. Phillips R, Rab P. Chromosome evolution in the Salmonidae (Pisces): an update. Biol Rev Camb Philos Soc. 2001;76:1–25.
13. Ocalewicz K, Woznicki P, Jankun M. Mapping of rRNA genes and telomeric sequences in Danube salmon (*Hucho hucho*) chromosomes using primed in situ labeling technique (PRINS). Genetica. 2008;134:199–203.
14. Symonova R, Majtanova Z, Sember A, Staaks GB, Bohlen J, Freyhof J, et al. Genome differentiation in a species pair of coregonine fishes: an extremely rapid speciation driven by stress-activated retrotransposons mediating extensive ribosomal DNA multiplications. BMC Evol Biol. 2013;13:42.
15. Fujiwara A, Abe S, Yamaha E, Yamazaki F, Yoshida MC. Chromosomal localization and heterochromatin association of ribosomal RNA gene loci and silver-stained nucleolar organizer regions in salmonid fishes. Chromosom Res. 1998;6:463–71.
16. Pieler T, Hamm J, Roeder RG. The 5S gene internal control region is composed of three distinct sequence elements, organized as two functional domains with variable spacing. Cell. 1987;48:91–100.
17. Wicke S, Costa A, Munoz J, Quandt D. Restless 5S: the re-arrangement(s) and evolution of the nuclear ribosomal DNA in land plants. Mol Phylogenet Evol. 2011;61:321–32.
18. Vierna J, Gonzalez-Tizon AM, Martinez-Lage A. Long-term evolution of 5S ribosomal DNA seems to be driven by birth-and-death processes and selection in Ensis razor shells (Mollusca: Bivalvia). Biochem Genet. 2009;47:635–44.
19. Cioffi MB, Martins C, Bertollo LA. Chromosome spreading of associated transposable elements and ribosomal DNA in the fish *Erythrinus erythrinus*. Implications for genome change and karyoevolution in fish. BMC Evol Biol. 2010;10:271.
20. Symonova R, Matjanova Z, Korinkova T, Jankun M, Dion-Coté A-M, Bernatchez L, et al. Chromosomal characteristics of rDNA genes in salmonid fishes: trends in their patterns and evolution. In: 20th International colloquium on animal cytogenetics and gene mapping, Cordoba, Spain 2012. Cordoba: Chromosom Res; 2012. p. 810.
21. Sember A, Bohlen J, Slechtova V, Altmanova M, Symonova R, Rab P. Karyotype differentiation in 19 species of river loach fishes (Nemacheilidae, Teleostei): extensive variability associated with rDNA and heterochromatin distribution and its phylogenetic and ecological interpretation. BMC Evol Biol. 2015;15:251.
22. Maneechot N, Yano CF, Bertollo LA, Getlekha N, Molina WF, Ditcharoen S, et al. Genomic organization of repetitive DNAs highlights chromosomal evolution in the genus *Clarias* (Clariidae, Siluriformes). Mol Cytogenet. 2016;9:4.
23. Rebordinos L, Cross I, Merlo A. High evolutionary dynamism in 5S rDNA of fish: state of the art. Cytogenet Genome Res. 2013;141:103–13.
24. Nieto Feliner G, Rossello JA, et al. Concerted evolution of multigene families and homeologous recombination. In: Plant Genome Diversity, vol. 1. Wien: Springer; 2012. p. 171–94.
25. Dover GA. Molecular drive: a cohesive mode of species evolution. Nature. 1982;299:111–7.
26. Eickbush TH, Eickbush DG. Finely orchestrated movements: evolution of the ribosomal RNA genes. Genetics. 2007;175:477–85.
27. Nei M, Rooney AP. Concerted and birth-and-death evolution of multigene families. Annu Rev Genet. 2005;39:121–52.
28. Pinhal D, Yoshimura TS, Araki CS, Martins C. The 5S rDNA family evolves through concerted and birth-and-death evolution in fish genomes: an example from freshwater stingrays. BMC Evol Biol. 2011;11:151.
29. Matyasek R, Renny-Byfield S, Fulnecek J, Macas J, Grandbastien MA, Nichols R, et al. Next generation sequencing analysis reveals a relationship between rDNA unit diversity and locus number in *Nicotiana diploids*. BMC Genomics. 2012;13:722.
30. Dubcovsky J, Dvorak J. Ribosomal RNA multigene loci–nomads of the Triticeae genomes. Genetics. 1995;140:1367–77.
31. Britton-Davidian J, Cazaux B, Catalan J. Chromosomal dynamics of nucleolar organizer regions (NORs) in the house mouse: micro-evolutionary insights. Heredity. 2012;108:68–74.
32. Raskina O, Belyayev A, Nevo E. Activity of the En/Spm-like transposons in meiosis as a base for chromosome repatterning in a small, isolated, peripheral population of *Aegilops speltoides* Tausch. Chromosom Res. 2004;12:153–61.
33. West C, James SA, Davey RP, Dicks J, Roberts IN. Ribosomal DNA sequence heterogeneity reflects intraspecies phylogenies and predicts genome structure in two contrasting yeast species. Syst Biol. 2014;63:543–54.
34. Wasko AP, Martins C, Wright JM, Galetti Jr PM. Molecular organization of 5S rDNA in fishes of the genus *Brycon*. Genome. 2001;44:893–902.
35. Kirtiklis L, Ocalewicz K, Wiechowska M, Boron A, Hliwa P. Molecular cytogenetic study of the European bitterling *Rhodeus amarus* (Teleostei: Cyprinidae: Acheilognathinae). Genetica. 2014;142:141–8.
36. Jankun M, Woznicki P, Dajnowicz G, Demska-Zakes K, Luczynski MJ, Luczynski M. Heterochromatin and NOR location in Northern Pike (*Esox lucius*). Aquat Sci. 1998;60:17–21.
37. Komiya H, Takemura S. Nucleotide sequence of 5S ribosomal RNA from rainbow trout (*Salmo gairdneri*) liver. J Biochem. 1979;86:1067–80.
38. Koukalova B, Moraes AP, Renny-Byfield S, Matyasek R, Leitch AR, Kovarik A. Fall and rise of satellite repeats in allopolyploids of Nicotiana over c. 5 million years. New Phytol. 2010;186:148–60.
39. Goecks J, Nekrutenko A, Taylor J, Team G. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol. 2010;11:R86.
40. Noe L, Kucherov G. YASS: enhancing the sensitivity of DNA similarity search. Nucleic Acids Res. 2005;33:W540–3.
41. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser. 1999;41:95–8.
42. Guo X, Han F. Asymmetric epigenetic modification and elimination of rDNA sequences by polyploidization in wheat. Plant Cell. 2014;26:4311–27.
43. Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics. 2003;19:2496–7.
44. Gruber AR, Lorenz R, Bernhart SH, Neubock R, Hofacker IL. The Vienna RNA websuite. Nucleic Acids Res. 2008;36:W70–4.
45. Miya M, Takeshima H, Endo H, Ishiguro NB, Inoue JG, Mukai T, et al. Major patterns of higher teleostean phylogenies: a new perspective based on 100 complete mitochondrial DNA sequences. Mol Phylogenet Evol. 2003;26:121–38.
46. Smirnov AV, Entelis NS, Krasheninnikov IA, Martin R, Tarassov IA. Specific features of 5S rRNA structure–its interactions with macromolecules and possible functions. Biochemistry (Mosc). 2008;73:1418–37.
47. Vierna J, Wehner S, Honer zu Siederdissen C, Martinez-Lage A, Marz M. Systematic analysis and evolution of 5S ribosomal DNA in metazoans. Heredity. 2013;111:410–21.
48. Prokopowich CD, Gregory TR, Crease TJ. The correlation between rDNA copy number and genome size in eukaryotes. Genome. 2003;46:48–50.
49. Stankova H, Hastie AR, Chan S, Vrana J, Tulpova Z, Kubalakova M, et al. BioNano genome mapping of individual chromosomes supports physical mapping and sequence assembly in complex plant genomes. Plant Biotechnol J. 2016;14:1523–31.

50. Emadzade K, Jang TS, Macas J, Kovarik A, Novak P, Parker J, et al. Differential amplification of satellite PaB6 in chromosomally hypervariable *Prospero autumnale* complex (Hyacinthaceae). Ann Bot. 2014;114:1597–608.
51. Gibbons JG, Branco AT, Godinho SA, Yu S, Lemos B. Concerted copy number variation balances ribosomal DNA dosage in human and mouse genomes. Proc Natl Acad Sci U S A. 2015;112:2485–90.
52. Hilder VA, Dawson GA, Vlad MT. Ribosomal 5S genes in relation to C-value in amphibians. Nucleic Acids Res. 1983;11:2381–90.
53. Grummt I, Pikaard CS. Epigenetic silencing of RNA polymerase I transcription. Nat Rev Mol Cell Biol. 2003;4:641–9.
54. Locati MD, Pagano JF, Ensink WA, van Olst M, van Leeuwen S, Nehrdich U, et al. Linking maternal and somatic 5S rRNA types with different sequence-specific non-LTR retrotransposons. RNA. 2016;4:446–56.
55. Venney CJ, Johansson ML, Heath DD. Inbreeding effects on gene-specific DNA methylation among tissues of Chinook salmon. Mol Ecol. 2016;25:4521–33.
56. Plohl M, Mestrovic N, Mravinac B. Satellite DNA evolution. Genome Dyn. 2012;7:126–52.
57. Heslop-Harrison JS, Schwarzacher T. Organisation of the plant genome in chromosomes. Plant J. 2011;66:18–33.
58. Arnheim N, Treco D, Taylor B, Eicher EM. Distribution of ribosomal gene length variants among mouse chromosomes. Proc Natl Acad Sci U S A. 1982;79:4677–80.
59. Roberts RJ, Carneiro MO, Schatz MC. The advantages of SMRT sequencing. Genome Biol. 2013;14:405.
60. Cloix C, Tutois S, Mathieu O, Cuvillier C, Espagnol MC, Picard G, et al. Analysis of 5S rDNA arrays in *Arabidopsis thaliana*: physical mapping and chromosome-specific polymorphisms. Genome Res. 2000;10:679–90.
61. Havlova K, Dvorackova M, Peiro R, Abia D, Mozgova I, Vansacova L, et al. Variation of 45S rDNA intergenic spacers in *Arabidopsis thaliana*. Plant Mol Biol. 2016;92:457–71.
62. Cazaux B, Catalan J, Veyrunes F, Douzery EJ, Britton-Davidian J. Are ribosomal DNA clusters rearrangement hotspots?: a case study in the genus *Mus* (Rodentia, Muridae). BMC Evol Biol. 2011;11:124.
63. Paule MR, White RJ. Survey and summary: transcription by RNA polymerases I and III. Nucleic Acids Res. 2000;28:1283–98.
64. Kapitonov VV, Jurka J. A novel class of SINE elements derived from 5S rRNA. Mol Biol Evol. 2003;20:694–702.
65. Kalendar R, Tanskanen J, Chang W, Antonius K, Sela H, Peleg O, et al. Cassandra retrotransposons carry independently transcribed 5S RNA. Proc Natl Acad Sci U S A. 2008;105:5833–8.
66. da Silva M, Barbosa P, Artoni RF, Feldberg E. Evolutionary dynamics of 5S rDNA and recurrent association of transposable elements in electric fish of the family Gymnotidae (Gymnotiformes): The case of *Gymnotus mamiraua*. Cytogenet Genome Res. 2016;149:297–303.
67. Cohen S, Agmon N, Sobol O, Segal D. Extrachromosomal circles of satellite repeats and 5S ribosomal DNA in human cells. Mob DNA. 2010;1:11.
68. Navratilova A, Koblizkova A, Macas J. Survey of extrachromosomal circular DNA derived from plant satellite repeats. BMC Plant Biol. 2008;8:90.
69. Hourcade D, Dressler D, Wolfson J. The amplification of ribosomal RNA genes involves a rolling circle intermediate. Proc Natl Acad Sci U S A. 1973;70:2926–30.
70. Jack CV, Cruz C, Hull RM, Keller MA, Ralser M, Houseley J. Regulation of ribosomal DNA amplification by the TOR pathway. Proc Natl Acad Sci U S A. 2015;112:9674–9.
71. Sevim V, Bashir A, Chin CS, Miga KH. Alpha-CENTAURI: assessing novel centromeric repeat sequence variation with long read sequencing. Bioinformatics. 2016;32:1921–4.
72. Kovarik A, Matyasek R, Lim KY, Skalicka K, Koukalova B, Knapp S, et al. Concerted evolution of 18-5.8-26S rDNA repeats in *Nicotiana* allotetraploids. Biol J Linn Soc. 2004;82:615–25.
73. Vinson C, Chatterjee R. CG methylation. Epigenomics. 2012;4:655–63.
74. Wang W, Ma L, Becher H, Garcia S, Kovarikova A, Leitch IJ, et al. Astonishing 35S rDNA diversity in the gymnosperm species *Cycas revoluta* Thunb. Chromosoma. 2016;125:683–99.
75. Iwasaki O, Tanaka A, Tanizawa H, Grewal SI, Noma K. Centromeric localization of dispersed Pol III genes in fission yeast. Mol Biol Cell. 2010;21:254–65.
76. Martins C, Ferreira IA, Oliveira C, Foresti F, Galetti Jr PM. A tandemly repetitive centromeric DNA sequence of the fish *Hoplias malabaricus* (Characiformes: Erythrinidae) is derived from 5S rDNA. Genetica. 2006;127:133–41.
77. Vittorazzi SE, Lourenco LB, Del-Grande ML, Recco-Pimentel SM. Satellite DNA Derived from 5S rDNA in *Physalaemus cuvieri* (Anura, Leiuperidae). Cytogenet Genome Res. 2011;134:101–7.
78. Kumke K, Macas J, Fuchs J, Altschmied L, Kour J, Dhar MK, et al. *Plantago lagopus* B Chromosome is enriched in 5S rDNA-derived satellite DNA. Cytogenet Genome Res. 2016;148:68–73.

Symonová et al. BMC Genetics (2017) 18:19
DOI 10.1186/s12863-017-0484-8

BMC Genetics

RESEARCH ARTICLE Open Access

Molecular cytogenetic differentiation of paralogs of Hox paralogs in duplicated and re-diploidized genome of the North American paddlefish (*Polyodon spathula*)

Radka Symonová[1,2*], Miloš Havelka[3], Chris T. Amemiya[4], William Mike Howell[5], Tereza Kořínková[1], Martin Flajšhans[3], David Gela[3] and Petr Ráb[1]

Abstract

Background: Acipenseriformes is a basal lineage of ray-finned fishes and comprise 27 extant species of sturgeons and paddlefishes. They are characterized by several specific genomic features as broad ploidy variation, high chromosome numbers, presence of numerous microchromosomes and propensity to interspecific hybridization. The presumed palaeotetraploidy of the American paddlefish was recently validated by molecular phylogeny and Hox genes analyses. A whole genome duplication in the paddlefish lineage was estimated at approximately 42 Mya and was found to be independent from several genome duplications evidenced in its sister lineage, i.e. sturgeons. We tested the ploidy status of available chromosomal markers after the expected rediploidization. Further we tested, whether paralogs of Hox gene clusters originated from this paddlefish specific genome duplication are cytogenetically distinguishable.

Results: We found that both paralogs HoxA alpha and beta were distinguishable without any overlapping of the hybridization signal - each on one pair of large metacentric chromosomes. Of the HoxD, only the beta paralog was unequivocally identified, whereas the alpha paralog did not work and yielded only an inconclusive diffuse signal. Chromosomal markers on three diverse ploidy levels reflecting different stages of rediploidization were identified: quadruplets retaining their ancestral tetraploid condition, semi-quadruplets still reflecting the ancestral tetraploidy with clear signs of advanced rediploidization, doublets were diploidized with ancestral tetraploidy already blurred. Also some of the available microsatellite data exhibited diploid allelic band patterns at their loci whereas another locus showed more than two alleles.

Conclusions: Our exhaustive staining of paddlefish chromosomes combined with cytogenetic mapping of ribosomal genes and Hox paralogs and with microsatellite data, brings a closer look at results of the process of rediploidization in the course of paddlefish genome evolution. We show a partial rediploidization represented by a complex mosaic structure comparable with segmental paleotetraploidy revealed in sturgeons (Acipenseridae). Sturgeons and paddlefishes with their high propensity for whole genome duplication thus offer suitable animal model systems to further explore evolutionary processes that were shaping the early evolution of all vertebrates.

Keywords: HoxA/D paralogs mapping, Sturgeon whole genome duplication, Ancient fish genome, Rediploidization

* Correspondence: radka.symonova@uibk.ac.at
[1]Laboratory of Fish Genetics, Institute of Animal Physiology and Genetics, Czech Academy of Sciences, 277 21, Liběchov, Czech Republic
[2]Research Institute for Limnology, University of Innsbruck, Mondseestr. 9, Mondsee, Austria
Full list of author information is available at the end of the article

Background

The North American (*Polyodon spathula*) and Chinese paddlefish (*Psephurus gladius*), i.e. fishes with paddle-like snout, are the only extant representatives of an early radiation of ray-finned fishes recognized as the family Polyodontidae within the order Acipenseriformes [1]. This ancient lineage, i.e. sturgeons, shovelnoses (Acipenseridae) and paddlefishes, represents an archaic group known to be at least as old as the early Jurassic (some 200–175 Mya). Polyodontids and acipenserids diverged from one-another in the Jurassic period ~180–140 Mya [2]. Paddlefishes form a distinct monophyletic clade within Acipenseriformes as evidenced by molecular phylogeny [3].

There is a consistent body of evidence on the chromosome number of the North American paddlefish but discordant data on its genome size, yet all authors consider this species palaeotetraploid. Namely, Dingerkus and Howell [4] based on karyotyping of 31 cells from two paddlefish males found the 2n = 120 and reported a karyotype apparently composed of twelve quadruplets of macro- and 72 microchromosomes and thus hypothesized a tetraploid origin in this species. Later, the 2n = 120 was reconfirmed and nuclear DNA content corresponding to genomes of sturgeons with 2n = 120 was reported by Zhang et al. [5]. Birstein et al. [6] re-examined nuclear DNA content of this species and confirmed its palaeotetraploid status. The Chinese paddlefish (*Psephurus gladius*) has 2n = 120 and nuclear DNA content corresponding to sturgeon genomes with 2n = 120 was reported [5]. All available data indicate a palaeotetraploid origin of the paddlefish lineage, similarly as in a number of sturgeon species [7].

The cytogenetic analyses of sturgeons and paddlefishes is challenging since, in addition to macrochromosomes, extremely small-sized microchromosomes typically constitute a substantial proportion of the acipenseriform karyotype and their morphology has long been characterized as indistinguishable [8]. However, a recent work [9] reports morphology of these small-sized chromosomes that would have been called "microchromosomes". Further details on sturgeon genetics and cytogenetics related to ploidy level were summarized by Havelka et al. [10] and Trifonov et al. [11].

Sturgeons generally exhibit a remarkable propensity for hybridization and polyploidization resulting in viable and even fertile highly polyploid individuals and interspecific (allopolyploid) hybrids [12–15]. Each such whole genome duplication (WGD) is followed by a rediploidization [16]. During this process, a partial retention of 4n features is observable, like chromosome numbers and their external morphology (e.g. [17]). Distinct signs of ongoing and to different extent advanced rediploidization can be found in other cytogenomic markers – e.g. HoxA/D sequences in paddlefish [12] and microsatellite studies in sturgeon [18, 19].

Recently, molecular analyses of paralogs of both HoxA and HoxD gene clusters demonstrated a specific WGD event in the paddlefish lineage dated about 42 million years ago (Mya) (for more details on dating [12]) and independent from the multiple WGDs in sturgeons [19]. Other comparable events among ray-finned fish are the salmonid specific WGD (called SR) dated about 88 Mya [20] and the teleost specific WGD (TSGD) dated about 320 Mya [21]. Paddlefish thus offers a model system complementary to others to study consequences of WGD by molecular cytogenetics in the light of clear indications of secondary rediploidization at the molecular level in HoxA/D genes clusters (e.g. [12]).

Hox genes clusters are expressed along the anteroposterior axis of all bilaterians and play a key role in animal development [22]. In vertebrates, four paralogous Hox clusters (Hox A, B, C, and D) originate from a single Hox cluster of the last common ancestor of vertebrates and cephalochordates ([23], [24] reviewed in [25]). These four paralogs arose by two ancestral consecutive WGD events known as R1 and R2 that occurred in the vertebrate stem lineage around 525 Mya [24, 26]. The paddlefish specific WGD gave rise to a further order of paralogs called here α and β, respectively, as introduced by [12].

In this study, using FISH (fluorescence in situ hybridization) we co-hybridized the HoxAα with HoxAβ and HoxDα with HoxDβ of these paddlefish specific paralogs (BAC DNA produced by Crow et al. [12]) to paddlefish chromosomes. This was combined with FISH with ribosomal (rRNA) genes, with conventional chromosome banding and microsatellite analysis to explore consequences of the paddlefish specific WGD and the subsequent rediploidization. The available markers and analyses, including those based on external chromosomal morphology, enabled us to assess the tetraploidy retention versus the extent of rediploidization on the level of coding regions, repetitive sequences, heterochromatin accumulations supplemented with microsatellite analysis.

Results

Karyotype and chromosome analysis

Karyotype analysis re-confirmed the diploid chromosome number 2n = 120 and enabled an arrangement of chromosomes into quadruplets, doublets and semi-quadruplets where possible (Fig. 1). This arrangement reflected the external morphological features of chromosomes and also information retrievable from the DAPI (i.e., AT-rich regions)/CMA_3 (Chromomycin A3, GC-rich regions specific) fluorescent staining. Firstly, a CMA_3^+ arm of a pair of small sub metacentric macrochromosomes (Fig. 1b). Secondly, $DAPI^+$

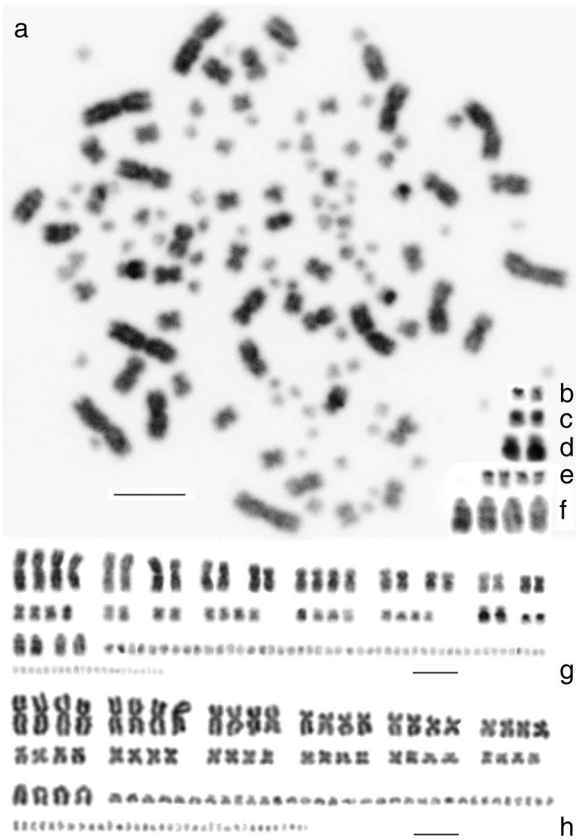


Fig. 1 Summary of karyological data. DAPI stained metaphase reversed to *black* and *white* (**a**), Chromomycin A_3 positive small submetacentric chromosomes, reversed to *black* and *white* (**b**), DAPI positive small submetacentric chromosomes (**c**), DAPI positive middle-sized subtelocentric chromosomes (**d**), AgNOR positive chromosomes (**e**), acipenseriform marker chromosomes - the largest acrocentric chromosomes, the first two of them with two distinct DAPI positive bands (**f**). Karyotype of the DAPI stained metaphase with macrochromosomes arranged into quadruplets, semi-quadruplet and duplets based on the DAPI signal (**g**). Karyotype of Giemsa stained chromosomes with macrochromosomes arranged only into quadruplets (**h**). Scale bars equal 10 μm

arm of a small submetacentric chromosomes (Fig. 1c). Thirdly, an interstitial $DAPI^+$ band on the larger arm of a pair of submetacentric chromosomes (Fig. 1d). DAPI staining demonstrated that there is no other pair of submetacentric chromosomes with comparable external morphology and banding pattern. Further, two pairs of the largest "acipenseriform" acrocentric chromosome markers were distinguishable based on their morphology in the tetraploid condition. However, the DAPI fluorescence revealed AT-rich signals accumulated in form of a double-band in only two of these four chromosomes (Fig. 1f). The $AgNO_3$ staining (visualizing chromosomes with the active ribosomal genes, the AgNORs) consistently yielded a quadruplet of transcriptionally active major ribosomal sites on four small metacentric macrochromosomes (Fig. 1e). The two CMA^{3+} signals (Fig. 1b) did not co-localize with the active NOR sites (Fig. 1e).

FISH experiments

The FISH with 28S ribosomal DNA (rDNA) showed up to 12 signals on mostly small-sized macrochromosomes or microchromosomes (Fig. 2a). The FISH with 5S rDNA revealed signals on three morphologically different pairs of chromosomes (Fig. 2b): i) small dot-like signals interstitially on one arm of a pair of large metacentric chromosomes; ii) large signals on a whole arm of a pair of small metacentric macrochromosomes, and iii) dot-like signals on a pair of small metacentric macrochromosomes. FISH with telomeric repeats (TTAGGG) n did not reveal any interstitial telomeric sites (not shown).

FISH with HoxAα/β and HoxDβ gene clusters, respectively, showed a clear 2n pattern of the respective paralogs (Fig. 2c-d). Namely, the HoxA α paralog was physically mapped to the telomeric region of a pair of large metacentric chromosomes (Fig. 2c, red signals), the HoxA β paralog mapped also to telomeric region of another pair of large metacentric chromosomes (Fig. 2c, green signals) with no overlapping signals. The HoxD β paralog mapped to the telomeric region of one pair of the largest acrocentric chromosomes (Fig. 2d, red

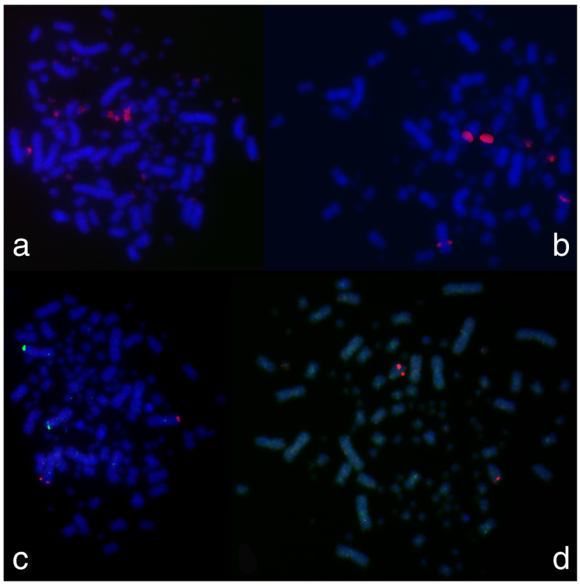


Fig. 2 FISH with rDNAs and Hox clusters. FISH with 28S rDNA (**a**, *red*), 5S rDNA (**b**, *red*) and FISH with BACs of the paralogs Hox-A alpha (**c**, *red*) and Hox-A beta (**c**, *green*) and of the paralogs Hox-D beta (**d**, *red*) and Hox-D alpha (**d**, *green* – the diffuse character of this signal causes a greenish coloration of almost all chromosomes). Scale bars equal 10 μm

signals). The HoxD α paralog yielded an inconclusive interspersed FISH signal and it was impossible to localize this region even with using competitor DNA (Fig. 2d, green and diffuse greenish signals on majority of chromosomes).

Integration of karyological and molecular cytogenetic markers

Integrating all analyzed markers, the following groups of chromosomes (labeled with abbreviations for later reference) were identified to compose the karyotype: M1 quadruplet of the largest metacentric chromosomes of equal size (according to external morphology); M2 semi-quadruplet of slightly smaller large metacentric chromosomes based on HoxA paralogs mapping; M3 semi-quadruplet of meta- to submetacentric chromosomes of intermediate size including one smaller and one larger doublet respectively based on size and 5S rDNA FISH signals; SM1 semi-quadruplet of submetacentric chromosomes of intermediate size; SM2 semi-quadruplet; SM3 semi-quadruplet; SM4-5 (semi-quadruplets/doublets of small meta- to submetacentric chromosomes, these chromosomes are difficult to classify into separate groups and pair with the exception of the AgNOR quadruplet; ST1 semi-quadruplet/doublet of a conspicuous pair of submeta- to subtelocentric chromosomes with a remarkable $DAPI^+$ band on the larger arm. This pair of chromosomes lacks any counterpart in its size and morphological category, however, it can be linked with a pair of small metacentric chromosomes with a distinct DAPI band on one arm; A1 semi-quadruplet of medium-size acrocentric chromosomes (the acipenseriform acrocentric markers). There is a $DAPI^+$ double-band on two of these chromosomes and the HoxD β paralog FISH signal on the other two chromosomes; m1 doublet of very small chromosomes (microchromosomes) with a clear CMA_3^+ band; m2 doublet of very small chromosomes (microchromosomes) with a clear $DAPI^+$ band. In this way, we identified 54 macrochromosomes (or microchromosomes with features enabling their identification) and 66 microchromosomes (i. e. small chromosomes without any cytogenetic features).

Contrary to these observations yielded by fluorescent stainings, AgNOR staining and C-banding, the Giemsa staining produced only rough and uniform external chromosomal morphology corresponding approximately to tetraploid condition. There were no apparent characteristics enabling any possibility to distinguish the aforementioned quadruplets, semi-quadruplets and duplets (Fig. 1h).

All cytogenetic data are visualized and summarized in the Fig. 3.

To sum up, the ancestrally palaeotetraploid paddlefish chromosomal complement (2n = 120) represents a complex mosaic structure consisting of characters with 1/ retained tetraploidy (AgNORs); 2/partially retained tetraploidy, i.e. apparent ancestral tetraploidy with clear signs of secondary diploidization (acipenseriform acrocentric marker, HoxA and HoxD loci); 3/diploidized characters ($DAPI^+$ and CMA_3^+ heterochromatic sites); and 4/characters with apparent chromosomal re-arrangements (5S rDNA sites). The situation in 5S rDNA sites did not match neither to ancestral tetraploidy nor to the diploidized condition.

Quantification of repetitive sequences in BAC clones used as FISH probes

Proportion of repeats in the BAC DNA used as FISH probes can be potentially crucial for functionality of

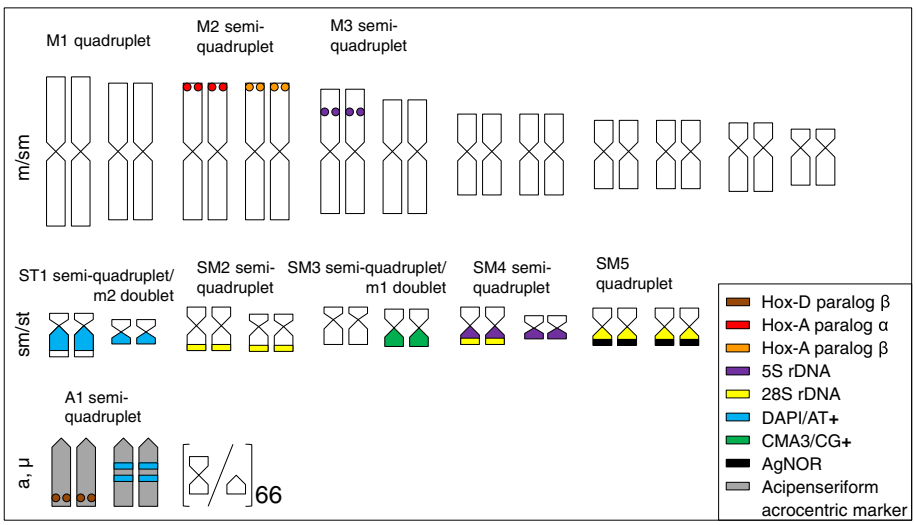


Fig. 3 Ideogram summarizing all chromosomal markers investigated in this study

FISH experiments since repeats interspersed throughout genome can yield signals unspecific for the BAC DNA under study. In Table 1, we summarize repeats proportion for each BAC DNA.

Microsatellite genotyping

From all together eleven tested sturgeon microsatellite markers, only the locus Afu 68 [27] displayed consistent amplification in *P. spathula*. This locus exhibited diploid allelic band pattern across all analyzed samples. Similarly, *P. spathula* specific loci (see Additional file 1, Supplementary Methods) had diploid allelic band patterns. An exception to this general pattern was shown at the locus Psp-29 [28] for which more than two alleles were observed (Fig. 4; Additional file 2, Supplementary Results).

Discussion

This study demonstrated an in situ localization of three of four available paralogs of the HoxA and HoxD gene clusters in *P. spathula*. Both HoxA paralogs were clearly distinguishable without any FISH signal overlaps showing that the sequence divergence was high enough to enable their unequivocal discrimination under high stringency conditions. The sequence divergence between the full coding regions varied among gene clusters ranging from 2.12% (Hox-A11) to 10.94% (Hox-D13) [12]. The proportion of repetitive sequences was identified to be approximately 10% where ascertainable (Table 1). The repetitive sequences, if specific for the Hox clusters, might have contributed to discrimination of Hox paralogs. On the other hand, the failure in localization of the HoxD α can also be ascribed to accumulation of repetitive sequences that were spread also throughout the rest of genome and thus prevented from an accurate localizing of this paralog. This result contributes as a comparative reference to the future attempts to localize paralogous regions and presents an example of linking genomic approach with molecular cytogenetics. Further, our results made it possible to better assess genome evolution after the paddlefish specific WGD on the finer-scale level in combination with other cytogenetic markers and to compare it with the outer chromosomal morphology based on conventional cytogenetics as shown by [4].

Ray-finned fishes provide an outstanding model to investigate WGDs and their consequences because of their complex evolutionary history involving among vertebrates unprecedented propensity for hybridization and polyploidization and their genome plasticity tolerating high variability in genome size and chromosome number [29].

Two basic ways of chromosomal evolution following a WGD event were documented among ray-finned fishes: 1/conserving of the chromosome numbers after WGD which exists in more than 50% of teleosts after the TSGD [30], in polyploid cyprinid lineages [31] and in most acipenseriforms [19], and 2/extensive chromosomal re-arrangements leading to diversified chromosome numbers, e.g. in salmonids (reviewed by [32]). Other lineages known to have experienced WGD [33] need to be cytogenetically documented.

Acipenserids, the sister lineage of polyodontids, underwent at least three rounds of lineage specific WGDs [19, 34, 35]. The first one occurred in their common (already extinct) ancestor with $2n = 60$ leading to a ~120-chromosomes lineage with presently unknown dating. The second lineage specific WGD took place separately in the Atlantic sturgeon lineage (~53 Mya) and in the Pacific lineage (~70 Mya). The third WGD is unique to the Shortnose sturgeon (*Acipenser brevirostrum*) dated ~35 Mya [2]. *Polyodon spathula* inhabits the same range as some species of the Atlantic sturgeon lineage. The *P. spathula* specific WGD is supposed to have occurred ~42 Mya [12] which is a timing somewhat similar to the WGD of the Atlantic sturgeon lineage [2]. Acipenserids split from polyodontids ~ 170 Mya and the divergence time between the Pacific and the Atlantic lineage appears as about 121 Mya [2]. Assuming WGD specific to *P. spathula* [12] and no presence of WGD specific to Atlantic lineage species with 120 chromosomes [2], the first specific WGD in Acipenseridae had to take place between split of polyodontids from acipenserids (~170 Mya) and split of Atlantic and Pacific lineage of Acipenseridae (~121 Mya; Fig. 5).

Chromosome numbers and their outer morphology in both acipenserids and polyodontids remain mostly conserved in the post-WGD situation [4]. Regarding the longevity estimated in the paddlefish WGD, this conservatism might be explained by two possible factors and/or by combination thereof: 1/significantly reduced rate

Table 1 Overview of BACs of Hox paralogs used as FISH probes

Clone ID *sensu* [12]	BAC	Paralogs	GenBank Accession Nr.	Size (bp)	Repeats (%)
BAC231C24	1816	Hox-D 8-13α	JX280946.1	22.134	5.57
BAC249G23	1817	Hox-D 8-13β	JX280945.1	33.595	9.17
BAC352P4	1818	Hox-A 1-13α	JX448769.1	131.867	10.66
BAC370N10	1819	Hox-A 1-13β	JX448770.1	139.159	10.48

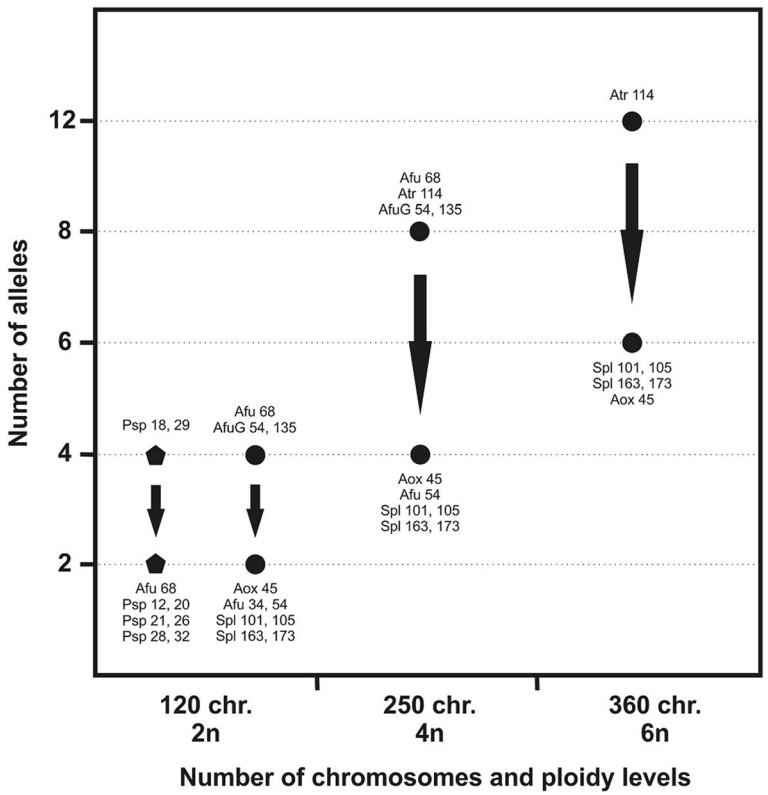


Fig. 4 Microsatellite analysis of ploidy level. Microsatellite data (number of alleles) in *Polyodon spathula* and comparison with sturgeons related to chromosomes numbers and ploidy levels provide evidence of partial genome rediploidization in *P. spathula* (⬟) and sturgeons (●). The figure is based on data from this study and previous studies [18, 28]

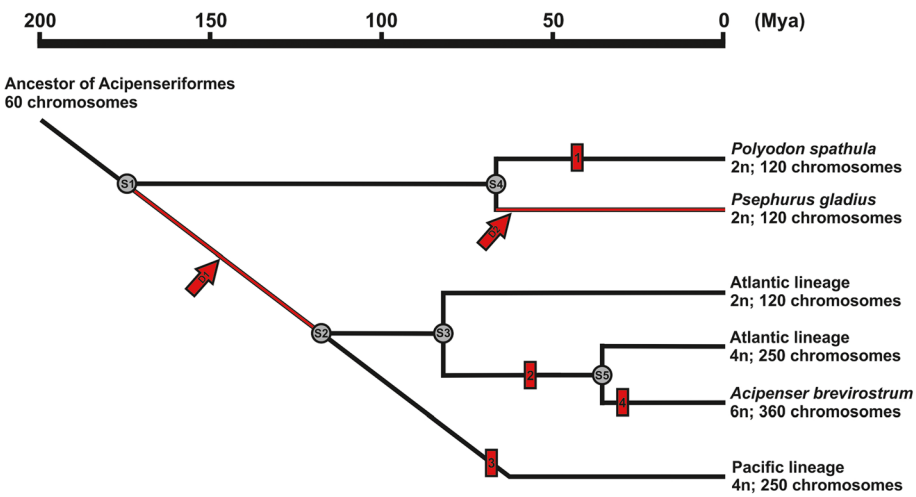


Fig. 5 Suggested main events in evolution of Acipenseriformes. S1 = split of Polyodontidae and Acipenseridae (~170 Mya); S2 = Split of Atlantic and Pacific lineage in Acipenseridae (~121 Mya); S3 = split of 2n and 4n species within Atlantic lineage (~80 Mya); S4 = Split of *Polyodon spathula* and *Psephurus gladius* (~68Mya); S5 = Split of *Acipenser brevirostrum* (~36 Mya). Whole genome duplication (WGD) events are: 1 = WGD specific to *Polyodon spathula* (60 ⟶ 120 chromosomes;~ 42 Mya [12]); 2 = WGD in Atlantic lineage (120 ⟶ 250 chromosomes; ~ 53 Mya); 3 = WGD in Pacific lineage (120 ⟶ 250 chromosomes; ~ 70 Mya); 4 = WGD specific to *Acipenser brevirostrum* (250 ⟶ 360 chromosomes ~ 35 Mya); D1 = the first WGD in Acipenseridae (60 ⟶ 120 chromosomes) had to take place between ~ 170 Mya and ~ 121 Mya; D2 = probable WGD (60 ⟶ 120 chromosomes) specific to *Psephurus gladius*. The data are based on study by Peng et al. [2] and Crow et al. [12]

of molecular evolution (both mitochondrial and nuclear) that was reported in both coding and non-coding sequences as well as in chromosomal evolution [36, 37]. It can be assumed that also the rate of secondary diploidization in paddlefish and acipenserids might have been influenced by this generally reduced rate of evolution; 2/ morphology of chromosomes undergoing WGD should be taken into account. Once exclusively or mostly metacentric chromosomes undergo a WGD (as in acipenseriforms and also in cyprinids), there is a narrower spectrum of potential chromosome rearrangements to be employed within the duplicated sets of chromosomes. This factor might have contributed to the situation shown here on paddlefish chromosomes. Whereas e.g. in salmonids, chromosomes of the hypothetical 2n ancestor might have been composed of mostly acrocentric chromosomes [38] which facilitates chromosomal evolution via Robertsonian fusions. Such mechanism was recently described in human acrocentrics harboring rDNA - under certain circumstances acrocentrics physically link via their centromeres and become fusogenic [39]. From our previous studies we know that salmonids possess amplified to extremely amplified rDNA sites mostly in centromeres of acrocentric chromosomes [40, 41]. These might have been the factors contributing to the numerous centric fusions described in all salmonid lineages [32] and at the same time contributing to the conservatism preserving the post-WGD chromosome number in the paddlefish. However, more detailed analyses on the finer scale as shown here by molecular cytogenetic approach or by sequencing of Hox genes paralogs [12] revealed advanced stages of diploidization in coding (HoxA/D gene clusters) and some non-coding regions (accumulation of AT^+- and GC^+-rich heterochromatic regions; microsatellites). This indicates dynamic processes on multiple levels in the acipenseriform lineage despite the chromosomal morphology prevented from Robertsonian fusions. On the other hand, as evidenced by the persisting of tetraploid state in AgNORs and four active nucleoli per nucleolus, it is advantageous to retain the elevated structural and functional ploidy level in rRNA genes. The multiplied number of non-active 28S rDNA sites on mostly small chromosomes suggests tendencies to spread these regions across chromosomes as evidence in other fishes (e.g. [41, 42]). Our findings of a complex mosaic of diploid, tetraploid and intermediary chromosomal features are in line with the concept of segmental paleotetraploidy revealed recently in the sterlet (*Acipenser ruthenus*) by chromosome painting [9].

In paddlefish, the re-diploidization process on the karyotype level might have occurred without a change of chromosome counts but by finer-scale re-arrangements of some chromosomes. The fact that some chromosomes and markers still retain the quadruplet nature (i.e. ancestral 4n = 120), supports the hypothesis that the diploid ancestor before the WGD event possessed 2n = 60 as proposed for the whole group of Acipenseriformes [4, 7].

For further comparisons of consequences of WGDs, more recent entirely polyploid lineages and families within Cypriniformes [33] as e.g. *Cyprinus carpio* (about 12 Mya; David et al. 2003) are available among ray-finned fish. This would be one of the most recent genome duplications among vertebrates indicating a higher incidence of WGD in fishes than in other vertebrate groups [43]. These authors report partially duplicated genome structures and disomic inheritance despite clearly tetraploid chromosome number (2n = 100) in *C. carpio* demonstrating the complexity of genome evolution in this group. There are also reports on recurrent allopolyploidization events within the *Carassius* complex [44] proving thus the suitability of ray-finned fish that provide almost a continuum in WGD events on the time scale. However, there are no comparable results available yet to assess the process of rediploidization.

Origin of the acipenseriform acrocentric chromosome marker and its relation to the ploidy level

One of the four Hox clusters - the HoxD β paralog - mapped to two of the four largest acrocentric chromosomes while both of the HoxAα/β paralogs to four large metacentric chromosomes (the M2 group). Therefore we assume that the largest acrocentric chromosomes originated from one of the large metacentric chromosomes that also represent the ancestral location of Hox clusters. These acipenseriform acrocentric marker chromosomes are known to reflect the ploidy level and used to ploidy level estimation (*sensu* [13]). They occur in all sturgeons (details on the online database, http://sveb.unife.it/it/ricerca-1/laboratori/geneweb). Presumably, the Hox clusters were residing on one pair of chromosomes before the 1R, i.e. the first vertebrate WGD (as shown by [45] in *Amphioxus*) and on two pairs before the 2R (the second vertebrate WGD). Hence, there were four pairs of Hox bearing chromosomes before the paddlefish specific WGD, which resulted in eight pairs (i.e. HoxA-D/α-β; here, we have investigated four of them and localized three of them). The largest acrocentric chromosomes might thus have arisen by fission of one pair of the metacentric chromosomes. This is in line with the assumption that the numerous metacentric chromosomes represent the ancestral chromosomal morphology. The few acrocentric chromosomes may represent rather exceptional derivatives of metacentrics of more recent origin. This scenario is supported by the fact that no interstitial telomeric sites were found in paddlefish and therefore nothing suggests the opposite way of origin of metacentric chromosomes by centric fusions as proposed by Birstein [46]. This

scenario finds its support in *Acipenser ruthenus*, where a single acrocentric pair shares FISH signals of a painting probe derived from the seventh pair of the large metacentrics [9]. This suggests that the fission of the ancestral pair of metacentric chromosomes might have happened already before the split of polyodontids and acipenserids or repeatedly in both lineages. These authors explicitly deny the presence of the second pair of the largest acrocentrics in individuals they investigated [9]. On the other hand, the other two largest acrocentric chromosome are clearly observable in *A. ruthenus* as published by other authors (see the above mentioned online database maintained by F. Fontana). Hence, there are apparent lineage specific trends and differences in the chromosomal evolution between the polyodontids and acipenserids.

Microsatellite analyses are in line with cytogenetic findings

Microsatellite data in Acipenseriformes were recently presented by Havelka et al. [18]. They identified sturgeon species in palaeotetraploid condition that were functionally diploid showing a diploid allelic band pattern in some microsatellite loci whereas residual tetraploid pattern in other ones. They further identified functionally tetraploid palaeooctaploids showing tetraploid patterns in some loci and residual octaploid patterns in other ones. Finally, a special situation was described in the functionally hexaploid palaeododecaploid *A. brevirostrum* which also showed hexaploid vs. residual dodecaploid patterns (for details [18]). Such observation of coexistence of diploid and tetraploid or tetraploid and octaploid allelic band patterns in one genome of sturgeon species might reflect functional rediploidization as an ongoing process in this fish lineage (e.g. [19]). This process is expected in polyploids until their complete rediploidization [16]. However, even in fully diploidized genomes, residual evidence for polyploid ancestry (e.g. residual polysomy) is occasionally observed (e.g. in salmonids, [47]). Since paleotetraploid acipenseriform species were considered to be basal group of Acipenseriformes [3], the process of rediploidization probably reaches further than in paleooctaploid species [19]. In light of all these facts, the observation of the duplicated locus Psp-29 in this study and coexistence of diploid and tetraploid allelic band patterns reported by Heist et al. [28] at several microsatellite loci of *P. spathula* supported our observation of partial rediploidization in *P. spathula* genome from the molecular point of view.

The estimation of locus ploidy by microsatellite genotyping may suffer from inbreeding of analyzed individuals. This might be the case of our microsatellite data as all analyzed individuals originated from a pet shop and we could not exclude their relatedness. Except the locus *Psp 18*, the estimated ploidy of analyzed loci was in accordance with Heist et al. [28] providing confidence for our conclusion based on microsatellite data. Heist et al. [28] suggested tetraploidy for the locus *Psp 18*, while we did not observe more than two alleles at the locus (Additional file 2, Supplementary Results). This inconstancy may be caused by close relatedness of analyzed individuals in our study.

Our study presented here and the recently published work [9] performed in sterlet (*Acipenser ruthenus*) represent the first steps towards a better understanding of processes involved in the rediploidization after WGD events in Acipenseriformes. Both works, although utilizing slightly different approaches of molecular cytogenetics, intersect in presenting a complex mosaic structure consisting of 2n and 4n chromosomes segments referred to as "segmental paleotetraploidy" by [9]. Both works thus show the complexity and at the same time the importance of this issue.

Conclusions

We have shown that the paddlefish *Polyodon spathula*, in which the chromosome numbers and chromosome morphology remain mostly conserved in the post-WGD situation, do nevertheless show signs of ongoing rediploidization. By combining karyological and molecular cytogenetic markers we were able to distinguish three diverse ploidy levels: tetraploidy (AgNORs), partially retained tetraploidy with secondary diploidization (acipenseriform acrocentric marker, HoxA and HoxD loci) and diploidized characters ($DAPI^+$ and CMA_3^+ heterochromatic sites). Accordingly, the altogether 120 chromosomes can be arranged into quadruplets, semi-quadruplets and doublets. We suggest that paddlefishes are similar to their relatives, sturgeons, in their propensity for genome duplication and subsequent rediploidization, and that both groups have a good prospect as vertebrate models for further exploration of these processes.

Methods

Material and metaphase chromosome preparation

Eight individuals of *P. spathula* of unknown sex examined cytogenetically in this study are summarized in the Table 2. They were sacrificed by overdose of anaesthetic 0.5% Phenoxyethanol (v/v, SIGMA), blood was taken for leucocytes cultivation and fin clips for microsatellite genotyping. Only a fraction of samples yielded metaphase spreads of sufficient quality to be used in molecular cytogenetic analyses.

All fish examined in this study, chromosome preparations, and DNA and tissue samples are deposited in the Laboratory of Fish Genetics, Institute of Animal Physiology and Genetics, Czech Academy of Science (LFG, IAPG, CAS), Liběchov as voucher specimens and reference samples.

Table 2 List of Polyodon spathula specimens analysed cytogenetically

Indivuals processed for chromosomes	Origin of specimens	Individuals used for molecular cytogenetics (# metaphases)
1–7/2008	University of South Bohemia in Ceske Budejovice	1 (15)/2008
1–6/2009	University of South Bohemia in Ceske Budejovice	1 (5), 5 (10)/2009
1–8/2012	Pet shop in Mlada Boleslav, originally from Hungary	1 (2), 4 (2), 5 (14)/2012
1–8/2013	Pet shop in Mlada Boleslav, originally from Hungary	7 (11), 8 (13)/2013

Additionally, fin clips were taken from twenty four individuals originating from a pet shop in 2014 and these were subsequently processed for microsatellite genotyping.

The leucocytes were cultured and chromosome spreads prepared according to the protocol [48] with some modifications described in [13]. To increase the chance of obtaining usable chromosome preparations, some individuals were processed using a direct method according to [13].

Chromosome staining and karyotype analysis

Paddlefish chromosomes were analysed by Giemsa staining and specific staining and banding methods combined with FISH mapping. Namely, buffered Giemsa (pH 7.0, 5%, 5 min) was performed to assess chromosome quality and provide comparison with earlier chromosomal studies. Subsequently, chromosomes were destained by incubation in fixative for 3 min at RT and briefly washed by distilled water. After air-drying, CMA_3 and DAPI fluorescent staining was performed sensu Sola et al. [49]. Finally, the $AgNO_3$ staining sensu Howell and Black [50] and C-banding sensu Haaf and Schmid [51] were performed. Separately DAPI- and CMA_3-stained chromosomes were converted into black and white images, inverted and arranged into karyotypes.

Fluorescence in situ hybridization (FISH) and signal detection

Whole genomic DNA (gDNA) was isolated from blood, using DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. The 28S rDNA and 5S rDNA were PCR amplified from gDNA according to [52] and [53]. The amplified fragments were purified using either the QIAquick PCR purification kit (Qiagen) or Qiagen Gel Extraction Kit, according to the manufacturer's instructions. Aliquots of the purified 28S rDNA were used in a sequencing reaction using BigDye® Terminator v1.1 Cycle Sequencing Kit (Life Technologies) and subsequently analysed at ABI Prism® 3700 Genetic Analyzer. Additionally, the same PCR products were sequenced by the Macrogen Inc. (Seoul, South Korea). Since a preliminary sequence screening has shown an intra-individual sequence variation of the 5SrDNA, purified products from three independent PCR runs were subjected to cloning using the Qiagen PCR Cloning Kit. The procedure followed the manufacturer's instructions, except for using a half of the recommended amount of the vector and competent cell: 20 ng of the PCR product was ligated to 25 ng of the pDrive cloning vector and a 1 μl aliquot of the obtained ligation-reaction mixture was used to transform 25 μl of Qiagen EZ Competent Cells. These were cultured for 12–24 h at 37 °C on LB agar plates containing Ampicillin (100 μg/ml). Randomly selected colonies from each transformed cell lineage were then transferred into the liquid LB medium and cultivated overnight on a shaking platform at 37 °C. Plasmids were isolated and 5SrDNA was extracted using the QIAprep Spin Miniprep Kit (Qiagen). The extract DNA was sequenced in the same way as previously described for 28S rDNA.

The nucleotide sequence data have been deposited in GenBank (under accession numbers KM103731-KM103735: KM103731 18S rDNA, KM103732 28S rDNA, KM103733- KM103735 5S rDNA).

DNA probes were indirectly labelled with biotin-16-dUTP and digoxigenin-11-dUTP (both Roche Diagnostics, Mannheim, Germany) through labelling PCR re-amplification of the previously sequenced PCR products. Reactions were performed in 50 μl total volume containing 1× reaction buffer, 2 mM $MgCl_2$, labelled dNTP nucleotide mix (dATP, dCTP, dGTP each 12.5 μM, dTTP 8.5 μM, dUTP conjugated with a hapten - biotin-16-dUTP or digoxygenin-11-dUTP (both Roche), final concentration 4 μM), forward and reverse primer (0.4 μM each), 1.25 U of Taq polymerase (all reagents from Top-Bio, Prague, Czech Republic) and approximately 100 ng of PCR product as template DNA. The dNTP nucleotide mix was prepared as a premix containing 5 μl dATP, dCTP, dGTP (each 2.5 nM) and 3.4 μl dTTP (2.5 nM), 4 μl dUTP (1 mM) conjugated with a hapten, 27.6 μl PCR water in total volume 50 μl.

FISH with telomeres were performed using the Star*-FISH Concentrated Human Chromosome Pan-Telomeric Painting Probe directly labelled with the Cy3 fluorescent dye (Cambio, Cambridge, UK) according to manufacturer's instruction.

The BAC clones of paralogous gene clusters for each of the HoxA and HoxD (Table 1, including accession numbers in GenBank) were produced by Crow et al.

(2012) [12] and provided for this study. The BAC DNA was labelled indirectly (biotin-16-dUTP and digoxigenin-11-dUTP) by nick translation using the Nick Translation Mix (Roche) according to the manufacturer's instructions. The precipitation and resuspension of the probe as well as aging (3–12 h at 37 °C and 30 min at 65 °C) and pepsin treatment of the chromosome preparations, hybridization and detection were performed as described in [54]. FISH with the BAC DNA of both paralogs of HoxA and HoxD were performed with and without competitor DNA derived from the Siberian sturgeon (*Acipenser baerii*) genomic DNA in excess 20–50 times of the probed BAC DNA.

The nomenclature of respective HoxA/D α and β paralogs was adopted from [12]. This nomenclature reflects the fact that these paralogs are products of the paddlefish specific WGD event.

Quantification of repetitive sequences in used BAC clones

DNA sequences of BAC clones used as FISH probes originating from [12] (Table 1) have been retrieved from the online "Nucleotide" NCBI database https://www.ncbi.nlm.nih.gov/nuccore according to their GenBank Accession Numbers (Table 1). These sequences have been subjected to screening for interspersed repeats and low complexity DNA sequences with the online tool RepeatMasker [55] using *Danio rerio* as the DNA source for sequence comparison, 'abblast' search engine, and otherwise default settings. Produced reports are summarized in this study (Table 1) and all complete reports are archived by RS.

Microscopy and image processing

Chromosome preparations were observed with the AX70 Olympus microscope equipped with a standard fluorescence filter set and captured with a black and white CCD camera separately for each fluorescent dye. Digital images were then pseudo-coloured (blue for DAPI, red for Rhodamine or Cy3, green for Fluorescein or FITC and CMA_3) and processed in Adobe Photoshop, version CS5. Karyotypes were produced using the IKAROS software (Metasystems, Germany).

Microsatellite analyses

Twenty four individuals were processed for microsatellite genotyping. The genomic DNA was extracted from fin clips stored in 96% molecular grade ethanol by the NucleoSpin®tissue kit (Macherey-Nagel, Germany) following manufacturer protocol. Nineteen microsatellite markers were tested for amplification using standard gradient – PCR. These markers are listed in Additional file 1 (Supplementary methods).

Markers, which consistently amplified, were selected for subsequent analyses. To avoid fluorescent labeling of each forward primer, forward primers within each of the primer sets possessed a 5′ prime end tail (M13R). During PCR, a fluorescently labelled primer (M13R) was added to the standard amplification reaction [56]. Detailed PCR protocol is listed in Additional file 1 (Supplementary methods). The level of genome reduplication/reduction was investigated as described by [18].

Additional files

Additional file 1: Supplementary Methods. Microsatellite markers tested for amplification in the present study and PCR protocol. (DOC 45 kb)

Additional file 2: Supplementary Results. Results of microsatellite genotyping, comparison of number of alleles, allele frequencies by locus. (PDF 83 kb)

Abbreviations

BAC (DNA): Bacterial artificial chromosome DNA; CMA3: Chromomycin A3 stain; Cy3: Cytochrome 3; DAPI: 4′, 6-diamidino-2-phenylindole stain; FISH: Fluorescence in situ hybridization; FITC: Fluorescein isothiocyanate; gDNA: Genomic DNA; HoxA/D: Homeobox gene clusters A/D; Mya: Millions of years ago; PCR: Polymerase chain reaction; rDNA: Ribosomal deoxyribonucleic acid; WGD: Whole genome duplication

Acknowledgments

We thank to Marie Rábová for her help during karyotypes preparation and to Jana Čechová, Šárka Pelikánová and Petra Šejnohová for their laboratory assistance.

Funding

The study was financially supported by the Ministry of Education, Youth and Sports of the Czech Republic projects CENAKVA No. CZ.1.05/2.1.00/01.0024, CENAKVAII No. LO1205 under the NPU I program and by projects of the Czech Science Foundation No.14-28375P and 14-02940S. Further, by the project EXCELLENCE CZ.02.1.01/0.0/0.0/15_003/0000460 OP RDE and with the institutional support RVO: 67985904. Granting agencies had no participation in the design of the study or interpretation of the results. This study was also supported by the University of Innsbruck.

Availability of data and materials

All fish examined in this study, chromosome preparations, and DNA and tissue samples are deposited in the Laboratory of Fish Genetics, Institute of Animal Physiology and Genetics, Czech Academy of Science (LFG, IAPG, CAS), Liběchov as voucher specimens and reference samples.

Authors' contributions

RS designed the study, prepared the manuscript, performed FISH experiments and the sequence analysis of BAC clones, MH performed the microsatellite analyses and co-drafted the manuscript, CTA provided isolated BACs of Hox paralogs, WMH and PR co-drafted the manuscript, TK took part in the FISH experiments and manuscript preparation, DG and MF provided fish for chromosome preparations, all authors participated on the preparation of the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Ethics approval and consent to participate

All experimental procedures involving fish were approved by the Institutional Animal Care and Use Committee of the IAPG AS CR, according with directives from the State Veterinary Administration of the Czech Republic, permit number 217/2010, and by the permit number CZ 00221 issued by Ministry of Agriculture of the Czech Republic. PR and RS have Certificate of competency according to §17 of the Czech Republic Act No. 246/1992 coll. on the

Protection of Animals against Cruelty (Registration numbers: CZU 955/06 and CZ 00832), provided by Central Commission for Animal Welfare, which authorizes animal experiments in the Czech Republic.

Author details
[1]Laboratory of Fish Genetics, Institute of Animal Physiology and Genetics, Czech Academy of Sciences, 277 21, Liběchov, Czech Republic. [2]Research Institute for Limnology, University of Innsbruck, Mondseestr. 9, Mondsee, Austria. [3]University of South Bohemia in České Budějovice, Faculty of Fisheries and Protection of Waters, South Bohemian Research Center of Aquaculture and Biodiversity of Hydrocenoses, 389 25 Vodňany, Czech Republic. [4]Benaroya Research Institute & University of Washington, Seattle, WA 98101, USA. [5]Department of Biological and Environmental Sciences, Samford University, 800 Lakeshore Drive, Birmingham, AL 35229, USA.

Received: 10 September 2016 Accepted: 11 February 2017
Published online: 02 March 2017

References

1. Nelson JS. Fishes of the world. 4th ed. Hoboken: John Wiley & Sons; 2006.
2. Peng Z, Ludwig A, Wang D, Diogo R, Wei Q, He S. Age and biogeography of major clades in sturgeons and paddlefishes (Pisces: Acipenseriformes). Mol Phylogenet Evol. 2007;42:854–62.
3. Krieger J, Hett AK, Fuerst PA, Artyukhin E, Ludwig A. The molecular phylogeny of the order Acipenseriformes revised. J Appl Ichthyol. 2008;24:36–45.
4. Dingerkus G, Howell WM. Karyotypic analysis and evidence of tetraploidy in the North American paddlefish, Polyodon spathula. Science. 1976;194:842–4.
5. Zhang S-M, Yang Y, Deng H, Wei QW, Wu QJ. Genome size and ploidy characters of several species of sturgeons and paddlefishes with comments on cellular evolution of Acipenseriformes. Acta Zool Sinica. 1999;45:200–6.
6. Birstein VJ, Poletaev AI, Goncharov BF. DNA content in Eurasian sturgeon species determined by flow cytometry. Cytometry. 1993;14:377–83.
7. Birstein VJ, Hanner R, Desalle R. Phylogeny of the Acipenseriformes: Cytogenetic and molecular approaches. Environ Biol Fishes. 1997;48:127–56.
8. Billard R, Lecointre G. Biology and conservation of sturgeon and paddlefish. Rev Fish Biol Fisher. 2001;10:355–92.
9. Romanenko SA, Biltueva LS, Serdyukova NA, Kulemzina AI, Beklemisheva VR, Gladkikh OL, et al. Segmental paleotetraploidy revealed in sterlet (*Acipenser ruthenus*) genome by chromosome painting. Mol Cytogen. 2015;8:90.
10. Havelka M, Kašpar V, Hulák M, Flajšhans M. Sturgeon genetics and cytogenetics: a review related to ploidy levels and interspecific hybridization. Folia Zool. 2011;60:93–103.
11. Trifonov VA, Romanenko SS, Beklemisheva VR, Biltueva LS, Makunin AI, Lemskaya NA, et al. Evolutionary plasticity of acipenseriform genomes. Chromosoma. 2016; doi: 10.1007/s00412-016-0609-2
12. Crow KC, Smith CD, Cheng JF, Wagner GP, Amemiya CA. An independent genome duplication inferred from Hox paralogs in the American paddlefish – a representative basal ray-finned fish and important comparative reference. Genome Biol Evol. 2012;4:937–53.
13. Symonová R, Flajšhans M, Sember A, Havelka M, Gela D, Kořínková T, et al. Molecular cytogenetics in artificial hybrid and highly polyploid sturgeons: an evolutionary story narrated by repetitive sequences. Cytogenet Genome Res. 2013;141:153–62.
14. Havelka M, Hulák M, Ráb P, Rábová M, Lieckfeldt D, Ludwig A, et al. Fertility of a spontaneous hexaploid male Siberian sturgeon, *Acipenser baerii*. BMC Genet. 2014;15:5.
15. Havelka M, Bytyuskyy D, Symonová R, Ráb P, Flajshans M. The second highest chromosome count among vertebrates is observed in cultured sturgeon and is associated with genome plasticity. Genet Sel Evol. 2016;48:12.
16. Wolfe KH. Yesterday's polyploids and the mystery of diploidization. Nat Rev Genet. 2001;2:333–41.
17. Fontana F, Congiu L, Mudrak VA, Quattro JM, Smith TIJ, Ware K, et al. Evidence of hexaploid karyotype in Shortnose sturgeon. Genome. 2008;51:113–9.
18. Havelka M, Hulák M, Bailie DA, Prodöhl PA, Flajšhans M. Extensive genome duplications in sturgeons: new evidence from microsatellite data. J Appl Ichthyol. 2013;29:704–8.
19. Ludwig A, Belfiore NM, Pitra Ch, Svirsky V, Jenneckens I. Genome duplication events and functional reduction of ploidy levels in sturgeon (*Acipenser*, *Huso* and *Scaphirhynchus*). Genetics. 2001;158:1203–15.
20. Macqueen DJ, Johnston IA. A well-constrained estimate for the timing of the salmonid whole genome duplication reveals major decoupling from species diversification. Proc R Soc B. 2014;281:20132881.
21. Vandepoele K, De Vos W, Taylor JS, Meyer A, Van de Peer Y. Major events in the genome evolution of vertebrates: paranome age and size differ considerably between ray-finned fishes and land vertebrates. Proc Natl Acad Sci U S A. 2004; 101:1638–43.
22. McGinnis W, Krumlauf R. Homeobox genes and axial patterning. Cell. 1992; 68:283–302.
23. Garcia-Fernandez J, Holland PWH. Archetypal Organization of the *Amphioxus* Hox Gene-Cluster. Nature. 1994;370:563–6.
24. Putnam NH, Butts T, Ferrier DEK, Furlong RF, Hellsten U, Kawashima T, et al. The *Amphioxus* genome and the evolution of the chordate karyotype. Nature. 2008;453:1064–71.
25. Martinez P, Amemiya CT. Genomics of the HOX gene cluster. Comp Biochem Physiol B: Biochem Mol Biol. 2002;133:571–80.
26. Van de Peer Y, Maere S, Meyer A. 2R or not 2R is not the question anymore. Nat Rev Genet. 2010;11:166.
27. May B, Krueger CC, Kincaid HL. Genetic variation at microsatellite loci in sturgeon: primer sequence homology in *Acipenser* and Scaphirhynchus. Can J Fish Aquat Sci. 1997;54:1542–7.
28. Heist EJ, Nicholson EH, Sipiroski JT, Keeney DB. Microsatellite markers for the American paddlefish (*Polyodon spathula*). Conserv Genet. 2002;3:205–7.
29. Venkatesh B. Evolution and diversity of fish genomes. Curr Opin Genet Dev. 2003;13:588–92.
30. Arai R. Fish Karyotypes. A Check List. Springer. 2011.
31. Glasauer SM, Neuhauss SC. Whole-genome duplication in teleost fishes and its evolutionary consequences. Mol Genet Genomics. 2014;289:1045–60.
32. Phillips R, Ráb P. Chromosome evolution in the Salmonidae (Pisces): an update. Biol Rev. 2001;76:1–25.
33. Mable BK, Alexandrou MA, Taylor MI. Genome duplication in amphibians and fish: an extended synthesis. J Zool. 2011;284:151–82.
34. Fontana F, Zane L, Pepe A, Congiu L. Polyploidy in Acipenseriformes: cytogenetic and molecular approaches. In: Pisano E, Ozouf-Costaz C, Foresti F, Kapoor BG, editors. Fish Cytogenetics. Enfield: Science Publisher, Inc.; 2007. p. 385–403.
35. Vasil'ev VP. Mechanisms of polyploid evolution in fish: Polyploidy in Sturgeons. In: Carmona R, Domezain A, García-Gallego M, Hernando JA, Rodríguez F, Ruiz-Rejón M, editors. Biology, Conservation and Sustainable Development of Sturgeons. The Netherlands: Springer Science; 2009. p. 97–117.
36. Birstein VJ, Vasil'ev VP. Tetraploid-octoploid relationships and karyological evolution in the order Acipenseriformes (Pisces): karyotypes, nucleoli, and nucleolus-organizer regions in four acipenserid species. Genetica. 1987;72:3–12.
37. Krieger J, Fuerst PA. Evidence for a slowed rate of molecular evolution in the order Acipenseriformes. Mol Biol Evol. 2002;19:891–7.
38. Ráb P. Karyotype evolution in fishes of the order Esociformes. Habilitation Thesis. Prague: Charles University in Prague (in Czech); 2004.
39. Stimpson KM, Sullivan LL, Kuo ME, Sullivan BA. Nucleolar organization, ribosomal DNA array stability, and acrocentric chromosome integrity are linked to telomere function. PLoS ONE. 2014;9(3), e92432.
40. Symonová R, Majtánová Z, Korínková T, Jankun M, Dion-C té A-M, Bernatchez L, Ráb P. Chromosomal characteristics of rDNA genes in salmonid fishes: trends in their patterns and evolution. 20th International Colloquium on Animal Cytogenetics and Gene Mapping, Cordoba, Spain. Chrom Res. 2012;20(6):810.
41. Symonová R, Majtánová Z, Sember A, Staaks GBO, Bohlen J, Freyhof J, et al. Genome differentiation in a species pair of coregonine fishes: an extremely rapid speciation driven by stress-activated retrotransposons mediating extensive ribosomal DNA multiplications. BMC Evol Biol. 2013;13:42.
42. Sember A, Bohlen J, Šlechtová V, Symonová R, Ráb P. Karyotype differentiation in nemacheilid loach fishes (Nemacheilidae, Cobitoidea, Cypriniformes): hidden variability uncovered by molecular cytogenetic markers in 19 species and its phylogenetic interpretation. BMC Evol Biol. 2015;15:251.
43. David L, Blum S, Feldman MW, Lavi U, Hillel J. Recent duplication of the common carp (*Cyprinus carpio* L.) genome as revealed by analyses of microsatellite loci. Mol Biol Evol. 2003;20:1425–34.
44. Knytl M, Kalous L, Symonová R, Rylková K, Ráb P. Chromosome studies of European cyprinid fishes: cross-species painting reveals natural allotetraploid origin of a *Carassius* female with 206 chromosomes. Cytogenet Genome Res. 2013;139:276–83.

45. Minguillón C, Gardenyes J, Serra E, Castro LF, Hill-Force A, Holland PW, et al. No more than 14: the end of the amphioxus Hox cluster. Int J Biol Sci. 2005; 1:19–23.
46. Birstein VJ. Phylogeny and evolution of Acipenseriformes: new molecular and genetic data create new puzzles. In: Gall YM, Kolchinsky EI, editors. Evolutionary biology: history and theory, vol. 3. St. Peterburg: Nauka; 2005. p. 231–69.
47. Allendorf FW, Thorgaard GH. Tetraploidy and the evolution of salmonid fishes. In: Turner BJ, editor. Evolutionary Genetics of Fishes. New York: Plenum Press; 1984. p. 1–46.
48. Fujiwara A, Hishida-Umehara C, Sakamoto T, Okamoto N, Nakyama I, Abe S. Improved fish lymphocyte culture for chromosome preparation. Genetica. 2001;111:77–89.
49. Sola L, Rossi AR, Iaselli V, Rasch EM, Monaco PJ. Cytogenetics of bisexual/unisexual species of *Poecilia*. II. Analysis of heterochromatin and nucleolar organizer regions in *Poecilia mexicana mexicana* by C-banding and DAPI, quinacrine, chromomycin A3 and silver staining. Cytogenet Cell Genet. 1992;60:229–35.
50. Howell WM, Black DA. Controlled silver-staining of nucleolus organizer regions with a protective colloidal developer: A 1-step method. Experientia. 1980;36:1014–5.
51. Haaf T, Schmid M. An early stage of ZW/ZZ sex chromosome differentiation in *Poecilia sphenops* var. *melanistica* (Poeciliidae, Cyprinodontiformes). Chromosoma. 1984;89:37–41.
52. Dayrat B, Tillier A, Lecointre G, Tillier S. New clades of euthyneuran gastropods (Mollusca) from 28S rRNA sequences. Mol Phylogenet Evol. 2001;19:225–35.
53. Komiya H, Takemura S. Nucleotide sequence of 5S ribosomal RNA from rainbow trout (*Salmo gairdnerii*) liver. J Biochem. 1979;86:1067–80.
54. Symonová R, Sember A, Majtánová Z, Ráb P. Characteriazation of Fish Genomes by GISH and CGH. In: Ozouf-Costaz C, Pisano E, Foresti F, De Almeida F, Toledo L, editors. Fish Cytogenetic Techniques. Ray-Fin Fishes and Chrondrichthyans. CRC Press. 2015. p. 118–31.
55. Smit AFA, Hubley R, Green P. Repeat Masker Open-3.0. 2010. www.repeatmasker.org
56. Schuelke M. An economic method for the fluorescent labeling of PCR fragments: a poor man's approach to genotyping for research and high-throughput diagnostics. Nat Biotechnol. 2000;18:233–4.

Cytogenetic and Genome Research

Cytogenet Genome Res 2013;141:153–162
DOI: 10.1159/000354882

Published online: September 14, 2013

Molecular Cytogenetics in Artificial Hybrid and Highly Polyploid Sturgeons: An Evolutionary Story Narrated by Repetitive Sequences

R. Symonová[a] M. Flajšhans[b] A. Sember[a] M. Havelka[b] D. Gela[b] T. Kořínková[a]
M. Rodina[b] M. Rábová[a] P. Ráb[a]

[a]Laboratory of Fish Genetics, Institute of Animal Physiology and Genetics, Academy of Sciences of the Czech Republic, Liběchov, and [b]Faculty of Fisheries and Protection of Waters, South Bohemian Research Centre of Aquaculture and Biodiversity of Hydrocenoses, Vodňany, Czech Republic

Key Words
Acipenser · GISH · Hybridization · Macrochromosomes · Microchromosomes · Polyploidy

Abstract
We applied comparative genomic hybridization (CGH) and genomic in situ hybridization (GISH) to examine genomes of artificially produced sturgeon hybrids between sterlet, *Acipenser ruthenus* female (~120 chromosomes) or Russian sturgeon, *A. gueldenstaedtii* female (~240 chromosomes) and a spontaneous triploid Siberian sturgeon *A. baerii* male (~360 chromosomes), respectively. The ploidy levels of progenies were analyzed by karyotyping and flow cytometry. We found that the species-specific regions were surprisingly identifiable only on some micro- and small(er) macrochromosomes in hybrid metaphases. We hypothesize that these distinguishable regions are represented by species-specific repetitive sequences driven by more dynamic molecular evolutionary mechanisms. On larger chromosomes, GISH faintly visualized only blocks of pericentromeric and telomeric repetitive sequences, remaining regions were equally shared by both parental species. We concluded that the interspecies hybridization producing viable and even fertile progeny is enabled by the fact that genomes of the species involved are likely divergent at the level of the repetitive sequences only and probably highly conserved in the coding sequences. These small differences of coding sequences are in concordance with previous estimations of relatedness of examined species producing artificial as well as natural hybrids. CGH and GISH represent a challenge in sturgeon cytogenetics as a valuable though technically not simple tool to discriminate chromosomes of parental species in hybrids. The potentials and drawbacks of CGH and GISH application in sturgeons are discussed and further experimental possibilities are proposed.

Sturgeons and paddlefishes (Acipenseriformes: Acipenseridae and Polyodontidae) are bony fishes representing very basal lineages of ray-finned fishes [Nelson, 2006]. They include species exhibiting 3 evolutionary ploidy levels: palaeotetraploidy (4n, ~120 chromosomes), palaeooctaploidy (8n, ~240 chromosomes) and palaeododecaploidy (12n, ~360 chromosomes); yet, they behave as normal diploid organisms with a diploid/haploid somatic/germinal cycle over generations [Bir-

KARGER

E-Mail karger@karger.com
www.karger.com/cgr

1424–8581/13/1413–0153$38.00/0

Radka Symonová
Laboratory of Fish, Academy of Sciences of the Czech Republic
Czech Academy of Sciences
Rumburská 89, CZ–27721 Liběchov (Czech Republic)
E-Mail symonova@natur.cuni.cz

stein et al., 1997; Birstein and DeSalle, 1998]. Acipenseriformes represent an ancient group still featuring very archaic characteristics (both genetic and anatomic) and therefore, being referred to as 'living fossils' [Gardiner, 1984]. Their propensity for whole-genome duplication and polyploidization expresses these archaic features [Ludwig et al., 2001; Crow et al., 2012]. Nowadays, only a fraction of their diversity (27 extant species) survives scattered throughout the Northern Hemisphere [Nelson, 2006]. The increasing anthropogenic impact onto the natural environment of Eurasian sturgeon species and populations steadily results in an elevated incidence of natural hybrids and dramatically declining natural populations of pure species during recent decades and their lower chance to reproduce [Pikitch et al., 2005; Ludwig, 2006; Tsekov et al., 2008; Dudu et al., 2011]. This poses another serious threat to the existence of pure species through extinction by hybridization of endangered populations (e.g. Ludwig et al. [2009]) together with the previous overexploiting of natural populations. Sturgeons are able to produce interspecies hybrids resulting in progeny with intermediate ploidy levels [Flajšhans and Vajcová, 2000; Vasil'eva et al., 2010; Havelka et al., 2011].

In this study, we present a molecular cytogenetic analysis of artificial hybrids of 3 *Acipenser* species, namely, between sterlet, *Acipenser ruthenus* female (~120 chromosomes) or Russian sturgeon, *A. gueldenstaedtii* female (~240 chromosomes) with a spontaneous triploid Siberian sturgeon *A. baerii* male (~360 chromosomes), respectively. The parent species are native to Eurasia. According to Birstein and Vasil'ev [1987] and Krieger et al. [2008], they belong to the Atlantic clade of the genus *Acipenser*, and at the same time, they are the most frequently bred and stocked sturgeon species in Europe.

Materials and Methods

Materials

The hybrid individuals resulting from the crossing experiments carried out in the hatchery of the Faculty of Fisheries and Water Protection, Vodňany [Havelka et al. in press] were used for molecular cytogenetic analyses. Among more crossing experiments, the following 2 batches of hybrids yielded chromosome preparations of sufficient quality to perform genomic in situ hybridization (GISH): the hybrid 4n ♀ *A. ruthenus* × 12n ♂ *A. baerii* and the hybrid 8n ♀ *A. gueldenstaedtii* × 12n ♂ *A. baerii*. In other experimental crossings, the most critical problems hampering molecular cytogenetic analyses were an insufficient chromosome spreading of metaphases with extremely high chromosome numbers (200–300 chromosomes).

Flow Cytometry

Prior to karyotyping, the ploidy level of specimens was verified as a relative DNA content in blood cells with DAPI according to Linhart et al. [2006], using Partec CCA I cytometer (Partec GmbH, Münster, Germany). Erythrocytes of the palaeotetraploid *A. ruthenus* gave a relative DNA content of 4n as the tetraploid standard.

Blood Culturing and Chromosome Preparation

The leucocytes were cultured and chromosome spreads prepared according to the protocol of Fujiwara et al. [2001] with some modifications. Briefly, 0.2–0.5 ml of blood was collected from an anesthetized fish by puncturing the vena cava caudalis using a heparinized syringe. The leucocyte-rich plasma was used to set up primary cultures in 5 ml of the Medium 199 (Sigma, St. Louis, Mo., USA) supplemented with 10% fetal bovine serum (FBS Superior, Biochrom, Berlin, Germany), 1% Antibiotic Antimycotic Solution (Sigma), LPS from *E. coli* (0.1 mg/ml of medium), PHA-P (18 μg/ml of medium; Remel, Lenexa, Kans., USA), Kanamycin (0.06 mg/ml of medium; Sigma) and 0.175 μl 10% mercaptoethanol (Sigma). After 120 h of incubation at 20°C, 5 ml of each culture was harvested using standard colchicine (2 drops of 0.1% colchicine per 5 ml of medium) and hypotonic (8 min) treatments followed by fixation in freshly prepared fixative (methanol: acetic acid 3:1, v/v), 3 times.

Some individuals were karyotyped using a direct method. These individuals were colchicinized (0.1% per 100 g of body weight, for 1 h) and sacrificed by overdose of 2-phenoxyethanol anesthetics agent (Sigma). The cephalic kidney was removed, minced and the resulting cells were hypotonized for 8 min in 0.075 M KCl before being fixed in freshly prepared fixative (methanol: acetic acid 3:1, v/v). Chromosome slides were air-dried and dehydrated in an ethanol series (70, 80 and 96%, 3 min each) and kept deep frozen until analysis. Specimens were not deposited as vouchers. This study was covered by the Valid Animal Use Protocols No. CZ 00221 at the Institute of Animal Physiology and Genetics issued by the Czech Ministry of Agriculture.

DNA Isolation

Whole genomic DNA (gDNA) of the pure species *A. ruthenus, A. baerii,* and *A. gueldenstaedtii* and of their hybrids were isolated from blood, muscles or fins using DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions.

Preparation of Probes for in situ Hybridizations

Labeling of the DNA probes was performed by nick translation using the Roche Nick Translation Mix (Roche Diagnostics, Mannheim, Germany; cat. No. 11745808910) according to the manufacturer's instructions. The gDNAs of parental species were labeled indirectly with biotin-16-dUTP (Roche, cat. No. 11093070910) or digoxigenin-11-dUTP (Roche, cat. No. 11093088910), respectively, with a reversal labeling scheme to confirm the observed results and exclude any influence of antibodies and/or fluorochromes. The hybridization signal was detected by Anti-Digoxigenin-Rhodamine (Roche, cat. No. 11207750910) and Streptavidin-FITC (Invitrogen Life Technologies, San Diego, Calif., USA; cat. No. 43-4311). Based on the fragment size of the high molecular gDNA used, the desired fragment size of probes (200–500 bp) was achieved after 45–90 min of nick translation. For most experiments, we used 1 μg of gDNA per species and experiment. In one set of experiments, we

also tried to follow the ratio of genomes in hybrid specimens (i.e. 2:3), but we did not observe any significant difference in the resulting signal patterns. Unlabeled competitor DNA for the suppression of unspecific hybridization was prepared by amplification of the total gDNA, since large quantities of DNA were required. DNA was isolated from hybrids used for chromosome preparation in the case of comparative genomic hybridization (CGH) or from the respective pure species in the case of single-color GISH experiments. Whole genomic amplification of the gDNA isolated was performed with Illustra GenomiPhi V2 DNA Amplification Kit (GE Healthcare, Buckinghamshire, UK, cat. No. 25-6600-31) and followed by sonication of amplified products (14–35 cycles, 10 pulses, 100% power) to the approximate fragment size of 100–200 bp, using the ultrasonic homogenizer Sonopuls HD 2070 (Bandelin Electronic, Berlin, Germany). To compare 2 different ways of competitor DNA fragmentation, we also cut competitor DNA by cyclic denaturation of the total hybrid gDNA at 97°C for 20 min followed by vortexing for 1 min (several cycles). However, after reaching the optimum fragment size of 100–200 bp, we recorded a rapid decrease of DNA concentration in the sample; therefore, the sonication procedure, despite also resulting in considerable losses of DNA, appeared to be a more efficient method. The competitor DNA was added with 50-fold excess of the concentration of the single gDNA probe. The probe cocktail contained 0.5 or 1 μg of labeled DNA of each species compared, 25 or 50 μg of unlabeled sonicated competitor from the corresponding hybrid DNA and 50 or 100 μg of sonicated salmon sperm DNA per slide (Sigma, cat. No. D9156) as a nonspecific competitor and DNA carrier. The hybridization mixture with DNA was precipitated in pure 96% ethanol, frozen at –20°C for at least 30 min (up to several hours), centrifuged at 13,000 RPM for 20 min, washed in 70% ethanol and air-dried at 37°C. The pellet was re-suspended in 35 μl of hybridization puffer as described by Neusser [2004].

Genomic in situ Hybridization

GISH was carried out according to the protocol used by Bi and Bogart [2006] with several modifications. After removal from the freezer and immediate dehydration in an ethanol series (70, 80 and 96%, 3 min each), slides were air-dried and then aged for 3 h at 37°C and for 30 min at 60°C. Prior to hybridization set up, chromosome preparations were incubated in 200 ng/ml DNAse-free RNAse A (Top Bio, Czech Republic, cat. No. D106) in 2× SSC for 90 min at 37°C in a humid chamber. The slides were then briefly washed in 2× SSC, dehydrated in an ethanol series and air-dried. To remove remnants of cytoplasm, the slides were incubated in PBS at room temperature for 5 min and then placed into a pepsin solution (0.005% pepsin in 0.01 N HCl, w/v) at 37°C for 3 min and again dehydrated in an ethanol series. After air-drying, the slides were denatured in 75% formamide in 2× SSC at 74°C for 3 min and then immediately cooled down and dehydrated in 70% (cold), 80 and 96% ethanol (room temperature). The hybridization mixture was denatured at 86°C for 6 min and then immediately placed on ice for at least 10 min. The hybridization was allowed to proceed for at least 3 days at 37°C in a dark humid chamber. The stringent washing was done twice in 50% formamide in 2× SSC, pH 7.0 at 42°C for 5 min, 3 times in 1× SSC at 42°C for 7 min, and finally in 2× SSC at room temperature for 20 s. To block nonspecific binding sites for Streptavidin and Anti-Digoxigenin, the slides were treated with 500 μl of blocking solution containing 3% BSA (Baria, Czech Republic, cat. No. SP-5050) in 4× SSC and 0.01% Tween 20 at 37°C for 20 min. Afterwards, the first probe was detected by adding the antibody solution (10 μl Anti-Digoxigenin-Rhodamine diluted in 150 μl 0.5% BSA in PBS or 2 μl Streptavidin-FITC diluted in 200 μl 10% FBS in PBS) and incubated at 37°C for 60 min in a dark humid chamber. The slides were then washed in 4× SSC and 0.01% Tween 4 times (7 min each) at 42°C. Afterwards, the second round of blocking took place and the second probe was detected with the respective antibody solution and then washed again as described above. Finally, chromosomes were counterstained with DAPI in mounting medium (Cambio, Cambridge, UK; cat. No. 1124-MD-50), added to the slide under a coverslip, sealed with nail polish, and stored cool and dark.

GISH and CGH results were completed by C-banding performed according to Haaf and Schmid [1984].

Image Analysis

Images were observed with an AX70 Olympus microscope equipped with a standard fluorescence filter set and captured with a black and white CCD camera separately for each fluorescent dye. Digital images were then pseudocolored (blue or red for DAPI, red for Rhodamine, green for FITC) and were processed in the Adobe Photoshop, version CS5. Chromosomes were classified according to Levan et al. [1964].

Results

Karyological Analyses and Chromosome Numbers in Hybrid Sturgeons

Representative Giemsa-stained metaphase chromosomes and respective karyotypes of the hybrids resulting from the 2 experimental crossings investigated are shown in figure 1. Namely, the hybrid 4n ♀ *A. ruthenus* × 12n ♂ *A. baerii* (fig. 1a) with 120 macrochromosomes and approximately 129 microchromosomes and the hybrid 8n ♀ *A. gueldenstaedtii* × 12n ♂ *A. baerii* (fig. 1b) with 150 macrochromosomes and approximately 186 microchromosomes. The numbers of 'acipenseriform' acrocentric chromosome markers (i.e. the group with the largest acrocentric chromosomes corresponding to the ploidy level, Symonová et al. [2010]) were identified to determine respective ploidy levels. The variation in the chromosome numbers of 3 individuals of each hybrid group examined are shown in figure 2a, b. The chromosome numbers ranged between 230–251 (fig. 2a) and 288–317 (fig. 2b). Several extremely low values of chromosome numbers were removed from the analyses, since they were incomplete.

Mapping of Parental Genomes in Hybrids Using CGH and GISH

To carry out CGH/GISH analyses in hybrid sturgeons, chromosome preparations of an extraordinary quality were essential; otherwise, no proper signal evaluation on

1

Cytogenet Genome Res 2013;141:153–162
DOI: 10.1159/000354882

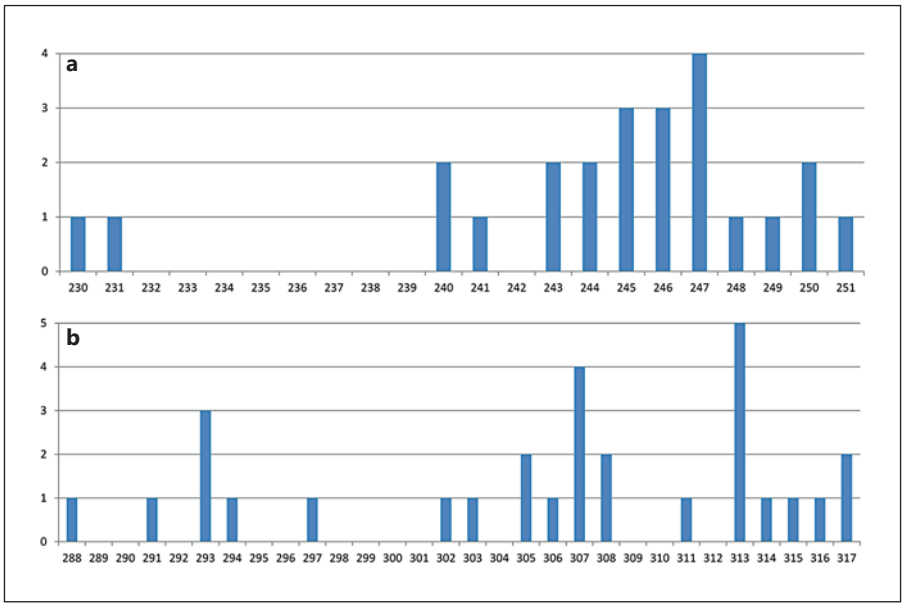


Fig. 2. Frequency of chromosome numbers in the hybrid 4n ♀ *A. ruthenus* × 12n ♂ *A. baeri* (**a**) and 8n ♀ *A. gueldenstaedtii* × 12n ♂ *A. baerii* (**b**). Three individuals scored in both groups. The apparently incomplete metaphases were not included as well as metaphases with numerous overlapping chromosomes preventing proper chromosome counting.

metaphases with such extreme chromosome numbers would be possible. Therefore, we could perform GISH only in the 2 aforementioned crossing groups of artificial hybrids with the chromosome spreads of the highest quality.

To begin with, we carried out several sets of CGH experiments with differentially labeled DNA of parental species to hybrid chromosomes with various concentrations (without competitor, 10× and 25× the concentration of the probe, respectively) of unlabeled competitor DNA derived from the tested hybrid genome. These experiments do not allow any interpretation, since a low concentration of competitor DNA derived from the hybrid genome. All chromosomes were stained with DNA of both genomes equally, and thus, the parent genomes were indistinguishable from each other in any of the hybrids (not shown). With the higher proportion of competitor DNA (up to 50× the concentration of a single genome probe) derived from the hybrid genome, some species-specific bands or whole microchromosomes of respective genomes were identifiable (fig. 3a, b).

In the second stage, we performed a single-color GISH to one of the available hybrids (the crossing group 11) using the DNA of one of the parental species as the probe and DNA of another species as unlabeled competitor (50×) and vice versa (i.e. 2 experiments, each visualizing one of the parent species). These results confirmed the previous ones and enabled a plausible explanation and interpretation of the dual-color GISH experiments. In this set of experiments with the hybrid *A. baerii* × *A. gueldenstaedtii* (~300 chromosomes), the probe derived from the *A. baerii* genome clearly hybridized to nearly the entire or substantial part of more than 50 microchromosomes and small macrochromosomes. To the remaining (approximately) 100 microchromosomes and small macrochromosomes, this probe hybridized at least partially, mostly to regions of pericentromeric heterochromatin. The large macrochromosomes remained mostly without any species-specific signals. The probe derived from the *A. gueldenstaedtii* genome clearly hybridized to the entire or substantial part of more than 25 microchromosomes and small macrochromosomes. To the remaining (approximately) 50 chromosomes, this probe hybridized at least partially, mostly to pericentromeric heterochromatin regions. Moreover, the probe of *A. gueldenstaedtii* also clearly hybridized to 2 middle-

Fig. 1. Metaphase and the respective karyotype of the hybrid 4n ♀ *A. ruthenus* × 12n ♂ *A. baeri* (**a**) resulting in an 8n individual (in this metaphase) with 249 chromosomes and the hybrid 8n ♀ *A. gueldenstaedtii* × 12n ♂ *A. baerii* (**b**) resulting in a 10n individual (in this metaphase) with 316 chromosomes. Chromosomes of each metaphase are arranged according to their sizes as macrochromosomes (upper lines) and microchromosomes (in a single block below). The asterisk indicates the 'acipenseriform' chromosome markers, i.e. the group of the largest acrocentric chromosomes. Scale bars are 10 μm and refer to the karyotypes; the size of the metaphases was reduced to fit the image.

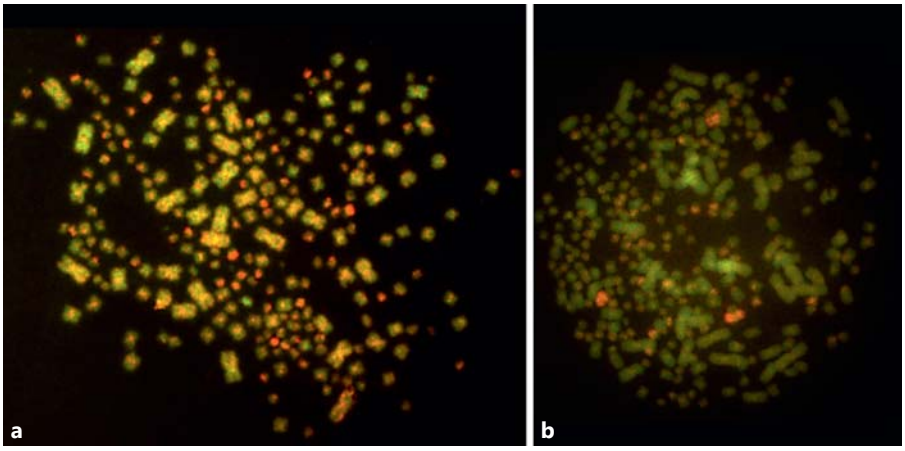


Fig. 3. CGH experiments to the metaphase spreads of the hybrid 4n ♀ *A. ruthenus* × 12n ♂ *A. baerii*; *A. ruthenus* gDNA was labeled in green and *A.baerii* gDNA in red (**a**) and 8n ♀ *A. gueldenstaedtii* × 12n ♂ *A. baerii*; *A. gueldenstaedtii* gDNA was labeled in red, *A. baerii* in green (**b**).

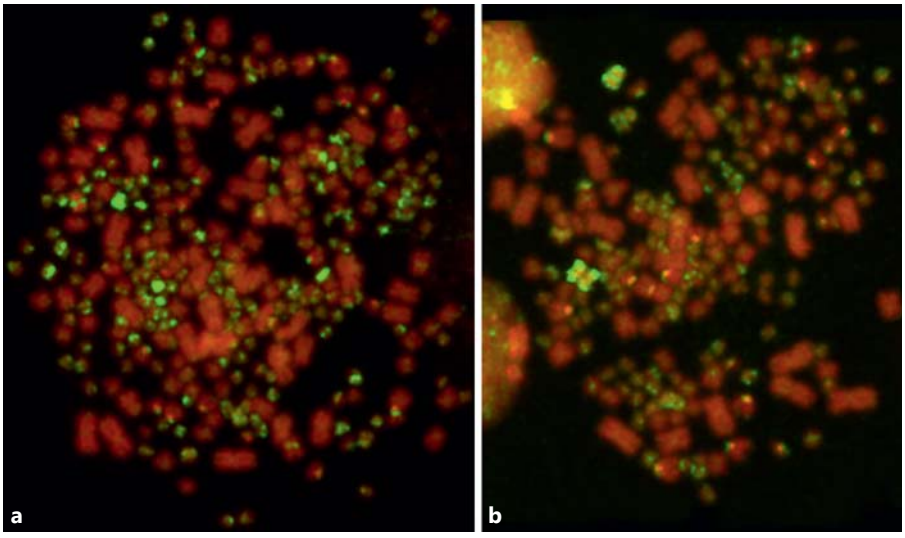


Fig. 4. GISH experiments to the metaphase spreads of a hybrid of the crossing 8n ♀ *A. gueldenstaedtii* × 12n ♂ *A. baerii*, with the *A. baerii* gDNA as probe (green signals) and the *A. gueldenstaedtii* gDNA as unlabeled competitor DNA (**a**) and the *A. gueldenstaedtii* gDNA as probe (green) and the *A. baerii* gDNA as unlabeled competitor (**b**). In both experiments, the black and white DAPI signals were visualized in the red channel to increase contrast.

sized metacentric macrochromosomes with distinctly DAPI positive (i.e. AT-rich) regions (fig. 4b, bright green signals). The remaining macrochromosomes also appeared without distinct species-specific signals as in the previous genome. Both probes hybridized faintly only to pericentromeric and telomeric regions of some macrochromosomes (fig. 4).

These results show that specific signals of the respective parental species were detectable only on microchromosomes and small macrochromosomes (and 2 AT-rich metacentric chromosomes of the genome *A. gueldenstaedtii*), and only these chromosomes could, thus, be assigned to the respective parents. The ratio between signals of each parental genome in their hybrid approximately corresponds to the (reduced) ploidy level of parental genomes entering the experimental crossing, i.e. 12n *A. baerii* × 8n *A. gueldenstaedtii*. After the meiotic reduction, this ratio results in 6n *A. baerii*: 4n *A. gueldenstaedtii*, with the observed approximate ratio of signals 3:2.

To better identify the signals on microchromosomes and small macrochromosomes, we performed C-banding to parental (pure) species used in our experiments. C-banding in the 8n *A. gueldenstadtii* (fig. 5) well documents the distribution of repetitive sequences on small macro- and microchromosomes and over the entire length of 4 medium-sized metacentric chromosomes. These metacentric chromosomes correspond to 2 (i.e. 4n, reduced set) chromosomes in the hybrid of 8n ♀ *A. gueldenstaedtii* × 12n ♂ *A. baerii* with distinct (green) *A. gueldenstaedtii* signal (fig. 4b).

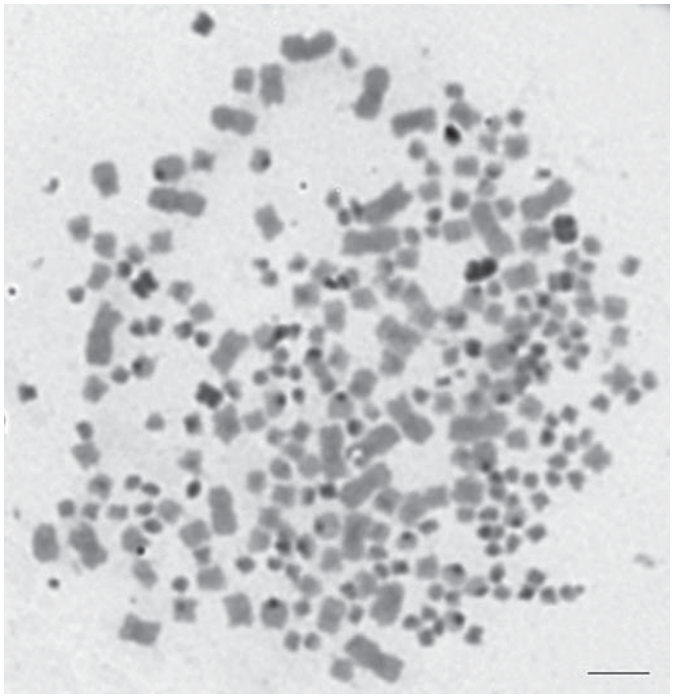


Fig. 5. C-banding pattern in a pure 8n *A. gueldenstaedtii*; one of the parental species used in experimental crossings in this study. Four medium-sized metacentric macrochromosomes are distinctly stained as well as numerous smaller macrochromosomes and other microchromosomes. Scale bar = 10 μm.

Discussion

This study represents the first stage of research aimed at the analysis of genomic composition and origin of naturally (likely) allopolyploid sturgeons by reticulate speciation as proposed by Vasil'ev [1999, 2009], particularly those with high chromosome numbers as e.g. *A. brevirostrum* and *A. mikadoi* [~360 and ~260 chromosomes, respectively; data on chromosome numbers have been collected by F. Fontana (http://www.unife.it/dipartimento/biologia-evoluzione/progetti/geneweb)] and to distinguish chromosomes of the parental species in hybrid chromosomal sets. Therefore, the initial experiments to hybrid genomes of known parental species and their ploidy level are essential prerequisites for the next stage of analysis.

Our observations indicate that the species-specific sequences on hybrid metaphases are mostly localized on micro- and small macrochromosomes. Due to the character of these signals (mostly bright spots) and their association with distinctly DAPI positive regions, we hypothesize that these fractions of DNA are represented by accumulations of repetitive sequences. On the other hand, the fact that the competitor DNA blocked the species-specific probe to macrochromosomes, another fraction of repetitive sequences can be expected also on large chromosomes. Little is known about the biological significance and DNA composition (genes vs. non-coding regions) of microchromosomes when compared with macrochromosomes in fish or, generally, in cold-blooded vertebrates. Up to date, microchromosomes have been mostly explored in birds, where they exhibit high GC-content, high genes density and higher mutation and recombination rates in comparison to macrochromosomes [Hillier et al., 2004; Axelsson et al., 2005; Kuraku et al., 2006; Matsubara et al., 2012]. On the other hand, microchromosomes in the lizard *Anolis carolinensis* do not share these characteristics [Alföldi et al., 2011]. Such features of microchromosomes in sturgeons have never been analyzed. Among the closest living relatives of sturgeons, occurrence of microchromosomes is known in gars [e.g. Ráb et al., 1999], the bowfin *Amia calva* [Ohno et al., 1969] and lampreys [Caputo et al., 2011], however, mostly without any further investigations.

The character of GISH signal distribution in hybrid sturgeons confirms the assumptions of a closer phylogenetic relationship among *A. baerii*, *A. gueldenstaedtii* and *A. ruthenus* proposed by Birstein and Vasil'ev [1987] and revised by Krieger et al. [2008]. Based on our results, it can also be speculated that only closely related species with low genome divergences, particularly in coding regions, can hybridize and form 'interspecies' viable and fertile hybrids. It this case, the efficacy of CGH and GISH might be considerably decreased. Moreover, the reduced rate of molecular evolution in sturgeons can also represent a factor preventing CGH and GISH from proper discrimination of species-specific sequences and chromosomes originating from the respective parental species. In sturgeons, the significantly reduced rate of molecular evolution has been recorded both in nuclear and mitochondrial genes on the level of gene and protein sequences [Krieger and Fuerst, 2002]. Further, de la Herrán et al. [2001] detected a reduced rate of molecular evolution also in an ancient satellite DNA family (otherwise rapidly evolving because of its lower functional constraints, Wichman et al. [1991]). Birstein and Vasil'ev [1987] evidenced a slow rate of karyological evolution in Acipenseriformes compared with teleost fishes. Similarly, a reduced rate of molecular evolution has also been recorded in other archaic groups (e.g. sharks and turtles) and has been

ascribed to some life-history traits shared by sturgeons (long generation time, large body size, ectothermy, low metabolic rate), summarized by Krieger and Fuerst [2002] and references therein. GISH has been routinely applied particularly in plants to discriminate parental genomes in interspecific hybrids and later modified for karyotyping plant chromosomes and phylogenetic applications based on semiquantitative analysis of GISH signals [Marková and Vyskot, 2009]. In plants, GISH utilizes genome-specific dispersed repetitive sequences, a prominent component of plant nuclear genomes [Bennetzen, 1998]. Recently, some first attempts in fishes have been performed, e.g. in a polyploid *Carassius,* [Knytl et al., 2013], diploid and polyploid *Cobitis* hybrids [Majtánová et al., 2012], in *Squalius* hybrids, [Rampin et al., 2012], and in cichlids [Valente et al., 2009]. In sturgeons, the application of GISH is highly desirable to explore the presumed allopolyploid origin of some species on the level of whole chromosome complements. However, at the same time, the application of this method is foreseen to be extremely difficult for several reasons, as opposed to e.g. *Carassius* or *Cobitis*. The high chromosome numbers preventing proper chromosome spreadings in metaphase preparations and prevailing small chromosome sizes (particularly numerous microchromosomes, frequently underspread and overlapping each other as well as macrochromosomes, result in a biased GISH signal pattern and intensity) make application of GISH extremely difficult and demanding with respect to the quality of chromosome preparations. Moreover, including specific competitor DNA in high concentrations is inevitable, and thus, an extremely high amount of gDNA is required per experiment.

Utilization of genome-specific satellite DNA as routinely used in plants to easily distinguish parent genomes in hybrids appears not to be so straightforward in sturgeons regarding our results and the present day knowledge on sturgeon genome composition [Lanfredi et al., 2001] as well as an observed reduced rate of molecular evolution of this type of DNA in sturgeons [de la Herrán et al., 2001]. Specifically, according to Lanfredi et al. [2001], there are substantial differences in the proportion and distribution pattern of the Hind III satellite DNA family on chromosomes of *A. gueldenstaedtii* (with the highest proportion of satellite DNA), *A. baerii* (with an intermediate proportion) and *A. ruthenus* (with the lowest proportion). Moreover, Lanfredi et al. [2001] have successfully used satellite DNA derived from a single species *(A. naccarii)* as a FISH probe to 7 other species of the genera *Acipenser* and *Huso*. These results indicate that at least some fraction of satellite DNA occurs unspecifically in more species and in various proportions and, therefore, could act as DNA blurring the genome-specific signal pattern on hybrid metaphases. In the light of results presented here, we can hypothesize that at least 2 classes of repetitive sequences may occur: one class, presumably more ancestral, with a reduced rate of molecular evolution as evidenced before and located on some micro- and on most macrochromosomes; and another class of repetitive sequences with a nonreduced, or even accelerated, rate of molecular evolution, presumably evolutionary derived and divergent in respective species. This second class of repetitive sequences may account for the species-specific differences and be located on some fraction of microchromosomes and smaller macrochromoses as evidenced by our single-color GISH experiment. This shows that repetitive sequences represent a clue to understanding genome evolution in sturgeons. Also in plants, the importance of repetitive sequences for the GISH efficacy is based on the fact that they generally evolve faster than unique sequences and genes and make it possible to differentiate chromosomes even from closely related species [Schwarzacher et al., 1989]. However, the repetitive sequences appear to be distributed predominantly in micro- and smaller macrochromosomes in sturgeons as opposed to plants. The crucial role of repetitive sequences in the fish genome was recently summarized in detail by Cioffi and Bertollo [2012] and Vicari et al. [2010], who both underline the importance of additional investigations in this field.

Therefore, further experiments will be desirable, beginning with CGH experiments between more and less related sturgeon species to 'calibrate' the method on species on a gradient of genome divergences, optimization in terms of stringency condition during detection and involving suppression subtractive hybridization [Lukyanov et al., 2007] methods to produce more species-specific probes. Moreover, a more detailed analysis of repetitive sequences would be highly desirable to extend our knowledge on types of repetitive sequences and their localization on micro- and macrochromosomes, e.g. by using Cot1-DNA and other fractions of DNA [Trifonov et al., 2009] as FISH probes, completed by sequencing. These steps will contribute to the successful application of GISH to naturally allopolyploid individuals and species with unknown parental genomes in the future.

Acknowledgements

This study was supported in part by the project Nos. 523/08/0824 and P506/11/P596 of the Grant Agency of the Czech Republic, and RVO 67985904. We extend our thanks to Dr. Francesco Fontana for maintaining and updating his web page on sturgeon chromosome literature 'The chromosomes of Acipenseriformes' and to Jana Čechová for her careful laboratory assistance. This study is a part of the series 'Molecular Cytogenetics of Ancient Fishes'.

References

Alföldi J, Di Palma F, Grabherr M, Williams C, Kong L, et al: The genome of the green anole lizard and a comparative analysis with birds and mammals. Nature 477:587–591 (2011).

Axelsson E, Webster MT, Smith NG, Burt DW, Ellegren H: Comparison of the chicken and turkey genomes reveals a higher rate of nucleotide divergence on microchromosomes than macrochromosomes. Genome Res 15:120–125 (2005).

Bennetzen JL: The structure and evolution of angiosperm nuclear genomes. Curr Opin Plant Biol 1:103–108 (1998).

Bi K, Bogart JP: Identification of intergenomic recombinations in unisexual salamanders of the genus *Ambystoma* by genomic in situ hybridization (GISH). Cytogenet Genome Res 112: 307–312 (2006).

Birstein VJ, DeSalle R: Molecular phylogeny of Acipenserinae. Mol Phylogenet Evol 9:141–155 (1998).

Birstein VJ, Vasil'ev VP: Tetraploid-octoploid relationships and karyological evolution in the order Acipenseriformes (Pisces): karyotypes, nucleoli, and nucleolus-organizer regions in four acipenserid species. Genetica 72:3–12 (1987).

Birstein VJ, Hanner R, DeSalle R: Phylogeny of the Acipenseriformes: cytogenetic and molecular approaches. Environ Biol Fish 48:127–155 (1997).

Caputo V, Giovannotti M, Cerioni PN, Splendiani A, Tagliavini J, Olmo E: Chromosomal study of a lamprey (*Lampetra zanandreai* Vladykov, 1955) (Petromyzonida: Petromyzontiformes): conventional and FISH analysis. Chromosome Res 19:481–491 (2011).

Cioffi MB, Bertollo LAC: Chromosomal distribution and evolution of repetitive DNAs in fish, in Garrido-Ramos MA (ed): Repetitive DNA. Genome Dyn 7:197–221 (2012).

Crow KC, Smith CD, Cheng JF, Wagner GP, Amemiya CA: An independent genome duplication inferred from Hox paralogs in the American paddlefish – a representative basal ray-finned fish and important comparative reference. Genome Biol Evol 4:937–953 (2012).

de la Herrán R, Fontana F, Lanfredi M, Congiu L, Leis M, et al: Slow rates of evolution and sequence homogenization in an ancient satellite DNA family of sturgeons. Mol Biol Evol 18: 432–436 (2001).

Dudu A, Suciu R, Paraschiv M, Georgescu SE, Costache M, Berrebi P: Nuclear markers of Danube sturgeons hybridization. Int J Mol Sci 12:6796–6809 (2011).

Flajšhans M, Vajcová V: Odd ploidy levels in sturgeons suggest a backcross of interspecific hexaploid sturgeon hybrids to evolutionarily tetraploid and/or octaploid parental species. Folia Zool 49:133–138 (2000).

Fujiwara A, Nishida-Umehara C, Sakamoto T, Okamoto N, Nakayama I, Abe S: Improved fish lymphocyte culture for chromosome preparation. Genetica 111:77–89 (2001).

Gardiner BG: Sturgeons as living fossils, in Eldredge N, Stanley SM (eds): Living Fossils, pp 148–152 (Springer, New York 1984).

Haaf T, Schmid M: An early stage of ZW/ZZ sex chromosome differentiation in *Poecilia sphenops* var. *melanistica* (Poeciliidae, Cyprinodontiformes). Chromosoma 89: 37–41 (1984).

Havelka M, Kašpar V, Hulák M, Flajšhans M: Sturgeon genetics and cytogenetics: a review related to ploidy levels and interspecific hybridization. Folia Zool 60:93–103 (2011).

Havelka M, Hulák M, Ráb P, Rábová M, Lieckfeldt D, et al: Fertility of a spontaneous triploid male Siberian sturgeon, *Acipenser baerii*. BMC Genetics, submitted.

Hillier LW, Miller W, Birney E, Warren W, Hardison RC, et al: Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature 432:695–716 (2004).

Knytl M, Kalous L, Symonová R, Rylková K, Ráb P: Chromosome studies of European cyprinid fishes: cross-species painting reveals natural allotetraploid origin of a *Carassius* female with 206 chromosomes. Cytogenet Genome Res 139:276–283 (2013).

Krieger J, Fuerst PA: Evidence for a slowed rate of molecular evolution in the order Acipenseriformes. Mol Biol Evol 19:891–897 (2002).

Krieger J, Hett AK, Fuerst PA, Artyukhin E, Ludwig A: The molecular phylogeny of the order Acipenseriformes revisited. J Appl Ichthyol 24 Suppl 1:36–45 (2008).

Kuraku S, Ishijima J, Nishida-Umehara Ch, Agata K, Kuratani1 S, MatsudaY: cDNA-based gene mapping and GC_3 profiling in the soft-shelled turtle suggest a chromosomal size-dependent GC bias shared by sauropsids. Chromosome Res 14:187–202 (2006).

Lanfredi M, Congiu L, Garrido-Ramos MA, de la Herrán R, Leis M, et al: Chromosomal location and evolution of a satellite DNA family in seven sturgeon species. Chromosome Res 9:47–52 (2001).

Levan AK, Fredga K, Sandberg AA: Nomenclature for centromeric position on chromosomes. Hereditas 52:201–220 (1964).

Linhart O, Rodina M, Flajšhans M, Mavrodiev N, Nebesářová J, et al: Studies on sperm of diploid and triploid tench (*Tinca tinca* L.). Aquac Int 14:9–25 (2006).

Ludwig A: A sturgeon view on conservation genetics. Eur J Wildlife Res 52:3–8 (2006).

Ludwig A, Belfiore NM, Pitra Ch, Svirsky V, Jenneckens I: Genome duplication events and functional reduction of ploidy levels in sturgeon *(Acipenser*, *Huso* and *Scaphirhynchus)*. Genetics 158:1203–1215 (2001).

Ludwig A, Lippold S, Debus L, Reinartz R: First evidence of hybridization between endangered sterlets *(Acipenser ruthenus)* and exotic Siberian sturgeons *(Acipenser baerii)* in the Danube River. Biol Invasions 11:753–760 (2009).

Lukyanov SA, Rebrikov D, Budzin AA: Suppression subtractive hybridization, in Buzdin A, Lukyanov S (eds): Nucleic Acids Hybridization: Modern Applications, pp 53–84 (Springer, Dordrecht 2007).

Majtánová Z, Choleva L, Symonová R, Ráb P: Molecular cytogenetic identification of parental genomes in triploid hybrid spined loaches (*Cobitis*, Cypriniformes). Chromosome Res 20:812 (2012).

Marková M, Vyskot B: New horizons of genomic in situ hybridization (GISH). Cytogenet Genome Res 126:368–375 (2009).

Matsubara K, Kuraku S, Tarui H, Nishimura O, Nishida Ch, et al: Intra-genomic GC heterogeneity in sauropsids: evolutionary insights from cDNA mapping and GC_3 profiling in snake. BMC Genomics 13:604 (2012).

Nelson JS: Fishes of the World, ed 4 (Wiley, New Jersey 2006).

Neusser M: Karyotypevolution, Genomorganisation und Zellkernarchitektur der Neuweltaffen. Thesis, München (2004).

Ohno S, Muramoto J, Stenius C, Christian L, Kittrell WA, Atkin NB: Microchromosomes in holocephalian, chondrostean and holostean fishes. Chromosoma 26:35–40 (1969).

Pikitch EK, Doukakis P, Lauck L, Chakrabarty P, Erickson D: Status, trends and management of sturgeon and paddlefish fisheries. Fish and Fisheries 6:233–265 (2005).

Ráb P, Rábová M, Reed KM, Phillips RB: Chromosomal characteristics of ribosomal DNA in the primitive semionotiform fish, longnose gar *Lepisosteus osseus*. Chromosome Res 7: 475–480 (1999).

Rampin M, Bi K, Bogart JP, Collares-Pereira MJ: Identifying parental chromosomes and genomic rearrangements in animal hybrid complexes of species with small genome size using Genomic in situ Hybridization (GISH). Comp Cytogenet 6:287–300 (2012).

Schwarzacher T, Leitch AR, Bennett MD, Heslop-Harrison JS: In situ localization of parental genomes in a wide hybrid. Ann Bot 64:315–324 (1989).

Symonová R, Flajšhans M, Gela D, Pelikánová Š, Rábová M, et al: Tetraploidy in paddlefish, *Polyodon spathula*? – 34 years later. Abstract 19th International Colloquium on Animal Cytogenetics and Gene Mapping. Chromosome Res 18:754 (2010).

Trifonov VA, Vorobieva NN, Rens W: FISH with and without COT1DNA, in Liehr T (ed): Fluorescence In Situ Hybridization (FISH) – Application Guide, pp 99–109 (Springer, Berlin 2009).

Tsekov A, Ivanova P, Angelov M, Atanasova S, Bloesch J: Natural sturgeon hybrids along Bulgarian Black Sea coast and in Danube River. Acta Zool Bulgar 60:311–316 (2008).

Valente GT, Schneider CH, Gross MC, Feldberg E, Martins C: Comparative cytogenetics of cichlid fishes through genomic in-situ hybridization (GISH) with emphasis on *Oreochromis niloticus*. Chromosome Res 17:791–799 (2009).

Vasil'ev VP: Polyploidization by reticular speciation in Acipenseriform evolution: a working hypothesis. J Appl Ichthyol 15:29–31 (1999).

Vasil'ev VP: Mechanisms of polyploid evolution in fish: polyploidy in sturgeons, in Carmona R, Domezain A, García-Gallego M, Hernando JA, Rodríguez F, Ruiz-Rejón M (eds): Biology, Conservation and Sustainable Development of Sturgeons, pp 97–117 (Springer, Dordrecht 2009).

Vasil'eva ED, Vasil'ev VP, Ponomareva EN, Lapukhin YA: Triple hybrids obtained by artificial hybridization of the Russian sturgeon *Acipenser gueldenstaedtii* with the hybrid of the starred sturgeon *A. stellatus* and the great sturgeon *A. huso* (Acipenseridae): the kind of inheritance of some morphological characters and fertility of the parental hybrid form. J Ichthyol 50:605–617 (2010).

Vicari MR, Nogaroto V, Noleto RB, Cestari MM, Cioffi MB, et al: Satellite DNA and chromosomes in Neotropical fishes: methods, applications and perspectives. J Fish Biol 76:1094–1116 (2010).

Wichman HA, Payne CT, Ryder OA, Hamilton MJ, Maltbie M, Baker RJ: Genomic distribution of heterochromatic sequences in equids: implications to rapid chromosomal evolution. J Hered 82:369–377 (1991).

Havelka *et al. Genet Sel Evol (2016) 48:12*
DOI 10.1186/s12711-016-0194-0

GSE Genetics Selection Evolution

RESEARCH ARTICLE Open Access

The second highest chromosome count among vertebrates is observed in cultured sturgeon and is associated with genome plasticity

Miloš Havelka[1,2*], Dmytro Bytyutskyy[1], Radka Symonová[3], Petr Ráb[4] and Martin Flajšhans[1]

Abstract

Background: One of the five basal actinopterygian lineages, the Chondrostei, including sturgeon, shovelnose, and paddlefish (Order Acipenseriformes) show extraordinary ploidy diversity associated with three rounds of lineage-specific whole-genome duplication, resulting in three levels of ploidy in sturgeon. Recently, incidence of spontaneous polyploidization has been reported among cultured sturgeon and it could have serious negative implications for the economics of sturgeon farming. We report the occurrence of seven spontaneous heptaploid (7n) Siberian sturgeon *Acipenser baerii*, which is a functional tetraploid species (4n) with ~245 chromosomes. Our aims were to assess ploidy level and chromosome number of the analysed specimens and to identify the possible mechanism that underlies the occurrence of spontaneous additional chromosome sets in their genome.

Results: Among 150 specimens resulting from the mating of a tetraploid (4n) *A. baerii* (~245 chromosomes) dam with a hexaploid (6n) *A. baerii* (~368 chromosomes) sire, 143 displayed a relative DNA content that corresponds to pentaploidy (5n) with an absolute DNA content of 8.98 ± 0.03 pg DNA per nucleus and nuclear area of 35.3 ± 4.3 μm^2 and seven specimens exhibited a relative DNA content that corresponds to heptaploidy (7n), with an absolute DNA content of 15.02 ± 0.04 pg DNA per nucleus and nuclear area of 48.4 ± 5.1 μm^2. Chromosome analyses confirmed a modal number of ~437 chromosomes in these heptaploid (7n) individuals. DNA genotyping of eight microsatellite loci followed by parental assignment confirmed spontaneous duplication of the maternal chromosome sets via retention of the second polar body in meiosis II as the mechanism for the formation of this unusual chromosome number and ploidy level in a functional tetraploid *A. baerii*.

Conclusions: We report the second highest chromosome count among vertebrates in cultured sturgeon (~437) after the schizothoracine cyprinid *Ptychobarbus dipogon* with ~446 chromosomes. The finding also represents the highest documented chromosome count in Acipenseriformes, and the first report of a functional heptaploid (7n) genome composition in sturgeon. To our knowledge, this study provides the first clear evidence of a maternal origin for spontaneous polyploidization in cultured *A. baerii*. To date, all available data indicate that spontaneous polyploidization occurs frequently among cultured sturgeons.

*Correspondence: havelm02@frov.jcu.cz
[1] Faculty of Fisheries and Protection of Waters, South Bohemian Research Center of Aquaculture and Biodiversity of Hydrocenoses, Research Institute of Fish Culture and Hydrobiology, University of South Bohemia in Ceske Budejovice, Zátiší 728/II, 389 25 Vodňany, Czech Republic
Full list of author information is available at the end of the article

Background

It is generally considered that the ancestral vertebrate genome underwent two rounds (1R and 2R) of whole-genome duplication (WGD) events [1–3] and that teleost fishes underwent an additional teleost-specific round of WGD (3R or TSGD) [4, 5]. Moreover, additional WGD events occurred independently in several teleostean lineages, e.g. 4R or SaGD in salmonids [6]. For most of the other vertebrates, there was no additional WGD since the 1R and 2R events [7, 8].

As a result of multiple rounds of lineage-specific WGD [9], the widest range of chromosome numbers among vertebrates is found in extant Acipenseriformes, which constitute an ancestral lineage of non-teleost ray-finned fishes [10]. Currently, three groups of acipenseriform species can be identified based on chromosome number, DNA content, and nucleus/cell size. The chromosome number reaches ~120, ~240, or ~360 [10, 11], and the genome size ranges from 2.44 in *Huso huso* [12] to 13.78 pg DNA per nucleus in *Acipenser brevirostrum* [13]. An increase in chromosome number is inherently associated with an increase in DNA content in the cell nucleus [14, 15], but an increase in nucleus/cell volume seems to correlate more with an increase in chromosome number than in DNA content [7, 16, 17].

Recent investigations were based on two scales of ploidy level in Acipenseriformes: (1) an evolutionary scale, which assumes tetraploid (4n)—octaploid (8n)—dodecaploid—(12n) relationships [18] and refers to ancient ploidy levels; and (2) a functional scale, which assumes diploid (2n)—tetraploid (4n)—hexaploid (6n) relationships [19] that originate from significant functional genome re-diploidization during the evolution of sturgeon [20, 21]. For clarity, in this study we relate all ploidy levels to the functional scale.

Genome plasticity in sturgeons, which display different ploidy levels and various chromosome numbers, combined with the ease with which different sturgeon species that differ in chromosome number can hybridize, result in hybrid individuals that have intermediate karyotypes compared to that of the parental species [22]. Moreover, hybridization can occur again between these hybrids and pure species [23].

Sturgeons are propagated in aquaculture, mainly for the production of black caviar and boneless meat. The high commercial value of sturgeon and the status of the wild sturgeon populations classified as critically endangered are conflicting issues. Overexploitation of wild populations for over 40 years has led to the listing of all sturgeon species in the Appendices to CITES (Convention on International Trade in Endangered Species of Wild Fauna and Flora) and also to the development of sturgeon aquaculture, originally for reintroduction, but more recently for caviar production [24]. Today, sturgeon farming is a rapidly growing branch of aquaculture, with China recognized as the leader in meat and caviar production, followed by Italy, France, Russia, and the USA [25]. To meet market demand for sturgeon products, aquaculture techniques make continuous progress, and commercial farms are increasingly using cultured broodstock.

The occurrence of sturgeon individuals with spontaneous modifications of ploidy levels, and hence DNA content, nucleus/cell size, and atypical chromosome numbers, has been reported in cultured sturgeon [14, 26–32]. Clearly, Acipenseriformes have a high tolerance for hybridization as well as for spontaneous doubling of chromosome sets (autopolyploidization). While the role of hybridization has been thoroughly explored in breeding of farm animals [33–35], including sturgeon [24, 35, 36], little research has been conducted on the influence of spontaneous polyploidy in cultured sturgeon.

In this study, we report the occurrence of seven Siberian sturgeon (*Acipenser baerii*) individuals with spontaneous heptaploidy (7n), *A. baerii* being a functional tetraploid species (4n) with ~245 chromosomes. These individuals originated from artificial crossbreeding between a hexaploid (6n ~368 chromosomes) *A. baerii* sire that was confirmed to be of spontaneous polyploid origin [31] and tetraploid (4n) *A. baerii* dam. The genetic predisposition of sturgeons for spontaneous polyploidization must be considered in aquaculture since it may represent a much more serious problem for sturgeon farming than currently believed. Since *A. baerii* is the most commonly cultured sturgeon species [24], our investigation is relevant to sturgeon aquaculture worldwide. The primary aims of our study were to assess ploidy level and chromosome number of assumed heptaploid (7n) *A. baerii* individuals and to identify the possible mechanisms that are responsible for the occurrence of spontaneous additional chromosome sets in their genome.

Methods

Sampling

This study was carried out in accordance with the Czech Law 246/1992 on animal welfare, for which the authors possess a certificate according to §17 of the law. Protocols underwent ethical review by the University of South Bohemia animal care committee (PP3/FROV/2012) and were approved by the University of South Bohemia animal care committee. Prior to handling, fish were anesthetized with 0.6 mL/L 2-phenoxyethanol (Merck Co., Darmstadt, Germany).

One-hundred-fifty specimens (with a mean body weight of 546 ± 78 g) that originated from artificial crossbreeding between a tetraploid (4n ~245 chromosomes) *A. baerii* dam with a hexaploid (6n ~368 chromosomes)

A. baerii sire were examined in 2012. Originally, crossbreeding was conducted to confirm fertility/sterility of the hexaploid (6n) *A. baerii* sire. Details on the parental fish and crossbreeding experiments are in Havelka et al. [31]. Peripheral blood was collected from the caudal vessel into a heparinized syringe [37]. Fin clips were taken from parental fish as well as from the 150 progeny and stored in 96 % molecular grade ethanol.

Flow cytometry

After sampling, 20 µL of blood was added to 1 mL of physiological solution and kept at 4 °C. Fifteen µL of this mixture were added to a 1 mL kit containing 4′,6-diamidino-2-phenylindole (DAPI; Partec GmbH, Görlitz, Germany). Blood from the diploid (2n) *A. ruthenus* was included as a standard in the same proportions for each sample. Samples were filtered through 30 µM nylon filters, maintained at ambient temperature in the dark for 15 min, and run on a CyFlow Cube 8 flow cytometer (Partec GmbH, Görlitz, Germany) using UV–LED light. Results were visualized by a single-parameter histogram that shows the relative fluorescence of the unknown sample and the standard. Ploidy level of each specimen was verified with respect to relative DNA content in erythrocyte nuclei. Results were presented as the mean of at least 3000 nuclei.

Feulgen image analysis

Feulgen image analysis densitometry was used for the quantification of absolute DNA. Blood smear slides were prepared using the flame tip method [38]. Diploid (2n) and induced triploid (3n) *Tinca tinca* (DNA content 2.02 and 3.10 pg DNA per nucleus, respectively) were used as internal standards [39]. Samples were stained using a DNA staining kit following Feulgen (Merck Co., Darmstadt, Germany). Feulgen image analysis densitometry as described in Hardie et al. [38] was conducted using a 3CCD Sony DXC-9100P camera coupled to an Olympus BX50 microscope (objective 100×) with OLYMPUS MICROIMAGE v. 4.0 image analysis software package (Olympus Corp., Tokyo, Japan) to measure integrated optical density (IOD) and area of erythrocyte nuclei. The IOD of 10 fish per ploidy level and a minimum of 100 nuclei per specimen were measured and compared to the IOD of two standards in order to calculate genome size. The erythrocyte nuclear area was also assessed in 10 fish per ploidy level and minimum of 100 nuclei per specimen.

Karyotyping

To verify the results on genome size, two individuals that were identified as heptaploid (7n) by cytometric examination were karyotyped. Metaphase chromosomes were prepared from peripheral blood leucocytes according to the protocol of Fujiwara et al. [40] and as described by Havelka et al. [31]. Representative Giemsa-stained metaphase chromosome plates were examined using an Olympus AX 70 microscope and recorded with an Olympus DP30VW digital camera. Well-spread metaphase chromosomes were arranged in karyotypes using Ikaros MetaSystems (Metasystems, Germany) software for chromosome quantification.

Microsatellite DNA genotyping

Microsatellite genotyping was performed to determine the origin of the additional sets of chromosomes that were observed in heptaploid (7n) individuals. Genomic DNA was extracted from fin clips of parental fish, seven progeny that were presumed to be heptaploid (7n), and seven randomly chosen specimens indicated as pentaploid (5n) by cytometric examination. After testing 14 microsatellite markers, the following eight markers, AciG 35 [41], Afu 68 [42], AfuG 54, AfuG 135 [43], Aox 45 [44], Spl 101, Spl 163, and Spl 173 [45] were selected based on the level of polymorphism between parents. Amplification was carried out according to the protocol described by Havelka et al. [21]. Fragment analysis for these microsatellites was performed on a 3500 ABI Genetic Analyzer (Applied Biosystems, TM) using a GeneScan LIZ 600 fluorescent size standard (Applied Biosystems, TM), and genotypes were scored with the GeneMapper v.5.0 software (Applied Biosystems, TM).

Origin of the spontaneous polyploidy

The origin of the spontaneous increase in ploidy level and chromosome number that was observed in the heptaploid (7n) individuals was investigated according to a slightly modified protocol reported by Gille et al. [32]. We determined the ratio of private dam and/or sire microsatellite alleles between the heptaploid (7n) individuals and their pentaploid (5n) full siblings. If an allele was unique to the sire or the dam, it was coded 1. Alleles present in both parental genotypes were not taken into account, since they were not informative for the analysis. All genotypes of the analysed progeny were assessed by this approach, and the number of dam and sire private alleles was determined for each microsatellite locus and each individual under study. The basic dataset included the number of private sire and dam microsatellite alleles at each locus in the seven heptaploid (7n) individuals and their pentaploid (5n) full siblings (see Additional file 1: Table S1). Subsequently, the total number of private sire and dam microsatellite alleles at all loci was calculated for each ploidy group and the following hypotheses were tested: (1) there is an increase in number of private dam alleles in the heptaploid (7n) individuals compared to

their pentaploid (5n) full siblings, which means that the part of the genome that was duplicated originated from the dam; (2) there is an increase in number of private sire alleles in the heptaploid (7n) individuals compared to their pentaploid (5n) full siblings, which means that the part of the genome that was duplicated originated from the sire; and (3) there is an increase in number of private sire and dam alleles in the heptaploid (7n) individuals compared to their pentaploid (5n) full siblings, which means that the part of the genome that was duplicated originated from both parental individuals. The significance of the increase in number of dam and/or sire private microsatellite alleles between the heptaploid (7n) individuals and their pentaploid (5n) full siblings was tested by paired Student's *t* test using the software Statistica v. 12 [46]. Prior to conducting the analysis, the assumption of normally distributed paired differences was examined by Shapiro–Wilk's test, which showed that the data were normally distributed. The paired Student's t test was performed separately for dam and sire private alleles. The number of private alleles was set as a measurement variable, and locus and ploidy were set as nominal variables. Each locus had one pair of observations for the measurement variable, one for number of private alleles in the heptaploid (7n) individuals and one for number of private alleles in their pentaploid (5n) full siblings. The level of significance was set at 0.05.

Results

We used a variety of methods for ploidy determination and analyzed absolute DNA content and erythrocyte nuclear area for 150 fish. While 143 specimens displayed relative DNA contents that corresponded to pentaploidy (5n) with an absolute DNA content of 8.98 ± 0.03 pg DNA per nucleus and nuclear area of 35.3 ± 4.3 μm^2, seven specimens exhibited relative DNA contents that corresponded to heptaploidy (7n), with an absolute DNA content of 15.02 ± 0.04 pg DNA per nucleus and nuclear area of 48.4 ± 5.1 μm^2.

Chromosome analyses confirmed that the number of chromosomes in the heptaploid (7n) individuals agrees with the level of ploidy. Nine countable metaphase chromosome spreads were analysed, with chromosome numbers ranging from 413 to 454 with a mean of 434. If the upper and lower extremes of the range of chromosome counts were omitted, a modal chromosome number of 430 ± 10 was found. This was demonstrated by analyzing a representative karyotype with 437 chromosomes from a heptaploid (7n) individual (Fig. 1). With the exception of the microchromosomes, all chromosomes could be grouped into heptaplets (Fig. 1). The recorded variation in total chromosome number was mainly due to variation in the number of small microchromosomes counted and to artefacts in the preparation (Fig. 1).

We found an unexpected increase in ploidy level, and hence genome size and chromosome number, in the genome of the heptaploid (7n) individuals. Microsatellite genotyping and subsequent parental assignment showed a significant increase in number of private dam alleles in the heptaploid (7n) individuals compared to their pentaploid (5n) full siblings (Paired Student t test, $t = -11.2$, 7 degrees of freedom, $P = 0.00001$). Conversely, there was no significant increase in number of private sire alleles between pentaploid (5n) and heptaploid (7n) full siblings (Paired Student t test, $t = 0$, 7 degrees of freedom, $P = 1$, see Fig. 2) The total number of private dam alleles was equal to 35 in the pentaploid (5n) group and 63 in the heptaploid (7n) group, which represents a 1.8-fold spontaneous increase (see Additional file 1: Table S1).

Discussion

Due to the uniqueness and rarity of the heptaploid (7n) individuals and to prevent loss of live individuals, only two were used for karyotyping. Metaphase chromosome spreads were prepared from cultured leukocytes that were obtained from blood samples. In accordance with Hardie and Hebert [13], since DNA contents were similar for all seven heptaploid (7n) individuals, we assumed that they had similar chromosome numbers.

Origin of known spontaneous polyploids

In this study, we report the second highest chromosome count among vertebrates after the schizothoracine cyprinid *Ptychobarbus dipogon* that has ~446 chromosomes [47] (see Additional file 2: Table S2). We showed that this very large number of chromosomes originates from spontaneous polyploidization. Generally, spontaneous polyploidization occurs via chromosome doubling, or production of unreduced gametes, or polyspermy. In animals, it is generally assumed to result from unreduced gamete formation [48, 49]. Among fish, production of unreduced oocytes via spontaneous duplication of the maternal chromosome set (SDM) is not rare [6]. Another mechanism of spontaneous polyploidy in fishes is the occurrence of polyspermic fertilization, which due to the presence of multiple micropyles in the oocytes of sturgeon, is theoretically more likely in this species compared to other fish taxa [50]. Our results confirmed that the observed increase in ploidy level, genome size, and chromosome number originated from the maternal genome by SDM. Because the *A. baerii* sire used in the experimental crossbreeding was hexaploid (6n) and thus, had triploid (3n) spermatozoa [31], dispermic or polyspermic fertilization can be excluded as the mechanism responsible for

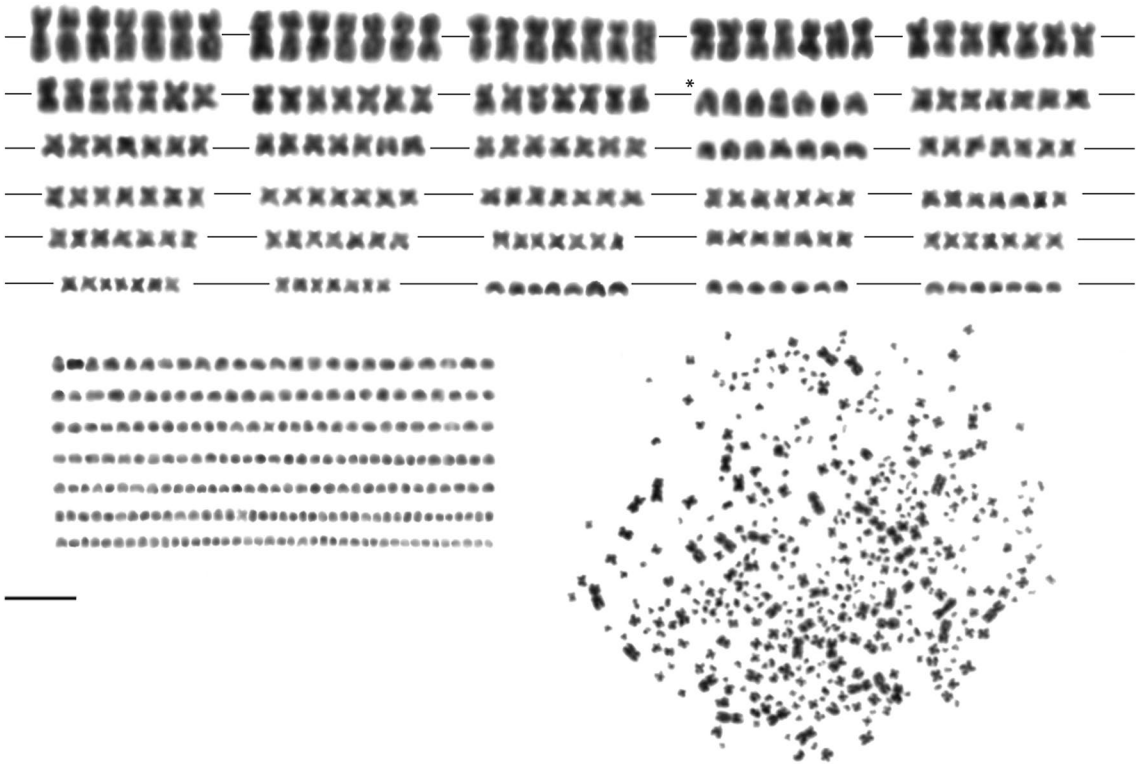


Fig. 1 Metaphase chromosome spread and corresponding karyotype derived from Giemsa-stained chromosomes of heptaploid Siberian sturgeon *Acipenser baerii*. This specimen has 437 chromosomes; asterisks denote acipenserine cytotaxonomic markers, the group with the largest acrocentric chromosomes; *bar* is equivalent to 10 µm

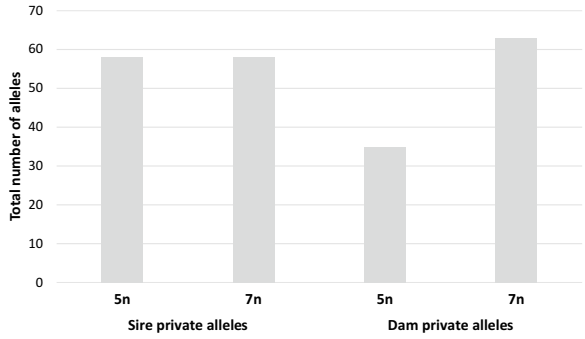


Fig. 2 Comparison of total number of private dam and sire microsatellite alleles in the pentaploid (5n) and heptaploid (7n) analysed individuals. See also Additional file 1: Table S1

the observed heptaploidy in the 7n individuals (Fig. 3). This, together with the significant increase in private dam alleles that was found for the heptaploid (7n) specimens, provides conclusive evidence that the heptaploid (7n) fish resulted from the fertilization of unreduced oocytes with SDM from a tetraploid (4n) female by triploid (3n) spermatozoa from the hexaploid (6n) male.

Several mechanisms can explain SDM including apomixis, premeiotic endomitosis, and retention of the second polar body in meiosis II. Apomixis and premeiotic endomitosis have been shown to provide unreduced clonal oocytes that are genetically identical to the dams [51–54]. Because the microsatellite genotypes of the spontaneous heptaploid (7n) individuals that were analyzed in this study were not identical to those of the dam (see Additional file 1: Table S1), both apomixis and premeiotic endomitosis can be excluded as the mechanisms of SDM during oocyte formation in this case. Because of the 1.8-fold spontaneous increase in private dam alleles and because genotypes of the heptaploid (7n) individuals were not fully identical to those of the dam (i.e. recombination occurred), retention of the second polar body is the most plausible mechanism to explain SDM in the oocytes and the resulting spontaneous heptaploid (7n) individuals. This agrees with the findings of Gille et al. [32] who identified spontaneous polyploid individuals in cultured *A. transmontanus* and suggested that this spontaneous autopolyploidy was most probably caused by failure of the segregation of the second polar body during meiosis II.

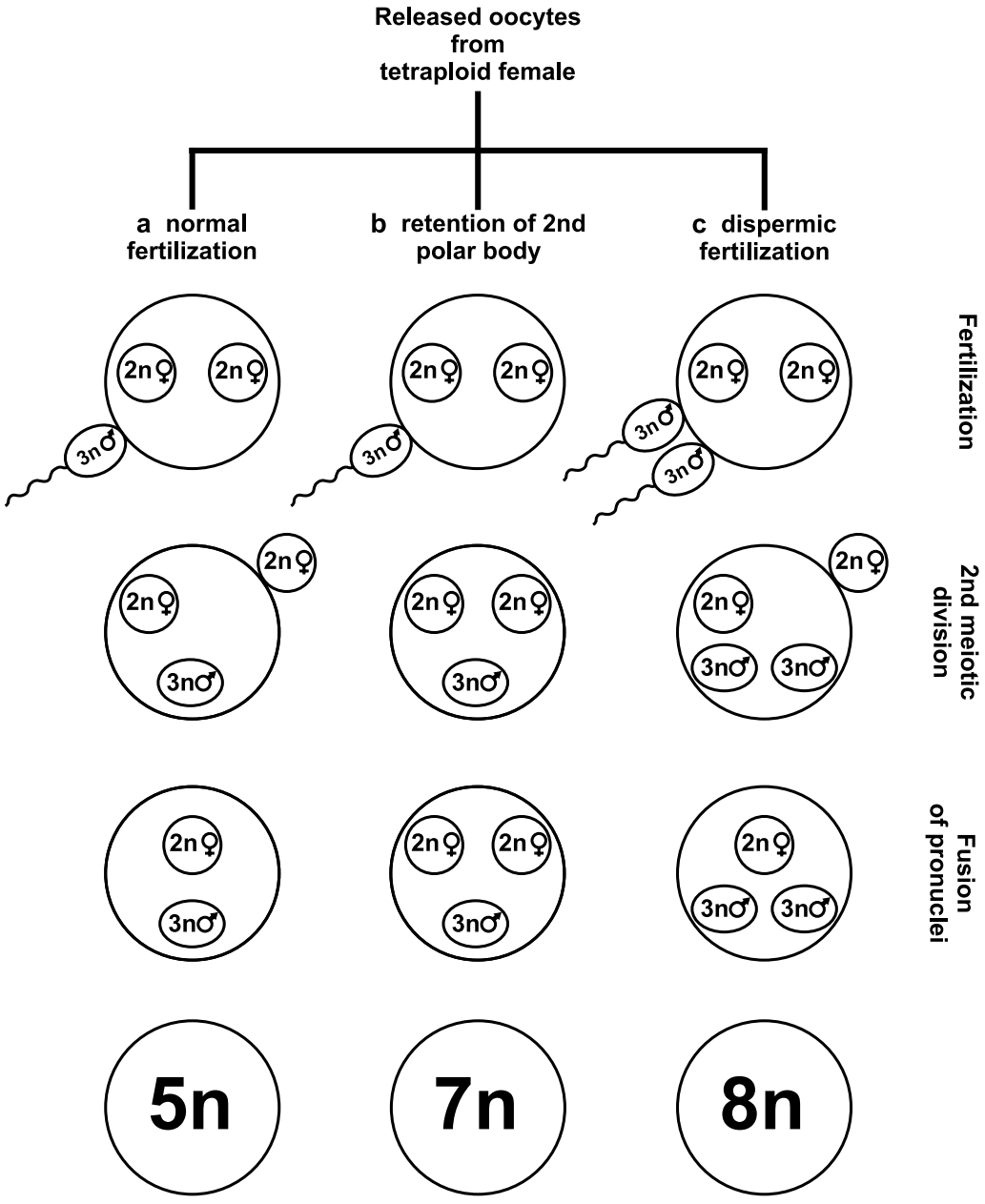


Fig. 3 Comparison of normal fertilization, retention of the second polar body, and dispermic fertilization of a tetraploid female (4n) by spermatozoa of hexaploid male (6n). **a** Normal fertilization of diploid (2n) oocyte by triploid (3n) spermatozoon. **b** Retention of the second polar body in meiosis II results in tetraploid (4n) unreduced oocytes. If such an oocyte is fertilized by a triploid (3n) spermatozoon, it results in heptaploidy ($4n♀ + 3n♂ = 7n$), which is the ploidy level observed in the seven individuals under study. **c** Dispermic fertilization results in an octaploid (8n) zygote. If a diploid (2n) oocyte is fertilized by two triploid (3n) spermatozoa from a hexaploid male (6n), it will lead to octaploidy ($2n♀ + 3n♂ + 3n♂ = 8n$)

Spontaneous polyploidy and its influence on sturgeon populations

Spontaneous polyploidy is a phenomenon that has been observed in a number of cultured fish species [55] including *Oncorhynchus mykiss* [56], *Tinca tinca* [57, 58], *Anguilla japonica* [59], *Oncorhynchus kisutch* [60], *Silurus glanis* [61], and *Salmo salar* [62], as well as in cultured sturgeon, e.g. hybrid (bester) sturgeon (*H. huso* × *A. ruthenus*) [26], *A. ruthenus* [30], *A. baerii* [31], *A. transmontanus* [27, 29, 32], *A. gueldenstaedtii* [15], *H. dauricus* and *A. mikadoi* [14]. In diploid (2n) species, the presence of an additional set of chromosomes results in

triploid (3n) individuals that are infertile or sub-sterile. In contrast, spontaneous polyploidization in tetraploid (4n) sturgeon species results in fertile hexaploid (6n) individuals, as reported for *A. baerii* [31] and *A. transmontanus* [29, 32]. Backcrossing of these spontaneous hexaploids (6n) to tetraploid (4n) individuals produces fully viable pentaploid (5n) progeny [29, 31, 32]. Although the reproductive potential of such pentaploid (5n) individuals is not confirmed, they are likely to present a significantly reduced fertility, since their chromosomes cannot pair during the zygotene stage of meiosis prophase I, due to the odd number of chromosome sets. Such impairment interferes with gonad development and gametogenesis, which is similar to what is observed for triploid (3n) individuals [63]. Currently, most cultured sturgeons originate from tetraploid (4n) species. The occurrence of fertile spontaneous polyploidy individuals among a tetraploid (4n) broodstock can negatively affect its reproductive capacity, and thus caviar production and the overall efficiency of sturgeon farms.

Prolongation of the period between ovulation, stripping, and fertilization may increase the incidence of retention of the second polar body in sturgeon artificial reproduction conditions [26, 32], as was observed for *O. mykiss* [56], *T. tinca* [58], and *A. japonica* [59]. To reduce the incidence of spontaneous polyploidy in cultured sturgeon and also in order to eliminate restocking of spontaneous polyploidy individuals into the wild populations, eggs should be stripped and fertilized immediately after ovulation, and the ploidy level of all fish should be determined before their inclusion in reproduction or reintroduction programs.

A maternal genetic predisposition for producing unreduced oocytes was suggested in *O. mykiss* [64, 65], *C. carpio* [66, 67], *T. tinca* [57], and *Misgurnus anguillicaudatus* [68], and was also discussed by Gille et al. [32] for *A. transmontanus.* Because polyploidization provides some genetic advantages [69], such genetic predisposition for the formation of unreduced gametes may be conserved in the genome of spontaneous polyploidy individuals and transmitted to the progeny, as was recently hypothesized by Mason and Pires [49]. This suggests that a higher incidence of spontaneous polyploidy may occur in cultured sturgeon than previously supposed, and hence could represent a greater issue for aquaculture.

The impact of spontaneous polyploidy on wild sturgeon populations has not been investigated. In plants, spontaneous polyploidization or production of unreduced gametes is known to occur in wild populations and the resulting individuals are able to survive and compete successfully in or near sites occupied by diploid populations [70]. Spontaneous polyploid individuals have also been observed in wild populations of amphibians, but at a very low frequency and the significance of this phenomenon for wild populations remains unclear [71]. In fish, spontaneous polyploidization has been reported in wild populations of several species e.g. [51, 52, 54], but the phenomenon appears to be a natural characteristic of a given species/population and its reproductive biology, rather than an unusual event [55]. To our knowledge, there is no evidence in the literature on the incidence of spontaneous polyploidization in wild sturgeon populations. However, sturgeon gametes can be exposed to stresses from unstable and rapidly fluctuating environmental conditions in natural spawning habitats, resulting in chromosome doubling, as suggested for fish in general [7]. In addition, since spontaneous polyploidy individuals have been reported in farmed populations and since reintroduction programs generate considerable interest, the presence of such spontaneous polyploid individuals that originate from release programs, is likely in wild sturgeon populations. Thus, ecological interactions may occur between spontaneous polyploids and wild individuals. As in captive sturgeon populations, fertile spontaneous polyploid individuals can spawn in the nature and result in progeny with reduced fertility. It may lead to a significant decrease in fitness of wild populations and thus have a negative impact on this critically endangered sturgeon species that has a very low abundance of natural spawners. Therefore, this is an important issue for reintroduction programs of this endangered sturgeon species that should be addressed in future studies.

Finally, formation of unreduced gametes has recently been suggested as a mechanism of speciation, contrary to the general interpretation that it is an evolutionary mishap [49]. It is interesting to note that most polyploid fish belong to the lower ray-finned fish and display a high incidence of hybridization [55]. Therefore, we hypothesize that the capacity of the sturgeon species for spontaneous polyploidization may contribute to explain their "living fossil" status, slow rate of genome evolution [72] and tolerance for hybridization.

Conclusions

This is the first study that describes a heptaploid (7n) genome composition in sturgeon based on seven individuals of *A. baerii*. It represents the highest documented chromosome count in Acipenseriformes and the second highest among all vertebrates. Spontaneous duplication of the maternal chromosome sets via retention of the second polar body in meiosis II was confirmed as the mechanism that underlies the formation of this high ploidy level and chromosome count. To the best of our knowledge, this represents the first evidence for a maternal origin of spontaneous polyploidization in *A. baerii*.

Additional files

Additional file 1: Table S1. Genotyping results with highlighted private dam and sire microsatellite alleles observed at all analysed loci for parental individuals, and for pentaploid (5n) and heptaploid (7n) full siblings.

Additional file 2: Table S2. Species with the highest chromosome count in vertebrate families and their nuclear DNA content. Asterisks denote references to DNA content.

Authors' contributions

MH performed fish handling and sampling, carried out molecular analyses, designed and performed parentage assignment, conducted statistical analyses, participated in the assessment of ploidy level by flow cytometry and wrote the manuscript. DB carried out image cytometry and flow cytometry analyses. RS performed karyotyping and provided karyological data. PR (co-senior author) described the karyological data, and performed data quality check. MF (senior author) participated in the evaluation of data that were obtained from flow cytometry and image cytometry and performed data quality check. All authors contributed to the design of the experiments and the preparation of the manuscript. All authors read and approved the final manuscript.

Author details

[1] Faculty of Fisheries and Protection of Waters, South Bohemian Research Center of Aquaculture and Biodiversity of Hydrocenoses, Research Institute of Fish Culture and Hydrobiology, University of South Bohemia in Ceske Budejovice, Zátiší 728/II, 389 25 Vodňany, Czech Republic. [2] Faculty of Fisheries Sciences, Hokkaido University, 3-1-1 Minato, Hakodate, Hokkaido 041-8611, Japan. [3] Research Institute for Limnology, University of Innsbruck, Mondseestraße 9, 5310 Mondsee, Austria. [4] Laboratory of Fish Genetics, Institute of Animal Physiology and Genetics, Czech Academy of Sciences, 277 21 Liběchov, Czech Republic.

Acknowledgements

Authors would like to acknowledge three anonymous reviewers for their helpful comments during revision of the manuscript. Special thanks belong to Reviewer 2 for an important contribution that improved statistical analysis of the data. This study was financially supported by the Ministry of Education, Youth and Sports of the Czech Republic projects CENAKVA (No. CZ.1.05/2.1.00/01.0024), CENAKVA II (No. LO1205 under the NPU I program), the Czech Science Foundation (No.14-28375P and 14-02940S) and by a grant from the Japan Society for the Promotion of Science under the International Research Fellow programme (ID No: P14751). Granting agencies had no participation in the design of the study or interpretation of the results. The Lucidus Consultancy is gratefully acknowledged for English correction and suggestions.

Competing interests

The authors declare that they have no competing interests.

Received: 24 August 2015 Accepted: 2 February 2016
Published online: 11 February 2016

References

1. Ohno S. Evolution by gene duplication. Berlin: Springer; 1970.
2. Lynch M. Genomics. Gene duplication and evolution. Science. 2002;297:945–7.
3. McLysaght A, Hokamp K, Wolfe KH. Extensive genomic duplication during early chordate evolution. Nat Genet. 2002;31:200–4.
4. Ventakhesh B. Evolution and diversity of fish genomes. Curr Opin Genet Dev. 2003;13:588–92.
5. Hoegg S, Brinkmann H, Taylor JS, Meyer A. Phylogenetic timing of the fish-specific genome duplication correlates with the diversification of teleost fish. J Mol Evol. 2004;59:190–203.
6. Braasch I, Postlethwait JH. Polyploidy in fish and the teleost genome duplication. In: Soltis PS, Soltis DE, editors. Polyploidy and genome evolution. Berlin: Springer; 2012. p. 341–83.
7. Mable BK, Alexandrou MA, Taylor MI. Genome duplication in amphibians and fish: an extended synthesis. J Zool. 2011;284:151–82.
8. Yang L, Sado T, Vincent Hirt M, Pasco-Viel E, Arunachalam M, Li J, et al. Phylogeny and polyploidy: resolving the classification of cyprinine fishes (*Teleostei: Cypriniformes*). Mol Phylogenet Evol. 2015;85:97–116.
9. Peng Z, Ludwig A, Wang D, Diogo R, Wei Q, He S. Age and biogeography of major clades in sturgeons and paddlefishes (*Pisces: Acipenseriformes*). Mol Phylogenet Evol. 2007;42:854–62.
10. Fontana F. A cytogenetic approach to the study of taxonomy and evolution in sturgeons. J Appl Ichthyol. 2002;18:226–33.
11. Havelka M, Kašpar V, Hulák M, Flajšhans M. Sturgeon genetics and cytogenetics: a review related to ploidy levels and interspecific hybridization. Folia Zool. 2011;60:93–103.
12. Birstein VJ, Poletaev AI, Goncharov BF. The DNA content in Eurasian sturgeon species determined by flow cytometry. Cytometry. 1993;14:337–83.
13. Hardie DC, Hebert PD. The nucleotypic effects of cellular DNA content in cartilaginous and ray-finned fishes. Genome. 2003;46:683–706.
14. Zhou H, Fujimoto T, Adachi S, Yamaha E, Arai K. Genome size variation estimated by flow cytometry in *Acipenser mikadoi*, *Huso dauricus* in relation to other species of Acipenseriformes. J Appl Ichthyol. 2011;27:484–91.
15. Bytyutskyy D, Srp J, Flajšhans M. Use of Feulgen image analysis densitometry to study the effect of genome size on nuclear size in polyploid sturgeons. J Appl Ichthyol. 2012;28:704–8.
16. Gregory TR. Genome size evolution in animals. In: Gregory TR, editor. The evolution of the genome. San Diego: Elsevier; 2005. p. 3–87.
17. Bytyutskyy D, Kholodnyy V, Flajšhans M. 3-D structure, volume, and DNA content of erythrocyte nuclei of polyploid fish. Cell Biol Int. 2014;38:708–15.
18. Birstein VI, Hanner R, DeSalle R. Phylogeny of the Acipenseriformes: cytogenetic and molecular approaches. Environ Biol Fish. 1997;48:127–55.
19. Fontana F, Zane L, Pepe A, Congiu L. Polyploidy in Acipenseriformes: cytogenetic and molecular approaches. In: Pisano E, Ozouf-Costaz C, Foresti F, Kapoor BG, editors. Fish cytogenetics. Enfield: Science Publisher; 2007. p. 385–403.
20. Ludwig A, Belfiore NM, Pitra C, Svirsky V, Jenneckens I. Genome duplication events and functional reduction of ploidy levels in sturgeon. (*Acipenser, Huso* and *Scaphirhynchus*). Genetics. 2001;158:1203–15.
21. Havelka M, Hulák M, Bailie DA, Prodöhl PA, Flajšhans M. Extensive genome duplication in sturgeons: new evidence from microsatellite data. J Appl Ichthyol. 2013;29:704–8.
22. Gorshkova G, Gorshkov S, Gordin H, Knibb W. Karyological study in hybrids of Beluga, *Huso huso* (L.) and the Russian sturgeon *Acipenser gueldenstaedtii* Brandt. Israel J Aquacult. 1996;48:35–9.
23. Vasil'eva ED, Vasil'ev VP, Ponomareva EN, Lapukhin YA. Triple hybrids obtained by artificial hybridization of the Russian sturgeon *Acipenser gueldenstaedtii* with the hybrid of the starred sturgeon *A. stellatus* and the great sturgeon *A. huso* (Acipenseridae): the kind of inheritance of some morphological characters and fertility of the parental hybrid form. J Ichthyol. 2010;50:605–17.
24. Bronzi P, Rosenthal H, Gessner J. Global sturgeon aquaculture production: an overview. J Appl Ichthyol. 2011;27:169–75.
25. Bronzi P, Rosenthal H. Present and future sturgeon and caviar production and marketing: a global market overview. J Appl Ichthyol. 2014;30:1536–46.
26. Omoto N, Maebayashi M, Adachi S, Arai K, Yamauchi K. The influence of oocyte maturational stage on hatching and triploidy rates in hybrid (bester) sturgeon, *Huso huso* × *Acipenser ruthenus*. Aquaculture. 2005;245:287–94.
27. Drauch Schreier A, Gille D, Mahardja B, May B. Neutral markers confirm the octaploid origin reveal spontaneous autopolyploidy in white sturgeon, *Acipenser transmontanus*. J Appl Ichthyol. 2011;27:24–33.
28. Zhou H, Fujimoto T, Adachi S, Abe S, Yamaha E, Arai K. Molecular cytogenetic study on the ploidy status in *Acipenser mikadoi*. J Appl Ichtyol. 2013;29:51–5.
29. Schreier AD, May B, Gille DA. Incidence of spontaneous autopolyploidy in cultured populations of white sturgeon *Acipenser transmontanus*. Aquaculture. 2013;416–417:141–5.

30. Havelka M, Hulák M, Rodina M, Flajšhans M. First evidence of autotriploidization in sterlet (*Acipenser ruthenus*). J Appl Genet. 2013;54:201–7.
31. Havelka M, Hulák M, Ráb P, Rábová M, Lieckfeldt D, Ludwig A, et al. Fertility of a spontaneous triploid Siberian sturgeon, *Acipenser baerii*. BMC Genet. 2014;15:5.
32. Gille DA, Famula TR, May B, Schreier AD. Evidence for a maternal origin of spontaneous autopolyploidy in cultured white sturgeon (*Acipenser transmontanus*). Aquaculture. 2015;435:467–74.
33. Swan AA, Kinghorn BP. Evaluation and exploitation of crossbreeding in dairy cattle. J Dairy Sci. 1992;75:624–39.
34. Mirkena T, Duguma G, Haile A, Tibbo M, Okeyo A, Wurzinger M, et al. Genetics of adaptation in domestic farm animals: a review. Livest Sci. 2010;132:1–12.
35. Zhang X, Wu W, Li L, Ma X, Chen J. Genetic variation and relationships of seven sturgeon species and ten interspecific hybrids. Genet Sel Evol. 2013;45:21.
36. Wei QW, Zou Y, Li P, Li L. Sturgeon aquaculture in China: progress, strategies and prospects assessed on the basis of nation-wide surveys (2007–2009). J Appl Ichthyol. 2011;27:162–8.
37. Pravda D, Svobodova Z. Haematology of fishes. Vet Haematol. 2003;268:381–97.
38. Hardie DC, Gregory TR, Hebert PD. From pixels to picograms: a beginners' guide to genome quantification by Feulgen image analysis densitometry. J Histochem Cytochem. 2002;50:735–49.
39. Bytyutskyy D, Flajšhans M. Use of diploid and triploid tench (*Tinca tinca*) blood as standards for genome size measurements. J Appl Ichthyol. 2014;30:12–4.
40. Fujiwara A, Nishida-Umehara C, Sakamoto T, Okamoto N, Nakayama I, Abe S. Improved fish lymphocyte culture for chromosome preparation. Genetica. 2001;111:77–89.
41. Börk K, Drauch A, Israel JA, Pedroia J, Rodzen J, May B. Development of new microsatellite primers for green sturgeon and white sturgeon. Conserv Genet. 2008;9:973–9.
42. Krueger CC, May B, Kincaid HL. Genetic variation at microsatellite loci in sturgeon: primer sequence homology in *Acipenser* and *Scaphirhynchus*. Can J Fish Aquat Sci. 1997;54:1542–7.
43. Welsh AB, Blumberg M, May B. Identification of microsatellite loci in lake sturgeon, *Acipenser fulvescens*, and their variability in green sturgeon, *A. medirostris*. Mol Ecol Notes. 2003;3:47–55.
44. King TL, Lubinski BA, Spidle AP. Microsatellite DNA variation in Atlantic sturgeon *Acipenser oxyrinchus oxyrinchus*: and cross-species amplification in the Acipenseridae. Conserv Genet. 2001;2:103–19.
45. McQuown EC, Sloss BL, Sheehan RJ, Rodzen J, Tranah GJ, May B. Microsatellite analysis of genetic variation in sturgeon (Acipenseridae): new primer sequences for *Scaphirhynchus* and *Acipenser*. Trans Am Fish Soc. 2000;129:1380–8.
46. StatSoft Inc. STATISTICA (data analysis software system), version 12; 2012. www.statsoft.com.
47. Yu XY, Yu XJ. A schizothoracin fish species, *Diptychus dipogon*, with very high number of chromosomes. Chrom Inform Serv. 1990;48:17–8.
48. Otto SP, Whitton J. Polyploidy: incidence and evolution. Annu Rev Genet. 2000;34:401–37.
49. Mason AS, Pires JCH. Unreduced gametes: meiotic mishap or evolutionary mechanism? Trends Genet. 2015;31:5–10.
50. Dettlaff TA, Ginsburg AS, Schmalhausen OI. Sturgeon fishes. Developmental biology and aquaculture. Berlin: Springer; 1993.
51. Kobayasi H. A cytological study on the maturation division in the oogenic process of the triploid ginbuta (*Carassius auratus langsdorfii*). Jpn J Ichthyol. 1976;76:234–40.
52. Yamashita M, Jiang J, Onozato H, Nakanishi T, Nagahama Y. A tripolar spindle formed at meiosis I assures the retention of the original ploidy in the gynogenetic triploid crucian carp, *Ginbuna Carassius auratus langsdorfii*. Dev Growth Differ. 1993;35:631–6.
53. Shimizu Y, Shibata N, Sakaizumi M, Yamashita M. Production of diploid eggs through premeiotic endomitosis in the hybrid medaka between *Oryzias latipes* and *O. curvinotus*. Zool Sci. 2000;17:951–8.
54. Arai K, Fujimoto T. Genomic constitution and atypical reproduction in polyploidy and unisexual lineages of the Misgurnus loach, a teleost fish. Cytogenet Genome Res. 2013;140:226–40.
55. Leggatt RA, Iwama GK. Occurrence of polyploidy in the fishes. Rev Fish Biol Fisher. 2003;13:237–46.
56. Aegerter S, Jalabert B. Effects of post-ovulatory oocyte ageing and temperature on egg quality and on the occurrence of triploid fry in rainbow trout, *Oncorhynchus mykiss*. Aquaculture. 2004;231:59–71.
57. Flajshans M, Kvasnicka P, Rab P. Genetic studies in tench (*Tinca tinca*)—high incidence spontaneous triploidy. Aquaculture. 1993;110:243–8.
58. Flajshans M, Kohlmann K, Rab P. Autotriploid tench *Tinca tinca* (L.) larvae obtained by fertilization of eggs previously subjected to postovulatory ageing in vitro and in vivo. J Fish Biol. 2007;71:868–76.
59. Nomura K, Takeda Y, Unuma T, Morishima K, Tanaka H, Arai K, et al. Post-ovulatory oocyte aging induces spontaneous occurrence of polyploids and mosaics in artificial fertilization of Japanese eel, *Anguilla japonica*. Aquaculture. 2013;404–405:15–21.
60. Devlin RH, Sakhrani D, Biagi CA, Eom KW. Occurrence of incomplete paternal-chromosome retention in GH-transgenic coho salmon being assessed for reproductive containment by pressure-shock-induced triploidy. Aquaculture. 2010;304:66–78.
61. Varkonyi E, Bercsenyi M, Ozouf-Costaz C, Billard R. Chromosomal and morphological abnormalities caused by oocyte aging in *Silurus glanis*. J Fish Biol. 1998;52:899–906.
62. Glover KA, Madhun AS, Dahle G, Sørvik AG, Wennevik V, Skaala Ø, et al. The frequency of spontaneous triploidy in farmed Atlantic salmon produced in Norway during the period 2007–2014. BMC Genet. 2015;16:37.
63. Piferrer F, Beaumont A, Falguiere JC, Flajšhans M, Haffray P, Colombo L. Polyploid fish and shellfish: production, biology and applications to aquaculture for performance improvement and genetic containment. Aquaculture. 2009;239:125–56.
64. Cuellar O, Uyeno T. Triploidy in rainbow trout. Cytogenetics. 1972;11:508–15.
65. Thorgaard GH, Gall GAE. Adult triploids in a rainbow trout family. Genetics. 1979;93:961–73.
66. Cherfas NB, Rothbard S, Hulata G, Kozinsky O. Spontaneous diploidization of maternal chromosome set in ornamental (koi) carp, *Cyprinus carpio* L. J Appl Ichthyol. 1991;7:72–7.
67. Cherfas NB, Gomelsky B, Ben-Dom N, Hulata G. Evidence for the heritable nature of spontaneous diploidization in common carp *Cyprinus carpio* L. eggs. Aquacult Res. 1995;26:289–92.
68. Itono M, Morishima K, Fujimoto T, Bando E, Yamaha E, Arai K. Premeiotic endomitosis produces diploid eggs in the natural clone loach, *Misgurnus anguillicaudatus* (Teleostei: Cobitidae). J Exp Zool A Comp Exp Biol. 2006;305:513–23.
69. Comai L. The advantages and disadvantages of being polyploid. Nat Rev Genet. 2005;6:836–46.
70. Lewis WH. Polyploidy in species populations. In: Lewis WH, editor. Polyploidy: biological relevance. New York: Plenum Press; 1980. p. 103–92.
71. Schmid M, Evans BJ, Bogart JP. Polyploidy in Amphibia. Cytogenet Genome Res. 2015;145:315–30.
72. Krieger J, Fuerst PA. Evidence for a slowed rate of molecular evolution in the order Acipenseriformes. Mol Biol Evol. 2002;2002(19):891–7.

MOLECULAR ECOLOGY

Molecular Ecology (2016) doi: 10.1111/mec.13816

SPECIAL ISSUE: THE MOLECULAR MECHANISMS OF ADAPTATION AND SPECIATION: INTEGRATING GENOMIC AND MOLECULAR APPROACHES

Standing chromosomal variation in Lake Whitefish species pairs: the role of historical contingency and relevance for speciation

ANNE-MARIE DION-CÔTÉ,*[1] RADKA SYMONOVÁ,† FABIEN C. LAMAZE,‡ ŠÁRKA PELIKÁNOVÁ,§ PETR RÁB§ and LOUIS BERNATCHEZ*

Département de Biologie, Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, 1030, Avenue de la Médecine, Québec, Québec, Canada G1V 0A6, †Research Institute for Limnology, University of Innsbruck, Mondseestraße 9, A-5310 Mondsee, Austria, ‡Ontario Institut for Cancer Research, MaRS Centre, 661 University Avenue, Suite 510, Toronto, Ontario, Canada M5G 0A3, §Laboratory of Fish Genetics, Institute of Animal Physiology and Genetics, AS CR, vvi, Liběchov 277 21, Czech Republic,

Abstract

The role of chromosome changes in speciation remains a debated topic, although demographic conditions associated with divergence should promote their appearance. We tested a potential relationship between chromosome changes and speciation by studying two Lake Whitefish (*Coregonus clupeaformis*) lineages that recently colonized postglacial lakes following allopatry. A dwarf limnetic species evolved repeatedly from the normal benthic species, becoming reproductively isolated. Lake Whitefish hybrids experience mitotic and meiotic instability, which may result from structurally divergent chromosomes. Motivated by this observation, we test the hypothesis that chromosome organization differs between Lake Whitefish species pairs using cytogenetics. While chromosome and fundamental numbers are conserved between the species ($2n$ = 80, NF = 98), we observe extensive polymorphism of subtle karyotype traits. We describe intrachromosomal differences associated with heterochromatin and repetitive DNA, and test for parallelism among three sympatric species pairs. Multivariate analyses support the hypothesis that differentiation at the level of subchromosomal markers mostly appeared during allopatry. Yet we find no evidence for parallelism between species pairs among lakes, consistent with colonization effect or postcolonization differentiation. The reported intrachromosomal polymorphisms do not appear to play a central role in driving adaptive divergence between normal and dwarf Lake Whitefish. We discuss how chromosomal differentiation in the Lake Whitefish system may contribute to the destabilization of mitotic and meiotic chromosome segregation in hybrids, as documented previously. The chromosome structures detected here are still difficult to sequence and assemble, demonstrating the value of cytogenetics as a complementary approach to understand the genomic bases of speciation.

Keywords: *Coregonus*, cytogenetics, polymorphism, salmonids, speciation, standing genetic variation

Received 30 November 2015; revision received 8 August 2016; accepted 11 August 2016

Correspondence: Anne-Marie Dion-Côté, Fax: (418) 656 7176; E-mail: anne-marie.dion-cote.1@ulaval.ca

[1]Present address: Molecular Biology and Genetics Department, Biotechnology building, Cornell University, 526 Campus Road, Ithaca, NY, 14853, USA

Introduction

Understanding the role of genetic and chromosomal changes associated with divergence is a major focus in evolutionary biology (Brown & O'Neill 2010; Marie Curie SPECIATION Network 2012). Thanks to the

advent of massive parallel sequencing technologies, considerable progress has been made in the past decade to decipher the genetic basis of adaptation and speciation (Seehausen *et al.* 2014). However, repetitive regions such as centromeres, constitutive heterochromatin and associated repetitive elements remain challenging to sequence, assemble and characterize (Kim *et al.* 2014), hindering our understanding of their role in population divergence and speciation. Cytogenetic techniques specifically targeting these regions are complementary to next-generation sequencing approaches. They can help to reveal how genomes are structurally organized into chromosomes, how they are shaped by the interplay of evolutionary forces and how chromosome structure changes contribute to speciation (Brown & O'Neill 2010; Faria & Navarro 2010). Such integrative approaches in plants, yeast, mammals and fishes begin to reveal that chromosome structure changes (in addition to well-documented inversions) are associated with divergence and reproductive isolation (Symonová *et al.* 2013; Charron *et al.* 2014).

Some heterochromatic and other poorly assembled repetitive regions are involved in gene expression regulation, meiotic recombination, chromosome segregation and genome stability (Grewal & Jia 2007; Hoskins *et al.* 2007; Brown & O'Neill 2010; Cioffi & Bertollo 2012; Altemose *et al.* 2014). Accordingly, these chromosome structures may greatly influence the genomic landscape of speciation, modulate recombination rate along chromosomes and impact hybrid fitness (Dernburg *et al.* 1996; Grewal & Jia 2007; Brown & O'Neill 2010). In addition, variations in heterochromatin and repetitive region distribution represent a substantial source of intraspecific variation (King 1993; Kidd *et al.* 2008; Britton-Davidian *et al.* 2012), which in turn may modulate the extent of reproductive isolation being achieved during the process of speciation (Cutter 2012). Therefore, more integrative studies are needed in order to understand the role of heterochromatic and repetitive regions in divergence and speciation.

Salmonids typically display substantial inter- and intraspecific chromosome rearrangement and polymorphism (Phillips & Ráb 2001; Sutherland *et al.* 2016), which may be the result of the plasticity of their genome conferred by their ancestral tetraploid state (Allendorf & Thorgaard 1984; Mable *et al.* 2011). Indeed, teleosts experienced a third whole genome duplication (3R – WGD) event preceding their diversification ~350 MYA and salmonids underwent an additional salmonid-specific WGD (4R) 60–90 MYA (Allendorf & Thorgaard 1984; Crête-Lafrenière *et al.* 2012; Macqueen & Johnston 2014). Cytogenetic studies have often revealed intraspecific polymorphism, typically resulting from Robertsonian fusion and fission of chromosomes, but also involving additions and deletions of heterochromatin (reviewed in Phillips & Ráb 2001). For example, the largest metacentric chromosome in *Coregonus* shows length polymorphism in the Lake Whitefish *C. clupeaformis* and the closely related European Whitefish *C. lavaretus*, possibly resulting from variable heterochromatin content (Ráb & Jankun 1992; Jankun *et al.* 1995; Phillips *et al.* 1996; Jankun & Ráb 1997). In Lake Trout (*Salvelinus namaycush*), large blocks of heterochromatin are heritable, as shown by inheritance studies, and polymorphic, in terms of presence/absence and band size (Phillips & Ihssen 1986). However, the role of these intrachromosomal changes has rarely been examined during the early stages of speciation.

The well-characterized phylogeography of the Lake Whitefish makes it a useful model to study potential chromosome changes in the context of divergence and speciation. During the Pleistocene glaciation, two Lake Whitefish lineages (the Atlantic and Acadian lineages) underwent geographical isolation ~60 000 YBP (or ~15–20 000 generations ago) in northeastern North America (Jacobsen *et al.* 2012). The Atlantic and Acadian lineages repeatedly came into secondary contact when they independently colonized newly formed lakes following the Laurentide ice sheet retreat ~12 000 YBP (3–4000 generations ago; Bernatchez & Dodson 1991). Competitive interactions and niche availability presumably contributed to the divergence of a derived, dwarf limnetic form from the ancestral normal benthic form in multiple lakes (Landry *et al.* 2007; Landry & Bernatchez 2010). Previous phylogeographic studies indicated that the dwarf form originated from the Acadian lineage, with variable levels of genetic divergence and admixture with the sympatric normal form of the Atlantic lineage (Lu & Bernatchez 1999; Renaut *et al.* 2012; Gagnaire *et al.* 2013). We subsequently refer to the normal and dwarf Lake Whitefish as distinct species, acknowledging the considerable level of divergence and reproductive isolation between them.

Earlier genetic and transcriptomic studies provided no evidence for differentially fixed mutations nor substantial gene expression differences between normal and dwarf Lake Whitefish (e.g. Campbell & Bernatchez 2004; Renaut *et al.* 2009; Gagnaire *et al.* 2013; Hébert *et al.* 2013; Dion-Côté *et al.* 2014). Yet pronounced postzygotic reproductive isolation has been documented, thus qualifying dwarf and normal Lake Whitefish as distinct species under a relaxed interpretation of the biological species concept. F1-hybrids and backcrosses suffer from a much higher embryonic mortality rate relative to pure parental forms (Lu & Bernatchez 1998; Rogers & Bernatchez 2006; Renaut *et al.* 2009). In backcrosses, hybrid breakdown involves the appearance of a characteristic malformed phenotype [including

reduction in head and eyes size and deformed tail, as described by Renaut & Bernatchez (2011)], gene expression deregulation and transposable element derepression (Renaut & Bernatchez 2011; Dion-Côté *et al.* 2014). Moreover, healthy and malformed backcrosses experience mitotic instability and meiotic breakdown respectively (Dion-Côté *et al.* 2015), suggesting a role for a chromosomal component to reproductive isolation in Lake Whitefish. Nevertheless, it remains unknown whether chromosome rearrangements or structure changes occurred between normal and dwarf Lake Whitefish, nor is it known whether these potential changes played a role in reproductive isolation, given that only modest genetic differences have been documented between them.

In this study, we report detailed cytogenetic characterization of three sympatric pairs of normal and dwarf Lake Whitefish, specifically targeting intrachromosomal markers associated with heterochromatin and repetitive sequences. We test the hypothesis that divergence in the Lake Whitefish system is accompanied by differentiation at the chromosomal and subchromosomal levels. Such structural divergence may result from ancestral allopatry, demographic processes associated with postglacial lake colonization such as founder effect, divergence in sympatry or a combination of these factors. While the basic karyotype remained stable among all populations examined, we observe that intrachromosomal accumulations of heterochromatin and repetitive regions are highly polymorphic. By applying multivariate analyses to these cytogenetic markers, we observe divergence between glacial lineages, among lakes and to some extent between species, and further identify markers associated with this divergence. We discuss how these observations support the presence of *standing chromosomal variation* at the time of lake colonization, and whether this type of polymorphism may be associated with reproductive isolation in this system.

Materials and methods

Sampling and chromosome suspension preparation

We sampled individuals from three lakes with sympatric populations of the St. John River basin: Cliff Lake (ME, USA), Témiscouata Lake (Québec, Canada) and East Lake (Québec, Canada), which are part of a long-term research programme on Lake Whitefish (Bernatchez *et al.* 2010). Glacial lineage assignment (Atlantic or Acadian) relies on earlier phylogeographic studies, which showed variable levels of admixture between glacial lineages within each lake (Lu *et al.* 2001; Pigeon *et al.* 1997). Species assignment (dwarf or normal) was done visually and any individual presenting an ambiguous phenotype was excluded from the study. In Cliff Lake, there is no admixture between dwarf and normal species and they are thus considered as being of pure Acadian and Atlantic origin, respectively. At the other end of the spectrum, normal individuals from Témiscouata Lake are of ~50% Acadian origin while dwarf individuals are of ~60% Acadian origin, likely resulting from a higher level of admixture following secondary contact. Recent analyses based on RADseq genotyping and historical demography inferences suggest that secondary contact also occurred in East Lake (Rougeux *et al.*, in preparation), where previous analyses of only a few available markers had predicted that these individuals originated from a single origin (Acadian). Yet we still labelled East Lake individuals as being of 'Acadian' origin, owing to excessively asymmetrical admixture. To summarize, in Cliff and Témiscouata Lakes dwarf and normal individuals are assigned to Acadian and Atlantic lineages, respectively, while dwarf and normal individuals from East Lake were both assigned to the Acadian lineage. In total, 29 individuals were sampled among the three lakes [see Table 1 for a summary and Table S1 (Supporting information) for more details].

Chromosome suspensions were prepared as described by Fujiwara *et al.* (2001) with some modifications (Dion-Côté *et al.* 2015). Between 0.2 and 2 mL of fresh blood was sampled with heparinized syringes and kept on ice for no more than 12 h. White blood cells were transferred to 5 mL of freshly prepared cell culture media [media 199 (Life technologies), 10% FBS (Sigma), 0.01% LPS (Sigma), 60 μg/mL kanamycin (Sigma), 18 μg/mL phytohemagglutinin (Sigma), 0.5× antibiotic antimycotic (Sigma) and 1.75 μL of 10% β-mercaptoethanol per 100 mL of media]. Cells were incubated for 6 days at 20 °C with gentle mixing every 24 h. Colchicine (25 μL of a 1% solution) was added to the cell suspension 45 min before collection. Cells were hypotonized for 20 min in 2 mL of 0.075 M KCl at room temperature and then fixed by the addition of an equal volume of fresh fixative (3:1 methanol:acetic acid). Three washes with fixative were performed before dropping the suspensions on slides (SuperFrost quality).

Giemsa, Chromomycin A_3 and C-band stainings

Metaphase spreads were stained for 10 min in 3% Giemsa-Romanowski (Dr. Kulich Pharma, Hradec Králové, Czech Republic) in phosphate buffer (pH 6.8–7.0) and then rinsed thoroughly with dH_2O. Chromosomes were sequentially stained with Chromomycin A_3 (CMA_3) and C-banding (with DAPI as a counter stain) according to Rábová *et al.* (2015). CMA_3 stains GC-rich

Table 1 Number of individuals analysed per lake and species with their average phenotypic characteristics (average ± standard deviation)

Lake	Species	Lineage	*n* individuals	Average weight (g)	Average length (cm)	Average Fulton	Average *n* gill rakers
Témiscouata	Dwarf	Acadian	5	82.00 ± 9.90	21.04 ± 0.90	0.88 ± 0.04	22.60 ± 3.51
	Normal	Atlantic	5	197.42 ± 111.41	26.56 ± 5.00	0.97 ± 0.06	22.80 ± 0.45
East	Dwarf	Acadian	4	48.98 ± 22.98	17.68 ± 2.18	0.84 ± 0.14	21.50 ± 2.12
	Normal	Acadian	4	310.38 ± 251.40	29.80 ± 6.49	1.01 ± 0.10	24.25 ± 2.50
Cliff	Dwarf	Acadian	5	138.68 ± 21.97	23.86 ± 1.61	1.02 ± 0.10	24.75 ± 1.26
	Normal	Atlantic	6	427.33 ± 25.42	33.97 ± 0.89	1.10 ± 0.03	24.00 ± 1.41

Fulton: Fulton condition index.

DNA regions, which often colocalize with heterochromatin and major rDNA genes, while C-banding stains constitutive heterochromatin (i.e. chromatin that remains compacted through interphase) presumably associated with repeats, including centromeres (Comings 1978). Chromosomes were examined using a Provis AX70 Olympus microscope, and images taken with a CCD camera (DP30W Olympus) equipped with standard filters. To reduce technical artefacts, at least 10 metaphases per individual were examined, and only consistent signals among metaphases were scored.

Fluorescent in situ hybridization (FISH)

We amplified the whole 5S rDNA and adjacent non-transcribed DNA segments (~170 bp) using previously published primers (5S-A: TACGCCCGATCTCGTCC GATC, 5S-B: CAGGCTGGTATGGCCGTAAGC, Pendás *et al.* 1995). A ~240-bp fragment of the 28S rDNA was also amplified using published primers (28S-C1: ACCCGCTGAATTTAAGCAT, Dayrat *et al.* 2001; 28S-D2: TCCGTGTTTCAAGACGGG, Chombard *et al.* 1998). PCR product identities were confirmed by Sanger sequencing (Macrogen Inc., the Netherlands). PCR products were purified on agarose gel and labelled with biotin-dUTP or digoxigenin-dUTP using the Roche Nick Translation kit according to the manufacturer's instructions (Roche, Mannheim, Germany). Chromosomes were prepared for hybridization according to Cremer *et al.* (2008) following a minimal ageing of 3 h at 37 °C. Labelled probes were hybridized for 24 h at 37 °C. Cy-3-Streptavidin (Invitrogen, San Diego, USA) and anti-digoxigenin–fluorescein (Roche, Mannheim, Germany) were used to detect biotin-dUTP- and digoxigenin-dUTP-labelled probes, respectively.

Multiple factor analyses

We used multivariate analyses to (i) test whether chromosome changes are associated with explanatory variables and (ii) identify chromosome markers associated with these explanatory variables. The input data set included categorical explanatory variables (Sex, Lineage, Lake, Species, Lake:Species), continuous explanatory variables (weight, length, Fulton's condition index [$K = W/L^3 \times 100$; where K = Fulton's condition index, W = weight in g, L = length in cm] and the number of gill rakers) and cytogenetic markers (Table S1, Supporting information). We included phenotypic variables because they may better reflect the proportion of dwarf or normal ancestry of each individual, considering a certain level of gene flow in all three lakes. Therefore, continuous explanatory variables were considered in the imputation of missing data.

As suggested by Dobigny *et al.* (2004), cytogenetic markers were transformed into a presence/absence matrix where a homozygote for absence is coded as '0', a heterozygote is '1' and a homozygote for presence is '2' (Table S1, Supporting information). Coding these variables as continuous better reflects their bi-allelic nature: the heterozygote is more closely related to each homozygote, while each homozygote is less similar. Since acrocentric chromosomes cannot be readily distinguished from one another, we counted the number of chromosomes that had CMA_3 and rDNA 28S (by FISH) signals (numbered 0–6). Similarly, we coded the acrocentric markers (aCMA and a28S) as continuous, which provides a better index of relative similarity between homo- and heterozygotes (e.g. aCMA = 5 is more similar to aCMA = 3 than to aCMA = 0). Missing or ambiguous data were coded as 'NA'. The marker names refer to (i) the number of the chromosome (but with indistinguishable acrocentric 'numbered' as 'a'), (ii) the chromosome arm on which it is found, (iii) the staining technique by which it was evidenced and (iv) the band number on the chromosome arm, where 1 is the closest to the centromere (see Table S1, Supporting information). Markers 1p and 10p refer to the length of the *p*-arm of chromosome 1 and 10, respectively, 0 being the homozygote for short form and 2 being the homozygote for long form. The final data set comprised 39 polymorphic markers (present in >1 individual) in

29 individuals. There were 143 missing data points ('NA') out of 1131 entries (12.6%, Table S1, Supporting information).

Missing data were imputed with the function imputeMFA() from the MISSMDA package (version 1.7.3, Josse & Husson 2012) in R (version 3.2.4, R Core Team 2012). This method imputes missing data (cytogenetic markers) based on the means of other known variables among *similar* individuals (length, weight, Fulton's condition index, gill raker numbers and cytogenetic markers). Without this preliminary step, downstream analyses would have replaced missing data by the mean of *all* of the individuals, thus potentially blurring signal. The imputation included four groups of variables (continuous and categorical): (i) explanatory categorical variables: glacial lineage (Atlantic or Acadian, *sensu* Bernatchez & Dodson 1991), lake (Cliff, East or Témiscouata), species (dwarf or normal) and the combination of lake and species (Lake:Species); (ii) explanatory continuous variables: weight, length, Fulton's condition index and the number of gill rakers; (iii) multichromosomic markers on indistinguishable acrocentric chromosomes; (iv) bi-allelic chromosome markers.

We used the function estim_ncpPCA() from the MISSMDA package (version 1.10, Josse & Husson 2016) to determine the number of principal components (PC) to be used to impute missing data, that is the number of PCs leading to the smallest mean square error of prediction. A function to estimate the number of PCs in MFA has not yet been implemented, but this function gives an approximate alternative (Julie Josse, personal communication). Three principal components were thus used for the imputation (ncp = 3) using the 'Regularized' method (see the MISSMDA package documentation for more details). The resulting complete data set (Table S2, Supporting information) was used for subsequent analyses.

We then performed a Multiple Factorial Analysis (MFA) on the chromosome markers using the MFA() function from the FACTOMINER package (version 1.32, Lê *et al.* 2008). Variables from groups 1 and 2 (explanatory categorical and continuous variables) were coded as supplementary, to test a potential relation with the chromosome markers, and thus did not contribute to defining the dimensions. The chromosome markers were coded in two groups: (i) multichromosomic aCMA3 and a28S markers ('a' stands for 'acrocentric', from 0 to 6 sites) and (ii) bi-allelic markers (from 0 to 2 for zygosity). The function dimdesc() from the FACTOMINER package was used to retrieve supplementary variables (factors and factor levels) significantly linked to dimensions constructed by the MFA. This function applies an ANOVA model with one factor for each dimension. *F*-tests were used to detect the association of each explanatory variable with the dimension (Sex, Lineage, Lake, Species, Lake:Species), followed by *t*-tests for each factor level (Atlantic, Acadian; Témiscouata, East, Cliff; Normal, Dwarf; Témiscouata:Dwarf, Témiscouata:Normal, East:Dwarf, etc.). The dimdesc() function also provides the estimate of the barycentre position (centroid) for each factor level with a significant association with MFA dimensions. The barycentre position can be interpreted as the average position of the individuals with this characteristic in the multivariate space. The function coord.ellipses() from the FACTOMINER package was used to calculate 95% confidence ellipses around the barycentre position.

Results

Karyotypes are stable among Lake Whitefish species pairs

Conventional Giemsa staining confirmed that the Lake Whitefish karyotype is of the salmonid type A *sensu* Phillips & Ráb (2001) and is conserved in all six Lake Whitefish populations from the three lakes ($2n = 80$, NF = 98). This karyotype includes 10 pairs of metacentric/submetacentric chromosomes, one pair of large acrocentric chromosomes and 29 pairs of subtelo/acrocentric chromosomes of decreasing size (Booke 1968; Phillips *et al.* 1996; Dion-Côté *et al.* 2015). Because chromosome 10 is a small submetacentric chromosome (almost subtelocentric) with variable length of the very small *p*-arm, we counted only two arms for this chromosome (and not four as for other submetacentric/metacentric chromosomes), following Phillips & Ráb (2001) guidelines, resulting in the NF = 98 instead of 100.

Subtle karyotype polymorphism was revealed and subsequently included in the multivariate analysis. As shown in Fig. 1, the length of the *p*-arm of chromosome 1 was polymorphic among individuals from all lakes. The short form was near fixation in dwarf fish compared to normal fish from Cliff Lake, while the opposite trend was found in fish from Témiscouata Lake (marker '1p', Table S1, Supporting information). The *p*-arm of chromosome 10 was also polymorphic among individuals from all three lakes (marker '10p', Table S1, Supporting information). In some instances, it was clearly submetacentric, while in others it was subtelomeric/acrocentric (Fig. 1). The *p*-arms of chromosome 1 and 10 are both heterochromatic, although of different nature as suggested by DAPI and C-banding staining, respectively, which may explain their length polymorphism.

In addition, we identified a B or supernumerary chromosome in one dwarf individual from Cliff Lake (Figure S1, Supporting information). B chromosomes,

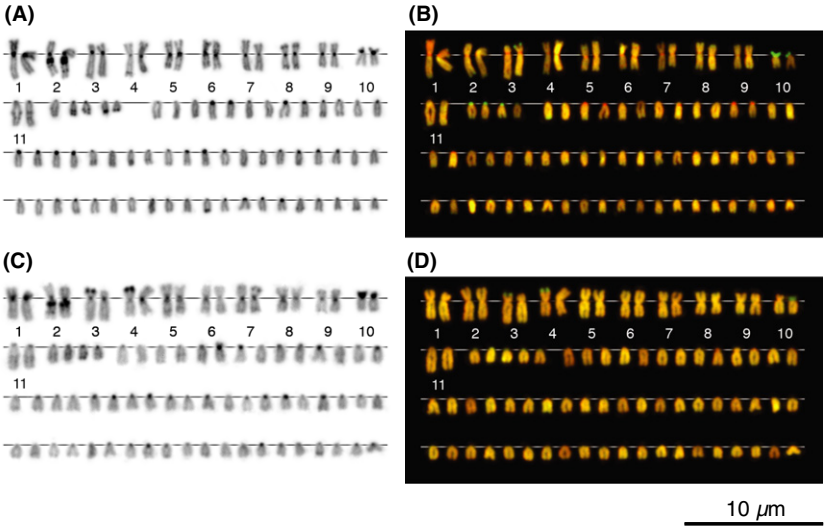


Fig. 1 Karyotypes of a normal and a dwarf individual from East Lake (upper and lower, respectively), sequentially stained with Giemsa (not shown), Chromomycin A_3 (CMA_3) and C-bands to exemplify polymorphism of heterochromatin blocks. Identifiable chromosomes are numbered, followed by acrocentric chromosomes with markers scored, and then remaining acrocentric chromosomes by decreasing size. (A) C-banded karyotype of a normal individual from East Lake (EN27). Note the long form of chromosome 1. (B) CMA_3 karyotype of the same normal individual from East Lake. (C) C-banded karyotype of a dwarf individual from East Lake (ED13). Note the short form of chromosome 1. (D) CMA_3 karyotype of the same dwarf individual from East Lake.

which are usually derived from A chromosomes, occur in some individuals of a population and do not segregate in a Mendelian fashion (Jones 1995; Camacho 2005). This bi-armed B chromosome was not present in all cells examined and was also never found in more than one copy. Additionally, it was positively stained with CMA_3 and C-bands, consistent with the presence of heterochromatin blocks and the presence of repeated elements, a common feature of B chromosomes (Ziegler *et al*. 2003; Camacho 2005; Valente *et al*. 2014).

Heterochromatin revealed by CMA_3 and C-banding

Polymorphic accumulations of heterochromatin and repetitive sequences were characterized among lakes and species pairs. Several chromosomes showed polymorphic CMA_3 banding patterns, indicative of GC-rich DNA regions (Fig. 1). Telomeric CMA_3 signals were present on chromosomes 3, 4, 5, 9 and 10 and on the centromere/*p*-arm of one to six acrocentric chromosomes. A strong telomeric CMA_3 signal on acrocentric chromosomes was found in only two normal individuals, one from Témiscouata Lake and the other from Cliff Lake (marker 'aCMA-telo', Table S1, Supporting information). A large CMA_3-positive block on the *p*-arm of chromosome 4 was also found in a dwarf and a normal individual from Cliff Lake. The same individuals were also the only ones showing CMA_3 band on the q-arm of chromosome 4. Although there was no fixed association between these variations and specific glacial lineage, species or lake, differential representations of these variants contributed to the resolution of significant sex, lake and glacial lineage clusters when combined with other chromosomal structures in subsequent multivariate analyses.

Monomorphic and polymorphic C-bands, indicating constitutive heterochromatin, were found. Most

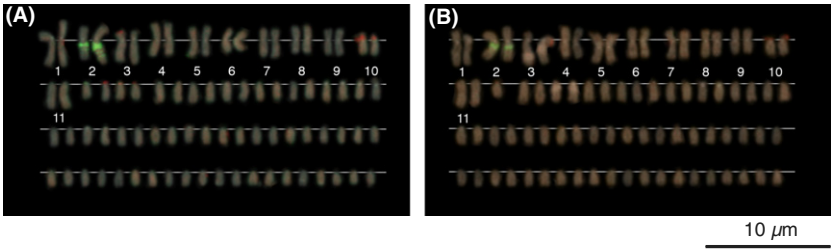


Fig. 2 Karyotypes of (A) a normal (EN17) and (B) a dwarf (ED13) individual from East Lake arranged from DAPI-stained chromosomes after FISH with 5S (green) and 28S (red) rDNA probes showing polymorphism of both rDNA sites.

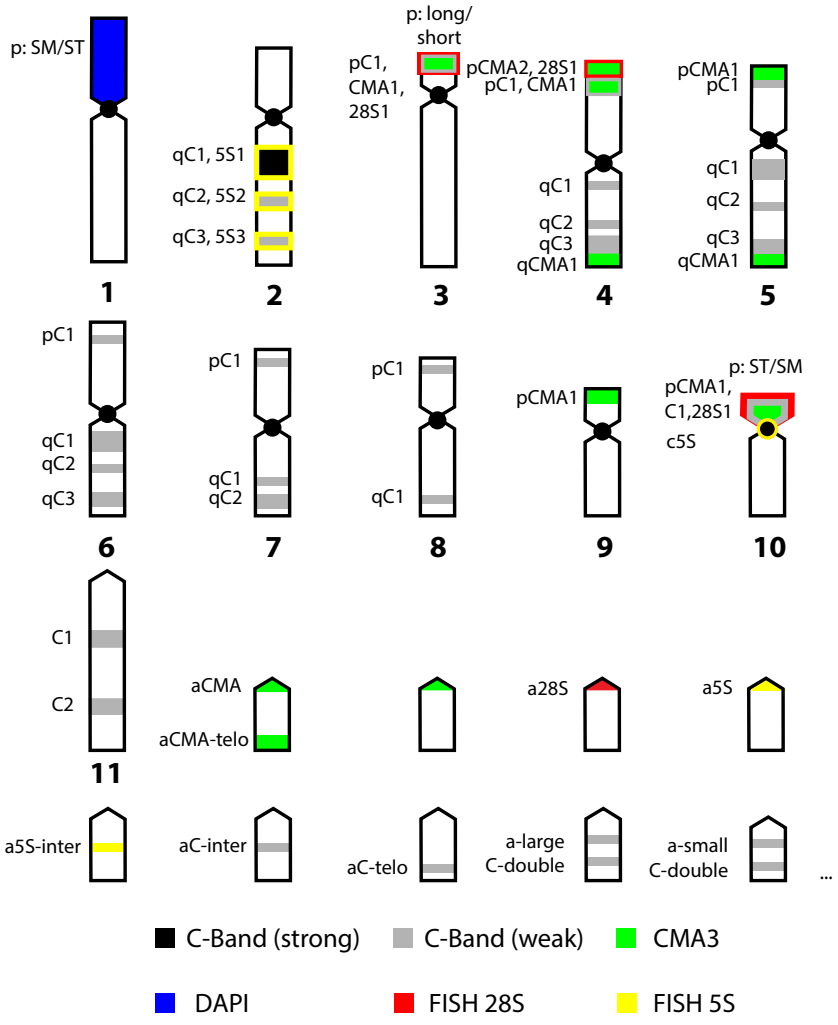


Fig. 3 Partial consensus ideogram for all three species pairs showing chromosome shape and all markers identified on chromosomes and scored. Markers are named according to the chromosome arm on which they are, technique used and distance from centromere (1 for the closest, then 2, etc.). The eleven readily identifiable chromosomes are numbered 1–11, followed by nine acrocentric chromosomes bearing markers (20 chromosomes missing).

chromosomes had centromeric C-bands in all individuals, although the staining was stronger in meta/submetacentric chromosomes compared to acrocentric chromosomes (Fig. 1). Several monomorphic heterochromatin blocks were found: (i) a large heterochromatin block on the q-arm of chromosome 2, close to the centromere; (ii) three bands on the q-arm of chromosomes 5 and 6, and (iii) a double interstitial C-band on the large acrocentric chromosome 11. The remaining bands were all polymorphic, and none were differentially fixed among sex, glacial lineages, lakes or species.

Polymorphism of 5S and 28S ribosomal RNA gene sites

Ribosomal RNA genes (rDNA) are organized as tandem repeats often associated with transposable elements (Cioffi *et al.* 2010; Symonová *et al.* 2013; Vergilino *et al.* 2013). Several sites located on different chromosomes hybridized with the 5S rDNA probe, most of which colocalized with C-bands (Fig. 2). There were three major 5S rDNA sites on chromosome 2: the two distal sites were polymorphic, similar to C-bands (Figs 1 and 2). The 5S rDNA signal was much weaker for other sites. One small interstitial band was found on chromosome 1, a centromeric signal on chromosome 10 (which almost colocalized with 28S rDNA), and centromeric and interstitial signals on different acrocentric chromosomes. Polymorphic 28S rDNA signals were also detected on the *p*-arms of chromosomes 3, 4 and 10. Finally, zero to six 28S rDNA signals were found on the *p*-arms/centromeres of acrocentric/subtelomeric chromosomes. These 28S rDNA sites tended to strongly colocalize with CMA_3 staining. All markers identified and described above are summarized in Fig. 3.

Multiple factorial analysis (MFA) reveals divergence between lineages and among lakes but no parallelism between species

To detect patterns of cytogenetic variation among all fish analysed, a multiple factorial analysis (MFA) was

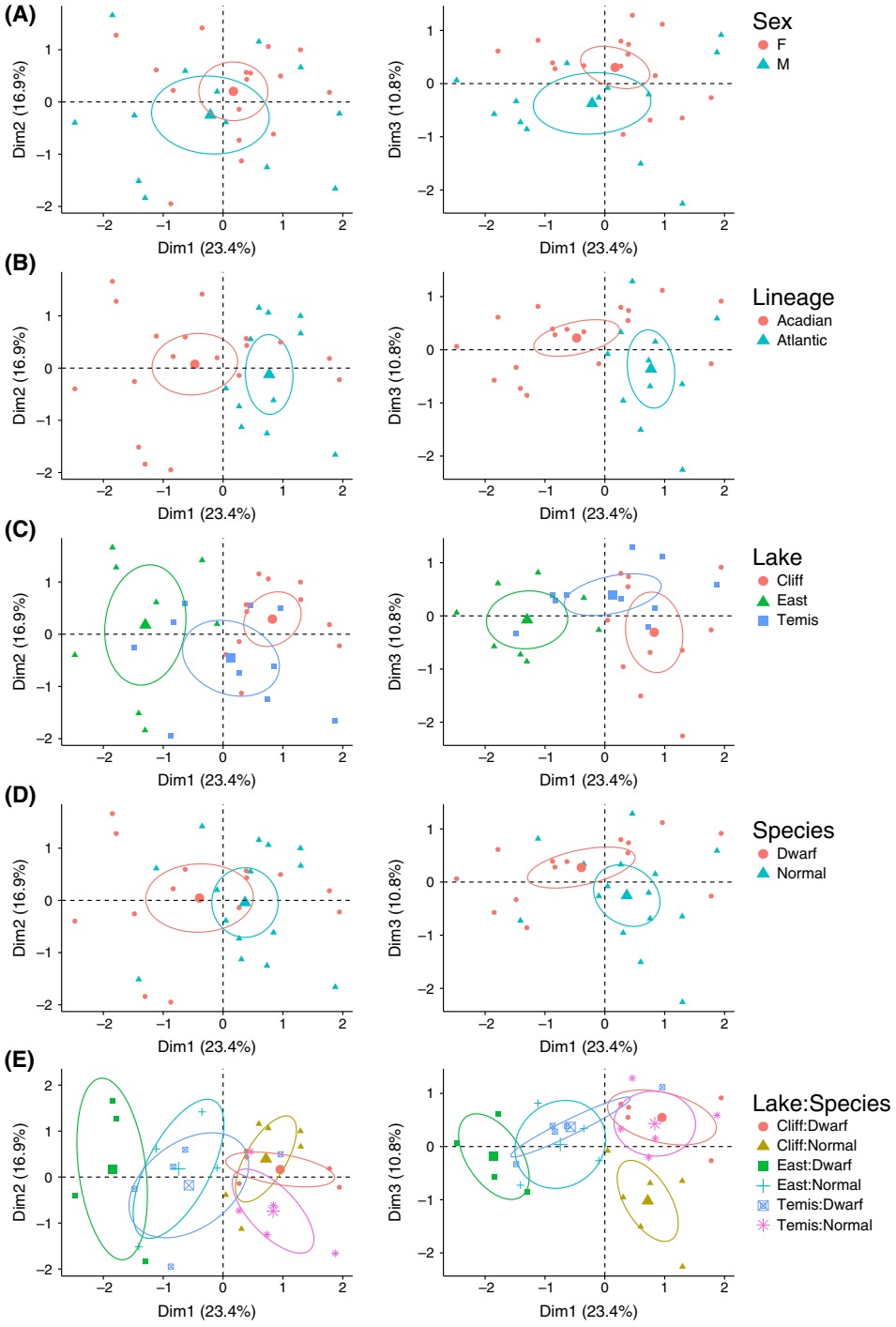


Fig. 4 Multiple factor analysis performed with FACTOMINER showing results for dimensions 1 and 2 (left) and 1 and 3 (right). Each data point represents a single individual in the multivariate space. (A) 95% confidence ellipses around the factor levels of 'Sex', (B) 95% confidence ellipses around the factor levels of 'Lineage', (C) 95% confidence ellipses around the factor levels of 'Species', (D) 95% confidence ellipses around the factor levels of 'Lake' and (E) 95% confidence ellipses around the factor levels of 'Lake:Species'.

applied to chromosome markers, after imputation of missing data (Table S2, Supporting information). Together, dimensions 1 and 3 (23.4% and 10.8% of the variance, respectively) revealed differentiation between glacial lineages, sex and among lakes (Fig. 4). Sympatric dwarf and normal whitefish within each lake also tended

Table 2 Factor effect on dimensions 1, 2 and 3 from the multiple factor analysis

Factor	R^2	*P*-value
Dimension 1		
Sex	0.0268	0.396
Lineage	0.2664	0.004
Lake	0.5352	<0.001
Species	0.1057	0.085
Lake:Species	0.7267	<0.001
Dimension 2		
Sex	0.0517	0.236
Lineage	0.0092	0.620
Lake	0.1148	0.205
Species	0.0019	0.822
Lake:Species	0.1474	0.564
Dimension 3		
Sex	0.1793	0.022
Lineage	0.127	0.058
Lake	0.1469	0.127
Species	0.1102	0.079
Lake:Species	0.5204	0.003

R^2 and *P*-values (*F*-test) were calculated by the dimdesc() function from the FACTOMINER package. The factor 'Lake:Species' denotes the combination of both characteristics (and not a statistical interaction). No factor had a significant effect on dimension 2.

to diverge (minimal overlap between ellipses), although not in parallel and not significantly. This can be seen by examining 95% confidence ellipses around the centroid position for each variable analysed (Fig. 4). The second dimension (16.9% of the variance) was not significantly associated with sex, glacial lineage, lake or species.

Dimension 1 correlated significantly with the variables 'Lake:Species', 'Lake' and 'Lineage' (R^2 = 0.73, 0.53, 0.27, respectively, *F*-test, all *P*-values <0.01; Table 2). Glacial lineage factor levels 'Atlantic' and 'Acadian' were significantly associated with dimension 1 (*t*-test, *P*-value <0.01; Table 3). Cliff and East Lakes were also differentiated by dimension 1 (*t*-test, *P*-value = 0.002 and <0.001, respectively; Table 3). East Lake and Cliff Lake dwarf whitefish were also significantly differentiated by dimension 1 (*t*-test, *P*-value <0.01 and = 0.046, respectively; Table 3). Finally, Fulton condition index was significantly correlated with dimension 1 (R^2 = 0.48, *q*-value = 0.05, *F*-test), but not length alone, weight alone or gill rakers number.

Dimension 3 correlated significantly with the variables 'Lake:Species' and 'Sex' (R^2 = 0.52, 0.18, respectively, *F*-test, *P*-values <0.05; Table 2). It is worth mentioning that the variables 'Lineage' and 'Species' almost reached significance but were more weakly cor-

Table 3 Barycentre (centroid) position estimates of the factor levels on dimensions 1, 2 and 3 from the multiple factor analysis

Factor level	Estimate	*P*-value
Dimension 1		
M	−0.1922	0.396
F	0.1922	0.396
Acadian	−0.6209	0.004
Atlantic	0.6209	0.004
Temis	0.246	0.673
East	−1.1847	<0.001
Cliff	0.9388	0.002
Dwarf	−0.3798	0.085
Normal	0.3798	0.085
Temis:Dwarf	−0.4615	0.244
Temis:Normal	0.9462	0.084
East:Dwarf	−1.7428	<0.001
East:Normal	−0.6338	0.182
Cliff:Dwarf	1.0649	0.046
Cliff:Normal	0.8271	0.098
Dimension 2		
M	−0.2271	0.236
F	0.2271	0.236
Acadian	0.0984	0.620
Atlantic	−0.0984	0.620
Dwarf	0.0434	0.822
Normal	−0.0434	0.822
Temis	−0.4621	0.075
East	0.175	0.570
Cliff	0.2871	0.235
Temis:Dwarf	−0.1776	0.672
Temis:Normal	−0.7397	0.071
East:Dwarf	0.177	0.716
East:Normal	0.18	0.711
Cliff:Dwarf	0.1654	0.698
Cliff:Normal	0.3948	0.293
Dimension 3		
M	−0.3382	0.022
F	0.3382	0.022
Acadian	0.2917	0.058
Atlantic	−0.2917	0.058
Dwarf	0.2638	0.079
Normal	−0.2638	0.079
Temis	0.394	0.052
East	−0.0796	0.766
Cliff	−0.3144	0.109
Temis:Dwarf	0.3391	0.271
Temis:Normal	0.4013	0.196
East:Dwarf	−0.2182	0.623
East:Normal	0.0115	0.916
Cliff:Dwarf	0.518	0.097
Cliff:Normal	−1.0516	<0.001

Position estimates and *P*-values (*t*-test) were calculated by the dimdesc() function from the FACTOMINER package. Note that no factor had a significant effect on dimension 2, although Témiscouata and Témiscouata:Normal were nearly significant.

related with dimension 3 (R^2 = 0.13, 0.11, respectively, *F*-test, *P*-value = 0.058 and 0.079, respectively; Table 2). The variable levels male ('M') and female ('F') had a significant effect on dimension 3 (*t*-test, *P*-value <0.05; Table 3). The variable level Cliff:Normal also had a significant effect on dimension 3 (*t*-test, *P*-value <0.001; Table 3).

Identification of chromosomal markers associated with divergence

To identify the chromosome markers that were most correlated to differentiation, we retrieved markers that were most correlated to dimensions 1, 2 and 3 (Table S3, Supporting information). Namely, signals of CMA_3 and 28S rDNA on acrocentric chromosomes (a28S and aCMA) were positively correlated with each other and with the dimension 1 (R^2 = 0.90, $P < 0.001$; R^2 = 0.82, *q*-value <0.001, respectively). This dimension also resolved glacial lineages (Table 3, Fig. 4). This means that individuals from the Atlantic lineage tend to have more CMA_3 and 28S rDNA sites on the centromeres of their acrocentric chromosomes. Twenty-two different markers out of 39 were significantly correlated with dimensions 1 and 3 after *P*-value adjustment for multiple testing (Table S3, Supporting information). Therefore, many markers covary with each other and are associated with divergence between glacial lineages, among lakes and between dwarf and normal whitefish within lakes.

Discussion

The main objective of this study was to test the hypothesis that divergence in the Lake Whitefish system is accompanied by chromosomal and subchromosomal structure changes. Towards this goal, we investigated the relationship between chromosomal polymorphism and the rapid genetic and phenotypic divergence among Lake Whitefish species pairs, targeting cytogenetic markers associated with heterochromatin and repetitive DNA. Basic karyotype remained stable among lineages, lakes and species, thus rejecting a contribution of large-scale chromosomal rearrangement to divergence in this system. However, by implementing a multivariate statistical framework we found that intrachromosomal changes were modelled by historical contingency, being primarily associated with earlier allopatric divergence among glacial lineages and recent interlake divergence. In addition, dwarf and normal fish showed a trend towards divergence within lakes, although not in parallel among lakes. Together, these observations support the hypothesis that polymorphic subchromosomal traits are influenced by historical contingency and associated with divergence in the Lake Whitefish system.

A statistical multivariate strategy to analyse chromosomal polymorphism

Our statistical multivariate strategy helped to resolve patterns in this highly polymorphic data set and identify markers associated with these patterns of divergence. Importantly, this approach allows the use of discrete and continuous data, such as presence/absence of a specific marker or the number of rDNA sites. In addition, it is possible to include supplementary phenotypic measures (e.g. length, weight) or environmental data (e.g. lake) to test their association with cytogenetic patterns. A similar method was recently published based on principal coordinates analysis (PCoA, Peruzzi & Altınordu 2014). However, this method is based on continuous variables such as total haploid chromosome length and centromeric asymmetry, which are difficult to implement in nonmodel systems. In addition, this method does not allow handling of discrete data, such as presence and absence of cytogenetic markers. To our knowledge, ours is the first study for which MFA is applied to cytogenetic data. We have made our code readily available to the community (see Data S1, Supporting information).

High chromosomal polymorphism in the Lake Whitefish system

The karyotypes described herein are consistent with previous work in other populations of Lake Whitefish and show no large-scale chromosomal (e.g. inversions) or Robertsonian (i.e. fusions and fission) rearrangements within this species complex (Booke 1968; Phillips *et al.* 1996; Dion-Côté *et al.* 2015). Yet substantial intrachromosomal polymorphism was identified, mainly involving labile or rapidly evolving structures associated with heterochromatin and repetitive DNA. Specific characteristics of the Lake Whitefish system may contribute to this high level of polymorphism.

First, pronounced karyotype and intrachromosomal polymorphism is common in salmonids, including *Coregonus* (Phillips & Ráb 2001). For example, the *p*-arm of chromosome 1 in *Coregonus* shows variable length (Phillips *et al.* 1996; Jankun & Ráb 1997), and rDNA sites are also polymorphic among lineages of *Salmo trutta* (Caputo *et al.* 2009) and *Coregonus albula* (Jankun *et al.* 2003). These observations are consistent with karyotype and genetic flexibility in fishes in general and especially in salmonids following another round of genome duplication (Phillips & Ráb 2001; Ravi & Venkatesh 2008; Mable *et al.* 2011).

Also, we hypothesize that gene flow between dwarf and normal Lake Whitefish may contribute to spread and diversify chromosomal polymorphisms in Lake Whitefish. We have previously shown that Lake Whitefish hybrids experience genomic instability at several levels. Introgressive hybridization, which occurs in Lake Whitefish natural populations (Gagnaire *et al.* 2013), can promote genome reorganization in fishes, even with the same basic karyotype structure (Pereira *et al.* 2014, 2015). Hybridization between dwarf and normal Lake Whitefish is also associated with transposable element derepression (or reactivation) in hybrids (Renaut *et al.* 2010; Dion-Côté *et al.* 2014), which can promote genome rearrangements (Levin & Moran 2011). Lastly, we have found that aneuploidy, a severe form of genome instability that can promote chromosome rearrangements (Santaguida & Amon 2015), occurs in Lake Whitefish backcrosses (Dion-Côté *et al.* 2015).

Overall, the intrinsic properties of salmonid genomes, ongoing introgressive hybridization and complex historical biogeography of the Lake Whitefish are all expected to influence genome lability and consequently polymorphism. The sampling of purely allopatric populations and other sympatric species pairs, including from European lineages, should help to further disentangle the relative contributions of ancestral chromosomal polymorphisms and *de novo* intrachromosomal reorganization in relation to divergence.

Historical contingency and divergence acting on standing chromosomal variation

Heterochromatin and rDNA polymorphisms are mostly shared among the three species pairs examined, supporting the idea that ancestral Lake Whitefish population had high levels of polymorphism, or *standing chromosomal variation*. Nonetheless, multivariate analyses revealed three nested levels of divergence based upon cytogenetic markers, which are consistent with well-documented population genetic structure in the system (Pigeon *et al.* 1997; Lu *et al.* 2001; Campbell & Bernatchez 2004; Bernatchez *et al.* 2010; Renaut *et al.* 2011): (i) between glacial lineages, (ii) among lakes and (iii) between sympatric species pairs within lakes (albeit to a lesser extent and not in parallel).

Geographical isolation between the Atlantic and Acadian lineages either promoted divergence among ancestral chromosomal variants (*standing chromosomal variation*) and/or allowed for *de novo* remodelling of subchromosomal traits within a glacial lineage. Three nonmutually exclusive hypotheses may explain divergence among lakes. First, population bottleneck associated with lake colonization may have led to stochastic differentiation of ancestral chromosomal variants (Mayr 1954). Second, there may also have been *de novo* remodelling following lake colonization, considering the markers used are associated with repetitive sequences contained in heterochromatin and are extremely labile. Such rapid remodelling (<15 000 years) of intrachromosomal structures associated with ecological divergence has been documented in another young *Coregonus* species pair in Europe (Symonová *et al.* 2013). Third, the shared chromosomal variation between sympatric dwarf and normal fish may also be enhanced by ongoing gene flow (Gagnaire *et al.* 2013).

Finally, species within lakes showed a trend towards chromosomal divergence. However, this was not significant in most cases and did not occur in parallel among lakes despite extensive parallelism at the phenotypic level (Landry *et al.* 2007). In other words, cytogenetic markers of dwarf fish do not diverge from cytogenetic markers of normal fish similarly in all lakes. The absence of parallelism and incomplete differentiation between species within lake can be explained by a combination of factors: (i) the short time since divergence (~12 000 YBP or ~3000–4000 generations, (ii) the unique interaction and relative intensity (e.g. genetic drift vs. gene flow) of evolutionary forces in each lake (Lu & Bernatchez 1999; Gagnaire *et al.* 2013) and (iii) possible *de novo* intrachromosomal reorganization. While *de novo* subchromosomal changes following postglacial lake colonization may have contributed to the high level of polymorphism observed, our data suggest that there was a rather high level of *standing chromosomal variation* segregating between both Lake Whitefish lineages, as most polymorphism is shared among lakes.

How the observed levels of shared polymorphism have been maintained before secondary contact and in contemporary diverging populations calls for further studies. A possibility is that balancing selection acts on certain of these polymorphic chromosome traits (da Cunha *et al.* 1950). In particular, it has been recently documented that antagonistic sexual selection, a form of balancing selection, was involved in maintaining polymorphism at an important fitness-related gene in Atlantic salmon, *Salmo salar* (Barson *et al.* 2015). In the same manner, it could hypothetically be that here, some of the observed polymorphic chromosomal traits could be selectively favoured in one sex but negatively favoured in the other, which would lead to the maintenance of variation. Interestingly, we observed some subchromosomal differences between sexes (but no differentially fixed markers), suggesting a possibly functional, perhaps fitness-related impact of subchromosomal variation between sexes.

Heterochromatin architecture divergence and reproductive isolation

Michalak (2009) suggested that heterochromatin structure and function may be involved in rapid divergence and hybrid breakdown. Heterochromatin plays pivotal roles in transcriptional regulation and chromosome segregation (Grewal & Jia 2007) and may influence crossover localization during meiosis (John & King 1985; Ramachandran *et al.* 1985; Cioffi & Bertollo 2012). Several indirect observations suggest that heterochromatin divergence between normal and dwarf whitefish, and disruption in their hybrids, may occur. We previously found that DNMT1, an enzyme involved in heterochromatin maintenance (DNA methylation specifically), is downregulated in malformed backcross embryos (Dion-Côté *et al.* 2014). We also reported global transcriptional deregulation, transposable element derepression and noncoding RNA upregulation, consistent with heterochromatin disruption in these malformed backcrosses (Dion-Côté *et al.* 2014). Moreover, we have previously shown that Lake Whitefish hybrids suffer from mitotic and meiotic instability (Dion-Côté *et al.* 2015). Here, we observed that cytogenetic markers associated with heterochromatin and repetitive DNA tend to differ between sympatric species pairs. As a first step to directly test the role of heterochromatin in the divergence of Lake Whitefish species pairs, we are currently studying DNA methylation patterns in dwarf and normal Lake Whitefish, and inheritance patterns in reciprocal hybrids.

In conclusion, by combining conventional and molecular cytogenetic techniques, we have uncovered extensive intrachromosomal polymorphism despite constant basic karyotype structure and chromosome number in three Lake Whitefish sympatric species pairs that evolved independently. Polymorphic chromosomal markers were correlated with geographical isolation, lake colonization and sympatric species divergence, albeit to a lower extent. Barbara McClintock predicted more than thirty years ago that rapid genome reorganization would accompany speciation (McClintock 1984). A large body of work has shown that natural selection drove ecological, morphological and physiological divergence in the Lake Whitefish system; polymorphic intrachromosomal traits may have contributed to consolidate reproductive isolation in the system. Clearly, cytogenetic tools can highlight important chromosome traits from an evolutionary standpoint, which cannot readily be evidenced even with the best sequencing techniques currently available. Consequently, we echo Valente *et al.* (2014) in claiming that such cytogenetics still have much to offer in the postgenomics era.

Acknowledgements

We thank Martin Laporte, Clément Rougeux, Anne C. Dalziel, Ben J.G. Sutherland and Daniel Zinshteyn for comments on an earlier version of this manuscript. Philipp Schlüter and three anonymous reviewers also provided careful and valuable feedback. We are grateful to Guillaume Côté and members of the Bernatchez laboratory for fieldwork, Alex Bernatchez for counting gill rakers, as well as Jana Čechová and Petra Šejnohová for technical help and support. This work was supported by a Natural Science and Engineering Research Council of Canada (NSERC) discovery grant and a Canadian Research Chair in Genomics and Conservation of Aquatic Resources to LB, a NSERC postgraduate scholarship to AMDC and a postdoctoral fellowship to FCL from the Fonds de Recherche du Québec – Santé (FRQS). RS and PR were supported by the project 14-02940S of the Czech Science Foundation. AMDC also received a stipend from the FRQ-NT for international training and from Québec-Océan. This is a contribution to the Québec-Océan research programme.

References

Allendorf FW, Thorgaard GH (1984) Tetraploidy and the evolution of salmonid fishes. In: *Evolutionary Genetics of Fishes* (ed. Turner BJ), 1–53. Plenum Press, New York.

Altemose N, Miga KH, Maggioni M, Willard HF (2014) Genomic characterization of large heterochromatic gaps in the human genome assembly. *PLoS Computational Biology*, **10**, e1003628: 1–14.

Barson NJ, Aykanat T, Hindar K *et al.* (2015) Sex-dependent dominance at a single locus maintains variation in age at maturity in salmon. *Nature*, **528**, 405–408.

Bernatchez L, Dodson JJ (1991) Phylogeographic structure in mitochondrial DNA of the Lake Whitefish (*Coregonus clupeaformis*) and its relation to Pleistocene glaciations. *Evolution*, **45**, 1016–1035.

Bernatchez L, Renaut S, Whiteley AR *et al.* (2010) On the origin of species: insights from the ecological genomics of Lake Whitefish. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **365**, 1783–1800.

Booke HE (1968) Cytotaxonomic studies of coregonine fishes of the Great Lakes, USA: DNA and karyotype analysis. *Journal of the Fisheries Research Board of Canada*, **25**, 1667–1687.

Britton-Davidian J, Cazaux B, Catalan J (2012) Chromosomal dynamics of nucleolar organizer regions (NORs) in the house mouse: micro-evolutionary insights. *Heredity*, **108**, 68–74.

Brown JD, O'Neill RJ (2010) Chromosomes, conflict, and epigenetics: chromosomal speciation revisited. *Annual Review of Genomics and Human Genetics*, **11**, 291–316.

Camacho JPM (2005) B chromosomes. In: *The Evolution of the Genome* (ed. Gregory TR), pp. 223–286. Academic Press, Burlington, MA.

Campbell D, Bernatchez L (2004) Generic scan using AFLP markers as a means to assess the role of directional selection in the divergence of sympatric whitefish ecotypes. *Molecular Biology and Evolution*, **21**, 945–956.

Caputo V, Giovannotti M, Nisi Cerioni P, Splendiani A, Olmo E (2009) Chromosomal study of native and hatchery trouts from Italy (*Salmo trutta* complex, Salmonidae): conventional and FISH analysis. *Cytogenetic and Genome Research*, **124**, 51–62.

Charron G, Leducq J-B, Landry CR (2014) Chromosomal variation segregates within incipient species and correlates with reproductive isolation. *Molecular Ecology*, **23**, 4362–4372.

Chombard C, Boury-Esnault N, Tillier S (1998) Reassessment of homology of morphological characters in tetractinellid sponges based on molecular data. *Systematic Biology*, **47**, 351–366.

Cioffi MB, Bertollo LAC (2012) Chromosomal distribution and evolution of repetitive DNAs in fish. *Genome Dynamics*, **7**, 197–221.

Cioffi MB, Martins C, Bertollo LA (2010) Chromosome spreading of associated transposable elements and ribosomal DNA in the fish *Erythrinus erythrinus*. Implications for genome change and karyoevolution in fish. *BMC Evolutionary Biology*, **10**, 1–9.

Comings DE (1978) Mechanisms of chromosome banding and implications for chromosome structure. *Annual Review of Genetics*, **12**, 25–46.

Cremer M, Grasser F, Lanctôt C *et al.* (2008) Multicolor 3D fluorescence *in situ* hybridization for imaging interphase chromosomes. *Methods in Molecular Biology (Clifton, NJ)*, **463**, 205–239.

Crête-Lafrenière A, Weir LK, Bernatchez L (2012) Framing the Salmonidae family phylogenetic portrait: a more complete picture from increased taxon sampling. *PLoS ONE*, **7**, e46662.

da Cunha AB, Burla H, Dobzhansky T (1950) Adaptive chromosomal polymorphism in *Drosophila willistoni*. *Evolution*, **4**, 212–235.

Cutter AD (2012) The polymorphic prelude to Bateson–Dobzhansky–Muller incompatibilities. *Trends in Ecology & Evolution*, **27**, 210–219.

Dayrat B, Tillier A, Lecointre G, Tillier S (2001) New clades of euthyneuran gastropods (Mollusca) from 28S rRNA sequences. *Molecular Phylogenetics and Evolution*, **19**, 225–235.

Dernburg AF, Sedat JW, Hawley RS (1996) Direct evidence of a role for heterochromatin in meiotic chromosome segregation. *Cell*, **86**, 135–146.

Dion-Côté A-M, Renaut S, Normandeau E, Bernatchez L (2014) RNA-seq reveals transcriptomic shock involving transposable elements reactivation in hybrids of young Lake Whitefish species. *Molecular Biology and Evolution*, **31**, 1188–1199.

Dion-Côté A-M, Symonová R, Ráb P, Bernatchez L (2015) Reproductive isolation in a nascent species pair is associated with aneuploidy in hybrid offspring. *Proceedings of the Royal Society B: Biological Sciences*, **282**, 20142862.

Dobigny G, Ducroz J-F, Robinson TJ, Volobouev V (2004) Cytogenetics and Cladistics. *Systematic Biology*, **53**, 470–484.

Faria R, Navarro A (2010) Chromosomal speciation revisited: rearranging theory with pieces of evidence. *Trends in Ecology & Evolution*, **25**, 660–669.

Fujiwara A, Nishida-Umehara C, Sakamoto T *et al.* (2001) Improved fish lymphocyte culture for chromosome preparation. *Genetica*, **111**, 77–89.

Gagnaire P-A, Pavey SA, Normandeau E, Bernatchez L (2013) The genetic architecture of reproductive isolation during speciation-with-gene-flow in Lake Whitefish species pairs assessed by RAD sequencing. *Evolution*, **67**, 2483–2497.

Grewal SIS, Jia S (2007) Heterochromatin revisited. *Nature Reviews Genetics*, **8**, 35–46.

Hébert FO, Renaut S, Bernatchez L (2013) Targeted sequence capture and resequencing implies a predominant role of regulatory regions in the divergence of a sympatric Lake Whitefish species pair (*Coregonus clupeaformis*). *Molecular Ecology*, **22**, 4896–4914.

Hoskins RA, Carlson JW, Kennedy C *et al.* (2007) Sequence finishing and mapping of *Drosophila melanogaster* heterochromatin. *Science (New York, NY)*, **316**, 1625–1628.

Jacobsen MW, Hansen MM, Orlando L *et al.* (2012) Mitogenome sequencing reveals shallow evolutionary histories and recent divergence time between morphologically and ecologically distinct European whitefish (*Coregonus* spp.). *Molecular Ecology*, **21**, 2727–2742.

Jankun M, Ráb P (1997) Multiple polymorphism of chromosome no. 1 in the karyotype of whitefish, *Coregonus lavaretus* (Salmonidae) from lake system Saimaa, Finland. *Caryologia*, **50**, 185–195.

Jankun M, Ráb P, Vuorinen J, Lucyznski M (1995) Chromosomal polymorphism in *Coregonus lavaretus* populations from two locations in Finland and Poland. *Archive Hydrobiology Special Issues Advance Limnology*, **46**, 1–11.

Jankun M, Ocalewicz K, Pardo BG *et al.* (2003) Localization of 5S rRNA loci in three coregonid species (Salmonidae). *Genetica*, **119**, 183–186.

John B, King M (1985) The inter-relationship between heterochromatin distribution and chiasma distribution. *Genetica*, **66**, 183–194.

Jones RN (1995) B chromosomes in plants. *New Phytologist*, **131**, 411–434.

Josse J, Husson F (2012) Handling missing values in exploratory multivariate data analysis methods. *Journal de la Société Française de Statistiques*, **153**, 79–99.

Josse J, Husson F (2016) missMDA: a package for handling missing values in multivariate data analysis. *Journal of Statistical Software*, **70**, 1–31.

Kidd JM, Cooper GM, Donahue WF *et al.* (2008) Mapping and sequencing of structural variation from eight human genomes. *Nature*, **453**, 56–64.

Kim KE, Peluso P, Babayan P *et al.* (2014) Long-read, whole-genome shotgun sequence data for five model organisms. *Scientific Data*, **1**, 140045.

King M (1993) *Species Evolution: The Role of Chromosome Change*. Cambridge University Press, Cambridge.

Landry L, Bernatchez L (2010) Role of epibenthic resource opportunities in the parallel evolution of Lake Whitefish species pairs (*Coregonus* sp.). *Journal of Evolutionary Biology*, **23**, 2602–2613.

Landry L, Vincent WF, Bernatchez L (2007) Parallel evolution of Lake Whitefish dwarf ecotypes in association with limnological features of their adaptive landscape. *Journal of Evolutionary Biology*, **20**, 971–984.

Lê S, Josse J, Husson F (2008) FactoMineR: an R package for multivariate analysis. *Journal of Statistical Software*, **25**, 1–28.

Levin HL, Moran JV (2011) Dynamic interactions between transposable elements and their hosts. *Nature Reviews Genetics*, **12**, 615–627.

Lu G, Bernatchez L (1998) Experimental evidence for reduced hybrid viability between dwarf and normal ecotypes of Lake Whitefish (*Coregonus clupeaformis* Mitchill). *Proceedings of the Royal Society of London B: Biological Science*, **265**, 1025–1030.

Lu G, Bernatchez L (1999) Correlated trophic specialization and genetic divergence in sympatric Lake Whitefish ecotypes (*Coregonus clupeaformis*): support for the ecological speciation hypothesis. *Evolution*, **53**, 1491–1505.

Lu G, Basley DJ, Bernatchez L (2001) Contrasting patterns of mitochondrial DNA and microsatellite introgressive hybridization between lineages of Lake Whitefish (*Coregonus clupeaformis*); relevance for speciation. *Molecular Ecology*, **10**, 965–985.

Mable BK, Alexandrou MA, Taylor MI (2011) Genome duplication in amphibians and fish: an extended synthesis. *Journal of Zoology*, **284**, 151–182.

Macqueen DJ, Johnston IA (2014) A well-constrained estimate for the timing of the salmonid whole genome duplication reveals major decoupling from species diversification. *Proceedings of the Royal Society B: Biological Sciences*, **281**, 20132881: 1–8.

Marie Curie SPECIATION Network (2012) What do we need to know about speciation? *Trends in Ecology & Evolution*, **27**, 27–39.

Mayr E (1954) Change of genetic environment and evolution.

McClintock B (1984) The significance of responses of the genome to challenge. *Science (New York, NY)*, **226**, 792–801.

Michalak P (2009) Epigenetic, transposon and small RNA determinants of hybrid dysfunctions. *Heredity*, **102**, 45–50.

Pendás AM, Moran P, Martinez JL, Garcia Vazquez E (1995) Applications of 5S rDNA in Atlantic salmon, brown trout, and in Atlantic salmon brown trout hybrid identification. *Molecular Ecology*, **4**, 275–276.

Pereira CSA, Aboim MA, Ráb P, Collares-Pereira MJ (2014) Introgressive hybridization as a promoter of genome reshuffling in natural homoploid fish hybrids (Cyprinidae, Leuciscinae). *Heredity*, **112**, 343–350.

Pereira CSA, Pazian MF, Ráb P, Collares-Pereira MJ (2015) Dynamics of Rex3 in the genomes of endangered Iberian Leuciscinae (Teleostei, Cyprinidae) and their natural hybrids. *Molecular Cytogenetics*, **8**, 1–10.

Peruzzi L, Altınordu F (2014) A proposal for a multivariate quantitative approach to infer karyological relationships among taxa. *Comparative Cytogenetics*, **8**, 337–349.

Phillips RB, Ihssen PE (1986) Inheritance of Q band chromosomal polymorphisms in lake trout. *Journal of Heredity*, **77**, 93–97.

Phillips RB, Ráb P (2001) Chromosome evolution in the Salmonidae (Pisces): an update. *Biological Reviews*, **76**, 1–25.

Phillips RB, Reed KM, Ráb P (1996) Revised karyotypes and chromosome banding of coregonid fishes from the Laurentian Great Lakes. *Canadian Journal of Zoology-Revue Canadienne De Zoologie*, **74**, 323–329.

Pigeon D, Chouinard A, Bernatchez L (1997) Multiple modes of speciation involved in the parallel evolution of sympatric morphotypes of Lake Whitefish (*Coregonus clupeaformis*, Salmonidae). *Evolution*, **51**, 196–205.

R Core Team (2012) R: a language and environment for statisti cal computing. Available from: http://www.R-project.org/.

Ráb P, Jankun M (1992) Chromosome studies of coregonine fishes: a review. *Polskie Archiwum Hydrobiologii*, **39**, 523–532.

Rábová M, Völker M, Pelikánová Š, Ráb P (2015) Sequential chromosome banding in fishes. In: *Fish Cytogenetic Techniques: Ray-Fin Fishes and Chondrichthyans* (eds Ozouf-Costaz C, Pisano E, Foresti F, de Almeida Toledo LF), pp. 66–73. CRC Press Inc., Enfield.

Ramachandran C, Brandenburg WA, Nijs den APM (1985) Infraspecific variation in C-banded karyotype and chiasma frequency in *Cucumis sativus* (Cucurbitaceae). *Plant Systematics and Evolution*, **151**, 31–41.

Ravi V, Venkatesh B (2008) Rapidly evolving fish genomes and teleost diversity. *Current Opinion in Genetics & Development*, **18**, 544–550.

Renaut S, Bernatchez L (2011) Transcriptome-wide signature of hybrid breakdown associated with intrinsic reproductive isolation in Lake Whitefish species pairs (*Coregonus* spp. Salmonidae). *Heredity*, **106**, 1003–1011.

Renaut S, Nolte AW, Bernatchez L (2009) Gene expression divergence and hybrid misexpression between Lake Whitefish species pairs (*Coregonus* spp. salmonidae). *Molecular Biology and Evolution*, **26**, 925–936.

Renaut S, Nolte AW, Bernatchez L (2010) Mining transcriptome sequences towards identifying adaptive single nucleotide polymorphisms in Lake Whitefish species pairs (*Coregonus* spp. Salmonidae). *Molecular Ecology*, **19**(Suppl. 1), 115–131.

Renaut S, Nolte AW, Rogers SM *et al.* (2011) SNP signatures of selection on standing genetic variation and their association with adaptive phenotypes along gradients of ecological speciation in Lake Whitefish species pairs (*Coregonus* spp.). *Molecular Ecology*, **20**, 545–559.

Renaut S, Maillet N, Normandeau E *et al.* (2012) Genome-wide patterns of divergence during speciation: the Lake Whitefish case study. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **367**, 354–363.

Rogers SM, Bernatchez L (2006) The genetic basis of intrinsic and extrinsic post-zygotic reproductive isolation jointly promoting speciation in the Lake Whitefish species complex (*Coregonus clupeaformis*). *Journal of Evolutionary Biology*, **19**, 1979–1994.

Santaguida S, Amon A (2015) Short- and long-term effects of chromosome mis-segregation and aneuploidy. *Nature Reviews Molecular Cell Biology*, **16**, 473–485.

Seehausen O, Butlin RK, Keller I *et al.* (2014) Genomics and the origin of species. *Nature Reviews Genetics*, **15**, 176–192.

Sutherland BJG, Gosselin T, Normandeau E *et al.* (2016) Salmonid chromosome evolution as revealed by a novel method for comparing RADseq linkage maps reveals chromosome evolution in salmonids. *BioRxiv*, 039164. doi: http://dx.doi.org/10.1101/039164.

Symonová R, Majtánová Z, Sember A *et al.* (2013) Genome differentiation in a species pair of coregonine fishes: an extremely rapid speciation driven by stress-activated retrotransposons mediating extensive ribosomal DNA multiplications. *BMC Evolutionary Biology*, **13**, 1–11.

Valente GT, Conte MA, Fantinatti BEA *et al.* (2014) Origin and evolution of B chromosomes in the Cichlid Fish *Astatotilapia latifasciata* based on integrated genomic analyses. *Molecular Biology and Evolution*, **31**, 2061–2072.

Vergilino R, Elliott TA, Desjardins-Proulx P, Crease TJ, Dufresne F (2013) Evolution of a transposon in Daphnia hybrid genomes. *Mobile DNA*, **4**, 1–14.

Ziegler CG, Lamatsch DK, Steinlein C *et al.* (2003) The giant B chromosome of the cyprinid fish *Alburnus alburnus* harbours a retrotransposon-derived repetitive DNA sequence. *Chromosome Research*, **11**, 23–35.

A.M.D.C., R.S., P.R. and L.B. designed the project, which is part of L.B.'s long-term research programme on Lake Whitefish speciation. A.M.D.C. and S.P. performed laboratory work. A.M.D.C., R.S. and F.C.L. analysed data. P.R. and L.B. supervised the project. A.M.D.C. wrote the manuscript in collaboration with R.S., F.C.L., P.R. and L.B.

Data accessibility

Complementary data including 5S and 28S rDNA sequencing files and raw microscopy images for all techniques used (Giemsa, C-bands, CMA_3 and FISH) are available on Dryad (doi: 10.5061/dryad.tg0mt).

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1 Summary of all individuals samples across three lakes, individual characteristics, phenotypic characteristics measured and markers scored.

Table S2 Summary of all individuals samples across three lakes, individual characteristics, phenotypic characteristics measured and imputed marker values.

Table S3 Chromosome markers significantly correlated (q-value < 0.05) to dimensions 1, 2 and 3 from the multiple factor analysis.

Fig. S1 Metaphase spreads of a dwarf individual from Cliff Lake (CD10) where a B chromosome was found (shown with arrow, also enlarged and framed). A) C-banding, B) CMA_3/DAPI staining and C) FISH with 28S (red) and 5S rDNA (green) probes.

Data S1 Sample code for Multiple Factor Analysis to be performed with FactoMiner.

PROCEEDINGS B

rspb.royalsocietypublishing.org

Research

Cite this article: Dion-Côté A-M, Symonová R, Ráb P, Bernatchez L. 2015 Reproductive isolation in a nascent species pair is associated with aneuploidy in hybrid offspring. *Proc. R. Soc. B* **282**: 20142862. http://dx.doi.org/10.1098/rspb.2014.2862

Received: 20 November 2014
Accepted: 15 December 2014

Subject Areas:
evolution, molecular biology

Keywords:
Coregonus, salmonid, speciation, cytogenetics, genome stability

Author for correspondence:
Anne-Marie Dion-Côté
e-mail: anne-marie.dion-cote.1@ulaval.ca

Electronic supplementary material is available at http://dx.doi.org/10.1098/rspb.2014.2862 or via http://rspb.royalsocietypublishing.org.

THE ROYAL SOCIETY PUBLISHING

Reproductive isolation in a nascent species pair is associated with aneuploidy in hybrid offspring

Anne-Marie Dion-Côté[1], Radka Symonová[2], Petr Ráb[2] and Louis Bernatchez[1]

[1]Département de Biologie, Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, Québec, Canada G1V 0A6
[2]Laboratory of Fish Genetics, Institute of Animal Physiology and Genetics, Czech Academy of Sciences, 277 21, Liběchov, Czech Republic

Speciation may occur when the genomes of two populations accumulate genetic incompatibilities and/or chromosomal rearrangements that prevent inter-breeding in nature. Chromosome stability is critical for survival and faithful transmission of the genome, and hybridization can compromise this. However, the role of chromosomal stability on hybrid incompatibilities has rarely been tested in recently diverged populations. Here, we test for chromosomal instability in hybrids between nascent species, the 'dwarf' and 'normal' lake whitefish (*Coregonus clupeaformis*). We examined chromosomes in pure embryos, and healthy and malformed backcross embryos. While pure individuals displayed chromosome numbers corresponding to the expected diploid number ($2n = 80$), healthy backcrosses showed evidence of mitotic instability through an increased variance of chromosome numbers within an individual. In malformed backcrosses, extensive aneuploidy corresponding to multiples of the haploid number ($1n = 40$, $2n = 80$, $3n = 120$) was found, suggesting meiotic breakdown in their F_1 parent. However, no detectable chromosome rearrangements between parental forms were identified. Genomic instability through aneuploidy thus appears to contribute to reproductive isolation between dwarf and normal lake whitefish, despite their very recent divergence (approx. 15–20 000 generations). Our data suggest that genetic incompatibilities may accumulate early during speciation and limit hybridization between nascent species.

1. Introduction

A fundamental goal in modern evolutionary biology is to characterize the barriers that promote and secure divergence between nascent species, thus resulting in reproductive isolation and ultimately speciation [1,2]. Pre-zygotic barriers have been shown to contribute more to total reproductive isolation than post-zygotic barriers between sympatric species pairs [3]. However, intrinsic post-zygotic reproductive barriers are thought to be permanent and contribute significantly to speciation in an irreversible fashion [4,5]. Among post-zygotic reproductive barriers, it is now clear that nucleotide divergence and genome re-organization through chromosomal rearrangements are intrinsically associated [6–9]. Although it is challenging to study these processes in non-model species, unravelling how nucleotide and chromosomal divergence accumulate and interact to lead to reproductive isolation is crucial to the understanding of speciation.

While the cytogenetic impact of interspecific hybridization has long been studied, it has only rarely been investigated between nascent species [10–12]. One notable exception is the study of chromosomal races in the house mouse (*Mus musculus* complex [13,14]), including a recent study showing that chromosome asynapsis between subspecies hybrids is responsible for infertility [15]. Given the scarcity of studies examining lineages in early stages of divergence, it is hard to draw any conclusions regarding the cytogenetic impact of hybridization and its role in reproductive isolation, and how it varies across

taxa or the divergence time of the system under scrutiny. Indeed, as divergence time increases, the initial genetic changes leading to reproductive isolation will be mixed with subsequent genetic changes that accumulate over time [16]. This will make it more difficult to detect causative mutations leading to reproductive isolation, including the role of chromosomal stability in early speciation [8]. To decipher the initial causes of divergence, and specifically the role of chromosome changes, it is thus necessary to look at the very first stages of speciation.

The geographical and ecological contexts under which divergence has occurred in the lake whitefish (*Coregonus clupeaformis*) system are well understood, thus making it an ideal system in which to study the early stages of speciation. The Acadian and Atlantic lake whitefish lineages were geographically separated approximately 60 000 YBP or approximately 12 000–15 000 generations ago [17,18], during which time, according to the Bateson–Dobzhansky–Muller model (BDM), they could freely accumulate genetic incompatibilities [5,19,20]. This geographical isolation was followed by secondary contact in newly formed lakes after the Laurentian ice sheet retreated approximately 12 000 YBP (approx. 3–4000 generations ago). Following secondary contact, the Acadian lineage evolved repeatedly by character displacement into a 'dwarf' limnetic form while the Atlantic lineage maintained the 'normal' benthic form [17,21]. Gene flow between these sympatric nascent species is still possible [22,23] despite the existence of hybrid incompatibilities leading to a dramatic reduction in embryonic survival in first and second generation hybrids [24–26]. This mortality is associated with the appearance of a 'malformed', slow-growing phenotype in approximately 30–50% of backcross individuals, with the remaining embryos developing normally ('healthy' phenotype) [26]. Consistent with predictions from a BDM model integrating transcriptional data [27], previous studies have documented a much higher variance in gene expression in malformed backcrosses compared with parental forms [26,28]. Transposable elements are also reactivated in both healthy adult backcrosses and malformed backcross embryos, potentially leading to genome instability [28,29]. While earlier studies suggested a primary role for gene expression dysregulation in the appearance of this malformed phenotype [26,28], the molecular basis remains unclear.

Importantly, mixed geographical modes of divergence (i.e. allopatry followed by secondary contact) are predicted to favour chromosome rearrangements between diverging lineages [30]. These chromosome changes can lead to chromosomal incompatibilities in hybrids, either because they will result in unbalanced gametes or disrupt meiosis [12]. Accumulating evidence shows that genetic and chromosomal incompatibilities among species lead to dysregulation involving gene expression, transposable element reactivation and epigenetic inconsistencies in hybrids, all of which also affect genome stability [31,32]. Intriguingly, many aneuploidy events (genome instability in the form of unbalanced segregation of chromosomes) in metazoans lead to similar phenotypes involving significant growth delays combined with malformations [33,34]. Hence, the fact that this malformed phenotype occurs only in post-F_1 lake whitefish hybrids, combined with extensive transcriptional dysregulation in backcrosses raises the question of whether aneuploidy might occur in the hybrid progeny of lake whitefish.

In this context, our goal was to test the hypothesis that genomic instability in the form of aneuploidy accompanies hybrid breakdown in the backcross progeny of dwarf and normal lake whitefish. We directly tested if 'healthy' and 'malformed' backcrosses display higher chromosomal instability compared with dwarf and normal lake whitefish by examining embryonic metaphase chromosomes. We reasoned that increased intra-individual variance in chromosome numbers would indicate increased mitotic instability, while increased inter-individual variance would be consistent with meiotic breakdown [35]. As predicted, we found that healthy backcrosses display higher chromosomal instability compared with pure embryos, and this effect is amplified in malformed backcrosses. Moreover, we found haploid, diploid and triploid individuals among malformed backcrosses, suggesting meiotic breakdown in their F_1 parents. Yet, conventional karyotyping of the parental forms did not reveal any chromosomal rearrangements. Thus, chromosomal instability occurred in hybrids despite the absence of any obvious chromosomal rearrangements between dwarf and normal genomes. Our results thus support the hypothesis that chromosomal instability in hybrids, possibly resulting from the accumulation of minute chromosomal or genetic divergence in allopatry, represents a strong post-zygotic reproductive barrier in this nascent species complex.

2. Material and methods

(a) Crosses and sampling

Dwarf lake whitefish (Acadian lineage) were caught on their spawning grounds in a tributary draining into Lake Témiscouata (47°41′ N, 68°47′ W) and normal lake whitefish (Atlantic lineage) were caught near Lake Aylmer (45°54′ N, 71°20′ W) during autumn 2011. Sperm and eggs were collected in the field and brought to the laboratory for artificial fertilization. Additionally, two laboratory-reared mature F_1-hybrid males (produced from a dwarf mother from Lake Témiscouata and a normal father from Lake Aylmer) from a previous study were used [36]. In total, eight partially half-sib backcross families (i.e. four half-sib families from the same F_1-hybrid father and four half-sib families from the other F_1-hybrid father) were produced and used in this study. Owing to the limited availability of sexually mature fish, it was impossible to create all complementary crosses (i.e. normal mother × dwarf father). However, previous work has documented similar mortality for both types of crosses [24,25]. Moreover, the malformed phenotype also occurs in the reciprocal backcross (i.e. with a F_1-hybrid female [26]). A complete description of the embryos sampled in this study can be found in table 1. It should be noted that the malformed phenotype was found in all backcross families, although only a subset was sampled.

All eggs were incubated in the same slowly flowing water system (4.5–5.5°C) and reared in a common environment at the LARSA (Laboratoire de recherche en sciences aquatiques, Université Laval).

(b) Chromosome preparation and microscopy

Healthy (pure dwarf, pure normal and backcross) and malformed (only found among backcrosses) individuals were sampled. It should be noted that the malformed phenotype segregates within all backcross families. As previously documented within the same long-term research programme, the malformed phenotype is easily identified by the strong deformities seen, including a curved tail and no visually detectable heartbeat. Malformed individuals still display characteristics of the phylotypic

Table 1. Individuals sampled in this study. The family name reflects the number of the parent and the direction of the cross (female × male). N, normal; D, dwarf.

type	family	*n* healthy	*n* malformed	total
N × N	N14 × N1	2	0	10
	N15 × N7	2	0	
	N17 × N8	2	0	
	N18 × N9	1	0	
	N19 × N13	3	0	
D × D	D55 × D79	4	0	11
	D57 × D80	3	0	
	D58 × D78	1	0	
	D59 × D75	3	0	
backcross	D63 × F1–2	3	3	20
	D64 × F1–1	2	0	
	D68 × F1–2	1	2	
	D72 × F1–2	0	1	
	N17 × F1–2	1	0	
	N18 × F1–1	0	2	
	N25 × F1–1	3	0	
	N28 × F1–1	0	2	
total		31	10	41

stage, including eye and dorsal line pigmentation, but still do not resemble any earlier stage of development in normally developing embryos (see [26] for more details).

Chromosome suspensions from embryos were prepared following a previously published method [37] using early-eyed stage embryos (approx. 150–180 degree-days, i.e. 30–36 days of development at 5°C). Chromosome suspensions from four wild dwarf individuals (from Lake Témiscouata, two males and two females) and four laboratory-reared normal individuals (from Lake Aylmer, undetermined sex) were prepared using leucocyte culture as described elsewhere [38]. Unfortunately, these individuals could not be sexed, as there is currently no sex marker available for *Coregonus* [39]. In addition, we did not detect a heteromorphic sex chromosome in either dwarf or normal individuals, a common situation in salmonids [40]. Admittedly, it cannot be determined whether we karyotyped normal males, normal females or a mix of them.

Metaphase chromosomes were stained with Giemsa–Romanowski dye (pH 6.8–7.0, Dr Kulich Pharma, Hradec Králové, Czech Republic) following standard protocols and examined using a Provis AX70 Olympus microscope. Images were captured with a CCD camera (DP30W Olympus). A total of 402 metaphase spreads from 41 embryos were examined, in addition to 64 metaphases from the eight adult fish (table 1). With the exception of one embryo for which only four observations could be made, at least five metaphases were examined per embryo, with an average of 9.8 observations per individual (table 2). In adults, eight metaphases were karyotyped per individual.

(c) Statistical analyses

All statistical analyses were performed using R v. 2.15.1 [41]. We first tested whether intra-individual variance of chromosome counts was dependent on the experimental group (i.e. pure dwarf, pure normal, backcross healthy or backcross malformed). We thus performed an ANOVA on log-transformed individual coefficient of variation of chromosome numbers. Coefficients of variation were used to control for the apparent correlation between chromosome number variance and ploidy level. The Tukey HSD test was then used to identify the comparisons responsible for the significant differences among groups.

We then tested the hypothesis that the variance in median chromosome numbers per individual is significantly different between groups. Specifically, we wanted to know if the variance in median chromosome counts per individual was higher in malformed backcrosses. We applied the Fligner–Kelleen test for homogeneity of variance on median chromosome counts per individual, first on all groups, and then using pairwise comparisons among all groups. A false-discovery rate (FDR) correction was applied to *p*-values using the function *p.adjust*.

3. Results

Adult Giemsa-stained karyotypes corresponded to previously described karyotypes [42]. Both have a diploid chromosome number of $2n = 80$, including 10 meta-/sub-metacentric chromosome pairs and 30 acrocentric chromosome pairs of gradually decreasing size, with the exception of one distinguishable large pair. No obvious differences were detected between the karyotypes of the two forms (figure 1*a*,*b*).

Summary statistics of chromosome number per group can be found in table 2. The complete summary statistics of chromosome number per individual are in electronic supplementary material, table S1. In pure dwarf and normal embryos, counts were centred on the expected diploid number ($2n = 80$, figure 2*a*,*b*; [42]). Dwarf embryos had a mean of 81.7 ± 11.7 chromosomes/metaphase and a median of 79 chromosomes/metaphase, whereas normal embryos had a mean of 78.3 ± 5.6 chromosomes/metaphase and a median of 79 chromosomes/metaphase (table 2). Among pure embryos, no counts exceeded 86 chromosomes, with the exception of a single suspected triploid dwarf individual (figure 3*a*, mean = 108.1 ± 13.5 chromosomes/metaphase, median = 111.5 chromosomes/metaphase). Pure embryos displayed some variance around the diploid number ($2n = 80$, figures 2*a*,*b* and 3), which was expected as chromosome suspensions from embryos are more difficult to spread than those from other tissues, and therefore more difficult to count ([43]; figure 1).

In healthy backcrosses (figure 2*c*), counts were also centred on $2n = 80$ (median = 78 chromosomes/metaphase) but with a lower mean (73.7 ± 13.3 chromosomes/metaphase) compared with pure embryos (table 2). Metaphases with 20–86 chromosomes were found, but all of these individuals seemed diploid (with a median chromosome number close to 80), although with increased variance in chromosome number.

In sharp contrast, a clear tri-modal distribution was found in malformed backcrosses (figures 2*d* and 3*a*), with chromosome numbers concentrated around multiples of the haploid number ($1n = 40$, $2n = 80$, $3n = 120$). The mean chromosome number was lower than all other groups (mean = 76.0 ± 34.7 chromosomes/metaphase), while the median was equal to healthy backcrosses (median = 78 chromosomes/metaphase, table 2). Malformed backcrosses could be separated according to their ploidy (figures 1*c*–*e* and 3*a*). Three malformed backcross individuals were clear

Table 2. Summary statistics of chromosome number per cross-type and group. N, normal; D, dwarf; BC, backcross.

group	*n* individuals	mean	s.d.	median	*n* metaphases
N × N	10	78.3	5.6	79	102
D × D	11	81.7	11.7	79	95
BC healthy	10	73.7	13.3	78	103
BC malformed	10	76.0	34.7	78	102

haploids ($1n = 40$), one individual was a triploid ($3n = 120$) and one individual was almost tetraploid (figure 3*e*, mean = 139 ± 20.8 chromosomes/metaphase, median = 145 chromosomes/metaphase). Metaphases with as few as 32 chromosomes to as many as 158 chromosomes were found in malformed backcrosses. Chromosome fragments were also found in malformed backcrosses, although these were relatively rare (figure 1*e*, arrowheads).

An ANOVA testing for differences in the intra-individual coefficient of variation of chromosome counts revealed a significant difference among groups ($F_{3,36} = 4.911$, $p = 0.0058$). There was a significant difference between healthy backcrosses and pure normal and dwarf embryos (Tukey HSD test, $p \leq 0.05$), but no comparison involving malformed backcrosses was significant (figure 3*b*). This is because there were three haploids among malformed backcrosses, which had very small variance of chromosome numbers. In addition, we found a significant difference in the variance of the median chromosome number among groups (Fligner–Killeen test, $\chi^2 = 12.0222$, d.f. = 3, $p < 2.20 \times 10^{-16}$). The variance of median chromosome number in malformed backcrosses was significantly different from pure normal, pure dwarf and healthy backcrosses, after correction for multiple testing (figure 3*c*, Fligner–Killeen test, FDR < 0.05, $p \leq 0.05$). Complete statistical analyses can be found in electronic supplementary material, table S4.

4. Discussion

In this study, we investigated the role of chromosomal instability in reproductive isolation between nascent lake whitefish species pairs by measuring the chromosome numbers of normal, dwarf, and healthy and malformed backcrosses. Increased intra-individual variance of chromosome number was found in healthy backcrosses. This strongly supports the hypothesis of mitotic chromosome segregation problems, resulting in extra or missing chromosomes after mitotic cell division [35]. However, malformed backcrosses did not display evidence for mitotic chromosome instability compared with pure embryos. This is likely because three stable haploid malformed backcrosses were sampled, thus reducing the variance within the group. Even more strikingly, higher inter-individual variance was found in malformed backcrosses than any other group, i.e. extensive aneuploidy, with an extra or missing haploid complement in the majority of individuals. This result is consistent with meiotic non-disjunction in their F_1-hybrid parent [35]. Yet, karyotypes from parental forms did not reveal any obvious differences at the whole-chromosome level. This suggests that aneuploidy in hybrids is caused by minute sub-chromosomal incompatibilities or genetic incompatibilities acting through mitotic and meiotic mechanisms. The accumulation of these incompatibilities may have been facilitated by the geographical isolation between the two pure forms for approximately 12 000–15 000 generations. Clearly, such incompatibilities cause substantial reproductive isolation between lake whitefish lineages, as 30–50% of post-F_1-hybrids are malformed and die during their early development [25,26]. Our results are especially noteworthy considering the very young age of these lineages on an evolutionary timescale [18].

Importantly, our results were collected from eight partially half-sib backcross families, arguing that these results are not only due to a 'family effect', but apply more generally to these populations. Moreover, the malformed phenotype associated with aneuploidy was observed in two different cohorts (crosses from [26] and this study). Finally, the malformed phenotype has also been observed in post-F_2 hybrids (A.-M.D.-C. and L.B. 2011, 2012, unpublished data). These independent observations strongly support the hypothesis of segregating sub-chromosomal or genetic incompatibilities between lake whitefish lineages leading to aneuploidy in their hybrid progeny, and reproductive isolation.

We note that this extensive aneuploidy would have been very difficult to interpret or even detect from whole genome sequence data alone, stressing the importance of cytogenetics in the post-genomic era. While these approaches have been largely neglected since the advent of modern sequencing techniques, we have shown here that they provide key information regarding genome organization and stability that are difficult to detect from sequence data.

(a) Potential mechanisms underlying chromosome segregation breakdown

We predicted that lake whitefish hybrids would display higher chromosomal instability compared with pure parental forms. Our data support this prediction and here, we discuss four, non-mutually exclusive, candidate mechanisms that are potentially responsible for chromosomal segregation breakdown, namely (i) chromosomal rearrangements, (ii) the mismatch repair (MMR) pathway, (iii) centromere divergence and (iv) heterochromatin decondensation.

The most parsimonious explanation for both mitotic and meiotic breakdown in backcrosses is that significant chromosomal rearrangements have occurred between the Atlantic and Acadian lake whitefish lineages, thus interfering with proper meiotic and mitotic chromosomal segregation [35]. However, karyotyping suggests that this is not the case, at least at the whole-chromosome scale as both karyotypes are essentially the same (figure 1*a*,*b*). Yet, our results cannot rule out the possibility that more subtle changes at the

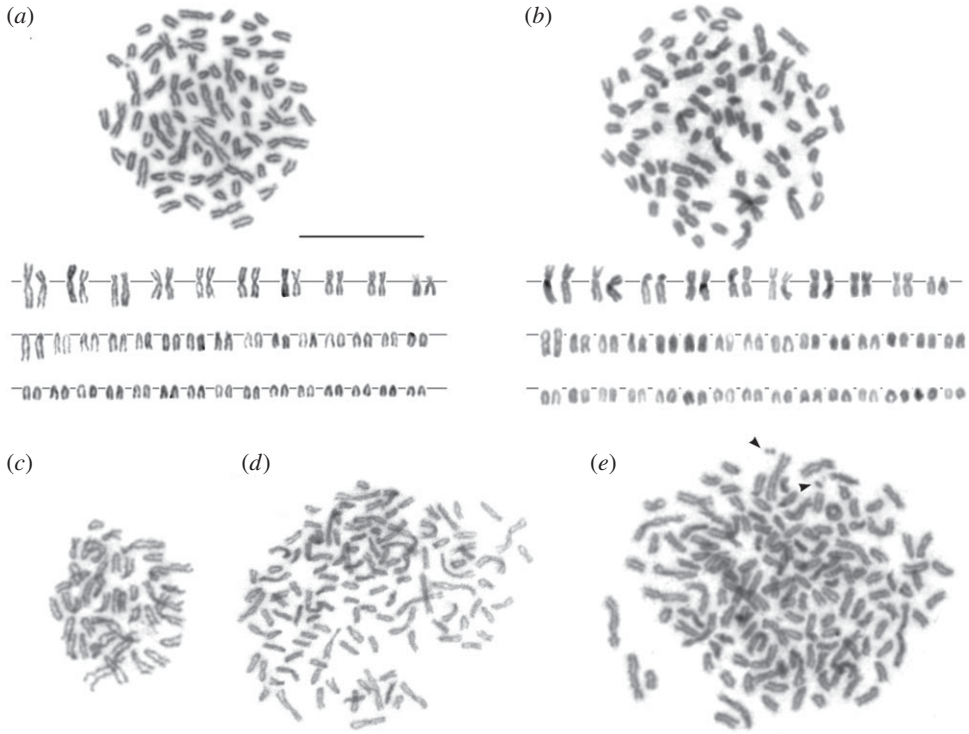


Figure 1. Karyotypes of pure parental forms and abnormal metaphases of malformed backcrosses. Pure karyotypes are composed of 10 metacentric pairs and 30 acrocentric pairs of decreasing size. (*a*) Normal individual from Lake Aylmer. (*b*) Dwarf individual from Lake Témiscouata. (*c*) Haploid metaphase from a malformed backcross. (*d*) Triploid metaphase from a malformed backcross. (*e*) Nearly tetraploid metaphase from a malformed backcross. Arrowheads denote chromosome fragments. Scale bar, 10 µm.

sub-chromosomal level might be involved, including heterochromatin and rDNA genes additions/deletions, with potential consequences for gene expression regulation. It is noteworthy that such sub-chromosomal changes have been recently detected in another *Coregonus* species pair from Europe, where no major karyotypic differences were found [44].

Alternatively, nucleotide divergence among ecotypes may prevent the proper functioning of DNA repair pathways, such as the highly conserved MMR DNA repair pathway [45], and result in meiotic breakdown. Indeed, this mechanism is responsible for induced aneuploidy between incipient species of yeast (*Saccharomyces*, [46]). Meiotic crossovers are critical for balanced chromosome segregation in meiosis as they maintain a tight connection between homologous chromosomes during meiosis I. When divergent chromosomes are combined in yeast hybrids, the MMR pathway prevents these crossovers, thus resulting in aneuploid progeny. Thus, the MMR pathway may underlie the meiotic breakdown in lake whitefish F_1-hybrids if nucleotide divergence is high enough in regions targeted by the meiotic recombination machinery. However, this mechanism does not provide an explanation for mitotic breakdown in healthy backcrosses.

A third candidate mechanism leading to increased chromosomal instability in hybrids was originally proposed by Henikoff *et al.* [47] based on the observation of concerted, rapid evolution of centromeres and their associated proteins. Centromeres are defined by repetitive sequences, including transposable elements. Centromeres are thus rapidly evolving due to their labile nature, despite their highly conserved and critical role in chromosome segregation. This could lead to chromosomal incompatibilities, even between allopatric populations of the same species [9,47]. Hence, it is plausible that aneuploidy in lake whitefish backcrosses may result from the disruption of the chromosome segregation machinery via centromere incompatibilities.

A fourth possibility is that aneuploidy results from heterochromatin decondensation due to mis-regulation in lake whitefish hybrids, which could significantly affect chromosome segregation in both mitosis and meiosis [48]. Indeed, accumulating studies support a role for heterochromatin regulation and associated proteins in reproductive isolation [31,49–51]. Therefore, heterochromatin deregulation in lake whitefish hybrids could also disrupt mitotic and meiotic chromosome pairing, inducing aneuploidy in backcrosses.

We cannot yet conclusively state which of these molecular mechanisms is responsible for the aneuploidy in lake whitefish backcrosses. However, the heterochromatin decondensation hypothesis is especially promising, as previous work in our system has found a massive reactivation of both transposable elements and non-coding RNAs in malformed backcrosses, consistent with heterochromatin disruption [26,28]. Also, we previously found that the Gene Ontology category 'chromosome condensation' was enriched among genes differentially expressed between dwarf and normal embryos, suggesting divergence in the regulation of chromosome compaction [28]. We also note that the reactivation of transposable elements may lead to aneuploidy via chromosomal rearrangements (mechanism 1). Further studies looking specifically at sub-chromosomal structure and heterochromatin regulation in dwarf and normal lake whitefish, as well as their hybrids, will help to disentangle these potentially non-mutually exclusive mechanisms.

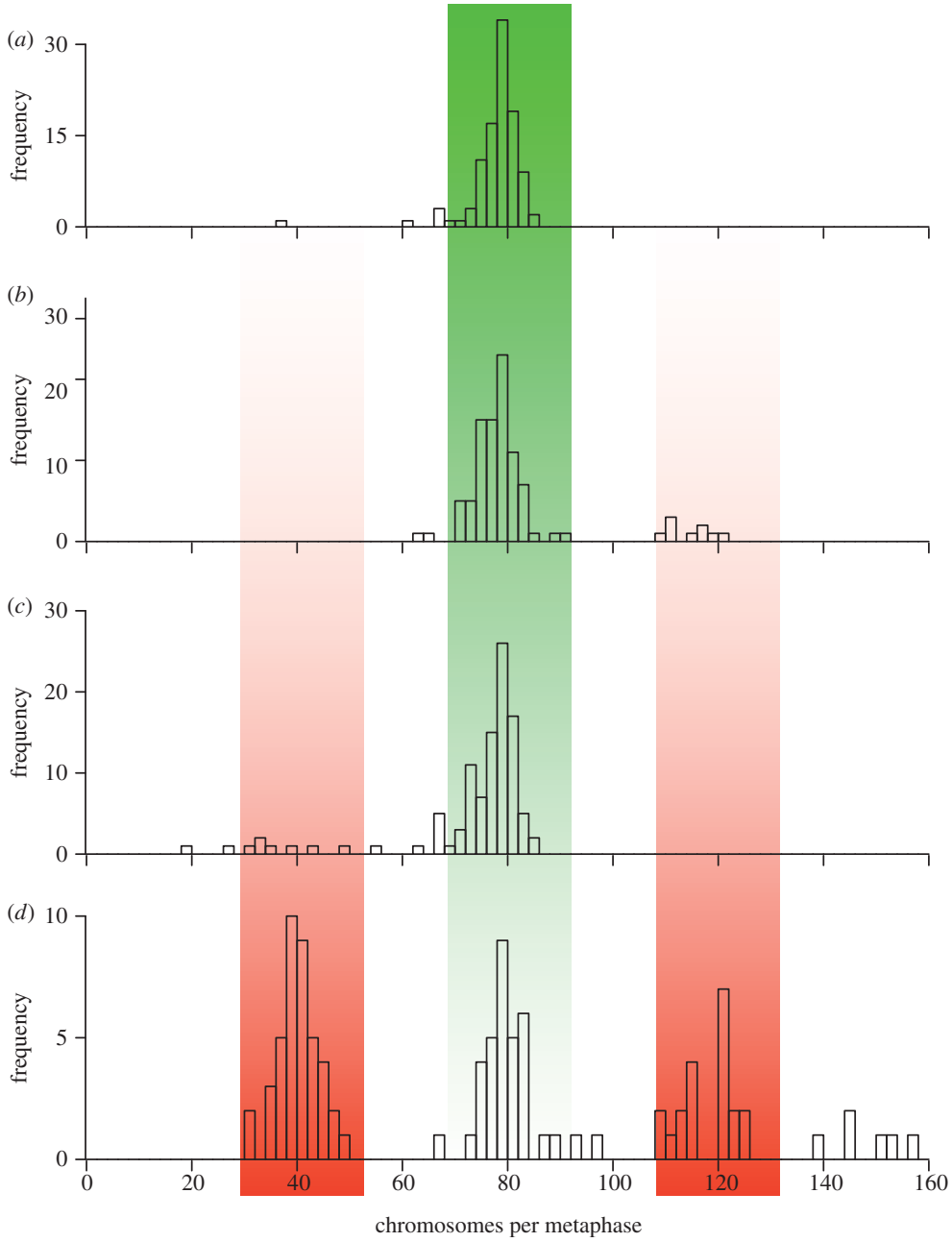


Figure 2. Meiotic breakdown in malformed backcrosses reflected by the analysis of mitotic chromosomes. Histogram showing the distribution of chromosome counts in lake whitefish embryos. (*a*) Normal fish ($n = 10$), (*b*) dwarf fish ($n = 11$), (*c*) healthy backcrosses ($n = 10$) and (*d*) malformed backcrosses ($n = 10$). (Online version in colour.)

(b) Development canalization and 'aneuploidy syndrome'

The fact that healthy backcrosses appear to develop normally and can eventually reproduce (A.-M.D.-C. and L.B. 2011, 2012, unpublished data) despite mitotic instability questions how such intra-individual variation is buffered through development. Indeed, Waddington [52] elegantly suggested that developmental pathways are under strong selective pressure (or canalized), thus buffering for genetic and environmental variations. Hence, the phenotype of healthy backcrosses appears canalized despite higher chromosomal variation compared with pure embryos.

However, this canalization is broken down in malformed backcrosses. The malformed phenotype occurs in conjunction with more variable cytogenetic backgrounds and also in putative diploid individuals (figure 3*a*). Moreover, malformed backcrosses did not show statistically significant mitotic instability ($p = 0.15$ versus dwarf embryos and $p = 0.09$ versus normal embryos, Tukey HSD post hoc test; electronic supplementary material, table S3). As explained above, this is because three haploid malformed backcrosses had a much smaller variance of chromosome number (within an individual) than other malformed backcrosses (figure 3*a*; electronic supplementary material, table S1). These haploid individuals may suffer from this 'aneuploidy syndrome', but not hybrid incompatibilities *per se* as they bear only one genome (likely the pure maternal).

How can one explain this consistent malformed phenotype despite high cytogenetic and transcriptional variability [26,28]? Lindsley *et al*. [33] found that different types of hyperploidy in *Drosophila* resulted in a common phenotype combining rough eyes, abnormal wings, bristle and abdomen, which they described as a 'hyperploidy syndrome'.

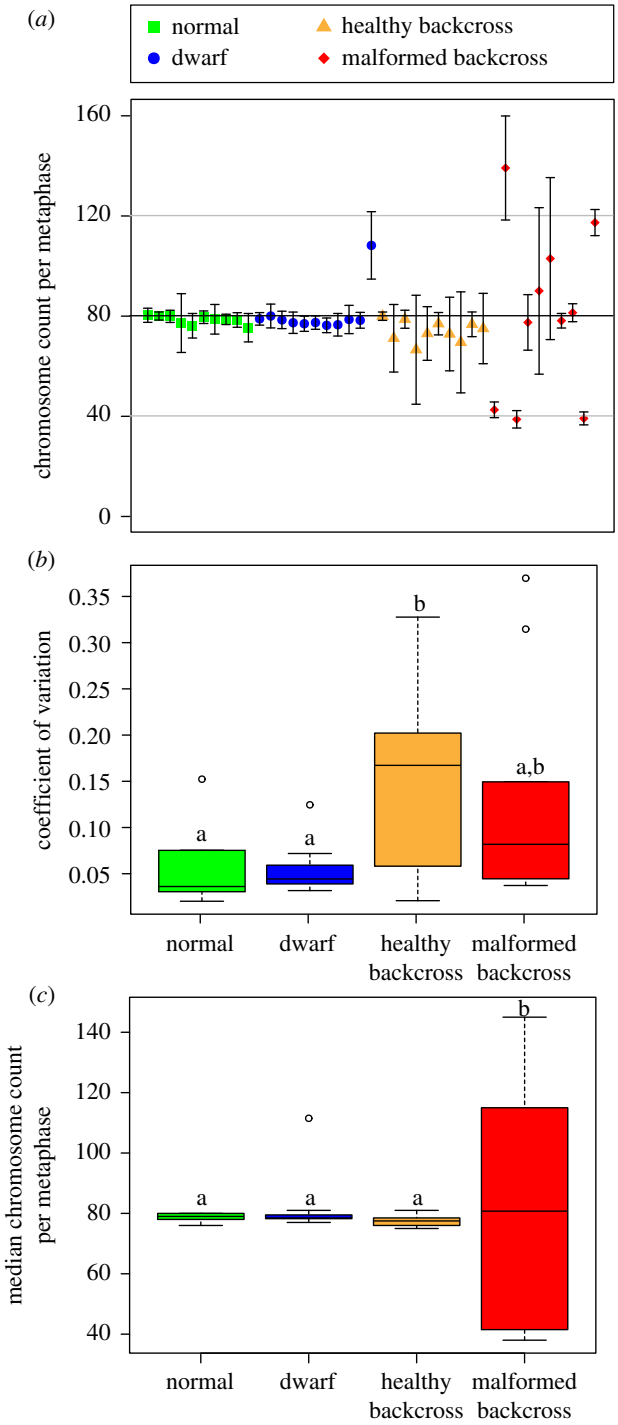


Figure 3. Mitotic and meiotic chromosomal instability occurs in backcrosses based on the analysis of mitotic chromosomes. (*a*) Mean chromosome counts and standard deviation per individual for each group. (*b*) Boxplot showing the distribution of individual coefficient of variation of chromosome number per group. Different letters indicate statistically significant differences (Tukey HSD test, $p \leq 0.05$). (*c*) Boxplot showing the distribution of median chromosome number per group. Different letters indicate statistically significant differences (Fligner–Killeen test, FDR < 0.05, $p \leq 0.05$). (Online version in colour.)

A recent study also found a consistent transcriptional profile within species that was independent of the specific chromosome aberration investigated [53]. In general, many organisms for which the aneuploidy effect has been studied were found to display developmental abnormalities, in addition to a transcriptional signature involving protein synthesis, inflammatory and stress responses [34,54]. Not surprisingly, this signature was also found in lake whitefish malformed backcrosses [28], in addition to a downregulation of essential developmental genes [26].

Aneuploidy may contribute to the mis-regulated transcriptional landscapes previously described in lake whitefish backcross embryos; alternatively, transcriptional mis-regulation in hybrids may lead to aneuploidy. While the causal relationship remains difficult to establish, it appears that the malformed phenotype and associated transcriptional response that we identified in lake whitefish hybrids mirror what has been found in other organisms. Developmental canalization breakdown in malformed backcrosses is thus associated with an 'aneuploidy syndrome' involving increased chromosomal instability and a distinctive transcriptional response.

(c) Implications for the study of speciation

Although cytogenetic studies looking at early diverging lineages are scarce, reproductive isolation through chromosomal instability has been observed in hybrids across several taxa. In yeast (*Saccharomyces paradoxus*), a recent study found that chromosomal differences lead to chromosomal instability in the progeny of diverging strains [55]. Importantly, the divergence of these strains occurred under a similar biogeographical context as the lake whitefish, i.e. a phase of geographical isolation followed by recent secondary contact between lineages. This further supports our interpretation that the conditions under which divergence occurred in lake whitefish have facilitated the accumulation of incompatibilities, which may be either of chromosomal or genetic nature, leading to chromosomal instability in hybrids. In the house mouse (*Mus musculus domesticus*), numerous studies have clearly shown that hybridization between certain chromosomal races leads to chromosomal mis-segregation and hence reduction in litter size [14,56]. Combined with our results, these studies suggest that chromosomal instability can occur in the hybrid progeny of early diverging lineages across a broad range of taxa.

However, it should be noted that, divergence time among these lineages is much greater than for the lake whitefish, given that mouse races have diverged several hundred thousands to million years ago [57] and that yeast produces multiple generations per year. Moreover, it should be stressed that the chromosomal instability we have documented appears to occur in the absence of detectable chromosomal rearrangement. To our knowledge, our study is thus the first to investigate the cytogenetic impact of hybridization among such recently diverged lineages (approx. 12–15 000 generations), at least in vertebrates. There are few, if any, examples of such striking incompatibilities in lineages as young as the lake whitefish, and it is possible that this is only because the cytogenetic consequences of hybridization have been overlooked.

As a consequence, chromosomal speciation models, including the cytogenetic impact of hybridization, have been somewhat neglected in the past decade. This is also due to the combination of the presumed small involvement of chromosome rearrangements to early speciation stages (with the exception of inversions, e.g. [6,7,58]) and theoretical issues concerning the fixation of strongly underdominant chromosomal rearrangements [8]. However, the conditions promoting the fixation of new chromosomal rearrangements

were present in the lake whitefish system, including: a mixed geographical mode of divergence [17], small effective population size ($Ne \sim 1000$, [59]), geographical isolation of lineages [60] and possibly meiotic drive [61]. Unfortunately, it is not yet possible to determine whether the chromosomal instability observed in lake whitefish hybrids is the result of genetic or chromosomal incompatibilities, and hence a case of chromosomal speciation. Yet, both genetic and subtle chromosomal changes may be involved, and future research should help to disentangle these alternative hypotheses.

Our results are critical to the understanding of how reproductive isolation has emerged in the lake whitefish system, and other nascent species. We show that genomic instability, through aneuploidy, transcriptional dysregulation and transposable reactivation, can interact and efficiently limit hybridization early in the divergence process, and thus contribute to speciation. Future work looking at systems where conditions promoting the appearance and fixation of chromosome rearrangements are found will help to draw conclusions regarding the generality of our observations.

Ethics statement. The Laval University animal care committee (CPAUL) revised and approved all experimental procedures (Protocol 82178).

Acknowledgements. We are grateful to Jean-Yves Masson, Anne C. Dalziel and Ben Sutherland for inspiring discussions, and Serge Higgins, Guillaume Côté and Jean-Christophe Therrien for their support during the crossing and rearing of embryos at the LARSA. We also thank Jana Čechová, Alain Goulet and Richard Janvier for their technical help during chromosome suspension preparation and microscopy. Glenn Yannic, Gaétan Daigle and Charles Bordet also contributed significantly to statistical approaches.

Funding statement. This work was supported by a Natural Science and Engineering Research Council of Canada (NSERC) discovery grant and Canadian Research Chair in Genomics and Conservation of Aquatic Resources to L.B. and a NSERC postgraduate scholarship to A.-M.D.-C. A.-M.D.-C. also received financial supports from Québec-Océan for a short training and a FRQNT international internship. This project is a contribution to the research programme of Québec-Océan. R.S. and P.R. were supported by the project 14–02940S of the Czech Science Foundation.

References

1. Coyne J, Orr HA. 2004 *Speciation*. Sunderland, MA: Sinauer Associates, Inc.
2. Seehausen O *et al*. 2014 Genomics and the origin of species. *Nat. Rev. Genet.* **15**, 176–192. (doi:10.1038/nrg3644)
3. Coyne J, Orr HA. 1997 "Patterns of speciation in *Drosophila*" revisited. *Evolution* **51**, 295–303. (doi:10.2307/2410984)
4. Muller HJ. 1939 Reversibility in evolution considered from the standpoint of genetics. *Biol. Rev.* **14**, 261–280. (doi:10.1111/j.1469-185X.1939.tb00934.x)
5. Muller HJ. 1942 Isolating mechanisms, evolution and temperature. *Biol. Symp.* **6**, 71–125.
6. Noor MA, Grams KL, Bertucci LA, Reiland J. 2001 Chromosomal inversions and the reproductive isolation of species. *Proc. Natl Acad. Sci. USA* **98**, 12 084–12 088. (doi:10.1073/pnas.221274498)
7. Rieseberg LH. 2001 Chromosomal rearrangements and speciation. *Trends Ecol. Evol.* **16**, 351–358. (doi:1016/S0169-5347(01)02187-5)
8. Faria R, Navarro A. 2010 Chromosomal speciation revisited: rearranging theory with pieces of evidence. *Trends Ecol. Evol.* **25**, 660–669. (doi:10.1016/j.tree.2010.07.008)
9. Lynch M. 2007 *The origins of genome architecture*. Sunderland, MA: Sinauer Associates Inc.
10. Pinney E. 1918 A Study of the relation of the behavior of the chromatin to development and heredity in teleost hybrids. *J. Morphol.* **31**, 225–291. (doi:10.1002/jmor.1050310202)
11. White M. 1969 Chromosomal rearrangements and speciation in animals. *Annu. Rev. Genet.* **3**, 75–98. (doi:10.1007/s10577-013-9377-5)
12. King M. 1993 *Species evolution: the role of chromosome change*. Cambridge, UK: Cambridge University Press.
13. Pialek J, Hauffe HC, Searle JB. 2005 Chromosomal variation in the house mouse. *Biol. J. Linn. Soc.* **84**, 535–563. (doi:10.1111/j.1095-8312.2005.00454.x)
14. Hauffe HC, Gimenez MD, Searle JB. 2012 Chromosomal hybrid zones in the house mouse. In *Evolution of the house mouse* (eds M Macholán, SJE Baird, P Munclinger, J Pialek), pp. 407–430. Cambridge, UK: Cambridge University Press.
15. Bhattacharyya T, Gregorova S, Mihola O, Anger M, Sebestova J, Denny P, Simecek P, Forejt J. 2013 Mechanistic basis of infertility of mouse intersubspecific hybrids. *Proc. Natl Acad. Sci. USA* **110**, E468–E477. (doi:10.1073/pnas.1219126110)
16. Via S. 2009 Natural selection in action during speciation. *Proc. Natl Acad. Sci. USA* **106**(Suppl. 1), 9939–9946. (doi:10.1073/pnas.0901397106)
17. Bernatchez L *et al*. 2010 On the origin of species: insights from the ecological genomics of lake whitefish. *Phil. Trans. R. Soc. B* **365**, 1783–1800. (doi:10.1098/rstb.2009.0274)
18. Jacobsen MW, Hansen MM, Orlando L, Bekkevold D, Bernatchez L, Willerslev E, Gilbert MTP. 2012 Mitogenome sequencing reveals shallow evolutionary histories and recent divergence time between morphologically and ecologically distinct European whitefish (*Coregonus* spp.). *Mol. Ecol.* **21**, 2727–2742. (doi:10.1111/j.1365-294X.2012.05561.x)
19. Bateson W. 1909 Heredity and variation in modern lights. In *Darwin and modern science* (ed. AC Seward), pp. 85–101. Cambridge, UK: Cambridge University Press.
20. Dobzhansky T. 1937 *Genetics and the origin of species*. New York, NY: Columbia University Press.
21. Bernatchez L. 2004 Ecological theory of adaptive radiation: an empirical assessment from coregonine fishes (Salmoniformes). In *Evolution illuminated: salmon and their relatives* (eds AP Hendry, SC Stearns), pp. 175–207. Oxford, UK: Oxford University Press.
22. Gagnaire P-A, Pavey SA, Normandeau E, Bernatchez L. 2013 The genetic architecture of reproductive isolation during speciation-with-gene-flow in lake whitefish species pairs assessed by RAD sequencing. *Evolution* **67**, 2483–2497. (doi:10.1111/evo.12075)
23. Renaut S, Nolte AW, Rogers SM, Derome N, Bernatchez L. 2011 SNP signatures of selection on standing genetic variation and their association with adaptive phenotypes along gradients of ecological speciation in lake whitefish species pairs (*Coregonus* spp.). *Mol. Ecol.* **20**, 545–559. (doi:10.1111/j.1365-294X.2010.04952.x)
24. Lu G, Bernatchez L. 1998 Experimental evidence for reduced hybrid viability between dwarf and normal ecotypes of lake whitefish (*Coregonus clupeaformis* Mitchill). *Proc. R. Soc. Lond. B* **265**, 1025–1030. (doi:10.1098/rspb.1998.0394)
25. Rogers SM, Bernatchez L. 2006 The genetic basis of intrinsic and extrinsic post-zygotic reproductive isolation jointly promoting speciation in the lake whitefish species complex (*Coregonus clupeaformis*). *J. Evol. Biol.* **19**, 1979–1994. (doi:10.1111/j.1420-9101.2006.01150.x)
26. Renaut S, Bernatchez L. 2011 Transcriptome-wide signature of hybrid breakdown associated with intrinsic reproductive isolation in lake whitefish species pairs (*Coregonus* spp. Salmonidae). *Heredity* **106**, 1003–1011. (doi:10.1038/hdy.2010.149)
27. Landry CR, Hartl DL, Ranz JM. 2007 Genome clashes in hybrids: insights from gene expression. *Heredity* **99**, 483–493. (doi:10.1038/sj.hdy.6801045)
28. Dion-Côté A-M, Renaut S, Normandeau E, Bernatchez L. 2014 RNA-seq reveals transcriptomic shock involving transposable elements reactivation in hybrids of young lake whitefish species. *Mol. Biol. Evol.* **31**, 1188–1199. (doi:10.1093/molbev/msu069)

29. Renaut S, Nolte AW, Bernatchez L. 2010 Mining transcriptome sequences towards identifying adaptive single nucleotide polymorphisms in lake whitefish species pairs (*Coregonus* spp. Salmonidae). *Mol. Ecol.* **19**(Suppl. 1), 115–131. (doi:10.1111/j.1365-294X.2009.04477.x)
30. Feder JL, Gejji R, Powell THQ, Nosil P. 2011 Adaptive chromosomal divergence driven by mixed geographic mode of evolution. *Evolution* **65**, 2157–2170. (doi:10.1111/j.1558-5646.2011.01321.x)
31. Ferree PM, Barbash DA. 2009 Species-specific heterochromatin prevents mitotic chromosome segregation to cause hybrid lethality in *Drosophila*. *PLoS Biol.* **7**, e1000234. (doi:10.1371/journal.pbio.1000234)
32. Brown JD, O'Neill RJ. 2010 Chromosomes, conflict, and epigenetics: chromosomal speciation revisited. *Annu. Rev. Genomics Hum. Genet.* **11**, 291–316. (doi:10.1146/annurev-genom-082509-141554)
33. Lindsley DL *et al*. 1972 Segmental aneuploidy and the genetic gross structure of the *Drosophila* genome. *Genetics* **71**, 157–184.
34. Torres EM, Williams BR, Amon A. 2008 Aneuploidy: cells losing their balance. *Genetics* **179**, 737–746. (doi:10.1534/genetics.108.090878)
35. Barbero JL. 2011 Sister chromatid cohesion control and aneuploidy. *Cytogenet. Genome Res.* **133**, 223–233. (doi:10.1159/000323507)
36. Renaut S, Nolte A, Bernatchez L. 2009 Gene expression divergence and hybrid misexpression between lake whitefish species pairs (*Coregonus* spp. Salmonidae). *Mol. Biol. Evol.* **26**, 925–936. (doi:10.1093/molbev/msp017)
37. Völker M, Rab P, Kullmann H. 2005 Karyotype differentiation in chromaphyosemion killifishes (Cyprinodontiformes, Nothobranchiidae). I: Chromosome banding patterns of *C. alpha*, *C. kouamense* and *C. lugens*. *Genetica* **125**, 33–41. (doi:10.1007/s10709-005-4267-1)
38. Fujiwara A, Nishida-Umehara C, Sakamoto T, Okamoto N, Nakayama I, Abe S. 2001 Improved fish lymphocyte culture for chromosome preparation. *Genetica* **111**, 77–89. (doi:10.1023/A:1013788626712)
39. Yano A, Nicol B, Jouanno E, Quillet E, Fostier A, Guyomard R, Guiguen Y. 2012 The sexually dimorphic on the Y-chromosome gene (sdY) is a conserved male-specific Y-chromosome sequence in many salmonids. *Evol. Appl.* **6**, 486–496. (doi:10.1111/eva.12032)
40. Davidson WS, Huang TK, Fujiki K, Schalburg von KR, Koop BF. 2009 The sex determining loci and sex chromosomes in the family Salmonidae. *Sex Dev.* **3**, 78–87. (doi:10.1159/000223073)
41. R Core Team. 2012 *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
42. Phillips RB, Reed KM, Rab P. 1996 Revised karyotypes and chromosome banding of coregonid fishes from the Laurentian Great Lakes. *Can. J. Zool.* **74**, 323–329. (doi:10.1139/z96-040)
43. Völker M, Rab P. In press. Direct chromosome preparation from embryos and larvae. In *Fish cytogenetic techniques* (*Chondrichthyans and Teleosts*) (eds C Ozouf-Costaz, E Pisano, F Foresti, L Foresti de Almeida Foresto). Boca Raton, FL: CRC Press.
44. Symonová R, Majtánová Z, Sember A, Staaks GB, Bohlen J, Freyhof J, Rábová M, Rab P. 2013 Genome differentiation in a species pair of coregonine fishes: an extremely rapid speciation driven by stress-activated retrotransposons mediating extensive ribosomal DNA multiplications. *BMC Evol. Biol.* **13**, 42. (doi:10.1007/BF02153623)
45. Harfe BD, Jinks-Robertson S. 2000 DNA mismatch repair and genetic instability. *Annu. Rev. Genet.* **34**, 359–399. (doi:10.1146/annurev.genet.34.1.359)
46. Greig D, Travisano M, Louis EJ, Borts RH. 2003 A role for the mismatch repair system during incipient speciation in *Saccharomyces*. *J. Evol. Biol.* **16**, 429–437. (doi:10.1046/j.1420-9101.2003.00546.x)
47. Henikoff S, Ahmad K, Malik HS. 2001 The centromere paradox: stable inheritance with rapidly evolving DNA. *Science* **293**, 1098–1102. (doi:10.1126/science.1062939)
48. Grewal SIS, Jia S. 2007 Heterochromatin revisited. *Nat. Rev. Genet.* **8**, 35–46. (doi:10.1038/nrg2008)
49. O'Neill MJ, Graves JA. 1998 Undermethylation associated with retroelement activation and chromosome remodelling in an interspecific mammalian hybrid. *Nature* **393**, 68–72. (doi:10.1038/29985)
50. Bayes JJ, Malik HS. 2009 Altered heterochromatin binding by a hybrid sterility protein in Drosophila sibling species. *Science* **326**, 1538–1541. (doi:10.1126/science.1181756)
51. Cattani MV, Presgraves DC. 2012 Incompatibility between X chromosome factor and pericentric heterochromatic region causes lethality in hybrids between *Drosophila melanogaster* and its sibling species. *Genetics* **191**, 549–559. (doi:10.1534/genetics.112.139683)
52. Waddington CH. 1942 Canalization of development and the inheritance of acquired characters. *Nature* **150**, 563–565.
53. Sheltzer JM, Torres EM, Dunham MJ, Amon A. 2012 Transcriptional consequences of aneuploidy. *Proc. Natl Acad. Sci. USA* **109**, 12 644–12 649. (doi:10.1073/pnas.1209227109)
54. Dürrbaum M, Kuznetsova AY, Passerini V, Stingele S, Stoehr G, Storchová Z. 2014 Unique features of the transcriptional response to model aneuploidy in human cells. *BMC Genomics* **15**, 139. (doi:10.1186/1471-2164-15-139)
55. Charron G, Leducq J-B, Landry CR. 2014 Chromosomal variation segregates within incipient species and correlates with reproductive isolation. *Mol. Ecol.* **23**, 4362–4372. (doi:10.1111/mec.12864)
56. Forejt J, Pialek J, Trachtulec Z. 2012 Hybrid male sterility in mouse subspecific crosses. In *Evolution of the house mouse* (eds M Macholán, SJE Baird, P Munclinger, J Pialek), pp. 482–503. Cambridge, UK: Cambridge University Press.
57. Macholán M, Baird SJE, Munclinger P, Pialek J (eds). 2012 *Evolution of the house mouse*. Cambridge, UK: Cambridge University Press.
58. Feder JL, Nosil P. 2009 Chromosomal inversions and species differences: when are genes affecting adaptive divergence and reproductive isolation expected to reside within inversions? *Evolution* **63**, 3061–3075. (doi:10.1111/j.1558-5646.2009.00786.x)
59. Campbell D, Bernatchez L. 2004 Generic scan using AFLP markers as a means to assess the role of directional selection in the divergence of sympatric whitefish ecotypes. *Mol. Biol. Evol.* **21**, 945–956. (doi:10.1093/molbev/msh101)
60. Bernatchez L, Dodson JJ. 1991 Phylogeographic structure in mitochondrial DNA of the lake whitefish (*Coregonus clupeaformis*) and its relation to Pleistocene glaciations. *Evolution* **45**, 1016–1035.
61. Gagnaire P-A, Normandeau E, Pavey SA, Bernatchez L. 2012 Mapping phenotypic, expression and transmission ratio distortion QTL using RAD markers in the lake whitefish (*Coregonus clupeaformis*). *Mol. Ecol.* **22**, 3036–3048. (doi:10.1111/mec.12127)

BMC Evolutionary Biology

RESEARCH ARTICLE Open Access

Genome differentiation in a species pair of coregonine fishes: an extremely rapid speciation driven by stress-activated retrotransposons mediating extensive ribosomal DNA multiplications

Radka Symonová[1*], Zuzana Majtánová[1], Alexandr Sember[1], Georg BO Staaks[2], Jörg Bohlen[1], Jörg Freyhof[2], Marie Rábová[1] and Petr Ráb[1]

Abstract

Background: Sympatric species pairs are particularly common in freshwater fishes associated with postglacial lakes in northern temperate environments. The nature of divergences between co-occurring sympatric species, factors contributing to reproductive isolation and modes of genome evolution is a much debated topic in evolutionary biology addressed by various experimental tools. To the best of our knowledge, nobody approached this field using molecular cytogenetics. We examined chromosomes and genomes of one postglacial species pair, sympatric European winter-spawning *Coregonus albula* and the local endemic dwarf-sized spring-spawning *C. fontanae*, both originating in Lake Stechlin. We have employed molecular cytogenetic tools to identify the genomic differences between the two species of the sympatric pair on the sub-chromosomal level of resolution.

Results: Fluorescence *in situ* hybridization (FISH) experiments consistently revealed a distinct variation in the copy number of loci of the major ribosomal DNA (the 45S unit) between *C. albula* and *C. fontanae* genomes. In *C. fontanae*, up to 40 chromosomes were identified to bear a part of the major ribosomal DNA, while in *C. albula* only 8–10 chromosomes possessed these genes. To determine mechanisms how such extensive genome alternation might have arisen, a PCR screening for retrotransposons from genomic DNA of both species was performed. The amplified retrotransposon *Rex1* was used as a probe for FISH mapping onto chromosomes of both species. These experiments showed a clear co-localization of the ribosomal DNA and the retrotransposon *Rex1* in a pericentromeric region of one or two acrocentric chromosomes in both species.

Conclusion: We demonstrated genomic consequences of a rapid ecological speciation on the level undetectable by neither sequence nor karyotype analysis. We provide indirect evidence that ribosomal DNA probably utilized the spreading mechanism of retrotransposons subsequently affecting recombination rates in both genomes, thus, leading to a rapid genome divergence. We attribute these extensive genome re-arrangements associated with speciation event to stress-induced retrotransposons (re)activation. Such causal interplay between genome differentiation, retrotransposons (re)activation and environmental conditions may become a topic to be explored in a broader genomic context in future evolutionary studies.

* Correspondence: radka.symonova@natur.cuni.cz
[1]Laboratory of Fish Genetics, Institute of Animal Physiology and Genetics, Czech Academy of Sciences, Rumburská 89, Liběchov 277 21, Czech Republic
Full list of author information is available at the end of the article

Background

Intra-lacustrine fish speciation as an example of ecological speciation is a much debated topic in evolutionary biology addressed by various experimental tools, mostly in complex systems with a number of species, in particular in ancient freshwater lakes [1]. In Europe, with its comparatively depauperate fish fauna, issues of adaptive radiation and ecological speciation in fishes are highly relevant in temperate postglacial lakes (originating after the last glaciation i.e. 12–15 kyrs BP). To assess potential modes of speciation in fishes, numerous model systems are available [2], among which one of the best groups with a robust knowledge on adaptive speciation and complex speciation patterns in postglacial lakes are coregonine fishes (Coregoninae, [3]) [4-6]. Within coregonines, their numerous sympatric species pairs and recent species flocks [7-9] are of particular importance [10]. In *Coregonus*, based on extensive genetic and population genetic [11], phylogenetic, biogeographic, morphological and eco-physiological data, six potential modes of speciation have been proposed [12]. However, none of these approaches utilized cytogenetic data despite salmonid fishes, to which coregonines belong, being one of the best karyologically studied fish groups in terms of the number of species, populations, individuals and material (adults and embryos) examined. Available cytogenetic data demonstrate that salmonids include two basic karyotypes – the high chromosome number 2n ~ 80 (type A and its derivatives) and the low chromosome number 2n ~ 60 (type B and its derivatives) – co-occurring in all recognized salmonid phylogenetic lineages (except graylings, Thymallinae), including whitefish, ciscoes and innconu (Coregoninae). Species with the type B karyotypes have in common either prominent anadromous behaviour and/or are found in lacustrine environments and are likely products of intra-lacustrine speciation (for review [13]). Such apparent parallelism might be explained by specific life history strategies leading in both types of environments to small effective population sizes, thus, enabling increased probability of fixation of genic or chromosomal mutations. Observed evolution of chromosome number in salmonids is likely affected by selection for increased or decreased genetic recombination rate as proposed by Quimseyh [14], explaining high variability in chromosome numbers in mammals based on fundamental numbers (NF, chromosome arms number).

In this study, we examined chromosomes and genomes of the sympatric species *Coregonus albula* and *C. fontanae* in the dimictic Lake Stechlin, northern Germany to test whether the above outlined parallelism on karyotype differentiation in intralacustrine species pairs can also be observed in incipient speciation processes in young postglacial lakes. Both species are pelagic zooplanktivores, but they differ considerably in their size, spawning time [15] and temperature-dependent metabolic physiological adaptations [16]. Up to now, *C. fontanae* has not yet been subjected to any cytogenetic analysis as opposed to *C. albula* (see [17] and references therein). The level of genetic differentiation between *C. albula* and *C. fontanea* tested by combined analyses of mitochondrial DNA and microsatellite loci showed a weak differentiation (F_{ST} = 0–0.008) between these two species when compared with another sympatric species pair *C. albula* and *C. lucinensis* [18]. Further population genetic analyses based on 1244 polymorphic AFLP loci demonstrated a lower differentiation between allopatric than sympatric populations of the *C. albula* complex and suggested a rather complex colonization history than simple sympatric speciation [6]. Therefore, we have employed a novel approach in this field to explore the up to now neglected aspects of genome evolution in this species pair and used different parts of ribosomal DNA of the 45S rDNA unit as cytotaxonomic markers.

At the first stage of this study, we have employed conventional methods of karyotype analysis (Giemsa and Ag staining, CMA_3 and DAPI fluorescence). At the second stage, we have performed molecular cytogenetic analyses (CGH and FISH with various rDNA fragments and non-LTR retrotransposons as probes) to identify any differences between chromosomal complements of these two species on the sub-chromosomal level of resolution since the karyotype analyses showed no significant differences. At the third stage, we performed molecular biological analyses of the 45S ribosomal RNA genes and the *Rex1* non-LTR retrotransposon. Furthermore, we discuss these results in the context of populations of small effective sizes under extreme stress conditions under which retrotransposons (re)activation could have contributed to accelerated speciation.The major cluster of ribosomal RNA genes is expressed as the 45S transcriptional unit (Figure 1). This unit consists of 18S, 5.8S and 28S rDNA genes, separated by internal transcribed spacers (ITS1, ITS2) and surrounded by external transcribed spacers (ETS). The 45S rDNA units are arranged in tandem repetitions with high copy numbers [19,20] therefore, they represent a useful cytotaxonomic marker. The individual units are separated by intergenic spacers (IGS) [21,22]. The structure and the order of genes within the unit are highly conserved among Eukaryota [23]. Different parts of the 45S transcriptional unit display different mutational rates. The most conserved region is the 18S rRNA gene and the most variable are ITSs [23].

Results

Karyotyping and comparative cytogenetics

Karyotypes of both examined ciscoes were very similar (2n = 80 in both species) and both belong to the

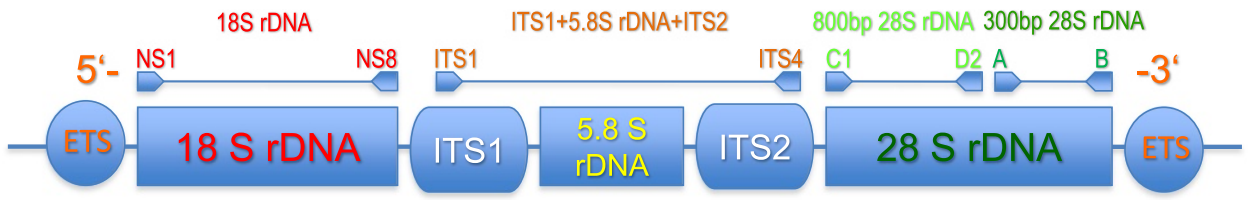


Figure 1 Schematic representation of the 45S rDNA unit and FISH probes construction including primers nesting sites. (not to scale).

karyotype category A *sensu* [13]. They both had 8 pairs of meta- (m) to submetacentric (sm) and 32 pairs of acrocentric (a) chromosomes (both sexes in *C. fontanae*, only males in *C. albula* were available), The NF was 96 in both species (Figure 2a-d). The sequential Chromomycin A_3 (CMA_3, particularly specific for CG rich regions) and DAPI (specific for AT rich regions) stainings revealed in both species a varying number of 6–8 sites with $CMA3^+/DAPI^-$ signals. The signals occurred at telomeric regions of 3–4 metacentric chromosomes and at pericentromeric regions of 3–4 acrocentric/submetacentric chromosomes (Figure 3a, b). In some nuclei, several other weakly CMA_3^+ regions not corresponding to $DAPI^-$ signals mostly with pericentromeric locations were observed (Figure 3a). This variability occurs on the inter-individual as well as on the intra-individual level.

Cytogenetic mapping of ribosomal DNA and comparative genomic hybridization (CGH)

Fluorescence *in situ* hybridization (FISH) with 28S ribosomal DNA (rDNA) probes derived from two non-overlapping regions of the 28S rRNA gene of both species (an 800 bp region adjacent towards the 5′-end of the 28S rDNA gene and a 300 bp region adjacent towards the 3′-end) showed strikingly different results. FISH using the shorter fragment as a probe revealed the presence of 6–10 chromosomes in both *C. albula* and *C. fontanae* bearing such sequences distributed similarly as the $CMA_3^-/DAPI^-$ (Figure 3c for *C. albula* only). FISH with the longer fragment revealed bright signals on 6–10 chromosomes in *C. albula* (shown in co-localization with *Rex1* retrotransposon, Figure 4c) but up to 40 signals (varying numbers) on chromosomes in *C. fontanae* (Figure 3d). Most of the signals of the 800 bp probe of the 28S rDNA in *C. fontanae* were localized in the AT rich (i.e. $DAPI^+$) centromeric or pericentromeric regions of acrocentric chromosomes. Two signals of the 800 bp 28S rDNA probe corresponded to the major NOR sites evidenced also by the 300 bp rDNA and the CMA_3/DAPI staining that were localized in telomeric regions of two large metacentric chromosomes (Figure 3d).

To verify these striking differences between *C. albula* and *C. fontanae*, we carried out a set of reciprocal comparative genomic hybridization (CGH) experiments. A mixture of the whole genome DNA (gDNA) of both *C. albula* and *C. fontanae* was hybridized simultaneously to both *C. fontanae* and *C. albula* chromosomes. This resulted in nearly no significant differences on *C. albula* chromosomes, i.e. a balanced hybridization of both gDNA probes was observed (Figure 3e). While signal of the *C. fontanae* gDNA when *in situ* compared with the

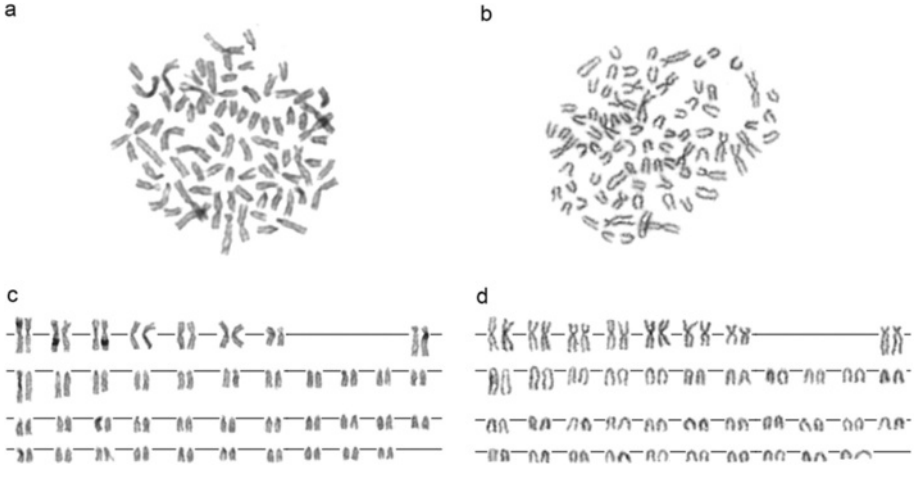


Figure 2 Giemsa-stained metaphase plates and corresponding karyogram of *C. albula* (a, c) and *C. fontanae* (b, d). Bar = 5 μm.

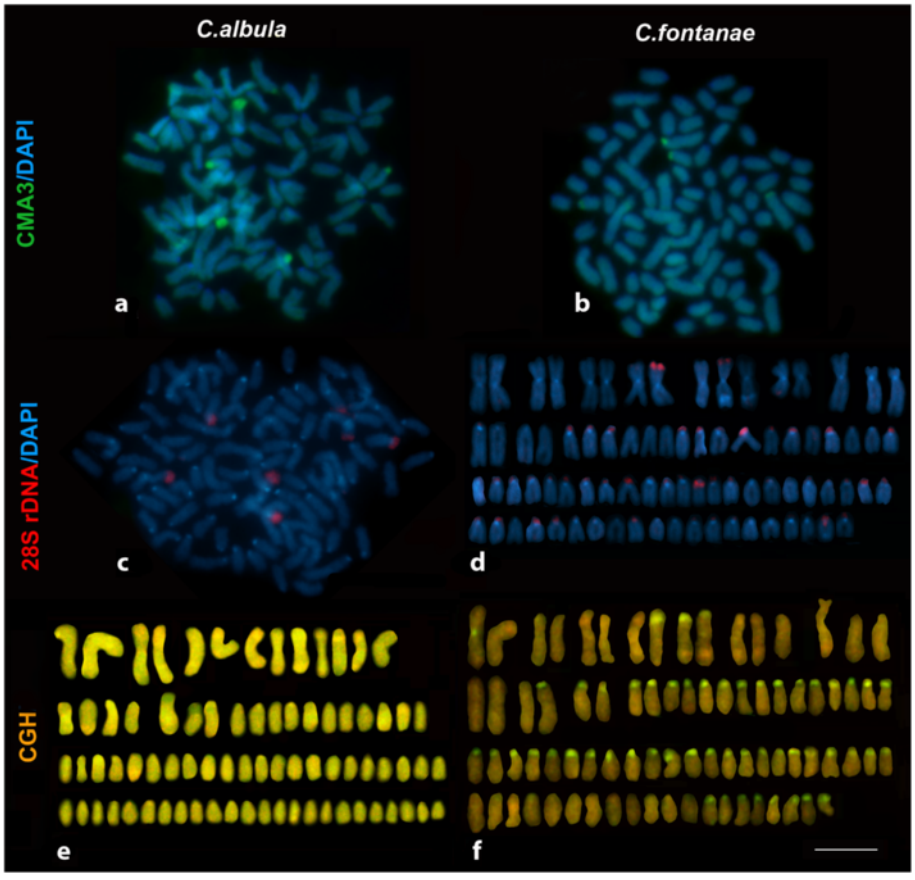


Figure 3 Metaphase plates and karyograms of *C. albula* and *C. fontanae* showing Chromomycin A_3/DAPI staining, FISH and CGH experiments. Chromomycin A_3 (CMA_3) fluorescent staining (green) and DAPI (blue) staining in *C. albula* (**a**) and *C. fontanae* (**b**). FISH with the 28S rDNA (300 bp probe) (red), DAPI counterstaining (blue) in *C. albula* (**c**). FISH with the 28S rDNA (800 bp probe) (red), DAPI counterstaining (blue) in *C. fontanae* (**d**). A set of reciprocal comparative genomic hybridization (CGH) experiments to *C. albula* chromosomes (**e**) and *C. fontanae* chromosomes (**f**). In both (**e**, **f**), the *C. albula* genomic DNA was labelled in red and the *C. fontanae* genomic DNA in green. Bar = 5 μm.

C. albula gDNA onto *C. fontanae* chromosomes was distinctly overrepresented in mostly pericentromeric regions of about 40 chromosomes (green signals in Figure 3f). This pattern corresponded to results of the FISH experiment with 800 bp 28S rDNA to *C. fontanae* chromosomes.

To assess quantitative differences in the distribution of the whole 45S rDNA unit in both species, a further set of comparative FISH experiments with a cocktail of the 18S rDNA and ITS1-ITS2 (including 5.8S rDNA) as probes amplified from both of the genomes were performed to *C. albula* and *C. fontanae* chromosomes. In the genome of *C. albula*, both the ITS1 and ITS2 were present in 6–12 signals with a varying number of signals (Figure 4a). In the genome of *C. fontanae*, both the ITS1 and ITS2 were multiplied to the same extent as the 800 bp 28S rDNA part, i.e. a varying number of approximately 40 signals (Figure 4b). The subsequent FISH experiment with the 18S rDNA in both species showed the number of 6–10 signals (Figure 4d for *C. fontanae* only). The typical chromosomes bearing rDNA signals in the unamplified condition (i.e. ITS1-ITS2 and 28S rDNA in *C. albula* and 18S rDNA in *C. fontanae*) are shown in Figures 4e-f. There is a reproducible difference in location of one of the 28S rDNA in *C. fontanae* (when compared with *C. albula*) related to a distinct $DAPI^+$ band on a large metacentric chromosome pair (Figure 4f). In *C. albula*, the rDNA signal was always located on the opposite arm than the $DAPI^+$ band occurred (Figure 4e). In *C. fontanae*, one signal is located on the same arm and one signal is on the opposite one (Figure 4f). The construction of the FISH probes used in this study is visualized in Figure 1.

Molecular characterization of multiplied rDNA sites

To determine mechanisms how such extensive multiplication of parts of rRNA genes in *C. fontanae* might have arisen, a PCR screening for non-LTR retrotransposons in genomic DNA of both species was performed. Retrotransposons of the *Rex* family are known to have

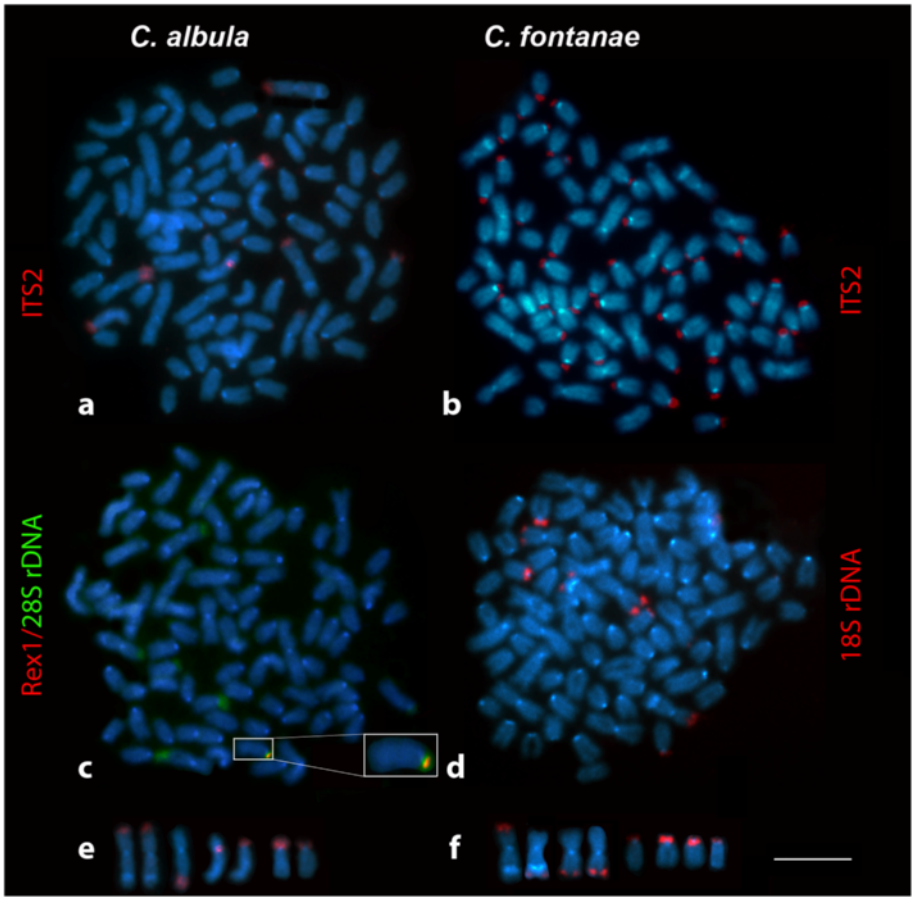


Figure 4 Metaphase plates and selected chromosomes of *C. albula* and *C. fontanae* showing FISH experiments. FISH with ITS2 (red) as probe hybridized to *C. albula* (**a**) and to *C. fontanae* (**b**), counterstained with DAPI. Double-FISH analysis with the *Rex1* retrotransposon (red) and the 800 bp 28S rDNA (green) to *C. albula* (**c**), detail of the chromosome with *Rex1* and 28S rDNA co-localization in inset. FISH with the 18S rDNA (red) to *C. fontanae* chromosomes (**d**) counterstained with DAPI. Chromosomes bearing the ITS2 (red) signal in *C. albula* (**e**). Chromosomes bearing the 18S rDNA signals (red) in *C. fontanae* (**f**). Chromosomes on (**f**) represent the hypothetically ancestral condition of rDNA distribution prior to its multiplication in *C. fontanae*. Bar = 5 μm.

invaded fish genomes in multiple lineages [24] and to also insert into rDNA, particularly in fishes [25]. Therefore, the retroelements *Rex1*, *Rex3* and *Rex6* were tested in this study. FISH with the *Rex3* and *Rex6* retroelements yielded inconclusive results. The *Rex1* element, as a probe hybridized to chromosomes of *C. albula* and *C. fontanae*, typically showed a dispersed pattern of signals on all chromosomes with a distinct accumulation in a pericentromeric region of one single acrocentric chromosome. Co-hybridization of the *Rex1* element with the 800 bp 28S rDNA probe in a double-FISH experiment showed co-localization of these two probes typically on one (exceptionally two to several), mostly acrocentric chromosomes in both *C. albula* and *C. fontanae* (Figure 4c for *C. albula* only, detail of the co-localization in inset). The *Rex1* signal with a distinctly weaker intensity occurred dispersed also on other sites corresponding to the NOR loci in both genomes. The sequences of the *Rex1* derived from the *C. albula* and *C. fontanae* genome were deposited in GenBank under the accession numbers JQ731754 and JQ731760, respectively.

Sequencing of the 18S and 28S rDNA, as well as ITS1 and ITS2 (deposited in GenBank under accession numbers JQ731749-JQ731753 and JQ731755-JQ731759) yielded no significant differences in these genes between *C. albula* and *C. fontanae*.

Discussion

Our findings of extensive genomic re-arrangements of a substantial fraction of the 45S rDNA unit in the *C. fontanae* genome when compared with the situation in *C. albula* are in strong contrast with previously reported low genetic differentiation between these two species [6,18].

Our results indicate that in the genome of *C. fontanae* next to the complete 6–10 NOR loci corresponding to similar number of NOR-bearing chromosomes in *C. albula*,

up to 30 supernumerary and incomplete NOR loci occur. This is supported by results of the sequential fluorescent staining (CMA_3 and DAPI), showing about 6–8 signals in karyotypes of both species, although on chromosomes of *C. fontanae* the signals were slightly weaker. However, these supernumerary sites in *C. fontanae* were not represented by repeating of the complete 45S rDNA unit (i.e. 18S rDNA, ITS1, 5.8S rDNA, ITS2, 28S rDNA)$_n$ but only by a part including probably complete ITS1 and ITS2 and a part of the 28S rDNA adjacent to the ITS2, i.e. the 5′ end of the 28S rDNA gene (the region of 5.8S rDNA was not investigated separately).

Most of the supernumerary signals of the 45S rDNA in chromosomes of *C. fontanae* were localized in the AT rich pericentromeric regions as well as the major accumulation of the *Rex1* retrotransposon on both *C. albula* and *C. fontanae* chromosomes. This is in accordance with findings of other authors describing accumulations of transposable elements in centromeric heterochromatin e.g. in genome of humans [26] and in a cichlid fish *Cichla kelberi* [27,28]. TEs in fishes generally tend to insert to heterochromatic areas of chromosomes ([29]; reviewed by [30]). There are also records of specific integration of some non-LTR retrotransposons at the rRNA genes found in most animal phyla (summarized by [31]), in insects *Drosophila melanogaster* and *Bombyx mori* [32] or in the fish *Erythrinus erythrinus*, where *Rex3* retrotransposons were found in the 5S rRNA genes [25].

In the above-mentioned *E. erythrinus* fish, a similar multiplication of rRNA genes was described [25]. In that case, four karyomorphs of *E. erythrinus* differ in their chromosomal number, karyotype, presence or absence of heteromorphic sex chromosomes and numbers of 5S rDNA loci. The karyomorph A in *E. erythrinus* showed only two 5S rDNA loci, while in the karyomorph D, 21–22 5S rDNA loci could be observed. All 5S rDNA sites co-localized with the *Rex3* retrotransposon. On the other hand, no changes in the heterochromatin and 18S rDNA patterns were found between these two karyomorphs [25]. Such two karyomorphs within a single species *E. erythrinus* may be seen as an incipient stage of a speciation event. This situation can thus represent an initial stage, later resulting in the condition observed in morphologically [15], ecologically and physiologically [33,34] diverged species pair *C. fontanae* and *C. albula* described in this study. A similar observation of extremely multiplied NOR sites (46 and 49 countable FISH signals), however, without any further detailed analysis, were reported in brook char *Salvelinus fontinalis* (Salmonidae) [35].

In salmonid fishes, TEs have been studied intensively [30,36,37]. Microarray studies showed that transcription of rainbow trout transposons is activated by external stimuli, such as toxicity, stress and bacterial antigens [38]. In the oligotrophic Lake Stechlin, the food availability for coregonines was extremely limited and the size at maturity and the maximal size of *C. albula* are far behind the other populations of this species in adjacent lakes in northern Germany [39]. *C. fontanae* is the smallest species of the genus *Coregonus* in Europe [8]. Raising both species in the laboratory demonstrated that both grew much larger if supported with unlimited food (unpublished obs., Freyhof). Therefore, it can be speculated that both species, especially *C. fontanae*, live in an extreme permanent starvation in the Lake Stechlin. It can be also hypothesized that the spring-spawning habit of *C. fontanae* might have originated simply by the shift of sexual maturity in the part of the population that has not been able to attain sexual maturity in autumn due to the lower food intake and hence environmental starvation stress.

Link between environmental stress and chromatin modification/regulation

Effects of stress on the genome can result in important perturbations creating new combinations better compatible with survival (summarized by [40]; more recently reviewed by [41]). After the discovery of transposable elements (TE) more than 50 years ago, their mutagenic effect had been increasingly viewed in association with rapid genome reorganizations by the creation of new regulation patterns and chromosome restructuring during last years [41]. Stress activated mobilization of these elements by failure of epigenetic silencing (the host defence model of repressing the movement of mobile elements; [42,43]) can lead to (re)activation of mobile elements and consequently to major and rapid genome alterations [40,41,44,45].

Barbara McClintock [46] already considered TE as a source of hypermutagenicity creating viable and fertile individuals from a stressed population under risk of extinction. Moreover, she originally named TE “controlling elements” due to their ability to alter gene activity and genome structure [47].

TE-mediated genome rearrangements as a factor in speciation

With growing evidence for the importance of TEs in the genome evolution, the role of TE-mediated genome changes in the speciation by their possible contribution to pre- and post-mating reproductive isolation formation has been increasingly taken into account and discussed generally in eukaryotes [48,49], *Drosophila* [50], fishes [51], mammals [52], and plants [53]. However, lack of experimental data makes it difficult to prove this possibility (reviewed by [51,54,55]). On the other hand, [41] provides an overview of TE transposition bursts

concomitant with radiation periods in seven cases. The same authors also discuss TE-induced rapid speciation associated with the ability of TEs to induce chromosomal rearrangements. Therefore, the sympatric species pair *C. albula* and *C. fontanae* in the context of other congeneric coregonine species and their variable evolutionary history in the Eurasian post-glacial lakes appears to be a suitable model system for exploring mechanisms of genomic differentiation and speciation with or without TE contribution.

In a very similar, but North American study system (lake whitefish species pairs, *Coregonus* spp.), [56] next generation sequencing (NGS) showed that TEs appeared to be highly expressed in hybrids between two recently diverged species. This may be potentially the mechanism responsible for post-zygotic reproductive isolation. Moreover, NGS can be viewed as a useful tool complementary with molecular cytogenetic approach presented in this study enabling confirmation of here documented results and search for other candidate groups of TEs involved in the genome re-arrangements and accelerated speciation.

Conclusion

In the sympatric species pair *C. albula* and *C. fontanae*, we encounter a complex situation involving several evolutionary phenomena and factors. Firstly, a rapid ecological speciation event with an unclear sympatric scenario, i.e. the derived species *C. fontanae* fully differentiated from *C. albula* physiologically, ecologically and morphologically within about 12 – 14 kyrs in the newly colonized Stechlin Lake after the last glacier retreated [15]. Secondly, genetic differentiation of these two species remained weak as the combined analyses of mtDNA and microsatellite loci [18] showed, as well as major karyotypic and chromosomal markers presented in this study. This is in contrast with extensive genome re-arrangements in a large proportion of the 45S rDNA cassette in *C. fontanae* when compared with its most likely ancestral species – *C. albula*. The genome re-arrangements are exhibited as a distinct loci number differences and relocation of variable number (about 30) AT rich pericentromeric regions in *C. fontanae*. The molecular mechanism behind these re-arrangements might be a retrotransposition of a part of the 45S rDNA unit mediated by retrotransposons. Retrotransposonal activity can be mobilized under certain conditions (stress, environmental changes) and cause rapid and extensive structural changes to the host genome. These structural genomic differences in *C. fontanae* accumulated to pericentromeric heterochromatin in almost half of the chromosome complement. This might then have been acting as a partial but permanent reproductive barrier by hampering recombination, thus, enabling and accelerating the morphological, ecological and physiological differentiation of *C. fontanae*. Moreover, interspecific hybridization between the old and the newly arising species might have activated retrotransposonal activity in hybrids resulting in hybrid sterility or unviability as reviewed by [51]. The population genetic parameters of this speciation event, favouring fixation of the re-arranged genomes, remain to be elucidated in detail, but small effective population size is a good hypothesis to be tested.

Methods

Materials

For this study, we had 12 individuals of *Coregonus albula* (Linnaeus, 1758) and 16 individuals of *C. fontanae* [15], both from Lake Stechlin (northern Germany, Brandenburg, 53° 10' N; 13° 02' E). All fish were raised in the laboratory under identical conditions as described by [33,34]. In *C. albula*, 3 individuals (samples alb 1, 2 and 5, males only) yielded metaphases usable for down-stream FISH and CGH experiments. In *C. fontanae*, 3 individuals also (samples font 2, 5 and 7, both males and females) yielded usable chromosome preparations. Of all studied individuals, we isolated genomic DNA from fin clips and muscles. All tissue and DNA samples, including cell suspensions and chromosome preparations, are deposited in the Laboratory of Fish Genetics of the Institute of Animal Physiology and Genetics (IAPG). This study was covered by the "Valid Animal Use Protocols" Nr. CZ 00221 at the IAPG issued by the Czech Ministry of Agriculture on 10 June 2009.

Chromosome preparations

Metaphases were prepared according to [57] with slight modifications. Briefly, the fish were injected with 0.1% colchicine solution (w/v, SIGMA), 1 ml/100 g body weight, for 45 minutes then sacrificed by overdose of anaesthetic 0.5% Phenoxyethanol (v/v, SIGMA). Kidneys were removed, dissected in 0.075 M KCl and the cell suspension free of tissue fragments was hypotonized for 8 min in 0.075 M KCl, fixed in methanol: acetic acid 3:1 (v/v) fixative, washed twice in fixative, and finally spread onto slides (Superfrost quality). Mitotic activity was not stimulated because these fish showed extremely high sensitivity to agents increasing mitotic rate. Simultaneously, the blood (around 0.5 ml) was collected from all analysed individuals by fine heparinized syringe for leukocyte culture according to the protocol of [58]. Briefly, partly washed leukocytes were cultivated in 5 ml of a complete medium composed of TC 199 (SIGMA, St. Louis, MO, USA), 10% FBS Superior (Biochrom, Berlin, Germany), 0.5% Antibiotic Antimycotic Solution (SIGMA), 1% Kanamycin monosulfate (SIGMA), 1% LPS (SIGMA), 0.2% PHA H15 (Remel, Lenexa, KS, USA) and 0.175μl Mercaptoethanol (SIGMA) at 19.5°C for 6–7 days,

then 2 drops of the 0.1% colchicine were added for 45 minutes at RT and cells harvested as for the direct preparation described above.

Fluorescence *in situ* hybridization (FISH) and comparative genomic hybridization (CGH)

Probes for *in situ* hybridization experiments were produced either by PCR (FISH probes) or directly from the genomic DNA (CGH probes). Probes were indirectly labelled with haptens (biotin and digoxigenin) by means of nick translation (whole genomic DNA and FISH probe longer than 600 bp) using the Roche Nick Translation Mix (Roche, Mannheim, Germany; Cat.No. 11745808910) according to the manufacturer's instructions. Shorter DNA fragments were labelled by PCR using the Roche PCR DIG Labeling Mix (Cat.No. 11585550910). The biotin-dUTP labelled probes (Roche, Cat. No. 11093070910) were detected by either the Invitrogen Cy™3-Streptavidin (Invitrogen, San Diego, CA, USA; Cat.No. 43–4315) or by the FITC-Streptavidin (Cat.No. 43–4311). The digoxigenin-dUTP labelled probes (Roche, Cat.No. 11093088910) were detected by either the Roche Anti-Digoxogenin-Fluorescein (Cat.No. 11207741910) or by the Anti-Digoxigenin-Rhodamin (Cat.No. 11207750910). An unlabelled DNA competitor for suppression of nonspecific hybridization of fragment size of 100–200 bp was added with 20-fold the concentration of the DNA probe in CGH experiments. The CGH DNA probe concentration was 1 ug per reaction for both genomes compared. An aging of chromosome preparations at 37°C for 3 hours was carried out prior to each of the hybridization experiment. Pepsinization, hybridization and detection were carried out under conditions as described by [59].

All rDNA FISH probes were constructed using published, mostly generally used PCR primer sets of the 45S rDNA unit to cover its major regions and to map them physically onto chromosomes.

Table 1 PCR primers used in this study

Name	Region of DNA/FISH probe	Primer sequence (5'to3')	Ref.
28S A	3′ end of the 28S rDNA involving the regions A and B, F primer	AAA CTC TGG TGG AGG TCC GT	[61]
28S B	Internally nested in the regions A and B of the 28S rDNA, R primer	CTT ACC AAA AGT GGC CCA CTA	[61]
28S C1	5′ end of the 28S rDNA adjacent to the ITS2, F primer	ACC CGC TGA ATT TAA GCA T	[62]
28S D2	Internally nested in the region C3 involving D2, C2, D1, C1, R primer	TCC GTG TTT CAA GAC GGG	[63]
ITS1	3′ end of the 18S rDNA adjacent to the ITS1, F primer	TCC GTA GGT GAA CCT GCG G	[64]
ITS2	3′ end of the 5.8S rDNA adjacent to the ITS2, R primer	GCT GCG TTC TTC ATC GAT GC	[64]
ITS3	5′ end of the 5.8S rDNA adjacent to the ITS1, F primer	GCA TCG ATG AAG AAC GCA GC	[64]
ITS4	5′ end of the 28S rDNA adjacent to the ITS2, R primer	TCC TCC GCT TAT TGA TAT GC	[64]
NS1	5′ end of the 18S rDNA F primer	GTA GTC ATA TGC TTG TCT	[64]
NS2	18 S rDNA R primer	GGC TGC TGG CAC CAG ACT TGC	[64]
NS3	18 S rDNA F primer	GCA AGT CTG GTG CCA GCA GCC	[64]
NS4	18 S rDNA R primer	CTT CCG TCA ATT CCT TTA AG	[64]
NS5	18 S rDNA F primer	AAC TTA AAG GAA TTG ACG GAA G	[64]
NS6	18 S rDNA R primer	GCA TCA CAG ACC TGT TAT TGC CTC	[64]
NS7	18 S rDNA F primer	GAG GCA ATA ACA GGT CTG TGA TGC	[64]
NS8	3′ end of the 18S rDNA R primer	TCC GCA GGT TCA CCT ACG GA	[64]
RTX1F1	Rex1 F primer	TTC TCC AGT GCC TTC AAC ACC	[28]
RTX1R3	Rex1 R primer	TCC CTC AGC AGA AAG AGT CTG CTC	[28]
RTX3F1	Rex3 F primer	TAC GGA GAA AAC CCA TTT CG	[65]
RTX3F2	Rex3 F primer	AAC ACC TTG GCT GCG CCT AG	[65]
RTX3F3	Rex3 F primer	CGG TGA YAA AGG GCA GCC CTG	[28]
RTX3R1	Rex3 R primer	AAA GTT CCT CGG TGG CAA GG	[65]
RTX3R2	Rex3 R primer	CCR GGG GTG GAT GAR RTC CGC CC	[65]
RTX3R3	Rex3 R primer	TGG CAG ACN GGG GTG GTG GT	[28]
RTX6F	Rex6 F primer	TAA AGC ATA CAT GGA GCG CCA C	[28]
RTX6R	Rex6 R primer	GGT CCT CTA CCA GAG GCC TGG G	[28]

PCR amplification of FISH probes and the analysis of the 45S rDNA unit

All primer sets used in this study are summarized in Table 1. Primers nesting within the 45S rDNA unit relevant for this study are shown in Figure 1. Thermal profiles were used according to references given in Table 1. FISH probes were constructed from PCR conducted on the respective species as they were later hybridized. All sequences used in this study as FISH probe or in the molecular-biological analyses of the 45S rDNA unit were deposited in the GenBank [60] under accession numbers JQ731749 - JQ731760.

Cloning, sequencing and sequences analysis

PCR products were cloned using the QIAGEN PCR Cloning Kit and QIAGEN EZ Competent Cells (Qiagen, Hilden, Germany); the plasmids were isolated from the cells with Qia PREP Spin Miniprep Kit according to the manufacturer's instructions. The primary PCR products were first sequenced on the ABI 3130 Genetic Analyzer (Applied Biosystems, Hitachi, Foster City, CA, USA) using the BigDye Terminator Cycle Sequencing Kit (Applied Biosystems). Furthermore, cloned DNA fragments that were later applied as FISH probes were commercially sequenced by Macrogen (Seoul, South Korea). The commercially obtained sequences were subjected to online megablast or discontiguous megablast [66] searches at the National Center for Biotechnology Information (NCBI) [67], where their similarity to the sequences deposited in the GenBank databases was checked.

Microscopy and image processing

Chromosome preparations were analysed with the Provis AX70 Olympus microscope equipped with standard fluorescence filter sets. Gray-scale hybridization signals on chromosomes and/or DAPI counterstained chromosomes were captured by the CCD camera (DP30W Olympus). Using the Olympus Acquisition Software, black and white images were pseudo-coloured and superimposed with the software MicroImage. The colour images have been analyzed and processed with Adobe Photoshop, Version CS5. The chromosomes were classified using the nomenclature proposed by [68]. Karyotypes based on the Giemsa-stained chromosomes were produced using the IKAROS (Metasystems) software. Chromosomal formulas were formed according to [69].

Abbreviations

AFLP: Amplified fragment length polymorphism; CGH: Comparative genomic hybridization; CMA_3: Chromomycin A_3; DAPI: 4', 6-diamidino-2-phenylindole; DAPI/$CMA_3^{+/-}$: DAPI/CMA_3 positive/negative signals; FISH: Fluorescence *in situ* hybridization; IAPG: Institute of animal physiology and genetics; kyrs BP: Thousand years before presence; NF: "Nombre fundamental" chromosome arm number; NGS: Next generation sequencing; NOR: Nucleolar organizer DNA; rDNA: Ribosomal DNA; TE: Transposable elements; 2n: Diploid chromosome number.

Competing interests

There are no competing interests to declare.

Authors' contributions

RS designed and performed experiments FISH and CGH and co-drafted the manuscript, ZM and AS performed FISH experiments, ZM performed PCRs and sequenced PCR products and co-drafted the manuscript, GS and JF collected and raised material, JB and RS prepared chromosome preparations, MR partly contributed to digital processing of images, PR designed the study and co-drafted the manuscript. All authors read and approved the final manuscript.

Acknowledgements

This study was supported in parts by the projects No. 523/08/0824 and No. P506/11/P596 of the Grant Agency of the Czech Republic, No. LC06073 Biodiversity Research of the Ministry of Education, Youth and Sports of the Czech Republic and No. IRP IAPG AV0Z 50450515.
This study is a part of the series "Chromosome evolution in Salmonidae".

Author details

[1]Laboratory of Fish Genetics, Institute of Animal Physiology and Genetics, Czech Academy of Sciences, Rumburská 89, Liběchov 277 21, Czech Republic. [2]Leibniz-Institut of Freshwater Biology and Inland Fisheries, Müggelseedamm 310, Berlin 12587, Germany.

Received: 29 May 2012 Accepted: 11 February 2013
Published: 14 February 2013

References

1. Cristescu ME, Adamowicz SJ, Vaillant JJ, Haffner DG: **Ancient lakes revisited: from the ecology to the genetics of speciation.** *Mol Ecol* 2010, **19:**4837–4851.
2. Schluter D: **Ecological speciation in postglacial fishes.** *Philos T Roy Soc B* 1996, **351:**807–814.
3. Nelson JS: *Fishes of the world*. 4th edition. Hoboken, NJ: John Wiley & Sons; 2006.
4. Bodaly RA, Vuorinen J, Ward RD, Luczynski M, Reist JD: **Genetic comparison of new and old world coregonid fishes.** *J Fish Biol* 1991, **38:**37–51.
5. Hudson AG, Vonlanthen P, Seehausen O: **Rapid parallel adaptive radiations from a single hybridogenetic ancestral population.** *P Roy Soc B-Biol* 2011, **278:**58–66.
6. Mehner T, Pohlmann K, Elkin C, Monaghan MT, Nitz B, Freyhof J: **Genetic population structure of sympatric and allopatric populations of Baltic ciscoes (*coregonus albula* complex, teleostei, coregonidae).** *BMC Evol Biol* 2010, **10:**85.
7. Douglas MR, Brunner PC, Bernatchez L: **Do assemblages of *coregonus* (teleostei: salmoniformes) in the central alpine region of Europe represent species flocks?** *Mol Ecol* 1999, **8:**589–603.
8. Kottelat M, Jörg F: *Handbook of European Freshwater Fishes*. Cornol, Switzerland: Publications Kottelat; 2007.
9. Vonlanthen P, Roy D, Hudson AG, Largiader CR, Bittner D, Seehausen O: **Divergence along a steep ecological gradient in lake whitefish (*coregonus* sp.).** *J Evolution Biol* 2009, **22:**498–514.
10. Taylor EB: **Species pairs of north temperate freshwater fishes: evolution, taxonomy, and conservation.** *Rev Fish Biol Fisher* 1999, **9:**299–324.
11. Vuorinen J, Himberg M, Lankinen P: **Genetic differentiation in *coregonus albula* (salmonidae) populations in Finland.** *Hereditas* 1981, **94:**113–121.
12. Hudson AG, Vonlanthen P, Müller R, Seehausen O: **Review: the geography of speciation and adaptive radiation in coregonines.** *Adv Limnol* 2007, **60:**111–146.
13. Phillips R, Ráb P: **Chromosome evolution in the salmonidae (pisces): an update.** *Biol Rev* 2001, **76:**1–25.
14. Quimseyh MB: **Evolution of number and morphology of mammalian chromosomes.** *J Hered* 1994, **85:**455–465.
15. Schulz M, Freyhof J: ***Coregonus fontanae*, a new spring-spawning Cisco from lake stechlin, northern Germany (salmoniformes: coregonidae).** *Ichthyol Explor Fresh* 2003, **14:**209–216.
16. Ohlberger J, Mehner T, Staaks G, Hoelke F: **Is ecological segregation in a pair of sympatric coregonines supported by divergent feeding efficiencies?** *Can J Fish Aquat Sci* 2008, **65:**2105–2113.

17. Jankun M, Martinez P, Pardo BG, Kirtiklis L, Rab P, Rabova M, Sanchez L: **Ribosomal genes in coregonid fishes (*coregonus lavaretus*, *C. Albula* and *C. Peled*) (salmonidae): single and multiple nucleolus organizer regions.** *Heredity* 2001, **87:**672–679.
18. Schulz M, Freyhof J, Saint-Laurent R, Østbye K, Mehner T, Bernatchez L: **Evidence for independent origin of two spring-spawning ciscoes (Salmoniformes: Coregonidae) in Germany.** *J Fish Biol* 2006, **68:**119–135.
19. Swarzacher HG, Wachtler F: **The nucleolus.** *Anat Embryol* 1993, **188:**515–536.
20. Wachtler F, Stahl A: **The nucleolus: a structural and functional interpretation.** *Micron* 1993, **24:**473–505.
21. Moss T, Stefanovsky VY: **Promotion and regulation of ribosomal transcription in eukaryotes by RNA polymerase I.** *Prog Nucleic Acids Res Mol Biol* 1994, **50:**25–66.
22. Wang S, Zhao M, Li T: **Complete sequence of the 10.3 kb silkworm attacus ricini rDNA repeat,determination of the transcriptional initiation site and functional analysis of the intergenic spacer.** *DNA Seq* 2003, **14:**95–101.
23. Hillis DM, Dixon MT: **Ribosomal DNA: molecular evolution and phylogenetic inference.** *Q Rev Biol* 1991, **66:**411–453.
24. Volff J-N, Körting C, Schartl M: **Multiple lineages of the non-LTR retrotransposon *Rex1* with varying success in invading fish genomes.** *Mol Biol Evol* 2000, **17:**1673–1684.
25. Cioffi MB, Martins C, Bertollo LA: **Chromosome spreading of associated transposable elements and ribosomal DNA in the fish *Erythrinus erythrinus*. Implications for genome change and karyoevolution in fish.** *BMC Evol Biol* 2010, **10:**271.
26. Wong LH, Andy Choo KH: **Evolutionary dynamics of transposable elements at the centromere.** *Trends Genet* 2004, **20:**611–616.
27. Teixeira WG, Ferreira IA, Cabral-de-Mello DC, Mazzuchelli J, Valente GT, Pinhal D, Poletto AB, Martins C: **Organization of repeated DNA elements in the genome of the cichlid fish *cichla kelberi* and its contribution to the knowledge of fish genomes.** *Cytogenet Genome Res* 2009, **125:**224–234.
28. Valente GT, Mazzuchelli J, Ferreira IA, Poletto AB, Fantinatti BEA: **Cytogenetic mapping of the retroelements *Rex1*, *Rex3* and *Rex6* among cichlid fish: new insights on the chromosomal distribution of transposable elements.** *Cytogenet Genome Res* 2011, **133:**34–42.
29. DaSilva C, Hadji H, Ozouf-Costaz C, Nicaud S, Jaillon O, Weissenbach J, Roest Crollius H: **Remarkable compartmentalization of transposable elemetns and pseudogenes in the heterochromatin of the *tetraodon nigroviridis* genome.** *Proc Natl Acad Sci U S A* 2002, **99:**1636–1641.
30. Koop BF, Davidson WS: **Genomics and the Genome Duplication in Salmonids**. In *Fisheries for Global Welfare and Environment*. Edited by Tsukamoto K, Kawamura T, Takeuchi T, Beard TD Jr, Kaiser MJ. Tokyo: Terrapub; 2008:77–86.
31. Kurzynska-Kokorniak A, Jamburuthugoda VK, Bibillo A, Eickbush TH: **DNA-directed DNA polymerase and strand displacement activity of the reverse transcriptase encoded by the R2 retrotransposon.** *J Mol Biol* 2007, **374:**322–333.
32. Jakubczak JL, Xiong Y, Eickbush TH: **Type I (RI) and type II (R2) ribosomal DNA insertions of *drosophila melanogaster* are retrotransposable elements close11 related to those of bombyx Mori.** *J Mol Biol* 1990, **212:**37–52.
33. Ohlberger J, Mehner T, Staaks G, Hölker F: **Temperature-related physiological adaptations promote ecological divergence in a sympatric species pair of temperate freshwater fish. Coregonus spp.** *Funct Ecol* 2008, **22:**501–508.
34. Ohlberger J, Staaks G, Petzoldt T, Mehner T, Hölker F: **Physiological specialization by thermal adaptation drives ecological divergence in a sympatric fish species pair.** *Evol Ecol Res* 2008, **10:**1173–1185.
35. Fujiwara A, Abe S, Yamaha E, Yamazaki F, Yoshida MC: **Chromosomal localization and heterochromatin association of ribosomal RNA gene loci and silver-stained nucleolar organizer regions in salmonid fishes.** *Chromosome Res* 1998, **6:**463–471.
36. Matveev V, Okadaa N: **Retroposons of salmonoid fishes (actinopterygii: salmonoidei) and their evolution.** *Gene* 2008, **434:**16–28.
37. De Boer JG, Yazawa R, Davidson WS, Koop BF: **Burst and horizontal evolution of DNA transposons in the speciation of pseudotetraploid salmonids.** *BMC Genomics* 2007, **8:**422.
38. Krasnov A, Koskinen H, Afanasyev A, Mölsä H: **Transcribed Tcl-like transposons in salmonid fish.** *BMC Genomics* 2005, **6:**107.
39. Helland IP, Harrod C, Freyhof J, Mehner T: **Co-existence of a pair of pelagic planktivorous coregonid fishes.** *Evol Ecol Res* 2008, **10:**373–390.
40. Arnault C, Dufournel I: **Genome and stresses: reactions against aggressions, behaviour of transposable elements.** *Genetica* 1994, **93:**149–160.
41. Rebollo R, Horard B, Hubert B, Vieira C: **Jumping genes and epigenetics: towards new species.** *Gene* 2010, **454:**1–7.
42. Bestor TH, Tycko B: **Creation of genomic methylation patterns.** *Nat Genet* 1996, **12:**363–367.
43. Yoder JA, Walsh CP, Bestor TH: **Cytosine methylation and the ecology of intragenomic parasites.** *Trends Genet* 1997, **13:**335–340.
44. Wichman HA, Van Den Bussche RA, Hamilton MJ, Baker RJ: **Transposable elements and the evolution of genome organization in mammals.** *Genetica* 1992, **86:**287–293.
45. Biémont C: **A brief history of the status of transposable elements: from junk DNA to major players in evolution.** *Genetics* 2010, **186:**1085–1093.
46. McClintock B: **The significance of responses of the genome to challenge.** *Science* 1984, **226:**792–801.
47. McClintock B: **Mutable loci in maize.** *Carnegie Inst Wash Yrbk* 1949, **48:**142–154.
48. Hurst GDD, Werren JH: **The role of selfish elements in eukaryotic evolution.** *Nat Rev Genet* 2001, **2:**597–606.
49. Hua-Van A, Le Rouzic A, Boutin TS, Filée J, Capy P: **The struggle for life of the genome's selfish architects.** *Biol Direct* 2011, **6:**19.
50. Vieira C, Lepetit D, Dumont S, Biémont C: **Wake up of transposable elements following *drosophila simulans* worldwide colonization.** *Mol Biol Evol* 1999, **16:**1251–1255.
51. Volff JN: **Genome evolution and biodiversity in teleost fish.** *Heredity* 2005, **94:**280–294.
52. Böhne A, Brunet F, Galiana-Arnoux D, Schultheis C, Volff J-N: **Transposable elements as drivers of genomic and biological diversity in vertebrates.** *Chromosome Res* 2008, **16:**203–215.
53. Noor MAF, Chang AS: **Evolutionary genetics: jumping into a New species.** *Curr Biol* 2006, **16:**R890–R892.
54. Hurst GDD, Schilthuizen M: **Selfish genetic elements and speciation.** *Heredity* 1998, **80:**2–8.
55. Kazazian HH Jr: **Mobile elements: drivers of genome evolution.** *Science* 2004, **303:**1626–1632.
56. Renaut S, Nolte AW, Bernatchez L: **Mining transcriptome sequences towards identifying adaptive single nucleotide polymorphisms in lake whitefish species pairs (*coregonus* spp. Salmonidae).** *Mol Ecol* 2010, **19**(Suppl.1):115–131.
57. Ráb P, Roth P: **Cold-blooded vertebrates**. In *Methods of chromosome analysis*. Edited by Balicek P, Forejt J, Rubes J. Brno: Cytogenet Sect Cs Biol Soc Publishers; 1988:115–124.
58. Fujiwara A, Nishida-Umehara C, Sakamoto T, Okamoto N, Nakayama I, Abe S: **Improved fish lymphocyte culture for chromosome preparation.** *Genetica* 2001, **111:**77–89.
59. Cremer M, Grasser F, Lanctôt C, Müller S, Neusser M, Zinner R, Solovei I, Cremer T: **Multicolor 3D Fluorescence *In Situ* Hybridization for Imaging Interphase Chromosomes**. In *The Nucleus: Volume I: Nuclei and Subnuclear Components, Methods in Molecular Biology™*, Volume Chapter 15, Volume 463. Edited by Hancock R.: Humana Press; 2008:205–239. Springer Protocols.
60. *Database GenBank*. http://www.ncbi.nlm.nih.gov/genbank/.
61. Zhang Q, Cooper RK, Tiersch TR: **Chromosomal location of the 28S ribosomal RNA gene of channel catfish by in situ polymerase chain reaction.** *J Fish Biol* 2000, **56:**388–397.
62. Dayrat B, Tillier A, Lecointre G, Tillier S: **New clades of euthyneuran gastropods (mollusca) from 28S rRNA sequences.** *Mol Phylogenet Evol* 2001, **19:**225–235.
63. Chombard C, Boury-Esnault N, Tillier S: **Reassessment of homology of morphological characters in tetractinellid sponges based on molecular data.** *Syst Biol* 1998, **47:**351–366.
64. White TJ, Bruns T, Lee S, Taylor JW: **Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics**. In *PCR Protocols: A Guide to Methods and Applications*. Edited by Innis MA, Gelfand DH, Sninsky JJ, White TJ. New York: Academic Press Inc; 1990:315–322.

65. Volff JN, Körting C, Meyer A, Schartl M: **Evolution and discontinuous distribution of Rex3 retrotransposons in fish.** *Mol Biol Evol* 2001, **18:**427–431.
66. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215:**403–410.
67. *National Center for Biotechnology Information*. http://www.ncbi.nlm.nih.gov/blast.
68. Levan AK, Fredga K, Sandberg AA: **Nomenclature for centromeric position on chromosomes.** *Hereditas* 1964, **52:**201–220.
69. Matthey R: **L'evolution de la formule chromosomale chez les vertébrés.** *Experientia* 1945, **1:**50–56.

doi:10.1186/1471-2148-13-42
Cite this article as: Symonová *et al.*: Genome differentiation in a species pair of coregonine fishes: an extremely rapid speciation driven by stress-activated retrotransposons mediating extensive ribosomal DNA multiplications. *BMC Evolutionary Biology* 2013 **13**:42.

Original Article

Cytogenetic and Genome Research

Cytogenet Genome Res 2013;139:276–283
DOI: 10.1159/000350689

Accepted: November 15, 2012
by M. Schmid
Published online: May 4, 2013

Chromosome Studies of European Cyprinid Fishes: Cross-Species Painting Reveals Natural Allotetraploid Origin of a *Carassius* Female with 206 Chromosomes

M. Knytl[a] L. Kalous[a] R. Symonová[b] K. Rylková[a] P. Ráb[b]

[a]Department of Zoology and Fisheries, Faculty of Agrobiology, Food and Natural Resources, Czech University of Life Sciences Prague, Prague, and [b]Laboratory of Fish Genetics, Institute of Animal Physiology and Genetics, AS CR v.v.i., Liběchov, Czech Republic

Key Words
Fish cytogenetics · Genome addition · GISH · Leaky gynogenetic reproduction · Polyploid cyprinids

Abstract
A single female with 206 chromosomes and another 26 females with 156 chromosomes identified as Prussian carp, *Carassius gibelio*, and 5 individuals with 100 chromosomes identified as crucian carp, *C. carassius,* were sampled during field survey in one locality in the upper Elbe River. To identify the origin of females with high chromosome numbers, comparative karyotype analysis, GISH, with whole *C. carassius* DNA as probe and phylogenetic positions of sampled individuals revealed by cytochrome *b* mitochondrial marker were performed. GISH showed consistently bright labeling of 50 chromosomal elements out of 206, corresponding to the haploid chromosome number of *C. carassius.* The position of these females with high chromosome numbers in a reconstructed phylogenetic tree was within the clade of *C. gibelio*, documenting its affiliation to *C. gibelio* mitochondrial, i.e. maternal lineage. Our findings indicated that the mother of the female with high chromosome numbers was a gynogenetically reproducing 156-chromosome *C. gibelio* female and the father a bisexually reproducing *C. carassius* male. We, therefore, hypothesized that the *C. gibelio* × *C. carassius* allopolyploid female with 206 chromosomes arose by a mechanism of sperm genome addition to an unreduced egg of the mother.

Interspecific hybridization and production of viable hybrid offspring is well known among lower vertebrates [Dawley and Bogart, 1989; Vrijenhoek, 1998; Neaves and Bauman, 2011]. The relatively higher frequency of cross-species breeding among fishes is caused by overall predominance of external fertilization in aquatic environment. Moreover, many cyprinid species share similar spawning grounds in the same time that indeed increases the probability of hybridization events [Wheeler and Easton, 1978]. Fishes of the genus *Carassius* are represented in Europe by 4 taxa: (1) native and highly endangered crucian carp (*C. carassius* L.) [Kottelat and

KARGER

1424–8581/13/1394–0276$38.00/0

E-Mail karger@karger.com
www.karger.com/cgr

Lukáš Kalous, Department of Zoology and Fisheries
Faculty of Agrobiology, Food and Natural Resources
Czech University of Life Sciences Prague
CZ–16521 Prague 6 (Czech Republic)
E-Mail kalous@af.czu.cz

Freyhof, 2007], (2) pan-globally distributed feral goldfish (*C. auratus* L.) [Szczerbowski, 2002], (3) recently found Japanese Ginbuna (*C. langsdorfii*, Temminck and Schlegel 1846) [Kalous et al., 2007], and (4) native Prussian carp (*C. gibelio*, Bloch 1782) [Lusková et al., 2010; Kalous et al., 2012]. Moreover, diploid-polyploid complexes within the genus *Carassius* exist throughout the vast territory of its occurrence including biotypes comprising individuals with approximately 150 chromosomes ('triploids') [Kalous and Knytl, 2011] that are often represented almost exclusively by females [Halačka et al., 2003]. These females are sperm dependent parthenogens (gynogens) that require sperm of another related species for their reproduction [Peňáz et al., 1979]. Interestingly, when the heterologous sperm enters the unreduced egg of a gynogenetic female, the biological function of a sperm is reduced to the triggering of egg development, and the resulting offspring is a clone of the mother with the same ploidy level [Golovinskaya and Romashov, 1947; Yamashita et al., 1993; Vrijenhoek, 1998; Gui and Zhou, 2010]. An appearance of a small amount of male genetic material in the genome of a gynogenetic offspring was described by Yi et al. [2003], and this phenomenon is known as paternal leakage [Tóth et al., 2005; Lamatsch and Stöck, 2009]. It was also hypothesized that such leakage can be the reason for sudden male appearance within the whole female gynogenetic populations, due to a possible interspecific transfer of sex-determining genes [Arai et al., 1995; Janko et al., 2007; Loewe and Lamatsch, 2008; Neaves and Bauman, 2011]. Hybridization between different *Carassius* species and biotypes with various ploidy levels appears to be quite common [Mezhzheryn et al., 2012]. Hybridization between *C. carassius* and *C. gibelio* in alluvium of the Thaya River, Danube River basin, was recently demonstrated by microsatellite analyses [Papoušek et al., 2008]. Similarly, a recent ongoing hybridization process between native crucian carp *C. carassius* and introduced goldfish *C. auratus* was described from the British Isles by Hänfling et al. [2005] and from Sweden by Wouters et al. [2012]. Even intergeneric hybridization between the fishes of the genera *Carassius* and *Cyrpinus* were revealed by several studies [Hänfling et al., 2005; Zhu and Gui, 2006; Liu, 2010]. On the other hand, the morphological recognition and identification of such hybrids is very difficult, due to high external similarities. Here, we report a discovery and identification of a natural allotetraploid female resulting from hybridization of bisexually reproducing *C. carassius* and gynogenetically reproducing *C. gibelio*.

Materials and Methods

Fish Sampling

A single female with 206 chromosomes and another 26 females with 156 chromosomes identified morphologically as Prussian carp, *C. gibelio*, and 5 individuals with 100 chromosomes identified as crucian carp, *C. carassius*, were collected during a field survey of ichthyofauna in alluvial ponds and old oxbows of the Elbe River close to the city of Lysá nad Labem (recognized as Byšičky, GPS: 50°10′33,7″ N, 14°46′25,4″ E). The specimens examined were not deposited as vouchers.

Chromosome Analysis

Mitotic activity was stimulated by intraperitoneal injection of 0.1% $CoCl_2$ (1 ml $CoCl_2$/100 g body weight) 24 h before chromosome preparation. Standard direct procedures for chromosome preparation from cephalic kidney followed Ráb and Roth [1988]. To arrest cell division in metaphase, 0.1% colchicine (1 ml colchicine/100 g body weight) by intraperitoneal injection was used. Valid animal use protocols were in force at IAPG and CULS during this study.

Microscopy and Image Processing

Metaphase chromosomes stained in 4% Giemsa-Romanowski solution in phosphate buffer (pH = 7) were observed with a microscope BX41TF equipped with a digital camera Olympus SP-350, and chromosomes were counted by PC software QuickPhoto Micro. Karyotypes were constructed using PC software Ikaros (karyotyping system) version V 3.4.0 and Adobe Photoshop version CS5. Chromosome morphology was determined according to Levan et al. [1964]. Analyzed slides with recorded coordinates of selected metaphases were distained in fixative (methanol and acetic acid; 3:1, v/v) for 3 min and stored at +4°C until the GISH experiment.

Isolation of Genomic DNA

Total genomic DNA was isolated from ethanol preserved tissue using DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany) according to manufacturer's protocol.

Genomic in situ Hybridization

Genomic DNA from *C. carassius* was indirectly labeled by a standard nick translation reaction using nick translation mix for in situ probes according to the manufacturer's instructions (Roche, Mannheim, Germany). Total 25 µl of hybridization mixture, containing nick translation mix, NT-dNTPs, labeled dUTPs, DNA template, and H_2O, was incubated for 90 min at 15°C. DNA was precipitated with salmon sperm (100 µg/ml), 3 M sodium acetate pH = 5.2 (25°C) and 96% ethanol. The biotin-dUTP labeled probes (Roche) were detected by either the Invitrogen (Karlsruhe, Germany) CyTM3-Streptavidin or by FITC-Streptavidin. The digoxigenin-dUTP labeled probes (Roche) were detected by either the Roche anti-digoxigenin-fluorescein or by anti-digoxigenin-rhodamin diluted according to manufacturer's instructions. Chromosome preparations were dehydrated through ethanol series (70, 80 and 96% for 3 min each) on ice and air-dried. Chromosome preparations were aged for 1 h at 37°C before and after pepsinization (3 min at 37°C 50 µl aliquot pepsin, 1 N HCl and distilled H_2O).

Hybridization and detection during GISH experiments were carried out as described by Cremer et al. [2008]. Slides were dehydrated through ethanol series, air dried and aged again for 45 min at 37°C. Chromosomal denaturation was carried out in 75% for-

Table 1. Material used in molecular analyses

Fish	n	Sex	Ploidy level	Chromosome number	Locality	GenBank number
C. carassius	5	–	2n	100	Byšičky, Elbe River, Czech Republic	JQ763597
C. gibelio	5	f	3n	156	Byšičky, Elbe River, Czech Republic	JQ763598
C. gibelio/ *C. carassius*	1	f	4n	206	Byšičky	JQ763599
C. gibelio (neotype)	1	f	2n	–	Český Těšín, Olza River, Czech Republic	JN402305*
C. auratus	1	–	–	–	Nanking, Yangtze River, China	EU663598**
Cyprinus sp. (outgroup)	1	–	–	–	Mekong River, Thailand	HM008692*

* Sequences from Kalous et al. [2012]; ** sequence from Rylková et al. [2010]; cytochrome *b* haplotypes are deposited under the corresponding number in the GenBank.

mamid in 2× SSC (pH = 7) for 3 min at 74°C and quickly dehydrated through –20°C ethanol series and air dried. Hybridization mixture with Salmon sperm was denaturated at 86°C for 6 min and then cohybridized to target slide with the denatured metaphases from the tetraploid *Carassius* female under a 24 × 50 mm coverslip. The slides were incubated for 48 h at 37°C in a dark room. After hybridization, slides were then washed for 2 × 10 min each in 50% formamid with 2× SSC, 3 × 7 min each in 1× SSC in water bath at 42°C and 1 × 20 s in 4× SSC at room temperature. After performing series of stringency washes, stop reaction was carried out with 2.5% BSA/4× SSC/Tween for 20 min at 37°C under a 24 × 50 mm coverslip. Chromosomes were counterstained with DAPI (4′,6-diamino-2-phenylindol) combined with a mounting media (Starfish, Cambio, Cambridge, UK).

Microscopy and Image Processing

GISH images were captured with a cooled CCD camera Olympus DP30BW (equipped with a B&W CCD-Chip Sony ICX285-AL) coupled to an epifluorescence microscope Olympus AX70 equipped with a set of 3 narrowband fluorescent filters. Micrographs were captured with the Olympus Acquisition Software, and B&W images were processed with the software MicroImage. The pseudocolor images were analyzed using the Adobe Photoshop Version CS5. Altogether, 30 metaphases for each specimen were analyzed.

Molecular Analysis

Detailed information about material used for molecular analysis is listed in table 1. The cytochrome *b* gene was amplified using the methods described in Rylková et al. [2010], with the forward primer Kai_F 5′-GAAGAACCACCGTTGTTATTC-3′ and reverse primer Kai_R 5′-ACCTCCRAYCTYCGGATTACA-3′ [Šlechtová et al., 2006]. PCR products were purified and sequenced by Macrogen Incooperation (Seoul, Korea).

The raw chromatograms were manually assembled and checked by eye for potential mistakes using computer software BioEdit 5.0.9. [Hall, 1999]; the same program was used to align the sequences using the ClustalW algorithm. Dataset was created for cytochrome *b* analysis, and the phylogenetic relationships were estimated using the methods of maximum parsimony in PAUP* version 4.0b10 [Swofford, 2000] and Bayesian analysis using the program MrBayes version 3.0 [Huelsenbeck and Ronquist, 2001].

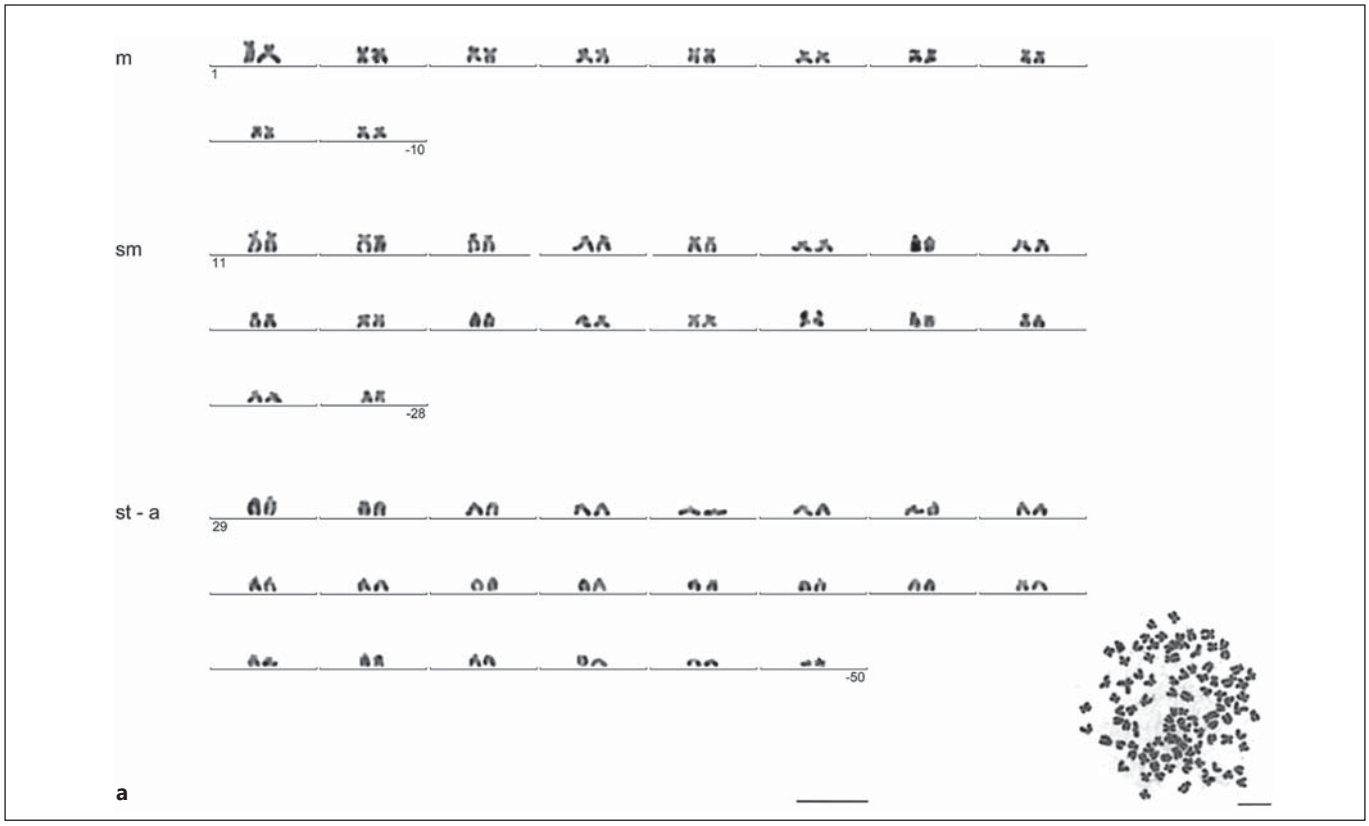
Results

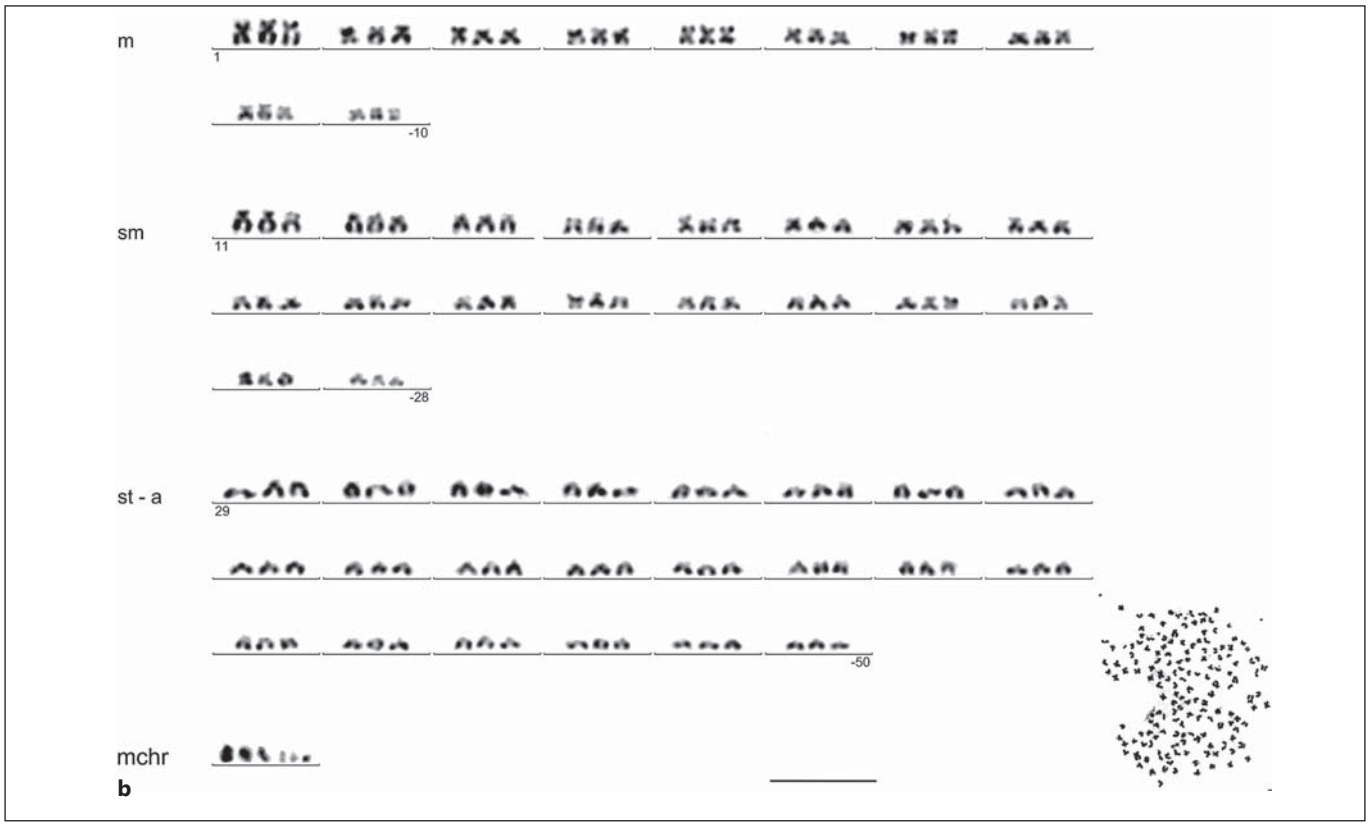
Chromosome Analysis

The analyzed individuals included 3 different categories. Individuals with 2n = 100 chromosomes were unambiguously identified as *C. carassius*, and their karyotypes consisted of 10 pairs of metacentric (m), 18 pairs of submetacentric (sm) and 22 pairs of subtelo- (st) to acrocentric (a) chromosomes (fig. 1a). All other fishes were identified as females of *C. gibelio*, where 26 individuals possessed 3n = 156 chromosomes with a karyotype composed of 30 m, 54 sm, 66 st to a, and 6 microchromosomes (fig. 1b), while one female had 4n = 206 and a karyotype composed of 40 m, 72 sm, 88 st to a, and 6 microchromosomes (fig. 1c).

Genomic in situ Hybridization

Biotin-labeled *C. carassius* genomic DNA hybridized to the chromosomes of tetraploid female and provided consistently intensive fluorescent signals on 50 chromosomes (fig. 2b) with distinctly pink fluorescence well discriminated from other blue-stained chromosomes. The positively hybridized chromosomes were graphically sep-





1

(For legend see page 280.)

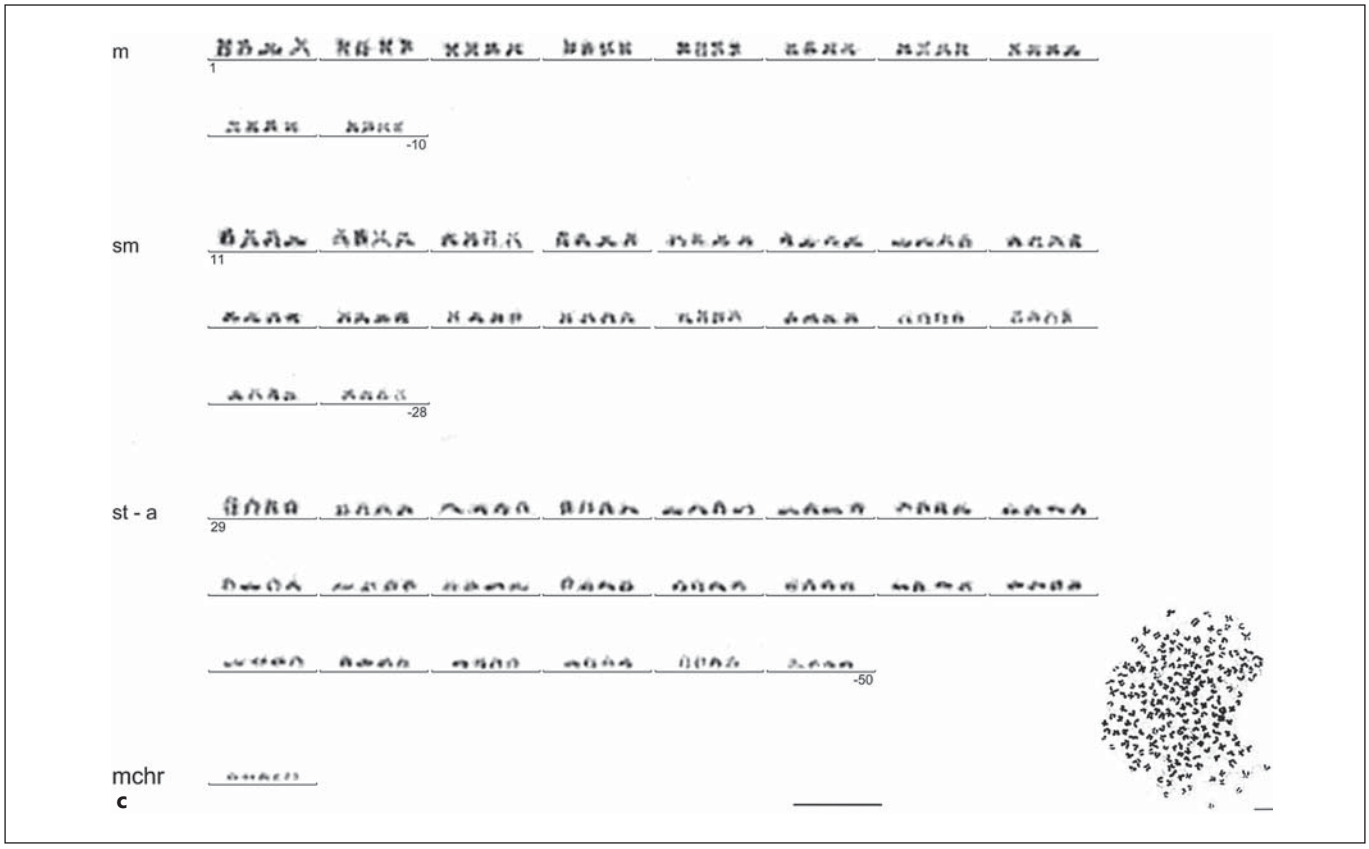


Fig. 1. Karyotype of male *C. carassius* (**a**), of female *C. gibelio* (**b**) and of female allotetraploid hybrid of *Carassius* (**c**), arranged from Giemsa-stained chromosomes (shown as inlay). m = Metacentric; sm = submetacentric; st = subtelocentric; a = acrocentric chromosomes; mchr = microchromosomes. Scale bar = 10 μm.

arated in the karyotype of tetraploid individual (fig. 3) and very likely corresponded to haploid chromosomes of *C. carassius* (fig. 1a).

Molecular Analyses

The final matrix of the cytochrome *b* sequences consisted of 1,110 bp containing 152 variable characters with 45 parsimony informative sites. Both employed methods have recovered topologies with high statistical supports. All 5 sequences of *C. carassius* from Byšičky showed one haplotype (GenBank accession number JQ763597); in case of *C. gibelio* from Byšičky one haplotype was found (GenBank accession number JQ763598). The position of *C. gibelio* × *C. carassius* hybrid (GenBank accession number JQ763599) in reconstructed phylogenetic tree (fig. 4) is within the clade of *C. gibelio* mitochondrial lineage indicating that the mother of the allotetraploid specimen was *C. gibelio*. Moreover, the hybrid and 5 individuals of *C. gibelio* from Byšičky shared the same haplotype.

Discussion

Our finding and subsequent genetic analyses of a natural allotetraploid female of the genus *Carassius* revealed its interspecific origin and combination of 3 chromosome sets of *C. gibelio* and a haploid one of *C. carassius* within one individual genome.

It was clearly demonstrated that the paternal chromosome haploid set detected by GISH analysis originated from *C. carassius*, while the maternal triploid set could be assigned by mtDNA markers to *C. gibelio*, but final confirmation must explore nuclear markers. This conclusion confirmed the previous assumptions because the males of *C. gibelio* were not found at the locality, and the local Prussian carp population consisted of triploid gynogenetic females only [Daněk et al., 2012].

Such occasional sperm genome additions in otherwise gynogenetically reproducing *Carassius* fishes are likely more common. Zhu and Gui [2006] reported on an inter-

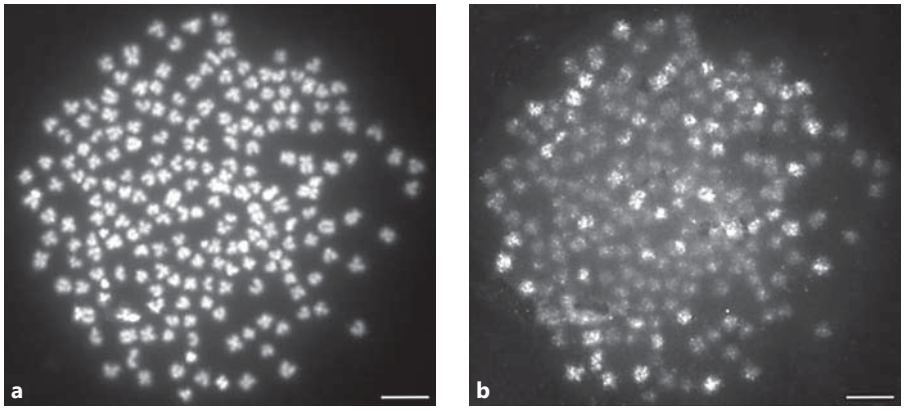


Fig. 2. Genomic in situ hybridization experiment using biotin-labeled *C. carassius* genomic DNA to the chromosomes of a tetraploid individual. Metaphase counterstained by DAPI shows all 206 chromosomes (**a**), metaphase image stained by biotin with Cy3 filter shows 50 chromosomes originating from *C. carassius* (**b**). Scale bar = 10 μm.

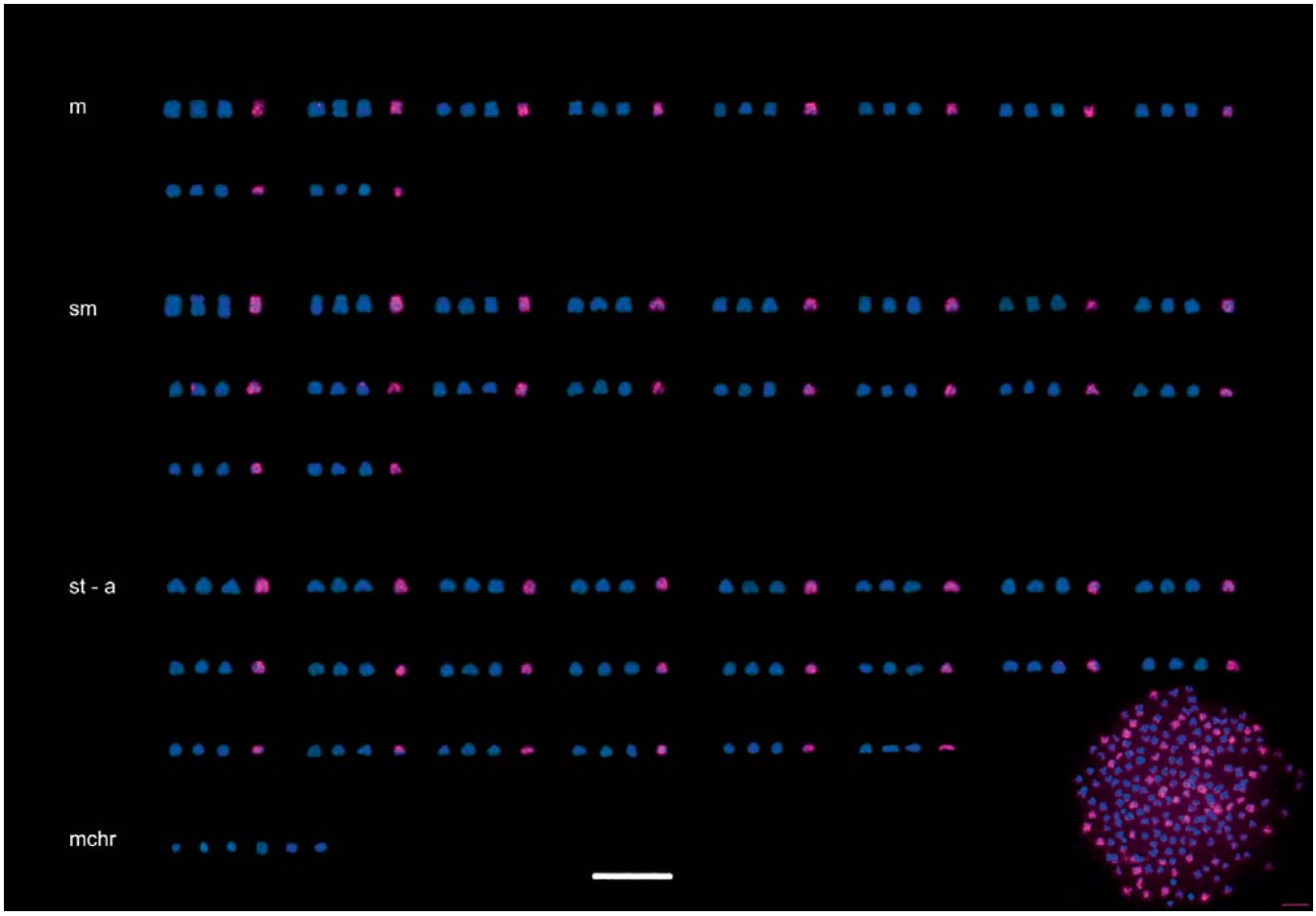


Fig. 3. Karyotype of an allotetraploid hybrid *Carassius* female with 206 pseudocolored chromosomes, DAPI (blue) and Cy3 filter (red). Scale bar = 10 μm.

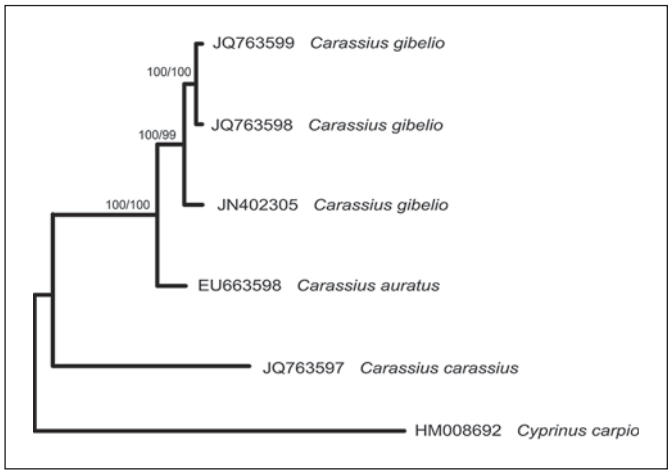


Fig. 4. Phylotree: reconstructed phylogeny of the cytochrome *b* haplotype sequences. The numbers at the nodes represent statistical supports for Bayesian and maximum parsimony analyses, respectively. As outgroup, the sequence of *Cyprinus carpio* was used.

generic hybrid from China and the cytogenetic analysis clearly showed the ratio 3 *Carassius*:1 *Cyprinus* genomes combination in the tetraploid fish. The ratio of representative genomes *(C. gibelio:C. carpio)* in genome of allotetraploid hybrids is the same as in our study *(C. gibelio:C. carassius)*, thus reproducing mechanism should be at least similar in both cases.

Besides the report of Zhu and Gui [2006], our finding proved that gynogenetic triploid females of *C. gibelio* have not only the capability to maintain all chromosomes, but also the ability to elevate ploidy level via sperm incorporation.

The molecular evidence for an occurrence of natural hybrids between *C. gibelio* and *C. carassius* was already described by Papoušek et al. [2008], but all of them were considered as diploids (i.e. with 100 chromosomes) with the assumption that only diploid biotypes of *C. gibelio* could have hybridized with *C. carassius* following the model of *C. auratus* and *C. carassius* hybrids in Great Britain [Hänfling et al., 2005].

The explanation of allopolyploidization in gynogenetic Prussian carp can be attributed to the dual reproduction modes, i.e. gynogenesis and sexual reproduction as it has been described in Chinese population of polyploid biotypes of the genus *Carassius*. These 2 modes are based on recognition of homologous and heterologous sperm by ovum. When heterologous sperm enters the egg, the gynogenesis takes place; the entered sperm does not decondense until the first cleavage and triggers embryogenesis. Contrary, when a homologous sperm enters the egg, the responding development mode is sexual reproduction; the entered sperm decondenses and forms a male pronucleus that fuses with the female pronucleus. The established zygote undergoes recombination and removes approximately half of the maternal chromosomes from the egg or dissolves them in cytoplasm [Zhou et al., 2000; Gui and Zhou, 2010].

It seems that the ability of sperm recognition and sperm genome elimination is not flawless, and possible mistakes can result in the presence of a certain number of tetraploid individuals within the triploid and diploid biotypes of the genus *Carassius*. Although the tetraploids are rare in the natural population, they are regularly recorded from various sites in Europe, e.g. Abramenko and Kravchenko [1998], Halačka and Lusková [2000], Halačka et al. [2003], Toth et al. [2005], Liasko et al. [2010], and Mezhzheryn et al. [2012]. This raises the question of what percentage of tetraploid Prussian carps can be of allopolyploid origin, since the genome composition is not usually investigated, and fish with different ploidy levels are morphologically indistinguishable [Vasileva, 1990].

It remains unknown whether such a genome addition mechanism is associated with phylogenetically closely related *Carassius* and *Cyprinus* genomes, or another mechanism is involved, for example that a certain amount of asexual females are available for genome addition as it was described by Choleva et al. [2012], following the idea that polyploidy is not a trigger of clonality, but rather a consequence.

Acknowledgements

We thank M. Rábová and M. Pokorná for their help with the preparation of karyotypes, and we are grateful to Tomáš Daněk for the valuable information about the locality. This study was supported by the S-grant MŠMT and the project No. P506/11/P596 of the Grant Agency of the Czech Republic.

References

Abramenko MI, Kravchenko OV: Chromosome mosaicism in somatic cells of fish from genus Carassius (Pisces: Cyprinidae). Cytogenet Cell Genet 81:123 (1998).

Arai K, Ikeno M, Suzuki R: Production of androgenetic diploid loach *Misgurnus anguillicaudatus* using spermatozoa of natural tetraploids. Aquaculture 137:131–138 (1995).

Choleva L, Janko K, De Gelas KD, Bohlen J, Šlechtová V, et al: Synthesis of clonality and polyploidy in vertebrate animals by hybridization between two sexual species. Evolution 66:2191–2203 (2012).

Cremer M, Grasser F, Lanctôt Ch, Müller S, Neusser M, et al: Multicolor 3D fluorescence in situ hybridization for imaging interphase chromosomes. Methods Mol Biol 463:205–239 (2008).

Daněk T, Kalous L, Veselý T, Krásová E, Reschová S, et al: Massive mortality of Prussian carp *Carassius gibelio* in the upper Elbe basin associated with herpesviral hematopoietic necrosis (CyHV-2). Dis Aquat Organ 102:87–95 (2012).

Dawley RM: An introduction to unisexual vertebrates, in Dawley RM, Bogart JP (eds): Evolution and Ecology of Unisexual Vertebrates, Bulletin 466, pp 1–28 (New York State Museum, Albany 1989).

Golovinskaya KA, Romashov DD: Investigations of the gynogenesis in *Carassius auratus gibelio,* Tr. Vsesoyuz. NII prud. ryb. khoz-va 4:73–113 (1947).

Gui JF, Zhou L: Genetic basis and breeding application of clonal diversity and dual reproduction modes in polyploid *Carassius auratus gibelio*. Sci China Life Sci 53:409–415 (2010).

Halačka K, Lusková V: Polyploidy in silver crucian carp *(Carassius auratus)* in lower reaches of Dyje river – determination according to erythrocyte nuclear sizes, in Mikešová J (ed): Sborník referátů IV. České ichthyologické konference, pp 110–113 (Jihočeská univerzita v Českých Budějovicích, Výzkumný ústav rybářský a hydrobiologický, Vodňany 2000).

Halačka K, Lusková V, Lusk S: *Carassius* '*gibelio*' in fish communities of the Czech Republic. Ecohydrol Hydrobiol 3:133–138 (2003).

Hall TA: BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series 41:95–98 (1999).

Hänfling B, Bolton P, Harley M, Carvalho GR: A molecular approach to detect hybridisation between crucian carp (*Carassius carassius*) and non-indigenous carp species (*Carassius* spp. and *Cyprinus carpio*). Freshwater Biol 50: 403–417 (2005).

Huelsenbeck JP, Ronquist F: MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17:754–755 (2001).

Janko K, Bohlen J, Lamatsch DK, Flajšhans M, Epplen JT, et al: The gynogenetic reproduction of diploid and triploid hybrid spined loaches (Cobitis: Teleostei), and their ability to establish successful clonal lineages–on the evolution of polyploidy in asexual vertebrates. Genetica 131:185–194 (2007).

Kalous L, Knytl M: Karyotype diversity of the offspring resulting from reproduction experiment between diploid male and triploid female of silver Prussian carp, *Carassius gibelio* (Cyprinidae, Actinopterygii). Folia Zool 60: 115–121 (2011).

Kalous L, Šlechtová V jr, Bohlen J, Petrtýl M, Švátora M: First European record of *Carassius langsdorfii* from the Elbe basin. J Fish Biol 70: 132–138 (2007).

Kalous L, Bohlen J, Rylková K, Petrtýl M: Hidden diversity within the Prussian carp and designation of a neotype for *Carassius gibelio* (Teleostei: Cyprinidae). Ichthyol Explor Freshwaters 23:11–18 (2012).

Kottelat M, Freyhof J: Handbook of European Freshwater Fishes (Publications Kottelat, Cornol, Freyhof, Berlin 2007).

Lamatsch DK, Stöck M: Sperm-dependent parthenogenesis and hybridogenesis in teleost fishes, in Schön I, Martens K, van Dijk P (eds): Lost Sex: The Evolutionary Biology of Parthenogenesis, pp 399–432 (Springer, Dordrecht 2009).

Levan A, Fredga K, Sandberg AA: Nomenclature for centromeric position on chromosomes. Hereditas 52:201–220 (1964).

Liasko R, Liousia V, Vrazeli P, Papiggioti O, Chortatou R, et al: Biological traits of rare males in the population of *Carassius gibelio* (Actinopterygii: Cyprinidae) from Lake Pamvotis (north-west Greece). J Fish Biol 77:570–584 (2010).

Liu SJ: Distant hybridization leads to different ploidy fishes. Sci China Life Sci 53:416–425 (2010).

Loewe L, Lamatsch DK: Quantifying the threat of extinction from Muller's ratchet in the diploid Amazon molly *(Poecilia formosa)*. BMC Evol Biol 8:88 (2008).

Lusková V, Lusk S, Halačka K, Vetešník L: *Carassius auratus gibelio* – the most successful invasive fish in waters of the Czech Republic. Russian Journal of Biol Invasions 1:176–180 (2010).

Mezhzheryn SV, Kokodyĭ SV, Kulysh AV, Verlat'iĭĭ DB, Fedorenko LV: Hybridization of crucian carp *Carassius carassius* (Linnaeus, 1758) in Ukrainian reservoirs and genetic structure of hybrids. Tsitol Genet 46:37–46 (2012).

Neaves WB, Bauman P: Unisexual reproduction among vertebrates. Trends Genet 27:81–88 (2011).

Papoušek I, Vetešník L, Halačka K, Lusková V, Humpl M, Mendel J: Identification of natural hybrids of gibel carp *Carassius auratus gibelio* (Bloch) and crucian carp *Carassius carassius* (L.) from lower Dyje River floodplain (Czech Republic). J Fish Biol 72:1230–1235 (2008).

Peňáz M, Ráb P, Prokeš M: Cytological analysis, gynogenesis and early development of *Carassius auratus gibelio*. Acta Sci Nat 13:1–33 (1979).

Ráb P, Roth P: Cold-blooded vertebrates, in Balíček P, Forejt J, Rubeš J (eds): Methods of Chromosome Analysis, pp 115–124. (Czechoslovak Biological Society Publishers, Brno 1988).

Rylková K, Kalous L, Šlechtová V, Bohlen J: Many branches, one root: first evidence for a monophyly of the morphologically highly diverse goldfish *(Carassius auratus)*. Aquaculture 302:36–41 (2010).

Šlechtová V, Bohlen J, Freyhof J, Ráb P: Molecular phylogeny of the Southeast Asian freshwater fish family Botiidae (Teleostei: Cobitoidea) and the origin of polyploidy in their evolution. Mol Phylogenet Evol 39:529–541 (2006).

Swofford DL: PAUP: phylogenetic analysis using parsimony (and other methods), Version 4 (Sinauer Associates, Sunderland 2000).

Szczerbowski JA: *Carassius auratus* (Linnaeus, 1758) 5–41(III), in: Banarescu PM, Paepke H-J (eds): The Freshwater Fishes of Europe, Cyprinidae 2/III and Gasterosteidae, pp 1–305 (AULA-Verlag, Wiesbaden 2002).

Tóth B, Várkonyi E, Hidas A, Meleg EE, Váradi L: Genetic analysis of offspring from intra- and interspecific crosses of *Carassius auratus gibelio* by chromosome and RAPD analysis. J Fish Biol 66:784–797 (2005).

Vasileva ED: On morphological divergence of gynogenetical and bisexual forms of *Carassius auratus* (Cyprinidae, Pisces). Zool Zh 69:97–110 (1990).

Vrijenhoek RC: Animal clones and diversity. Are natural clones generalists or specialists? BioScience 48:617–628 (1998).

Wheeler A, Easton K: Hybrids of chub and roach *(Leuciscus cephalus* and *Rutilus rutilus)* in English rivers. J Fish Biol 12:167–171 (1978).

Wouters J, Janson S, Lusková V, Olsén KH: Molecular identification of hybrids of the invasive gibel carp *Carassius auratus gibelio* and crucian carp *Carassius carassius* in Swedish waters. J Fish Biol 80:2595–2604 (2012).

Yamashita M, Jiang J, Onozato H, Nakanishi T, Nagahama Y: A tripolar spindle formed at meiosis I assures the retention on the original ploidy in the gynogenetic triploid crucian carp, Ginbuna *Carassius auratus langsdorfii*. Develop Growth Differ 35:631–636 (1993).

Yi MS, Li YQ, Liu JD, Zhou L, Yu QX, Gui JF: Molecular cytogenetic detection of paternal chromosome fragments in allogynogenetic gibel carp, *Carassius auratus gibelio* Bloch. Chromosome Res 11:665–671 (2003).

Zhou L, Wang Y, Gui JF: Genetic evidence for gonochoristic reproduction in gynogenetic silver crucian carp (*Carassius auratus gibelio* Bloch) as revealed by RAPD assays. J Mol Evol 51:498–506 (2000).

Zhu H-P, Gui JF: Identification of genome organization in the unusual allotetraploid form of *Carassius auratus gibelio*. Aquaculture 265: 109–117 (2006).

Sember et al. BMC Evolutionary Biology (2015) 15:251
DOI 10.1186/s12862-015-0532-9

BMC Evolutionary Biology

RESEARCH ARTICLE Open Access

Karyotype differentiation in 19 species of river loach fishes (Nemacheilidae, Teleostei): extensive variability associated with rDNA and heterochromatin distribution and its phylogenetic and ecological interpretation

Alexandr Sember[1,2*], Jörg Bohlen[1], Vendula Šlechtová[1], Marie Altmanová[1,3], Radka Symonová[1,4] and Petr Ráb[1]

Abstract

Background: Loaches of the family Nemacheilidae are one of the most speciose elements of Palearctic freshwater ichthyofauna and have undergone rapid ecological adaptations and colonizations. Their cytotaxonomy is largely unexplored; with the impact of cytogenetical changes on this evolutionary diversification still unknown. An extensive cytogenetical survey was performed in 19 nemacheilid species using both conventional (Giemsa staining, C- banding, Ag- and Chromomycin A_3/DAPI stainings) and molecular (fluorescence in situ hybridization with 5S rDNA, 45S rDNA, and telomeric $(TTAGGG)_n$ probes) methods. A phylogenetic tree of the analysed specimens was constructed based on one mitochondrial (*cytochrome b*) and two nuclear (*RAG1*, *IRBP*) genes.

Results: Seventeen species showed karyotypes composed of 2n = 50 chromosomes but differentiated by fundamental chromosome number (NF = 68–90). *Nemachilichthys ruppelli* (2n = 38) and *Schistura notostigma* (2n = 44–48) displayed reduced 2n with an elevated number of large metacentric chromosomes. Only *Schistura fasciolata* showed morphologically differentiated sex chromosomes with a multiple system of the XY_1Y_2 type. Chromomycin A_3 (CMA_3)- fluorescence revealed interspecific heterogeneity in the distribution of GC-rich heterochromatin including its otherwise very rare association with 5S rDNA sites. The 45S rDNA sites were mostly located on a single chromosome pair contrasting markedly with a pattern of two (*Barbatula barbatula*, *Nemacheilus binotatus*, *N. ruppelli*) to 20 sites (*Physoschistura* sp.) of 5S rDNA. The cytogenetic changes did not follow the phylogenetic relationships between the samples. A high number of 5S rDNA sites was present in species with small effective population sizes.

(Continued on next page)

* Correspondence: alexandr.sember@seznam.cz
[1]Laboratory of Fish Genetics, Institute of Animal Physiology and Genetics, Czech Academy of Sciences, Rumburská 89, Liběchov 277 21, Czech Republic
[2]Department of Genetics and Microbiology, Faculty of Science, Charles University in Prague, Viničná 5, 128 44 Prague 2, Czech Republic
Full list of author information is available at the end of the article

(Continued from previous page)

Conclusion: Despite a prevailing conservatism of 2n, Nemacheilidae exhibited a remarkable cytogenetic variability on microstructural level. We suggest an important role for pericentric inversions, tandem and centric fusions in nemacheilid karyotype differentiation. Short repetitive sequences, genetic drift, founder effect, as well as the involvement of transposable elements in the dispersion of ribosomal DNA sites, might also have played a role in evolutionary processes such as reproductive isolation. These remarkable dynamics of their genomes qualify river loaches as a model for the study of the cytogenetic background of major evolutionary processes such as radiation, endemism and colonization of a wide range of habitats.

Keywords: Fish cytotaxonomy, Karyotype variability vs. 2n uniformity, FISH, 45S - 5S ribosomal genes, Chromosome banding, Pericentric inversion, Robertsonian translocation, Effective population size

Background

Cypriniformes, the largest order of freshwater fishes globally, is composed of two highly diverse Palearctic superfamilies – Cyprinoidea and Cobitoidea [1, 2]. Cobitoidea, or "loaches", are a group of small benthic fishes which are one of the most common elements of Eurasian freshwater ichthyofauna. To date, Cobitoidea includes about 1100 species, currently recognized in ten families [3], and yet only representatives of Cobitidae, Botiidae, Catostomidae and Vaillantellidae have so far been studied cytogenetically. Several cases of highly diverse karyotypes and polyploidy have been discovered in the first three families, although not in Vaillantellidae [4]. The Botiidae family consists of two subfamilies differing in ploidy levels (one diploid and one tetraploid) [5]. In Cobitidae, several independent polyploidization events occurred [6, 7], in some cases after hybridization, leading to an asexual mode of reproduction [8–10]. From these limited data we can see that cytogenetic changes might have played an important role in the evolution of loaches and it remains an open question as to whether this is also true for the remaining cobitoid lineages.

With nearly 600 recognized species in 46 genera [3], Nemacheilidae, or "river loach", represents the most diverse family of loach fishes, as well as being the most widespread with a distribution area ranging continuously from Portugal to Japan, and from most Siberian rivers to Java [11]. Importantly, river loaches are also very abundant within this enormous distribution area, occurring in virtually all rivers in Europe and Asia. On the other hand, their distribution pattern varies considerably; while some species are geographically very restricted, others are widely distributed, a feature often found even within the same genus, e.g., *Schistura* [3]. Additionally, Nemacheilidae have colonized an unusual variety of habitats including standing swamps, torrential rapids, major rivers, small forest streams, caves and lakes. Their ecological diversity is further illustrated by them being both the highest (above sea level) and the lowest (below ground level) freshwater fish in the world [3]. All these characteristics make Nemacheilidae a vital model for evolutionary study and our candidate group with which to evaluate the impact of cytogenetic changes on their diversity.

Despite the vast biodiversity within Nemacheilidae, the cytogenetics and cytotaxonomy of this group remain poorly explored. Giemsa-stained chromosomes have been studied in only 24 species [7, 12–14] and banding techniques were performed solely in the single species *Barbatula barbatula* [15] while no molecular cytogenetics had previously been applied. From this limited data, karyotypes of most analysed species display the stable diploid chromosome number 2n = 50, while interspecific karyotype variability in the number of chromosomal arms (Nombre Fundamental, NF) is apparent (see, e.g., [15–17]). In some species, intraspecific variability in 2n and karyotype composition has also been documented [17–20]. Polyploidy has been recorded only in one species *B. 'barbatula'* (2n = 3x = 75) [21]. The scarce available data does indicate the extensive but unexplored cytogenetic diversity of nemacheilid loaches.

The aim of this study is to assess cytogenetic variability within the Nemacheilidae family using conventional and molecular chromosome markers and to evaluate these data with regards to the evolutionary processes behind morphological and ecological diversification. A representative sampling of 19 species from eleven genera were used to investigate karyotypes, heterochromatin distribution and chromosomal characteristics of both rDNA classes and (in some cases) the telomeric sequence motif $(TTAGGG)_n$. All cytogenetic characteristics were mapped onto a phylogenetic tree based on molecular analyses of one mitochondrial and two nuclear genes.

Methods

Animals

Fifty-two individuals belonging to 19 different nemacheilid species were analysed (Table 1). Their distribution areas are specified in Fig. 1 and references for taxonomic identification are given in Additional file 1: Supplementary Methods 1. All analysed specimens were obtained from

Table 1 Species under study, their sex, origin and geographical distribution

Species	Individuals	Source (country, province, river basin)	Distribution
Barbatula barbatula (Linnaeus, 1758)	3	Czech Republic, Středočeský kraj, Elbe	widespread (Europe, Asia)
Lefua costata (Kessler, 1876)	2♀	Republic of Korea, Gangwon, Geojin	widespread (Korea, China)
Mesonoemacheilus guentheri (Day, 1867)	1♂, 1♀	Ornamental fish trade	moderately widespread (southern India)
Nemacheilus binotatus (Smith, 1933)	1♂, 1♀	Ornamental fish trade	moderately widespread (Thailand)
Nemachilichthys ruppelli (Sykes, 1839)	1♂, 1♀	Ornamental fish trade	moderately widespread (southern India)
Paracanthocobitis pictilis (Kottelat, 2012)	2♀	Ornamental fish trade	endemic to Ataran river (Myanmar)
Paracanthocobitis zonalternans (Blyth, 1860)	1 ♂1♀	Myanmar, no details known	widespread (Bangladesh to Malaysia)
Petruichthys brevis (Boulenger, 1893)	1♂, 1♀ + 1	Ornamental fish trade	endemic to Inle Lake (Myanmar)
Physoschistura elongata (Sen & Nalbant, in Singh, Sen, Bănărescu & Nalbant, 1982)	2	Ornamental fish trade	endemic to Shilling county (northeast India)
Physoschistura sp.	2	Myanmar, Shan, Salween	endemic to surrounding of Inle Lake (Myanmar)
Pteronemacheilus lucidorsum (Bohlen & Šlechtová, 2011)	1♂, 1♀	Myanmar, Shan, Irrawaddy	endemic to upper Myitnge river basin (Myanmar)
Schistura bolavenensis (Kottelat, 2000)	3	Laos: Champasak, Mekong	moderately spread (Bolaven plateau, Laos)
Schistura corica (Hamilton, 1822)	1♂, 3♀	Ornamental fish trade	widespread (northern India, Bangladesh)
Schistura fasciolata (Nichols and Pope, 1927)	2♂, 1♀	Ornamental fish trade	widespread (southern China and northern Vietnam)
Schistura hypsiura (Bohlen, Šlechtová & Udomritthiruj, 2014)	1♂, 1♀ +3	Ornamental fish trade	endemic to southern Rakhine state (Myanmar)
Schistura notostigma (Bleeker, 1863)	6	Ornamental fish trade	endemic (Sri Lanka)
Schistura pridii (Vidthayanon, 2003)	2	Ornamental fish trade	local endemic (northern Thailand)
Schistura savona (Hamilton, 1822)	3	Ornamental fish trade	widespread (northern India, Bangladesh)
Seminemacheilus lendlii (Hankó, 1924)	1♂, 1♀	Turkey, Anatolia, no details known	endemic to southeast Anatolia (Turkey)

ornamental fish trade, from a commercial fish farm or from private aquarium fish breeders. All experimental procedures involving fishes were approved by the Institutional Animal Care and Use Committee of the IAPG AS CR, according with directives from the State Veterinary Administration of the Czech Republic, permit number 217/2010, and by permit number CZ 02386 from the Ministry of Agriculture of the Czech Republic. Voucher specimens are deposited to the fish collection of the Laboratory of Fish Genetics, IAPG, CAS, Liběchov.

Chromosome preparation and analysis of constitutive heterochromatin

Mitotic chromosomes were obtained from regenerating fin tissue by the technique described by Völker et al. [22] and Völker and Ráb [23], with slight modifications (see Additional file 1: Supplementary Methods 2). For conventional cytogenetic analysis, chromosomes were stained with 5 % Giemsa solution (pH 6.8) (Merck, Darmstadt, Germany). Selected slides were destained in methanol:acetic acid fixation (see above) and re-used for the other techniques. For fluorescence in situ hybridization (FISH), slides were dehydrated in an ethanol series (70, 80 and 96 %, 3 min each) and stored in a freezer (−20 °C). Visualization of the constitutive heterochromatin was done by C-banding according to Haaf and Schmid [24] using 4′,6-diamidino-2-phenolindole (DAPI) (Sigma, St. Louis, MO, USA) counterstaining. Fluorescence staining was performed sequentially or in separate experiments by GC-specific fluorochrome Chromomycin A_3 (CMA_3) (Sigma-Aldrich) [25] and AT-specific fluorochrome DAPI (Sigma-Aldrich) [26], following Mayr et al. [27] and Sola et al. [28]. In *P. elongata*, a silver staining technique was employed according to Howell and Black [29]. At least ten metaphases per specimen were analysed, in some cases sequentially. In a few cases, metaphases with incomplete 2n were selected (see Figs. 6i and 7a; Additional file 2: Figure S1K), but were sufficient enough to present the required features. Chromosome morphology was classified according to Levan et al. [30], but modified as m – metacentric, sm – submetacentric, st – subtelocentric, a – acrocentric, where st and a chromosomes were scored as uniarmed, together in one category.

DNA isolation and probe preparation

Whole genomic DNA was extracted from fin tissue using the conventional phenol-chloroform-izoamylalcohol method [31] using PhaseLock Eppendorf tubes (5PRIME,

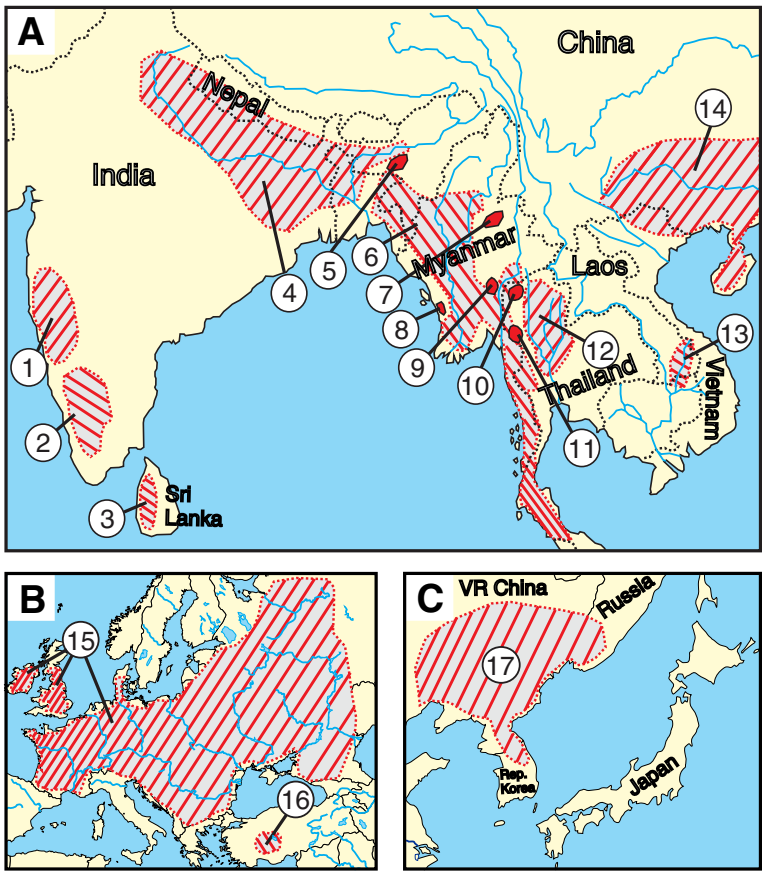


Fig. 1 Distribution areas of the investigated species of Nemacheilidae. **a** Asia, **b** Europe, **c** China. 1 – *N. ruppelli*, 2 – *M. guentheri*, 3 – *S. notostigma*, 4 – *S. corica* and *S. savona*, 5 – *P. elongata*, 6 – *P. zonalternans*, 7 – *P. lucidorsum*, 8 – *S. hypsiura*, 9 – *P. sp.* and *P. brevis*, 10 – *S. pridii*, 11 – *P. pictilis*, 12 – *N. binotatus*, 13 – *S. bolavensis*, 14 – *S. fasciolata*, 15 – *B. barbatula*, 16 – *S. lendlii*, 17 – *L. costata*

Gaithersburg, USA) to prevent protein contamination, or the Qiagen DNAeasy Blood & Tissue Kit (Qiagen, Hilden, Germany). rDNA fragments were obtained by polymerase chain reaction (PCR) using previously described primers (see Additional file 3: Table S1; for PCR conditions see Additional file 1: Supplementary Methods 3). The resulting PCR products were purified using QIAquick PCR purification Kit (Qiagen), with multiple bands being electrophoresed in 0.8 % agarose gels and purified using QIAquick Gel Extraction Kit (Qiagen). DNA fragments were cloned to pDrive Cloning Vector (Qiagen) and transformed into QIAGEN EZ Competent Cells (Qiagen). Selected recombinant plasmids were isolated by QIAprep Spin Miniprep Kit (Qiagen) and sequenced in both strands by Macrogen (South Korea, Netherlands). Chromatograms of obtained sequences were verified and assembled using SeqMan Pro 10.1.2 (LaserGene, DNASTAR, Madison, Wl.). The resulting consensus sequences were confirmed using NCBI BLAST/N analysis [32] and selected clones used to construct FISH probes.

Probes were labelled by PCR with biotin-16-dUTP (Roche, Mannheim, Germany) or digoxigenin-11-dUTP (Roche). For each slide 200 ng of 5S rDNA, 200 ng of 45S rDNA and 25 μg of sonicated salmon sperm DNA (Sigma-Aldrich) were added and the resulting probe precipitated in 96 % ethanol, washed in 70 % ethanol, air-dried and re-dissolved in hybridization buffer (50 % formamide, 10 % dextran sulphate, 2× SSC, 0.04 M $NaPO_4$ buffer, 0.1 % SDS, Denhardt reagens, see [33]) to give a final concentration of 25 ng/μl for each rDNA probe.

For telomeric FISH, non-templated PCR with primers $(TTAGGG)_5$ and $(CCCTAA)_5$ was carried out according to Ijdo et al. [34]. The amplified product was labelled using Nick Translation Mix (Abbot Molecular, Illinois, USA) with biotin-16-dUTP, taking 3–4 h to reach optimal probe size (100–500 bp).

FISH analysis

FISH was carried out according to Cremer et al. [35] with several modifications. Briefly, dehydration in an

ethanol series (70, 80 and 96 %, 3 min each) was followed by thermal aging for 1–2 h at 37 °C and 30 min at 60 °C. Prior to hybridization, the chromosomes were treated with RNase A (200 μg/ml in 2× SSC) (Sigma-Aldrich) for 90 min at 37 °C in a humid chamber and digested with pepsin (50 μg/ml in 10 mM HCl, 3 min, 37 °C). Slides were subsequently denatured in 75 % formamide (pH 7.0) (Sigma-Aldrich) in 2× SSC at 74 °C for 3 min, and then immediately cooled and dehydrated in 70 % (cold), 80 % and 96 % (RT) ethanol. The hybridization mixture was denatured at 86 °C for 6 min and immediately chilled on ice for 10 min. 10–20 μl of probe mixture was applied to a denatured slide and hybridization was performed overnight at 37 °C in a dark humid chamber. Post-hybridization washes were done twice in 50 % formamide in 2× SSC (pH 7.0) at 42 °C for 5 min and three times in 1× SSC at 42 °C (7 min each) before equilibration washing in 2× SSC at RT for 20 s. Prior to probe detection 500 μl of 3 % BSA (Vector Labs, Burlington, Canada) in 4× SSC in 0,01 % Tween 20 was dropped onto the slide (at 37 °C for 20 min) as a blocking treatment. Probes were detected by Anti-Digoxigenin-Rhodamine (Roche) and Streptavidin-FITC (Invitrogen Life Technologies, San Diego, CA, USA) along with Anti-Digoxigenin-Fluorescein (Roche) and Streptavidin-Cy3 (Invitrogen Life Technologies) to exclude any artificial results (influenced e.g., by the type of applied antibody). Experiments with altered labelling (biotin for 45S and digoxigenin for 5S rDNA) were included to verify the observed patterns. All rDNA FISH pictures presented here are pseudocoloured in red for the 45S rDNA probe and in green for the 5S rDNA.

The slides were incubated with antibodies at 37 °C for 60 min in a dark humid chamber, washed four times (7 min each) in 4× SSC in 0.01 % Tween (pH 7.0) at 42 °C and the chromosomes then counterstained with DAPI in mounting medium (Cambio, Cambridge, United Kingdom), covered and sealed with a coverslip.

To enhance telomeric FISH signals, tyramid signal amplification (TSA) was performed using a kit with tyramide conjugated with Alexa 488 fluorochrome (Invitrogen Life Technologies).

After image processing FISH slides selected for fluorescence banding and/or C-banding were washed in 4× SSC in 0.01 % Tween (pH 7.0) and dehydrated in an ethanol series.

Microscopy and image analysis

Giemsa-stained chromosomes and FISH images were inspected using a Provis AX70 Olympus microscope with a standard fluorescence filter set. FISH images were captured under immersion objective 100× with a black and white CCD camera (DP30W Olympus) for each fluorescent dye using Olympus Acquisition Software. The digital images were then pseudocoloured (blue for DAPI, red for Rhodamine or Cy3, green for FITC or Alexa488) and superimposed with MicroImage software (Olympus, version 4.0). FISH karyotype images were optimized and arranged using Adobe Photoshop, version CS6. Karyotypes from Giemsa-stained and C-banded images were arranged in IKAROS (Metasystems) software.

Phylogenetic analyses

Phylogenetic hypothesis was based on the analyses of three molecular markers: mitochondrial *cytochrome b* (*cyt b*), *recombination-activating gene 1* (*RAG1*) and *interphotoreceptor retinoid-binding protein* (*IRBP*). The primers and PCR reaction protocols for *cyt b* and *RAG1* followed Šlechtová et al. [5, 36], and Chen et al. [37] for the *IRBP* amplification (for details, see Additional file 1: Supplementary Methods 4). The same sets of PCR primers were used for sequencing (summarized for all genes in Additional file 3: Table S1). All three genes were sequenced for each of the 39 analysed specimens of Nemacheilidae.

Chromatograms were edited and assembled using SeqMan Pro 10.1.2 (LaserGene, DNASTAR). The sequences were aligned in BioEdit 7.0.5.3 [38] and evaluated based on their amino acid translation.

Prior to the phylogenetic analyses, the congruence among the three gene partitions was assessed using the incongruence length difference (ILD) test [39] with 1000 replication as implemented in PAUP 4.0b10 [40]. Since the test did not reveal any significant conflict (see the Results), all three datasets could be concatenated into a single matrix.

Alignments of all three genes were concatenated into a single 2998 bp dataset (1124 bp of *cyt b*, 974 bp of *RAG1* and 900 bp of *IRBP*) and 40 individuals (39 Nemacheilidae plus 1 outgroup). All sequences but one (*cyt b* sequence of *Botia lohachata*) are original data and were deposited in GenBank [41] under the accession numbers [KP738491 - KP738609] (see Additional file 4: Table S2).

Phylogenetic analysis of the concatenated dataset was performed using the partitioned Bayesian inference in MrBayes 3.2.2 [42]. The dataset was partitioned by genes and codon positions, involving in total nine partitions. The analysis was set to six Metropolis Coupled Markov Chains Monte Carlo (MCMCMC) with default heating conditions, searching the tree space for 5 milion generations under the GTR + G + I settings for each partition, in two runs, starting with random trees and a sampling frequency of each 100 generations. The log-likelihood score distribution was examined to determine the burn-in values. The first 1000 trees were discarded as burn-in and the remaining

ones were used to build a 50 % majority rule consensus tree and statistical support of clades was assessed by posterior probabilities.

Results

Sequence analysis of *RAG1*, *IRBP* and *cyt b*

The *RAG1*, *cytochrome b* and *IRBP* datasets consisted of 974 (30 % of variable positions), 1124 (44 % v.p.) and 900 bp (35 % v.p.), respectively. The ILD test did not reject the null hypothesis about the homogeneity of any of the analysed datasets: $P = 0.94$ for *RAG1* vs. *cyt b*, $P = 0.71$ for *cyt b* vs. *IRBP* and $P = 0.14$ for *RAG1* vs. *IRBP*. Therefore the data were concatenated into a single dataset for the further analysis, altogether providing a dataset of 2998 bp.

In the final phylogeny all analysed species were identified as monophyletic and well-separated lineages. The topology shows a prominent basal split into one major clade that contains *Nemacheilus binotatus* from northern Thailand plus all samples from Myanmar, India, Sri Lanka and Turkey and a second major clade that is composed from all samples from China, Laos, Europe and Korea. Within the first major clade, four subclades are visible: the first containing *N. binotatus*, the second *Schistura savona* and both species of *Paracanthocobitis*, the third solely *Nemachilichthys ruppelli* and the fourth containing all remaining samples from the genera *Mesonoemacheilus, Schistura, Physoschistura, Seminemacheilus, Pteronemacheilus* and *Petruichthys*. Within the second major clade, three subclades are visible: the first containing *Lefua costata* from Korea, the second *B. barbatula* from Europe and the third with two species of *Schistura* from Laos and China.

Sequence analysis of 5S and 28S rDNA

PCR amplification of 28S rDNA resulted consistently in a fragment 300 bp in size, containing partial sequence of 28S rRNA coding region. Sequences for *P. elongata*, *S. bolavenensis, S. corica, S. fasciolata* as well as for *Botia almorhae* (from related family Botiidae) were deposited in GenBank [41] (see Additional file 5: Table S3). For 5S rDNA, a high degree of variability, both in length as well as in number of putative 5S rDNA fragments was observed among the analysed species, so sequenced fragments from *Esox lucius* (300 bp) and *B. almorhae* (500 bp) were used for constructing the FISH probe. The sequence of 5S rDNA fragment from *E. lucius* was verified in GenBank [EF514228]. The sequence of 5S rDNA from *B. almorhae* (deposited in GenBank; see Additional file 5: Table S3) contained a partial sequence of the 5S rDNA coding region (83 bp) and a putative NTS (non-transcribed spacer). For detailed analysis of nemacheilid 5S rDNA, we selected 200 and 600 bp PCR fragments from two specimens of *S. pridii*. Thirteen clones were sequenced and verified in BLAST/N and also searched against the Repbase database at the Genetic Information Research Institute (GIRI) [43] for the presence of transposable elements (TEs) or other repetitive sequences. Indeed, each cloned sequence contained - next to the 71 bp of the 5S rRNA gene coding region - a putative NTS (85 bp or 475 bp) containing a fragment (54 bp) of L1-2_DR non-long terminal repeat (non-LTR) retrotransposon (RTE) at the 3′end (Additional file 6: Figure S2). The differences between both PCR fragments were thus in the length of the putative NTS and in the distance of the RTE fragment from the 5S rRNA coding region. No such association between TEs and rDNA loci was observed in the 5S rDNA of *B. almorhae* or in the 28S rDNA fragments characterized in this study.

Cytogenetic characteristics

Figure 2 summarizes 2n, karyotype structure, NF and rDNA phenotypes (i.e., number and position of both major and minor rDNA sites) within the phylogenetic tree context analysis. Seventeen out of 19 species displayed karyotypes with uniform $2n = 50$, but with a marked variability in NF values (68–90) (Figs. 2, 3 and 5; Additional file 7: Figure S3). In the remaining two species, karyotypes with reduced 2n were observed: *N. ruppelli* ($2n = 38$) (Fig. 4a), *S. notostigma* ($2n = 44$ or 48) (Additional file 8: Figure S4A, C, E). Two different karyomorphs occured in examined individuals of the latter species – with $2n = 44$ (five individuals, Additional file 8: Figure S4A, C) and with $2n = 48$ (a single individual, Additional file 8: Figure S4E). Karyotypes of both species exhibited a significantly higher number of large m chromosomes compared to karyotypes with $2n = 50$. Except for the large m chromosomes in *N. ruppelli* (six pairs) and *S. notostigma* (one or two pairs), karyotypes in all other species were composed of comparatively small chromosomes, gradually decreasing in size. Very tiny chromosomes were observed in *L. costata*, *P. pictilis*, *P. zonalternans*, *S. hypsiura* and *S. savona.* Centromere positions often gradually differed making it difficult to establish strict borderlines between formal chromosomal categories.

In almost all species no intraspecific numerical or structural polymorphisms between males and females that might indicate the presence of sex chromosomes were detected, although only females were examined in *L. costata* and *P. pictilis* and unsexed specimens in 7 other species – see Table 1. However, in *S. fasciolata*, males exhibited $2n = 51$ chromosomes with a karyotype composed of (9 m + 20sm + 22st-a) while a female presented a $2n = 50$ (8 m + 20sm + 22st-a), suggesting the presence of a multiple XY_1Y_2 sex chromosome system (Fig. 5a, c).

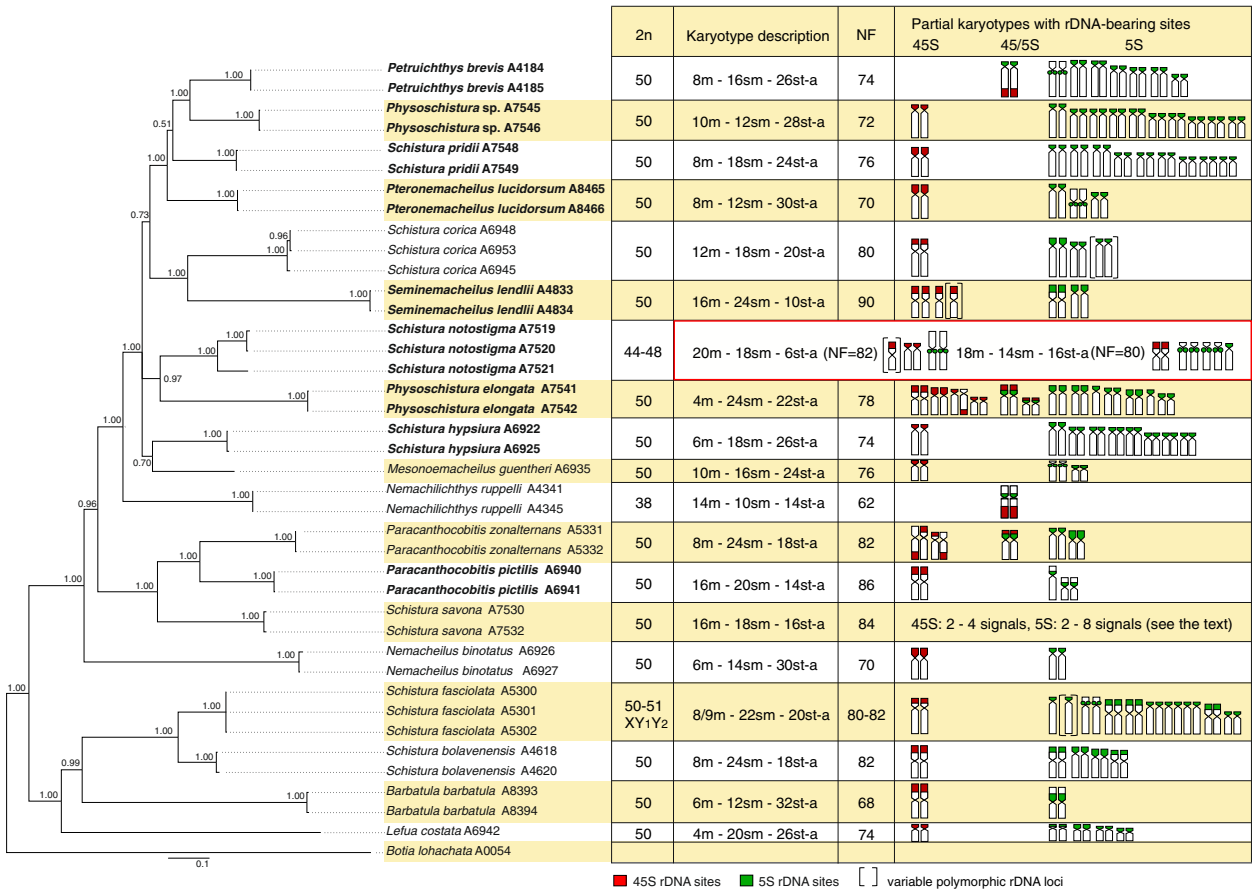


Fig. 2 Phylogenetic relationships and karyotype characteristics of inspected nemacheilids. 2n, karyotype characteristics, FN, 45S and 5S rDNA patterns are plotted onto phylogenetic tree obtained by Bayesian analysis based on the mitochondrial (*cytochrome b*) and nuclear (*RAG1*, *IRBP*) genes. Idiograms represent partial karyotypes with chromosomes bearing 45S rDNA (*red signals*) and 5S rDNA (*green signals*). Polymorphic rDNA sites are in brackets. Note higher numbers of 5S rDNA sites in the majority of endemic species (whose taxonomic names are in bold italics). Note: in *P. elongata*, only the karyotype version from one individual is presented, to avoid confusion due to the high number and variability of rDNA sites in this species

Heterochromatin distribution and composition

The distribution of constitutive heterochromatin was studied by CDD (CMA_3/DAPI) banding in all species and C-banding in a subset of 10 species (Additional file 9: Table S4). The C-banding and DAPI patterns were usually congruent with the exceptions observed in *N. ruppelli* and *S. bolavenensis*, where also some CMA_3-positive (CMA_3^+) regions (NOR-associated) were slightly positively heteropycnotic after C-banding at the same time (Fig. 4c; Additional file 2: Figure S1L). In the remaining species, the CMA_3^+ regions did not match the C-bands. With the exception of *N. ruppelli* and *S. lendlii* (Fig. 4c; Additional file 2: Figure S1Q) all other species displayed generally low or moderate levels of AT-rich C-heterochromatin. In almost all species, its predominant location was in the pericentromeric regions of some or all chromosomes, except for *S. corica*, where only a few interstitial bands and two whole-arm heterochromatic regions (p-arms, sm) were apparent (Additional file 2: Figure S1M). In one or more chromosomal pairs of m-sm type in *M. guentheri, P.* sp., *S. bolavenensis*, *S. hypsiura* and *S. lendlii* (Additional file 2: Figure S1D, I, L, N, Q) the heterochromatin encompasses a substantial part or even the entire arm of the chromosome. These regions were adjacent to 5S or 45S rDNA only in *S. bolavenensis*, *S. corica* and *S. lendlii*. Few heterochromatic p-arms of st-a chromosomes were observed in *L. costata, M. guentheri, N. ruppelli*, *P. zonalternans* and *S. pridii* (Fig. 4c, Additional file 2: Figure S1B, C, D, G, O). Huge heterochromatic regions were found flanking the primary constrictions of m chromosomes in *N. ruppelli* (six pairs, Fig. 4c), *S. notostigma* (one or two pairs, Additional file 8: Figure S4C), *S. savona* (one pair, Additional file 2: Figure S1P) as well as in one (male) or two (female) st chromosomes in *S. fasciolata* (Fig. 5d). In the latter species, compared to Giemsa-stained karyotypes,

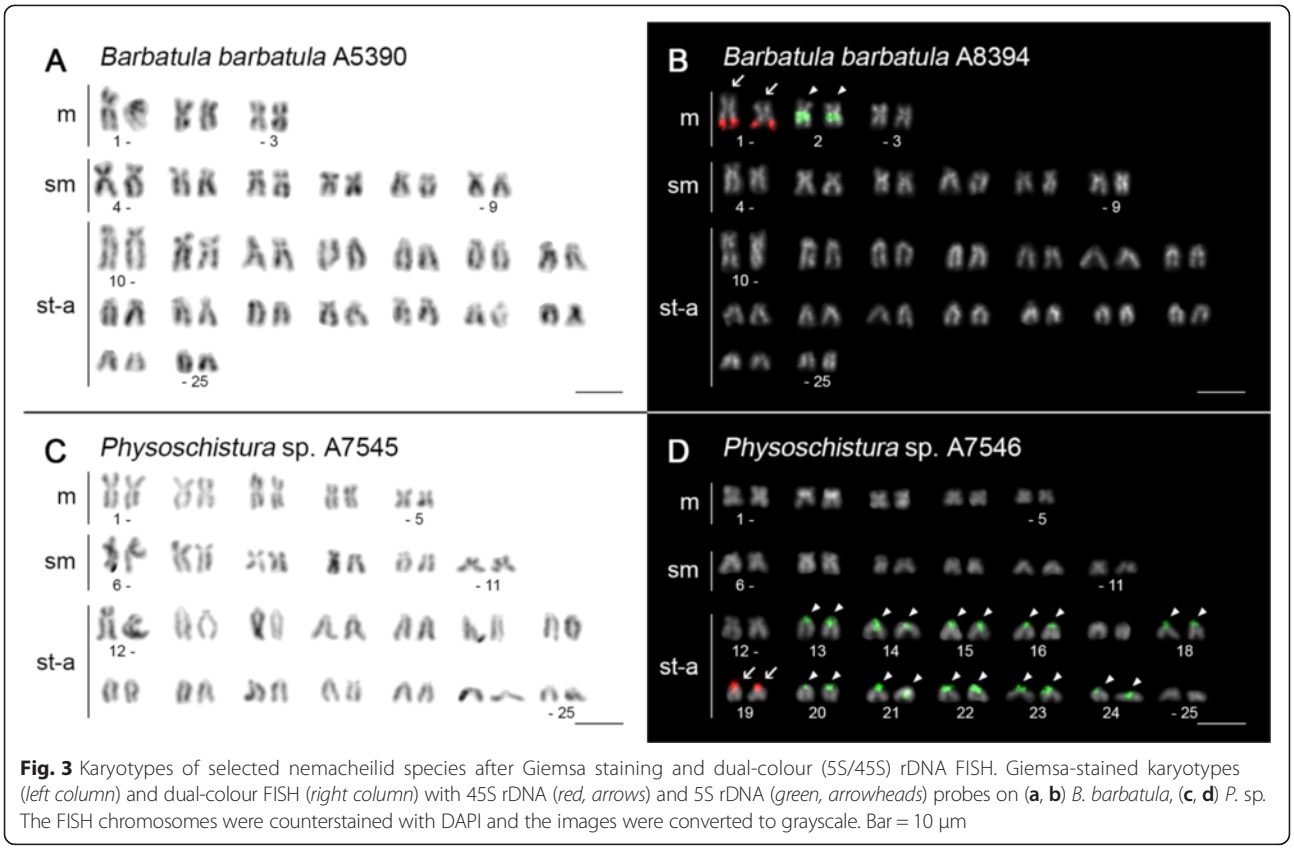


Fig. 3 Karyotypes of selected nemacheilid species after Giemsa staining and dual-colour (5S/45S) rDNA FISH. Giemsa-stained karyotypes (*left column*) and dual-colour FISH (*right column*) with 45S rDNA (*red, arrows*) and 5S rDNA (*green, arrowheads*) probes on (**a**, **b**) *B. barbatula*, (**c**, **d**) *P.* sp. The FISH chromosomes were counterstained with DAPI and the images were converted to grayscale. Bar = 10 μm

this heterochromatic region was confined to the st chromosome present on one homologue in males and on both homologues in females. Also noticeable was the C-heterochromatic block on the male-specific single large m chromosome. Furthermore, intercalar DAPI-positive bands were clearly visible after C- or CDD banding in a subset of sm/st chromosomes (from one to four pairs) in *B. barbatula*, *N. ruppelli*, *N. binotatus*, *S. bolavenensis*, *S. corica* and *S. notostigma*, often appearing as dot-like sites located proximally on the q arms. Finally, a polymorphic AT-rich p-arm was observed in one homologue of pair 18 in *M. guentheri,* but only in the male karyotype (Additional file 7: Figure S3D).

CMA_3 labelled only GC-rich regions associated exclusively with NORs in seven species (*L. costata*, *P. brevis*, *P.* sp., *S. fasciolata*, *S. pridii*, *S. savona* and *S. lendlii*), but also with 5S rDNA regions in seven other nemacheilids (Additional file 9: Table S4). More specifically, species with only some 5S rDNA sites being CMA_3^+ (e.g., *M. guentheri*, *P. zonalternans*, *P. elongata*) (Fig. 6b, g; Additional file 10: Figure S5B) and others with all of them (e.g., *B. barbatula*, *N. ruppelli*, *S. notostigma*) (Figs. 4b and 6a, i). In six out of seven species, we observed the 5S rDNA/CMA_3^+ pattern directly by sequential application of CDD banding and rDNA FISH (Fig. 6b, g-i) In *S. corica* (Fig. 6f), however, a similar conclusion was based on observation of remarkably high number of CMA_3^+ sites and their distribution in centromeres and chromosomal p-arms, similarly to 5S rDNA sites. In *P. pictilis*, *P. lucidorsum* and *S. hypsiura*, association of CMA_3^+ and 5S rDNA sites was inconclusive (Fig. 6d; Additional file 10: Figure S5D, F). In *N. binotatus* and *S. bolavenensis*, CMA_3 labelled NORs and some other regions non-related to 5S rDNA (Fig. 6c, h). A more complicated pattern was observed in *P. elongata* (with a subset of CMA_3^+ 5S rDNAs and additional CMA_3^+ regions) and *S. notostigma* (where all 5S rDNAs were CMA_3^+ and other CMA_3^+ regions also appeared) (Fig. 6g, i; Additional file 10: Figure S5G). Finally, *S. corica* displayed an extensive dispersal of CMA_3^+ regions with locations in all centromeres, some p-arms and along the single pair of NOR (Fig. 6f).

rDNA phenotypes

All karyotypes resulting from the rDNA FISH experiments are shown in Figs. 3, 4 and 5; Additional file 7: Figure S3, Additional file 8: Figure S4B, D, F and Additional file 11: Figure S6B, D and partial idiograms showing rDNA phenotypes in the phylogenetic context are summarized in Fig. 2. In most species, the 28S rDNA probe (i.e., corresponding to the NOR-associated major ribosomal cluster 45S rDNA, which

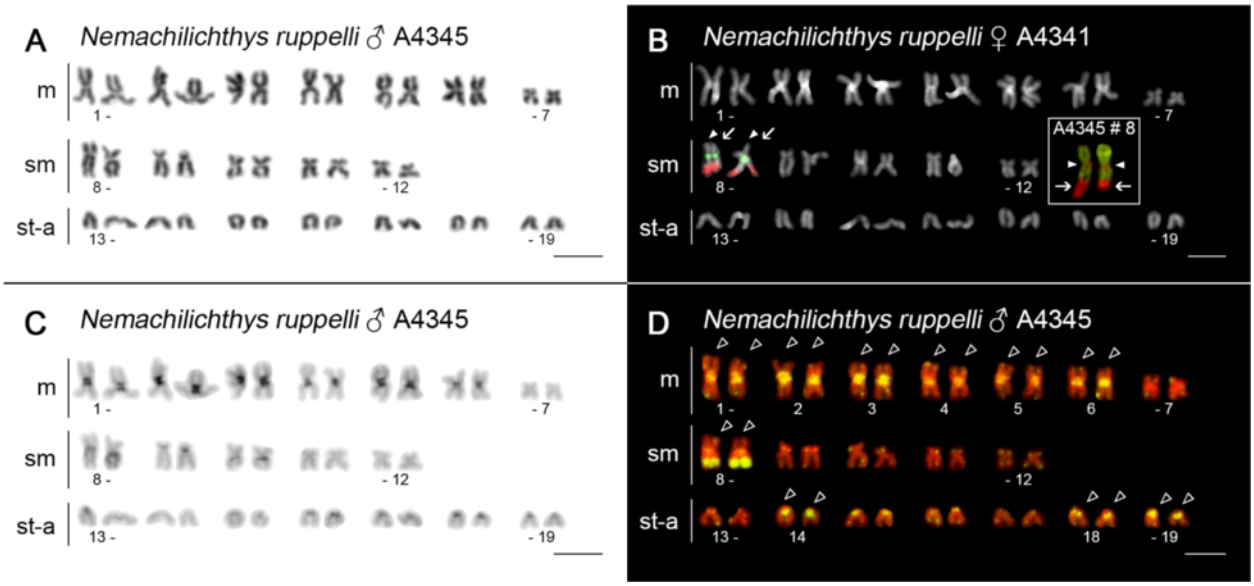


Fig. 4 Karyotypes of *N. ruppelli* after different cytogenetic protocols. **a** conventional Giemsa staining, **b** dual-colour FISH with 45S rDNA (*red, arrows*) and 5S rDNA (*green, arrowheads*) probes, **c** C-banding and **d** FISH with telomeric $(TTAGGG)_n$ probe. For better contrast, pictures were pseudocoloured in green (telomeric probe) and red (DAPI). Inset (**b**) – chromosome pair 8 bearing CMA_3^+ sites coinciding with both 45S (*arrow*) and 5S rDNA (*arrowhead*) sites. For better contrast, pictures were pseudocoloured in red (CMA_3^+) and green (DAPI). Note the prominent pericentromeric heterochromatin in metacentric chromosome pairs 1–6 (**b**) and an almost equivalent intensity of percentromeric ITSs (*open arrowheads*) in the same subset of chromosomes (**d**). Remaining ITSs (*open arrowheads*) are confined to a 45S rDNA region on chromosome pair 8 (compare pics. **b** and **d**) and to p-arms of st-a chromosomes. Finally, compare chromosome pair 8 on pics. **b**, **c** and **d**; entire q-arms bearing 45S rDNA/ITS are weakly C-positive after C-banding procedure. Bar = 10 μm

codes for 28S, 5,8S and 18S rRNA genes) showed only one pair of NOR-bearing chromosomes located in CMA_3^+ sites. NOR phenotypes with two or more loci were observed in *P. zonalternans* (Additional file 11: Figure S6B), *S. savona* (not shown; see later in the text), *S. lendlii* (Additional file 7: Figure S3V) and especially in *P. elongata* with the number of sites ranging from 12 (Additional file 11: Figure S6D) to 14 (Fig. 6g). In the latter, not more than six NORs were stained also by $AgNO_3$ impregnation (data not shown). The 45S rDNA sites were located exclusively terminally (m) or covered entire p-arms of a particular st-a chromosome pair. By contrast, we found a considerable variability in the number of 5S rDNA sites, ranging from two (*B. barbatula*, *N. binotatus*, *N. ruppelli*, *S. notostigma*; Figs. 3b and 4b; Additional file 7: Figure S3F and Additional file 8: Figure S4B, D) to 20 (*Physoschistura* sp.; Fig. 3d). The 5S rDNA clusters were mainly located in pericentromeric regions or distributed in the entire p-arms of some st-a chromosomes, but location in/nearby centromeres of m-sm chromosomes was also observed (*B. barbatula*, *N. ruppelli*, *P. lucidorsum*, *S. notostigma*; Figs. 3b and 4b; Additional file 7: Figure S3L and Additional file 8: Figure S4B, D). In two species we observed one pair of chromosomes with a syntenic association of both rDNA classes (*P. brevis* – pair 13, Additional file 7: Figure S3J; *N. ruppelli* – pair 8, Fig. 4b) and another two species displayed direct co-localization of them (*P. zonalternans* – pair 12; *P. elongata* – pairs 4 and 12 – Additional file 11: Figure S6B, D). In the latter species there is an intraspecific variability in the number of both rDNA clusters as well as the number of their co-localization sites, based on observation of 5S rDNA ranging between 14 and 16 sites (Fig. 6g and Additional file 11: Figure S6D) and even six co-localized rDNA sites in some metaphases (Fig. 6g). Here, we further observed intraspecific variability in 1) size polymorphism, especially in 45S rDNA (best seen on FISH karyotypes of *S. bolavenensis* and *S. corica* – Additional file 7: Figure S3N, P) 2) polymorphism in the presence/absence of homologous rDNA sites (*P. pictilis* – pair 10; *S. lendlii* – pair 3 and *S. notostigma* – pairs 12 and 22, Additional file 7: Figure S3H, V, Additional file 8: Figure S4D, F), 3) number of rDNA sites (*S. corica* - pair 17; *S. hypsiura* – pair 18, Additional file 7: Figure S3P, R; *S. notostigma* - compare Additional file 8: Figure S4B and F), 4) heterozygosity for inversion involving rDNA loci (*P. zonalternans* – pair 10; *P. elongata* – pair 5, Additional file 11: Figure S6B, D), and 5) linkage of the 5S rDNA locus to a putative sex chromosome (*S. fasciolata* – Fig. 5b). Interestingly, a conspicuous difference in the 5S rDNA phenotype was discovered between two

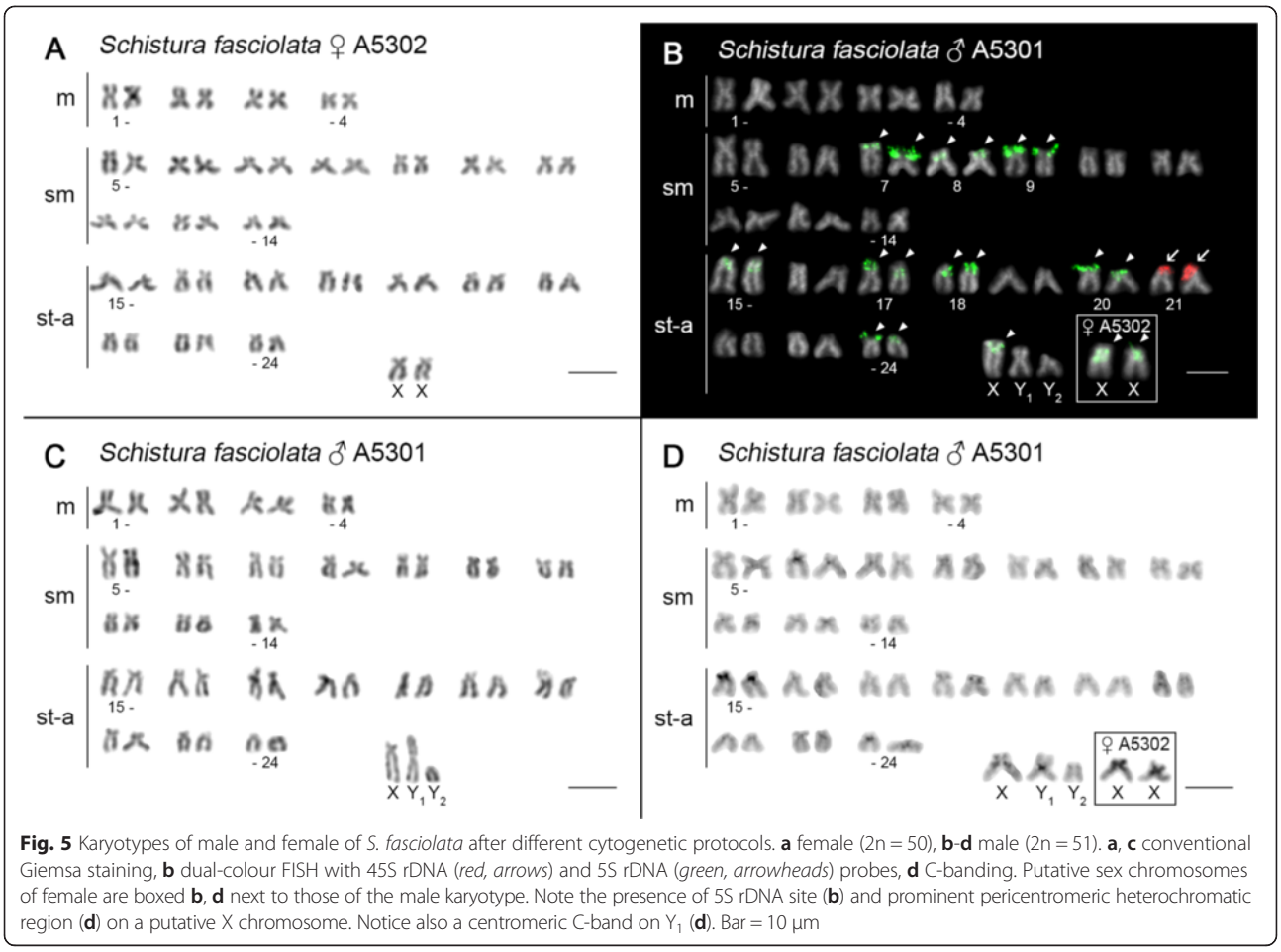


Fig. 5 Karyotypes of male and female of *S. fasciolata* after different cytogenetic protocols. **a** female (2n = 50), **b**-**d** male (2n = 51). **a**, **c** conventional Giemsa staining, **b** dual-colour FISH with 45S rDNA (*red, arrows*) and 5S rDNA (*green, arrowheads*) probes, **d** C-banding. Putative sex chromosomes of female are boxed **b**, **d** next to those of the male karyotype. Note the presence of 5S rDNA site (**b**) and prominent pericentromeric heterochromatic region (**d**) on a putative X chromosome. Notice also a centromeric C-band on Y_1 (**d**). Bar = 10 μm

karyomorphs of *S. notostigma*. While the karyomorph with 2n = 48 (1 specimen) exhibited five sites of 5S rDNA (all in pericentromeric regions of the st-a chromosomes, Additional file 8: Figure S4F), the karyomorph with 2n = 44 displayed only two of them, adjacent to centromeres of large-sized m chromosomes (5 specimens) (Additional file 8: Figure S4B, D). In *S. savona*, we observed considerable intraspecific variability being shown from two to four signals of 45S and from two to eight signals of 5S rDNA cluster (data not shown). In this species, however, it was not possible to conclusively distinguish whether this was the result of high intraspecific and intra-individual variability or whether it was artificial due to the limited visibility of the hybridization signals on such extraordinarily small chromosomes and so, these results are not discussed further.

Telomeric FISH

In order to document interstitial telomeric sites (ITSs) as remnants of chromosomal rearrangements, we employed FISH with conserved vertebrate telomeric $(TTAGGG)_n$ repeat [44] in a subset of seven species (*L. costata*, *N. binotatus*, *N. ruppelli*, *P. brevis*, *P. elongata*, *S. corica* and both karyomorphs of *S. notostigma*). As expected, the telomeric probe labelled the ends of all chromosomes, and no ITSs were revealed in five out of seven species (Fig. 7b, c and Additional file 12: Figure S7A-D). Clear ITSs, however, were observed consistently on ten metaphases of *N. binotatus* (Fig. 7a) and 15 metaphases of *N. ruppelli* (Fig. 4d). In *N. binotatus*, a single pair of ITSs occurred proximally on the q-arms of the largest chromosome in the karyotype (pair 11). These ITSs co-localized with sequentially heterogeneous AT/GC-rich heterochromatic regions. In *N. ruppelli*, three pairs of extensive and three pairs of faint pericentromeric ITSs were observed in large-sized m chromosomes (Fig. 4d). These six ITSs were coincident with AT-rich C-heterochromatin (Fig. 4c). Moreover, in this species additional large ITSs were also scattered all along the region of the single pair of 45S rDNA (compare Fig. 4b and d). The high intensity of some ITSs signals resulted in very limited visibility of natural telomeric signals on

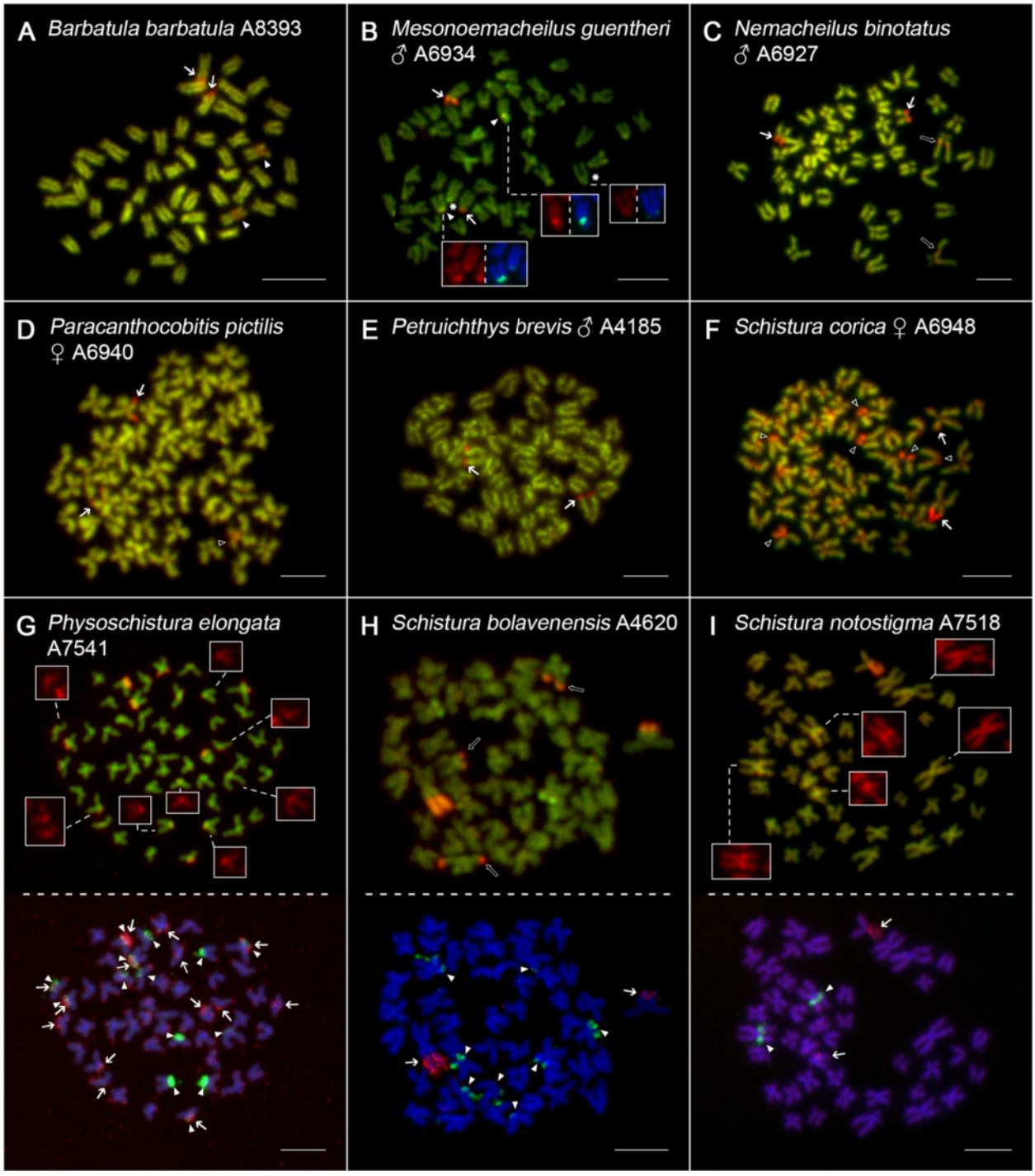


Fig. 6 (See legend on next page.)

(See figure on previous page.)
Fig. 6 Mitotic metaphases of selected nemacheilid species after CDD banding. **a**, **c**, **d**, **e**, **f** single metaphases; **b** metaphase arranged with boxes showing particular chromosomes sequentially after CDD banding and dual-colour rDNA FISH. **g**-**i** whole metaphases arranged sequentially – after CDD banding (*upper row*) and corresponding dual-colour FISH showing locations of 45S rDNA and 5S rDNA (*lower row*). **a** *B. barbatula*, **b** *M. guentheri*, **c** *N. binotatus*, **d** *P. pictilis*, **e** *P. brevis*, **f** *S. corica*, **g** *P. elongata*, **h** *S. bolavenensis*, **i** *S. notostigma*. For better contrast, CDD-banded pictures were pseudocoloured in red (for CMA_3) and green (for DAPI). FISH metaphases follow the same colour scheme as in Figs. 2, 3, 4 and 5. Arrows show CMA_3^+/45S rDNA sites, arrowheads show CMA_3^+/5S sites, open arrowheads show a putative CMA_3^+/5S sites and open arrows show CMA_3^+ regions non-related to rDNAs and minor/putative CMA_3^+ sites. In the particular case of *M. guentheri* (**b**), note the CMA_3-negative 5S rDNA sites (*denoted by asterisk*), while the remaining boxes clearly show CMA_3^+/5S rDNA sites. In non-sequential metaphases (**a**-**f**), considering the number and location of CMA_3^+ signals in comparison to respective FISH karyotypes (Fig. 2 and Additional file 7: Figure S3), the association between 45S rDNA and CMA_3^+ sites is clearly apparent from pics. and the same is true also for some or all 5S rDNA sites in (**a**, **d** and **f**). Due to the close proximity of 5S rDNA sites to centromeres (which are usually AT-rich and display bright fluorescence), some CMA_3^+/5S rDNA sites are not clearly apparent from the pictures, therefore they are boxed with a separate channel for CMA_3 (*red*) (**b**, **g**, **i**). Note the significant spreading of CMA_3^+ regions in centromeres of *S. corica* (**f**) and CMA_3-positive ITSs in *N. binotatus* (**c**). Bar = 10 μm

the chromosomal ends. Finally, in *N. ruppelli* and *S. notostigma*, some p-arms of small or medium-sized st-a were entirely covered by telomeric repeats.

Discussion

Topology of the phylogenetic tree

In our phylogenetic reconstruction the seven analysed species of *Schistura* do not form a monophyletic lineage, but appear as polyphyletic. This result reflects the massive flaws in the present taxonomy of this genus as already formerly stated by several taxonomists who referred to *Schistura* as 'a provisional, polyphyletic assemblage' [45], 'polyphyletic' and 'waste-basket name' [46] or 'possibly not monophyletic' [47]. The observed polyphyly of *Schistura* is therefore not surprising, but most likely reveals the true natural relationships between the analysed taxa. The two analysed species of *Physoschistura* turn out not to be closely related, supporting the former opinion that *P. elongata* is not closely related to the Burmese species of *Physoschistura* [47].

Karyotype differentiation and evolution

Karyotypes of *B. barbatula*, *L. costata*, *P. pictilis*, *S. fasciolata* and *S. savona* were revisited, whereas the remaining 14 species were examined for the first time. Our study thus increased the number of karyologically described river loaches to 38. Comparison of nemacheilid karyotypes reported in former studies with results presented here (Fig. 2) showed a different degree of congruence. While in *B. barbatula* our karyotype description matched the previous report of Vasil'ev [48] only, the karyotype of *P. pictilis* (formerly as *Acanthocobitis botia*) differed slightly in morphological classification from that recorded by Rishi et al. [49]. Also, we evaluated the karyotype of *L. costata* as having a higher number of biarmed elements than in Kim et al. [50]. Moreover, the karyotype of *S. fasciolata* described here is not consistent with that reported by Yu et al. [19], where no sex chromosomes were found, but one specimen of *S. fasciolata* with 44 chromosomes was included. Finally, karyotypes of *S. savona* reported in Khuda-Bukhsh et al. [18] consisted of 36 chromosomes while

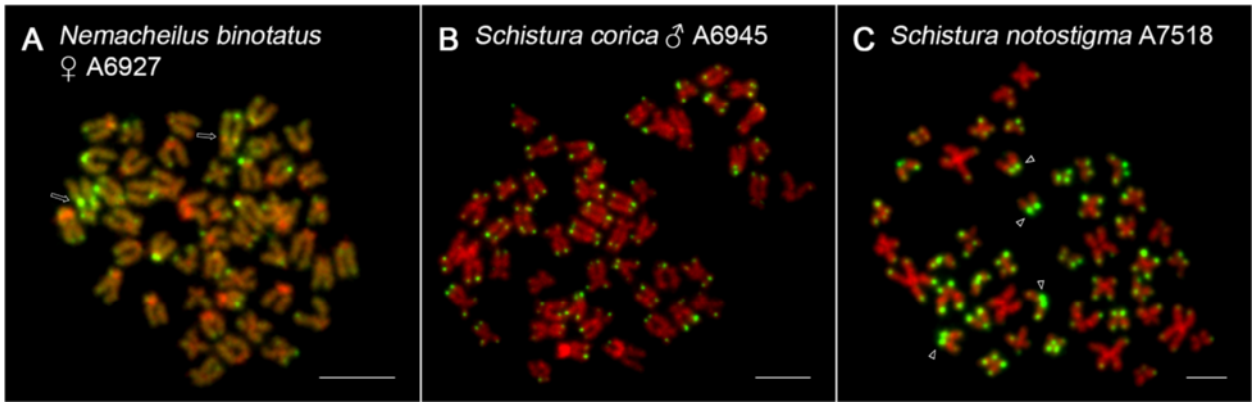


Fig. 7 Mitotic metaphases of selected nemacheilid species after TSA FISH with telomeric $(TTAGGG)_n$ probe. **a** *N. binotatus*, **b** *S. corica*, **c** *S. notostigma* (karyomorph with 44 chromosomes). Chromosomes with the telomeric repeat probe (*green colour*) are counterstained with DAPI, pseudocoloured in red colour for better contrast. Arrows point to the chromosomes with ITSs (**a**). Arrowheads show telomeric probe covering entire p-arm of st chromosomes (**c**). Bar = 10 μm

our results showed uniformly 2n = 50 in all three examined specimens. All these discrepancies may have resulted - besides the differences in chromosomal morphology classification due to difficulties described in the previous section – from the description of chromosomally different populations or by the misidentification of some species in the earlier studies.

Seventeen out of 19 analysed species showed conserved karyotypes with 2n = 50 (Figs. 2, 3 and 5; Additional file 7: Figure S3 and Additional file 11: Figure S6). This 2n has been already documented for the majority of previously surveyed river loaches [7, 12–14] as well as in some other loach families [4] and cyprinid fishes [7, 51]. Similar karyotypes (with either 48 or 50 chromosomes) were found in more than 50 % teleost species, thus indicating high conservativeness of this 2n [52]. Additionally, the 2n = 48 with exclusively monoarmed chromosomes is hypothesized to be ancestral for all Teleostei [53, 54].

Despite the generally stable karyotype macrostructure, the river loaches analysed here varied greatly in the proportion of chromosome types reflected by the increase or decrease of the NF value. The occurrence of species with similar karyotypes did not correspond with their phylogenetic relationships. Changes of NF without changes of the 2n are strong indicatives that nemacheilid chromosomes have evolved by diverse intrachromosomal rearrangements, such as various types of centromeric shifts.

We further recorded two species with karyotypes exhibiting reduced 2n, namely *N. ruppelli* (2n = 38) and *S. notostigma* (2n = 44 or 48). In the latter, our sample included two different karyomorphs. A single individual with 48 chromosomes did not show any significant differences in morphology and in sequences of *IRBP*, *RAG1* and *cyt b* in comparison to individuals with 2n = 44. Since we do not know the exact localities of analysed specimens, we cannot conclude, whether this result indicate the interpopulational variability.

In both species, reduction in 2n was accompanied by an increased number of large m chromosomes, implying their origin via one or several centric fusions of Robertsonian (Rb) type. Based on comparison of 2n and NF [55] and with respect to prevailing 2n = 50 in examined nemacheilids, *N. ruppelli* most likely underwent six Rb translocations, while karyotype differentiation in *S. notostigma* probably involved one Rb translocation (in karyomorph 2n = 48) and two Rb translocations, one tandem fusion and one para/pericentric inversion (in karyomorph 2n = 44), respectively. According to our phylogenetic analysis (Fig. 2), *N. ruppelli* and *S. notostigma* are not closely related, therefore the reduction of 2n in these species apparently represents independent events. Furthermore, the combined results from C-banding and telomeric FISH suggest a slightly different scenario of karyotype changes in both species (see below).

Besides our study, the evidence of reduced 2n among Nemacheilidae has already been documented for *Nemacheilus selangoricus* (2n = 40) [17], *Paracobitis potanini* (2n = 48) [19], *S. fasciolata* (2n = 44) [19], *S. savona* (2n = 36) [18] and *Triplophysa siluroides* (2n = 48) [20]. A different bias towards an increased number of either mono- or biarmed elements was apparent in these species. Some nemacheilid species (or at least representatives from some subpopulations) thus tend to reduce their 2n via centric or tandem fusions. Except for the studied males of *S. fasciolata* here and the report on triploidy [21], karyotypes of river loaches analysed to date did not exceed 2n = 50 ([7, 12–14], this study).

Our data show that karyotypes in nemacheilid loaches have diversified mainly via centric or tandem fusions and pericentric inversions. In general, such chromosomal rearrangements can act as an efficient barrier for gene flow (by suppressing recombination in the affected region) and thus can contribute to speciation and/or local adaptation processes [56–58].

Distribution and sequence composition of constitutive heterochromatin

Heterochromatin is an important source of karyotype diversification in several fish groups (e.g., [27]) and its unusual distribution may sometimes correspond to remnants of particular chromosome rearrangements [59]. As we present here, the karyotypes of river loaches differ greatly in their distribution of AT-rich C-heterochromatin (Additional file 9: Table S4), and contain some noticeable common patterns. We especially emphasise a) the dot-like intercalary heterochromatic bands on the q-arms of sm or st chromosomes, very close to the centromere (e.g., in *B. barbatula*, *N. binotatus*, *S. bolavenensis* and *S. notostigma*) and b) the presence of entirely heterochromatic arms in some elements (e.g., in *M. guentheri*, *S. bolavenensis*, *S. corica*, *S. hypsiura* and *S. lendlii*). Both observed patterns might be related to pericentric inversions (heterochromatinization of short arms are usually the result of this kind of rearrangement) and/or heterochromatin block addition [60]. The dot-like intercalary sites could be also explained by tandem fusions [60, 61], but considering the constant 2n in the majority of species under study, it would only be a plausible explanation for *S. notostigma*. Interestingly, although the presence of large biarmed chromosomes with entirely (or almost entirely) heterochromatic arms was shared by five nemacheilid species (*M. guentheri*, *P.* sp., *S. hypsiura*, *S. corica* and *S. lendlii*), these regions were adjacent to rDNA clusters only in two of them (*S. corica* and *S. lendlii*). Whether these chromosomes are homeologous among some of the mentioned species and

whether the heterochromatic blocks contribute to the dynamics of rDNA clusters remains inconclusive. Also, polymorphism regarding the addition of AT-rich heterochromatic p-arms was observed in one pair of chromosomes in *M. guentheri*. While the male was heterozygous for the presence of a prolonged heterochromatic arm, the female possessed only the short variants. Comparable results have been previously documented in some other fishes ([62] and references therein) and may be explained by an unequal crossing-over or by transposition/amplificaton processes involving a DAPI-rich centromeric region [22]. The last intriguing feature was the presence of large blocks of AT-rich heterochromatin in the pericentromeric region of the largest m chromosomes of *N. ruppelli* and *S. notostigma*. These regions are possibly remnants of pericentromeric heterochromatin of previously monoarmed elements. Similar feature displayed also *S. savona* on one m chromosome pair, however, with unreduced 2n = 50.

In fishes, GC-rich DNA segments labelled by CMA_3 are almost exclusively associated with NORs [63, 64], with some exceptions in sturgeons [65]. NOR regions were usually not visualized after C-banding, thus most likely suggesting that GC-rich sequences were inserted into the intergenic spacers (IGSs) of the 45S rDNA arrays [63, 66]. Additionally, nearly half of the species analysed showed further CMA_3^+ sites restricted to 5S rDNA regions – a feature that up to now has only been found among fishes in some Polypteriformes [67] and Perciformes, namely in Centrarchidae [68], Pomacanthidae [69] and Gobiidae [70]. Deiana et al. [68] attributed this feature to the presence of GC-rich repeats in NTS. Particularly interesting was the observation of all centromeres being CMA_3^+ in *S. corica* – a similar feature as, for instance, in Gobiidae [71, 72] and Polypteriformes [73]. Also in the genus *Cobitis*, high number of CMA_3^+ regions were recorded which were non-related to NORs (together with CMA_3-negative NOR sites) [74]. In a recent study, some CMA_3^+ regions non-related to NORs were observed also in *P. elongata*, *S. bolavenensis* and *S. notostigma*. Therefore, our results represent another example that CMA_3-staining and 45S rDNA FISH do not always correspond and that CDD banding itself is not sufficient for the proper identification of NORs in fishes (discussed in [75, 76]).

The scattered occurrence of non-45S rDNA GC-rich sites does not appear to imply any correlation with phylogenetic relationships. However, the phylogenetically most derived species (*P. brevis* and *Physoschistura sp.*) apparently lack GC-rich 5S signals (Fig. 6e and Additional file 10: Figure S5C). The evolutionary significance of this type of variability is still under debate. For instance, the sequence composition of heterochromatin can be associated with the different success of recombination processes and with a propensity to some kind of chromosome rearrangements. Here, the GC-rich regions were involved in two Rb translocations in *S. notostigma* and one of the resulting fusion points also involved a 5S rDNA site, because karyomorph with 2n = 48 display higher number of exclusively terminally located 5S rDNA and GC-rich sites, while karyomorph with 2n = 44 exhibit reduced number of such regions, with some of them being apparently relocated to the pericentromeric region of large m chromosomes (Fig. 6i; Additional file 8: Figure S4B, F and Additional file 10: Figure S5G). Hence, centric fusion is very likely partly responsible for the reduction of 5S rDNA sites from five (karyomorph with 2n = 48, st chromosomes, Additional file 8: Figure S4F) to two (karyomorph with 2n = 44, m chromosomes, Additional file 8: Figure S4B, D). A similar scenario could also explain the largest sm pair in *N. ruppelli* (no. 8, Fig. 4b), with GC-rich 5S rDNA in the centromeric region. However, considering the other six pairs of large m chromosomes with marked large pericentromeric heterochromatin and ITSs (as evidence of Rb translocation; Fig. 4c, d), there is no space for additional fusions since 2n = 38 had already been reached. Therefore, two alternative explanations for this discrepancy can be hypothesized: 1) the occurrence of conspicuous pair 8 in *N. ruppelli*, with syntenic association of both rDNAs, may be the result of Rb translocation only in the case of parallel fission of some other previously metacentric pair (resulting possibly in st-a pairs 18 and 19, with a markedly strong telomeric signal on the p-arms, Fig. 4d) or 2) synteny of both clusters on chromosome pair 8 has been caused by another type of translocation event, non-affecting the 2n.

GC-rich regions are more prone to high recombination rates [77]. In a similar way, GC-rich centromeres have been hypothesized to be favoured or even essential in the process of Rb translocations in some gobiid fishes [71, 72]. On the other hand, the majority of Rb translocations in *N. ruppelli* originated from elements containing AT-rich centromeres, and therefore it appears that more mechanisms exist for Rb translocations besides involvement of GC-rich regions. These findings contrast with those studies, but are consistent with results observed in killifishes [22] and *Mus musculus domesticus* [78].

Due to the number of reports evidencing 5S rDNA in the centromeres of fused chromosomes are gradually increasing in fishes [79, 80], it raises the question whether the 5S rDNA region could contribute in some way to the fusion process or it is only a consequence of it. It has been suggested that 5S rDNA can serve as breakpoints for the fusion due to its intensive activity and chromatin decondensation [80] but further data supporting this hypothesis would be required.

In our study, GC-rich sequences may be involved in the dispersion and homogenization of GC-rich/5S rDNA sequences as well as 45S rDNA sites and 5S/45S co-localized sites in the genome of *P. elongata* by ectopic recombination, similarly as observed in Gobiidae [70]. However, other nemacheilid species bearing GC-rich/5S rDNA regions do not display such extensive dispersion of 5S rDNA. Thus, other factors such as transposition together with stochastic processes in isolated populations may have been involved in the dynamics of GC-rich/5S rDNA sites. Similarly, a combination of transposition and unequal crossing-overs could have contributed to the dispersion of GC-rich centromeres in *S. corica*.

Our results from C- and CDD- banding further reinforced our initial hypotheses about the roles of pericentric inversions and centric/tandem fusions as the main processes underlaying the karyotype differentiation of examined river loaches. Collectively, our data point to a substantial heterogeneity both in heterochromatin distribution and composition among the analysed river loaches, resulting probably from intense dynamics at chromosomal and genomic levels.

Sex chromosomes

While the majority of analysed species lacked morphologically differentiated gonosomes, we identified a putative multiple sex chromosome system XY_1Y_2 in *S. fasciolata*. The two Y chromosomes in males (m and st) possibly arose from a double-strand break (or fission) in one proto-Y chromosome, followed perhaps by intrachromosomal rearrangements, such as pericentric inversions, in the larger element. Interestingly, the FISH results showed a pericentromeric 5S rDNA site on a putative X chromosome – a situation previously observed e.g., in rainbow trout [81]. In general, about 10 % of fish species cytogenetically examined to date exhibit morphologically differentiated gonosomes [82] and within them, only a few cases of the multiple system XY_1Y_2 have been reported (e.g., [83–85]), with apparently phylogenetically independent origins among genera and families. Our finding is the first observed in river loaches. However, because our sample was rather small, we can not exclude the possibility that we are still dealing with a polymorphism instead of a sex chromosome system. Therefore our conclusions should be further confirmed using comparative genomic hybridization (CGH) [86] and analyses of meiotic chromosomes on a larger sample base.

Telomeric FISH pattern

Tandemly-arrayed telomeric $(TTAGGG)_n$ repeats are usually present at the ends of vertebrate chromosomes, ensuring their stability and integrity. However, they also occasionally appear in non-telomeric locations (ITSs), possibly as putative markers of previous chromosomal rearrangements, transpositions or as the result of DNA repair mechanisms ([87–89] and references therein). In *L. costata, P. brevis, P. lucidorsum* and *S. corica* and both karyomorphs of *S. notostigma*, the telomeric signals were restricted to the chromosome ends. Although some metaphases displayed putative intercalar telomeric sites, the lack of a second terminal signal on the particular chromosome suggests that these signals label natural telomeres. ITSs were therefore only found in *N. ruppelli* and *N. binotatus*. In the latter species, the single prominent ITS located interstitially on the long arm of the largest st pair may indicate a pericentric inversion or a tandem fusion event. Since the ancestral diploid chromosome number (2n = 50) remained unchanged, the observed pair of ITSs is rather a relic of a previous pericentric inversion, although such types of rearrangement are not frequently associated with retained telomeric repeats in vertebrates ([90, 91]). The intense telomeric signal may be the result of an additional amplification of telomeric repeats either before or after the rearrangement, or, in the case of *N. binotatus,* could have originated from a relatively recent pericentric inversion. The failure to detect ITSs in the majority of the remaining species does not necessarily mean, that inversions did not occur as it is possible that the residual traces of ITSs have been lost or reduced to such a low copy number as to be undetectable by FISH analyses. The telomeric FISH also provided the interesting evidence that the mechanism of Rb translocations differs significantly between species with reduced 2n (*N. ruppelli* and *S. notostigma*). While *N. ruppelli* possessed several huge pericentromeric ITSs, none were found in *S. notostigma*. As described by Slijepcevic [92], the mechanism of Rb translocations can be either 1) associated with a loss of telomeric sequences prior to fusion or 2) with their preservation in otherwise inactivated telomeres. Moreover, there is also the possibility that 3) some degenerate telomere-like sequences may become part of the centromeric heterochromatin and subsequently expand along this region as a result of the action of a variety of amplification mechanisms [92, 93]. We suggest that a combination of scenarios 2) and 3) apply in *N. ruppelli*, while *S. notostigma* followed the first scenario, hence residual telomeric sequences were absent at the fusion points. The assumption of amplified centromeric ITSs in *N. ruppelli* is based on their remarkably stronger signal compared to native telomeres and on their C-positive character. ITSs often co-localize with heterochromatin blocks [87, 88, 94] and large, mostly centromeric ITSs similar to those in *N. ruppelli* (sometimes referred to as heterochromatic ITSs, or "het-ITSs"), have previously been described in other fishes as well as in a variety

of other organisms [88, 89]. Interestingly, additional ITSs were found to be associated with the 45S rDNA cluster in *N. ruppelli* as confirmed by FISH and CDD banding. This association has been previously described in Anguilliformes, Mugiliformes, Salmonifromes and Syngnathiformes [89] and was believed to play a role in the silencing of additional 45S rDNA copies [95]. This seems unlikely in *N. ruppelli*, however, as the telomeric repeats perfectly match with the entire region of the only pair of 45S rDNA. Alternatively, the mechanism of rDNA silencing could be more complex or prone to leakage of rDNA expression in some way. Finally, large ITS-like blocks covering entire p-arms of some monoarmed chromosomes as observed in *N. ruppelli* and *S. notostigma* bring another example of enormous nemacheilid cytogenetic variability.

Genomic organization and distribution of rDNA clusters

Mapping of tandemly-arrayed repetitive sequences has proven to be an important tool for karyotype analysis [59] and this is especially true for ribosomal RNA genes. The rDNA phenotypes are often species-specific and have been used as cytotaxonomic markers [96]. However, a number of reports demonstrating extensive inter/intra-specific variability of these markers is still growing in fishes [97–100], other animal groups [101] and plants [102]. Here, we point to the conservative NOR phenotype, presented by one pair bearing 45S rDNA signals in 15 out of 19 nemacheilid species. Although the possibility of interspecific homeology of NOR-bearing chromosomes is rather unlikely, definitive proof based, for instance, on the approach described by Milhomem et al. [103] is missing. From all our samples we documented multiple 45S rDNA sites only in *P. elongata* (Fig. 2). Subsequent analysis made by silver staining detecting only NORs actively transcribed in the preceding metaphase [104] revealed not more than six loci (data not shown), thus, some extra loci are either nonfunctional or silenced.

We observed a conservative NOR phenotype of one major rDNA bearing pair – a pattern found in more than 70 % of examined fish species to date [76]. In contrast, we detected a considerable variability in the pattern of 5S rDNA ranging from two to 20 sites (Fig. 2). The presence of a single pair of both rDNA clusters is thought to be the plesiomorphic condition in teleost fishes, whereas two or more chromosome pairs bearing either 45S or 5S rDNA sites represent a derived condition [76, 105]. In our study, only *B. barbatula* and *N. binotatus* exhibited karyotypes with the ancestral 2n = 50 together with one pair of both rDNA clusters and in *B. barbatula*, our results confirmed the previously reported NOR phenotype based on silver staining [15]. In the remaining two species exhibiting the characteristic teleost rDNA phenotype the karyotypes were derived (*N. ruppelli* and *S. notostigma*). A variable 5S rDNA pattern in combination with a conservative NOR phenotype has been observed in some fish groups [70] while other fish groups have shown the opposite situation (variable 45S and conservative 5S rDNA; [106]).

The 45S rDNA site has a predominantly terminal position on the different chromosomes of the analysed species, while the 5S rDNA is located almost exclusively in the pericentromeric regions or it covers entire p-arms of monoarmed chromosomes. Pericentromeric or, more generally, interstitial position of 5S rDNA appears to be universal among fishes [107].

In fishes, the chromosome locations of both rDNA multigene families are usually on different chromosomes, perhaps due to 1) the elimination of possible rearrangements between both multigene families and 2) to allow rDNA clusters to evolve independently [105, 108]. On the other hand, exceptions with syntenic location or direct co-localization of both rDNA clusters (or their linkage to other multigene families) has already been documented in a variety of vertebrates [109, 110], including reports from several fish groups [111] as well as in loaches of the family Cobitidae [98], a sister lineage to nemacheilids. This pattern is rather patchily distributed across the phylogenetic trees and was also evidenced in our study. In *N. ruppelli*, the 5S rDNA loci occupied the pericentromeric region of a big m chromosome while the 45S rDNA was situated terminally on the q-arm of the same chromosome. In *P. brevis*, a similar association was observed in one pair of big st elements. Moreover, direct co-localization of rDNA clusters was observed in two species: one pair in *P. zonalternans* and from four to six co-localized sites in *P. elongata*. Such a rare situation has probably no evolutionary advantage as both classes of rRNA genes are transcribed by different RNA polymerases [109]. Therefore, this constitution is a possible consequence of recent genome instability and reshuffling as typically observed in hybridization events [100].

In all species analysed here, a size polymorphism in the homologous 45S rDNA sites was apparent. Such an observation is relatively common among fishes and is attributed to the processes of unequal crossing-over or the amplification of adjacent heterochromatin [112]. We also observed an intraspecific polymorphism in terms of the number of rDNA sites present (*S. corica, S. hypsiura, S. notostigma*) and a polymorphism in the presence/absence of rDNA site on one homologous chromosome in *S. notostigma* (5S rDNA and 45S rDNA), *S. lendlii* (female, 45S rDNA) as well as for both females of *P. pictilis* (5S rDNA). Unfortunatelly, the limited number of specimens available in our sample is insufficient to conclusively determine either

fixation or heterogeneity of this feature in the population. Similar heterozygous constitutions of rDNA FISH signals have been commonly observed in several species of fishes (e.g., [112, 113]) including some from Cobitidae [74, 97, 98]. The lack of signal on one of the homoloques may be a direct consequence of sequence elimination due to unequal crossing-overs, often related to a process of concerted evolution in tandemly-repeated genes [114] or by the activities of repetitive DNA such as TEs [59, 115]. Finally, we also observed the polymorphism caused by rDNA loci inversion in *P. zonalternans* and *P. elongata*. This feature, present in Cobitidae [98], suggests a strikingly similar dynamics of rDNA loci in these closely related loach families as well as another clue about the contribution of inversions to the karyotype differentiation of river loaches.

Our study has revealed an extensive dispersion of multiplied sites of 5S rDNA and also of 45S rDNA in nemacheilids. The dominance of the ancestral 2n = 50 karyotype in Nemacheilidae refutes chromosomal rearrangements as the trigger mechanism for this dispersion, but amplification and dispersion of 5S rDNA clusters may also be caused by transposition and unequal crossing-over or ectopic recombination between various tandemly-arrayed sequences in adjacent heterochromatin [102, 115, 116]. Thus, rDNA clusters themselves can provide a substrate for non-homologous recombination, thereby promoting chromosomal rearrangements [101]. A significant fraction of the rDNA units in animals are interrupted by TEs highly specialized for insertion into conserved sites within the rRNA genes [114, 117] and recent studies suggested that they might cause rDNA mobility [118–120]. Co-localization of non-LTR RTEs of the Rex family with rDNA followed by a subsequent expansion of rDNA sites have been uncovered by FISH analyses for 5S [94] and 45S rDNA [121, 122]. It is tempting to hypothesize that a similar mechanism could cause the amplification of 5S/45S rDNA in other fish species with documented extensive rDNA dispersion. In our study, we have found the non-LTR retrotransposon L1-2_DR element, from the Tx1 clade (L1 lineage) – inserted close to a coding region of 5S rDNA in *S. pridii*. This element has been previously described in zebrafish [123].

Since the karyotype of *S. pridii* exhibit a large number of 5S rDNA loci (18), the L1-2_DR may have been inserted into the NTS of both analysed 5S variants and subsequently retrotransposed to other chromosomal loci. RTEs of this L1 family preferentialy jump into AT-rich regions [123], therefore the AT-rich pericentromeric heterochromatin of *S. pridii* located adjacent to 5S rDNA could serve as a primary location for this mechanism. Alternatively and/or as secondary consequences, the L1-2_DR (or other TEs) could provide the substrate for non-homologous (ectopic) recombination between centromeres of several chromosomes in *S. pridii* yielding to a dispersion of 5S rDNA to other sites. Both hypotheses deserve further investigation regarding the localization of L1-2_DR elements on the chromosomes of *S. pridii* and also the investigation of possible rDNA/TEs association in other river loaches through FISH analysis. However, we can not rule out the hypothesis, that L1-2_DR elements are just following the spread of 5S rDNA and not driving it (for additional note, see Additional file 1: Supplementary Discussion).

The variation observed in the distribution of 5S rDNA sites implies a complex microevolutionary mechanism behind the organization of nemacheilid genomes. The final questions are: wheter or not a dispersion of 5S rDNA is only a byproduct of rapid genomic change, is there any possible contribution to the host genome worth maintaining such a high number of copies, or are the excessive copies most likely sentenced to pseudogenization and elimination? Could an extensively elevated number of rDNA loci somehow contribute to the speciation process? We are still far from understanding this but some indications have come from studies on 45S rDNA in notothenoid fishes [111] and humans [124]. According to Pisano and Ghigliotti [111] the differential pattern of the rDNA phenotype could have a possible adaptive significance in subzero temperatures. Furthermore, the study of Gibbons et al. [124] shows that 45S rDNA dosage is correlated with mitochondrial DNA abundance and with the expression of some chromatin modifiers thereby affecting mitochondrial-related processes and changes in global gene expression. However, whether a similar correlation is true also for 5S rDNA dosage remains an open question (but see [125]). Thus, such an explanation does not yet fit our hypotheses about the mechanisms behind nemacheilid radiation success, although, it does suggest a frame in which to evaluate the contribution of multiple 5S rDNA to adaptation and speciation processes.

Conclusively, our data suggest frequent changes of 5S rDNA phenotypes in contrast to the stable pattern of 45S rDNA. Extensive variability of 5S rDNA loci may be regarded as an indicator of significant intragenomic processes [115, 116] and thus can be viewed in the context of an incipient stage of speciation, where evolutionary changes driven by the dynamics of repetitive DNA are currently in action [59]. This process can be also related to extreme ecological conditions possibly resulting in (re)activation of TEs [122]. As documented in several animal and plant species, elevated activity of TEs may contribute to adaptation to a new environment [126, 127]. Furthermore, the processes of transposition and/or ectopic recombination were not likely

restricted only to regions of 5S rDNA. Numerous studies have documented the involvement of TEs in chromosomal rearrangements [59, 115, 126, 128]. We therefore conjecture that TEs might also contribute to the dynamics of nemacheilid genomes in this way.

Phylogenetic and ecological inferences

We have used a phylogenetic tree to show the relationships between the analysed individuals. When mapped on this tree, the observed cytogenetic characteristics did not reflect the phylogenetic pattern, suggesting that certain cytogenetic character stages, like a lowered number of chromosomes, did not occur in closely related, but in non-related species. Therefore the parallel occurrence of cytogenetic character stages in two species is not the result of a single evolutionary event, but of convergence or parallel evolution. Our study has revealed a high variability in cytogenetic characters with almost none of them producing a phylogenetic signal. Therefore, a vast number of independent events with no general direction must have happened to cause the observed cytogenetic variability. The frequent occurrence of independent cytogenetic changes as revealed by the phylogenetic reconstruction further emphasises the high mutational activity of the nemacheilid genome at the cytogenetic level.

In contrast to the general observation of independent cytogenetic events, one of the variable cytogenetic characters did show an interesting pattern. The highest numbers of 5S rDNA loci (up to 20 sites) were almost exclusively observed in local endemics or inhabitants of small, fragmented habitats (*P. brevis*, *P. elongata, P.* sp., *S. hypsiura* and *S. pridii* – see Fig. 2). This produces a comparably small effective population size and therefore a small gene pool for the species, encouraging the establishment of new chromosomal patterns through genetic drift, meiotic drive and inbreeding [60]. In *P. brevis* the actual population size is quite large, but as it occurs only in a single lake, it can be assumed that the species has undergone through a serious bottleneck during the colonization of this area.

Conclusions

Our data provides important information regarding the karyotype differentiation trends in Nemacheilidae. The majority of surveyed species showed the karyotype characteristics common for teleost fishes – e.g., 2n = 50 chromosomes with a slightly changing centromere position, a single pair of NOR and its association with GC-rich blocks of heterochromatin. However, a number of deviations were also apparent – e.g., reduced 2n in two species, atypical locations of GC-rich heterochromatin (e.g., in 5S rDNA sites), cases of multiple rDNA sites and the presence of putative sex chromosomes. While conventional staining showed prevailing uniformity of the nemacheilid karyotypical macrostructure, analysis at the molecular-cytogenetic level revealed much more variability and greater diversity than previously expected. An increased number of 5S rDNA sites were observed, especially in species with a small effective population size. The mechanisms responsible for such intense dynamics can possibly be attributed to the presence of repetitive sequences and could contribute to enormous success of Nemacheilidae in their colonization and exploitation of new niches, as well as with their adaptation processes. Our study presents river loaches as a new attractive model fish group for investigating the dynamics of cytogenetic markers in association with evolutionary and ecological questions. Importantly, we have also introduced a new non-invasive technique for obtaining chromosome spreads for molecular-cytogenetics protocols.

Availability of supporting data

All the supporting data are included as additional files.

Additional files

Additional file 1: Supplementary Methods 1. Taxonomic identification of nemacheilids. **Supplementary Methods 2.** Preparation of chromosomes from regenerating fin tissue. **Supplementary Methods 3.** PCR conditions of 5S and 45S rDNA amplification. **Supplementary Methods 4.** PCR conditions of *RAG1*, *IRBP* and *cyt b* amplification. **Supplementary Discussion**. Possible functional consequences of excessive 5S rDNA copies. (PDF 181 kb)

Additional file 2: Figure S1. Mitotic metaphases of selected nemacheilid species after C-banding or DAPI-staining. (A,D,C,F.G,H,I,J,M,O,P,Q) DAPI staining; (D,E,K,L,N) C-banding improved with DAPI counterstaining. Metaphases from both methods are converted to inverted pictures. (A) *B. barbatula*, (B) *L. costata,* (C,D) *M. guentheri* (E) *N. binotatus*, (F) *P. pictilis*, (G) *P. zonalternans*, (H) *P. brevis*, (I) *P.* sp., (J) *P. elongata,* (K) *P. lucidorsum*, (L) *S. bolavenensis*, (M) *S. corica*, (N) *S. hypsiura,* (O) *S. pridii*, (P) *S. savona,* (Q) *S. lendlii*. Arrows depicts whole-armed heterochromatin, arrowheads denote interstitial heterochromatin, the asterisk indicates C-positive NORs (as rare feature among species under study). For comparison of banding patterns between both methods, see pics. C and D. Note that several species share marked interstitial heterochromatic sites indicating the remnants of putative chromosomal rearrangements (e.g., pericentric inversion) (A,E,L,M,Q). Of particular interest are the completely heterochromatic arms in m-sm chromosomes occuring in a subset of species (C,D,L,M,P,Q). Note also heterochromaic p-arms in some st-a chromosomes (C,D,E,O,P). Bar = 10 μm. (ZIP 4276 kb)

Additional file 3: Table S1. Primer sequences used in this study. (PDF 221 kb)

Additional file 4: Table S2. GenBank accession numbers of *cyt b*, *IRBP* and *RAG1* sequences of nemacheilids and one botiid species (*B. lohachata*). (XLSX 17 kb)

Additional file 5: Table S3. GenBank accession numbers of 5S and 28S rRNA sequences of four nemacheilids (*P. elongata, S. bolavenensis, S. corica, S. fasciolata*) and one botiid species (*B. almorhae*). (XLSX 14 kb)

Additional file 6: Figure S2. Sequence alignment of cloned 5S rDNA fragments from *S. pridii*. Nucleotide sequences (5′-3′) obtained from both specimens (A7548, A7549) corresponding to the short (A) and long (B) variant of 5S rDNA, containing partial 5S rDNA coding sequence (green), partial sequence of L1-2_DR non-LTR retrotransposon (in red) and a

putative non-transcribed spacer (NTS) (rest of the sequence). In the short fragment (A), the consensus sequence is shown for specimen no. A7548 and only base changes according to this sequence are shown for the specimen no. A7549. Dots indicate the upper consensus sequence. Sequence of the long fragment (B) was assembled only from specimen no. A7548. (PDF 111 kb)

Additional file 7: Figure S3. Karyotypes arranged from Giemsa-stained chromosomes and dual-colour FISH showing 5S and 45S rDNA sites. Giemsa-stained karyotypes (left column) and dual-colour FISH (right column) with 45S rDNA (red, arrows) and 5S rDNA (green, arrowheads) probes on (A,B) *L. costata*, (C,D) *M. guentheri*, (E,F) *N. binotatus*, (G,H) *P. pictilis*, (I,J) *P. brevis*, (K.L) *P. lucidorsum*, (M,N) *S. bolavenensis*, (O,P) *S. corica*, (Q,R) *S. hypsiura*, (S,T) *S. pridii*, (U,V) *S. lendlii*, (W) *S. savona*. The FISH chromosomes were counterstained with DAPI and the images were converted to grayscale. Inset (D) – chromosome pair 18 from *M. guentheri* female showing absence of heterochromatic p-arm in contrast to a single homologue in the male karyotype. Inset (V) depicts the absence of a 45S rDNA site on one homologue in female (pair 3). In *P. brevis* (J), note the syntenical association of both rDNAs on pair 13. Note also the intense size polymorphism in *S. bolavenensis* (pair 1) (N) and *S. corica* (pair 7) (P). Additional polymorphic rDNA sites from the other specimen are boxed for *S. corica* (pairs 7 and 17) (P) and *S. hypsiura* (pair 18) (R). Bar = 10 μm. (ZIP 2482 kb)

Additional file 8: Figure S4. Karyotypes of *S. notostigma* after different cytogenetic protocols. (A-D) karyomorph with 44 chromosomes, (E,F) karyomorph with 48 chromosomes. (A,E) conventional Giemsa staining, (C) C-banding, (B,D,F) dual-colour FISH with 45S rDNA (red, arrows) and 5S rDNA (green, arrowheads) probes. Arrangement of st-a chromosome pairs 18 and 19 (E,F) demonstrates a putative origin (centric fusion) of metacentric chromosome pair 2 (A-D). Note also chromosome pairs heterozygous for presence/absence of 45S rDNA site (pair 12) (D) or 5S rDNA site (pair 22) (F). Finally, notice conspicuous regions of constitutive heterochromatin located in centromeres of metacentric pairs 1 and 2 and those located intercalarly on q-arms of sm chromosome pairs 11, 13, 14, 15, 17. Bar = 10 μm. (TIF 540 kb)

Additional file 9: Table S4. Distribution of AT- and GC-rich sites and its relation to constitutive heterochromatin and rDNA regions in nemacheilid genomes as inferred from C-banding, DAPI- and CMA3-stainings. Species order reflects their phylogenetic relationships. (XLSX 13 kb)

Additional file 10: Figure S5. Mitotic metaphases of selected nemacheilid species after CDD banding. (A) *L. costata*, (B) *P. zonalternans*, (C) *P.* sp., (D) *P. lucidorsum*, (E) *S. fasciolata*, (F) *S. hypsiura*, (G) *S. notostigma* (karyomorph with 48 chromosomes) (H) *S. pridii*, (I) *S. savona*, (J) *S. lendlii*. Pictures were pseudocoloured in red (for CMA_3) and green (for DAPI). Bar = 10 μm. (TIF 2549 kb)

Additional file 11: Figure S6. Karyotypes of *P. zonalternans* and *P. elongata* after Giemsa staining and dual-colour (5S/45S) rDNA FISH. 45S rDNA (red, arrows) and 5S rDNA (green, arrowheads) probes on (A,B) *P. zonalternans* and (C,D) *P. elongata*. Insets – chromosomes showing co-localization of 45S and 5S rDNA – *P. zonalternans*, pair 12 (B); *P. elongata* pairs 4, 12 (D). For clarity, chromosomes are arranged as separated images for each rDNA probe. Note also the heterozygosity for inverted 45S rDNA locus – *P. zonalternans*, pair 10 (B); *P. elongata,* pair 5 (D). The asterisk denotes a missing chromatid in one homologue of chromosome pair 25 in *P. elongata* (D). Bar = 10 μm. (TIF 522 kb)

Additional file 12: Figure S7. Mitotic metaphases of selected nemacheilid species after TSA FISH with telomeric $(TTAGGG)_n$ probe. (A) *L. costata*, (B) *P. brevis*, (C) *P. lucidorsum,* (D) *S. notostigma* (karyomorph with 48 chromosomes). Chromosomes with the telomeric repeat probe (green) are counterstained with DAPI, pseudocoloured in red for better contrast. Bar = 10 μm. (TIF 785 kb)

Abbreviations

AT: Adenine-thymine; CMA_3: Chromomycin A_3; BSA: Bovine serum albumin; *cyt b*: Cytochrome b; DAPI: 4′, 6-diamidino-2-phenylindole; DAPI/$CMA_3^{+/-}$: DAPI/CMA_3 positive/negative signals; FISH: Fluorescence in situ hybridization; FITC: Fluorescein isothiocyanate; GC: Guanine-cytosine; GIRI: Genetic information research institute; het-ITSs: Heterochromatic interstitial telomeric sequences; IAPG: Institute of Animal Physiology and Genetics; IGS: Intergenic spacer; *IRBP*: Interphotoreceptor retinoid-binding protein; ITSs: Interstitial telomeric sequences; LTR: long terminal repeat; NF: Nombre fundamental; NOR: Nucleolar organizer region; NTS: Non-transcribed spacer; PCR: Polymerase chain reaction; *RAG1*: Recombination-activating gene 1; rDNA: Ribosomal DNA; RTEs: Retrotransposable elements; SSC: Standard saline buffer; TEs: Transposable elements; TSA: Tyramid signal amplification; 2n: Diploid chromosome number.

Competing interests

There are no competing interests to declare.

Authors' contributions

AS prepared chromosome preparations, performed all C-banding, CDD-banding procedures, designed and performed probe preparations and FISH experiments and co-drafted the manuscript. JB managed, raised and identified specimens material and contributed to chromosome preparations. VŠ managed the specimens material and performed all phylogenetic analyses. MA performed digital processing of all images. RS partly contributed to 5S rDNA probe preparation, design of FISH experiments and digital processing of images. PR designed the study and co-drafted the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We would like to thank J. Freyhof, R. Hoyer, G. Ott, I. Seidel and K. Udomritthiruji for their help obtaining specimens. We are also grateful to K. Ocalewicz and M.B. Cioffi for useful discussions and inspiring comments and further to A.-M. Dion-Côté and C. S. A. Pereira for technical suggestions. We are thankful also to J. Čechová, P. Šejnohová and Š. Pelikánová for laboratory assistance and to C. Johnson for reading the manuscript and language corrections. The study was carried out with the financial support from the Czech Science Foundation (grant no. 206/08/0637) and with the institutional support RVO: 67985904.

Author details

[1]Laboratory of Fish Genetics, Institute of Animal Physiology and Genetics, Czech Academy of Sciences, Rumburská 89, Liběchov 277 21, Czech Republic. [2]Department of Genetics and Microbiology, Faculty of Science, Charles University in Prague, Viničná 5, 128 44 Prague 2, Czech Republic. [3]Department of Ecology, Faculty of Science, Charles University in Prague, Viničná 7, 128 44 Prague 2, Czech Republic. [4]Research Institute for Limnology, University of Innsbruck, Mondseestraße 9, A-5310 Mondsee, Austria.

Received: 9 June 2015 Accepted: 4 November 2015
Published online: 14 November 2015

References

1. Nelson JS. Fishes of the world. 4th ed. Hoboken: Wiley; 2006.
2. Saitoh K, Sado T, Doosey MH, Bart Jr HL, Inoue JG, Nishida M, et al. Evidence from mitochondrial genomics supports the lower Mesozoic of South Asia as the time and place of basal divergence of cypriniform fishes (Actinopterygii: Ostariophysi). Zool J Linn Soc. 2011;161:633–62.
3. Kottelat M. Conspectus cobitidum: an inventory of loaches of the world (Teleostei: Cypriniformes: Cobitoidei). Raffles Bull Zool. 2012;Suppl 26:1–199.
4. Bohlen J, Völker M, Rábová M, Ráb P. Note on the banded karyotype of the enigmatic South Asian loach *Vaillantella maassi* (Cypriniformes: Vaillantellidae). Ichthyol Res. 2008;55:82–4.
5. Šlechtová V, Bohlen J, Freyhof J, Ráb P. Molecular phylogeny of the Southeast Asian freshwater fish family Botiidae (Teleostei: Cobitoidea) and the origin of polyploidy in their evolution. Mol Phylogenet Evol. 2006;39:529–41.
6. Saitoh K, Chen W-J, Mayden RL. Extensive hybridization and tetrapolyploidy in spined loach fish. Mol Phylogenet Evol. 2010;56:1001–10.
7. Arai R. Fish karyotypes: a check list. 1st ed. Tokyo: Springer Japan; 2011.
8. Vasil'eva ED, Vasil'ev VP. Sibling species in genus *Cobitis* (Cobitidae). *Cobitis rossomeridionalis* sp. nova. J Ichthyol. 1998;38:580–90.
9. Bohlen J, Ritterbusch D. Which factors affect sex ratio of spined loach (genus *Cobitis*) in lake Müggelsee? Environ Biol Fishes. 2000;59:347–52.

10. Bohlen J, Ráb P. Species and hybrid richness in spined loaches of the genus *Cobitis* (Teleostei: Cobitidae), with a checklist of European forms and suggestions for conservation. J Fish Biol. 2001;59(Suppl A):75–89.
11. Bănărescu P. Zoogeography of fresh waters, Distribution and dispersal of freshwater animals in North America and Eurasia, vol. 2. Wiesbaden: AULA-Verlag GmbH; 1992.
12. Gaffaroglu M, Karasu M, Unal S. Karyotype of river loach *Turcinoemacheilus kosswigi* Bănărescu and Nalbant, 1964 (Cypriniformes, Balitoridae) from the Euphrates river, Turkey. J Agric Sci Technol. 2012;14:821–6.
13. Kumar R, Sahoo PK, Vishwanath W, Barat A. Karyotype of a new loach *Schistura obliquofascia* and a mahseer *Puntius chelynoides* from Kumaun Hills of Himalaya. Cytologia. 2014;79:243–6.
14. Esmaeili HR, Pirvar Z, Ebrahimi M, Geiger MF. Karyological and molecular analysis of three endemic loaches (Actinopterygii: Cobitoidea) from Kor River basin, Iran. Mol Biol Res Commun. 2015;4:1–13.
15. Boroń A. Chromosome banding studies of *Noemacheilus barbatulus* (Linnaeus, 1758) from Poland. Caryologia. 1995;48:239–46.
16. Mazik EY, Toktosunov AT. The karyotypes of three species of loaches of the genus *Nemachilus* Hasselt (Cypriniformes, Cobitidae). Tsitologiya. 1984;26:960–2.
17. Suzuki A. Chromosome and DNA studies of eight species in the family Cobitidae (Pisces, Cypriniformes). La Kromosomo II. 1992;67/68:2275–82.
18. Khuda-Bukhsh AR, Chanda T, Barat A. Karyomorphology and evolution in some Indian hillstream fishes with particular reference to polyploidy in some species. In: Uyeno T, Arai R, Taniuchi T, Matsuura K, editors. Indo-Pacific fish biology. Tokyo: Ichthyological Society of Japan; 1986. p. 886–98.
19. Yu XJ, Zhou T, Li YC, Li K, Zhou M, editors. Chromosomes of Chinese freshwater fishes. Beijing: Science Press; 1989 (In Chinese).
20. Cui J, Ren X, Yu Q. Nuclear DNA content variation in fishes. Cytologia. 1991;56:425–9.
21. Collares-Pereira MJ, Madeira JM, Ráb P. Spontaneous triploidy in the stone loach *Noemacheilus barbatulus* (Balitoridae). Copeia. 1995;2:483–4.
22. Völker M, Sonnenberg R, Ráb P, Kullmann H. Karyotype differentiation in *Chromaphyosemion* killifishes (Cyprinodontiformes, Nothobranchiidae). II: cytogenetic and mitochondrial DNA analyses demonstrate karyotype differentiation and its evolutionary direction in *C. riggenbachi*. Cytogenet Genome Res. 2006;115:70–83.
23. Völker M, Ráb P. Direct chromosome preparation from regenerating fin tissue. In: Ozouf-Costaz C, Pisano E, Foresti F, de Almeida Toledo LF, editors. Fish cytogenetic techniques: Ray-Fin fishes and chondrichthyans. Enfield: CRC Press, Inc.; 2015. p. 37–41. ISBN 9781482211986.
24. Haaf T, Schmid M. An early stage of ZZ/ZW sex chromosomes differentiation in *Poecilia sphenops* var. *melanistica* (Poeciliidae, Cyprinodontiformes). Chromosoma. 1984;89:37–41.
25. Behr W, Honikel K, Hartmann G. Interaction of the RNA polymerase inhibitor chromomycin with DNA. Eur J Biochem. 1969;9:82–92.
26. Kapuściński J, Szer W. Interactions of 4′,6-diamidine-2-phenylindole with synthetic polynucleotides. Nucleic Acids Res. 1979;6:3519–34.
27. Mayr B, Ráb P, Kalat M. Localisation of NORs and counterstain-enhanced fluorescence studies in *Perca fluviatilis* (Pisces, Percidae). Genetica. 1985;67:51–6.
28. Sola L, Rossi AR, Iaselli V, Rasch EM, Monaco PJ. Cytogenetics of bisexual/unisexual species of *Poecilia*. II. Analysis of heterochromatin and nucleolar organizer regions in *Poecilia mexicana mexicana* by C-banding and DAPI, quinacrine, chromomycin A_3, and silver staining. Cytogenet Cell Genet. 1992;60:229–35.
29. Howell WM, Black DA. Controlled silver-staining of nucleolus organizer regions with a protective colloidal developer: a 1-step method. Experientia. 1980;36:1014–5.
30. Levan A, Fredga K, Sandberg AA. Nomenclature for centromeric position on chromosomes. Hereditas. 1964;52:201–20.
31. Graham DE. The isolation of high molecular weight DNA from whole organisms or large tissue masses. Anal Biochem. 1978;85:609–13.
32. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215:403–10. http://blast.ncbi.nlm.nih.gov/blast. Accessed 23 Apr 2014.
33. Neusser M. Karyotypevolution, genomorganisation und zellkernarchitektur der neuweltaffen, PhD Thesis. München: Ludwig-Maximilians-Universität; 2004.
34. Ijdo JW, Wells RA, Baldini A, Reeders ST. Improved telomere detection using a telomere repeat probe $(TTAGGG)_n$ generated by PCR. Nucleic Acids Res. 1991;19:4780.
35. Cremer M, Grasser F, Lanctôt C, Müller S, Neusser M, Zinner R, et al. Multicolor 3D fluorescence *in situ* hybridization for imaging interphase chromosomes. Methods Mol Biol. 2008;463:205–39.
36. Šlechtová V, Bohlen J, Tan HH. Families of Cobitoidea (Teleostei; Cypriniformes) as revealed from nuclear genetic data and the position of the mysterious genera *Barbucca*, *Psilorhynchus*, *Serpenticobitis* and *Vaillantella*. Mol Phylogenet Evol. 2007;44:1358–65.
37. Chen W-J, Miya M, Saitoh K, Mayden RL. Phylogenetic utility of two existing and four novel nuclear gene loci in reconstructing Tree of Life of ray-finned fishes: the order Cypriniformes (Ostariophysi) as a case study. Gene. 2008;423:125–34.
38. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/96/NT. Nucleic Acids Res. 1999;41:95–8. http://www.mbio.ncsu.edu/BioEdit/bioedit.html. Accessed 21 Apr 2014.
39. Farris JS, Källerjö M, Kluge AG, Butt C. Testing significance of incongruence. Cladistics. 1994;10:315–9.
40. Swofford DL. PAUP*. Phylogenetic analysis using parsimony (*and other methods), v. 4.0 b10. Sunderland: Sinauer Associates; 2002.
41. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, et al. GenBank. Nucleic Acids Res. 2012;41:D36–42. http://www.ncbi.nlm.nih.gov/genbank/. Accessed 9 May 2015.
42. Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 2003;19:1572–4.
43. Jurka J, Kapitonov VV, Pavlíček A, Klonowski P, Kohany O, Walichiewicz J. Repbase update, a database of eukaryotic repetitive elements. Cytogenet Genome Res. 2005;110:462–7. www.girinst.org. Accessed 23 Apr 2014.
44. Meyne J, Ratliff RL, Moyzis RK. Conservation of the human telomere sequence $(TTAGGG)_n$ among vertebartes. Proc Natl Acad Sci U S A. 1989;86:7049–53.
45. Bănărescu P, Nalbant TT. A generical classification of Nemacheilinae with description of two new genera (Teleostei: Cypriniformes: Cobitidae), Travaux du Museum d'Histoire Naturelle "Grigore Antipa", vol. 35. 1995. p. 429–96.
46. Prokofiev AM. Morphological classification of loaches (Nemacheilinae). J Ichthyol. 2010;50:827–913.
47. Kottelat M. Indochinese Nemacheilines, a revision of nemacheiline loaches (Pisces: Cypriniformes) of Thailand, Burma, Laos, Cambodia and southern Viet Nam. München: Pfeil; 1990. p. 1–262.
48. Vasil'ev VP. Evolutionary karyology of fishes. Moscow: Nauka Press; 1985 (In Russian).
49. Rishi KK, Sharma MP, Mankotia R. Somatic chromosomes of three Indian teleosts. Matsya. 1977;3:6–9.
50. Kim IS, Lee EH, Son YM. Morphological variation and geographic distribution of two species of Nemacheiline loaches (Pisces, Cobitidae) from Korea. Korean J Zool. 1988;31:283–94 (In Korean with English abstract).
51. Klinkhardt M, Tesche M, Greven H. Database of fish chromosomes. Magdeburg: Westarp Wissenschaften; 1995.
52. Mank JE, Avise JC. Phylogenetic conservation of chromosome numbers in Actinopterygiian fishes. Genetica. 2006;127:321–7.
53. Ohno S. Evolution by gene duplication. New York: Springer; 1970.
54. Kohn M, Högell J, Vogel W, Minich P, Kehrer-Swatzki H, Graves JAM, et al. Reconstruction of a 450-My-old ancestral vertebrate protokaryotype. Trends Genet. 2006;22:203–10.
55. Matthey R. L' evolution de la formule chromosomiale chez les vertebrees. Experientia. 1945;1:50–6. 78–86.
56. Hoffmann AA, Rieseberg LH. Revisiting the impact of inversions in evolution: from population genetic markers to drivers of adaptive shifts and speciation? Annu Rev Ecol Evol Syst. 2008;39:21–42.
57. Faria R, Navarro A. Chromosomal speciation revisited: rearranging theory with pieces of evidence. Trends Ecol Evol. 2010;25:660–9.
58. Guerrero RF, Kirkpatrick M. Local adaptation and the evolution of chromosome fusions. Evolution. 2014;68:2747–56.
59. Cioffi MB, Bertollo LAC. Chromosomal distribution and evolution of repetitive DNAs in fish. In: Garrido-Ramos MA, editor. Repetitive DNA, Genome Dyn, vol. 7. 2012. p. 197–221.
60. King M. Species evolution: the role of chromosome change. Cambridge: University Press; 1993.
61. Hale DW, Greenbaum IF. Synapsis of a chromosomal pair heterozygous for a pericentric inversion and the presence of a heterochromatic short arm. Cytogenet Cell Genet. 1988;48:55–7.
62. Völker M, Sonnenberg R, Ráb P, Kullmann H. Karyotype differentiation in *Chromaphyosemion* killifishes (Cyprinodontiformes, Nothobranchiidae). III:

extensive karyotypic variability associated with low mitochondrial haplotype differentiation in *C. bivittatum*. Cytogenet Genome Res. 2007;116:116–26.
63. Schmid M, Guttenbach M. Evolutionary diversity of reverse (R) fluorescent chromosome bands in vertebrates. Chromosoma. 1988;97:101–14.
64. Ráb P, Rábová M, Reed KM, Phillips RB. Chromosomal characteristics of ribosomal DNA in the primitive semionotiform fish, longnose gar *Lepisosteus osseus*. Chromosome Res. 1999;7:475–80.
65. Fontana F, Tagliavini J, Congiu L. Sturgeon genetics and cytogenetics: recent advancements and perspectives. Genetica. 2001;111:359–73.
66. Pendás AM, Morán P, García-Vázquez E. Ribosomal RNA genes are interspersed throughout a heterochromatic chromosome arm in Atlantic salmon. Cytogenet Cell Genet. 1993;63:128–30.
67. Morescalchi MA, Liguori I, Rocco L, Stingo V. Karyotypic characterization and genomic organization of the 5S rDNA in *Erpetoichthys calabaricus* (Osteichthyes, Polypteridae). Genetica. 2007;131:209–16.
68. Deiana AM, Cau A, Salvadori S, Coluccia E, Cannas R, Milia A, et al. Major and 5S ribosomal sequences of the largemouth bass *Micropterus salmoides* (Perciformes, Centrarchidae) are localized in GC-rich regions of the genome. Chromosome Res. 2000;8:213–8.
69. Affonso PRAM, Galetti Jr PM. Chromosomal diversification of reef fishes from genus *Centropyge* (Perciformes, Pomacanthidae). Genetica. 2005;123:227–33.
70. de Lima-Filho PA, Bertollo LAC, Cioffi MB, Costa GWWF, Molina WF. Karyotype divergence and spreading of 5S rDNA sequences between genomes of two species: darter and emerald gobies (*Ctenogobius*, Gobiidae). Cytogenet Genome Res. 2014;142:197–203.
71. Caputo V, Marchegiani F, Sorice M, Olmo E. Heterochromatin heterogeneity and chromosome variability in four species of gobiid fishes (Perciformes: Gobiidae). Cytogenet Cell Genet. 1997;79:266–71.
72. Ene AC. Chromosomal polymorphism in the goby *Neogobius eurycephalus* (Perciformes: Gobiidae). Mar Biol. 2003;142:583–8.
73. Morescalchi MA, Stingo V, Capriglione T. Cytogenetic analysis in *Polypterus ornatipinnis* (Actinopterygii, Cladistia, Polypteridae) and 5S rDNA. Mar Genomics. 2011;4:25–31.
74. Rábová M, Ráb P, Boroń A, Bohlen J, Janko K, Šlechtová V, et al. Cytogenetics of bisexual species and their asexual hybrid clones in european spined loaches, genus *cobitis*. I. Karyotypes and extensive polymorphism of major ribosomal sites in four parental species. Abstracts of 16th european colloqium on animal cytogenetics and gene mapping. Cytogenet Genome Res. 2004;106:1–24.
75. Gromicho M, Ozouf-Costaz C, Collares-Pereira MJ. Lack of correspondence between CMA_3-, Ag-positive signals and 28S rDNA loci in two Iberian minnows (Teleostei, Cyprinidae) evidenced by sequential banding. Cytogenet Genome Res. 2005;109:507–11.
76. Gornung E. Twenty years of physical mapping of major ribosomal RNA genes across the teleosts: a review of research. Cytogenet Genome Res. 2013;141:90–102.
77. Hastie ND, Allshire RC. Human telomeres: fusion and interstitial sites. Trends Genet. 1989;5:326–31.
78. Redi CA, Garagna S, Della Valle G, Bottiroli G, Dell'Orto P, Viale G, et al. Differences in the organization and chromosomal allocation of satellite DNA between the European long tailed house mice *Mus domesticus* and *Mus musculus*. Chromosoma. 1990;99:11–7.
79. Molina WF, Galetti Jr PM. Robertsonian rearrangements in the reef fish *Chromis* (Perciformes, Pomacentridae) involving chromosomes bearing 5 s rRNA genes. Genet Mol Biol. 2002;25:373–7.
80. Rosa KO, Ziemniczak K, de Barros AV, Nogaroto V, Almeida MC, Cestari MM, et al. Numeric and structural chromosome polymorphism in *Rineloricaria lima* (Siluriformes: Loricariidae): fusion points carrying 5S rDNA or telomere sequence vestiges. Rev Fish Biol Fish. 2012;22:739–49.
81. Morán P, Martínez JL, Garcia-Vásquez E, Pendás AM. Sex chromosome linkage of 5S rDNA in rainbow trout (*Oncorhynchus mykiss*). Cytogenet Cell Genet. 1996;75:145–50.
82. Devlin RH, Nagahama Y. Sex determination and sex differentiation in fish: an overview of genetic, physiological, and environmental influences. Aquaculture. 2002;208:191–364.
83. Bertollo LAC, Born GG, Dergam JA, Fenocchio AS, Moreira-Filho O. A biodiversity approach in the neotropical Erythrinidae fish, *Hoplias malabaricus*. Karyotypic survey, geographic distribution of cytotypes and cytotaxonomic considerations. Chromosome Res. 2000;8:603–13.
84. Centofante L, Bertollo LAC, Moreira-Filho O. Cytogenetic characterization and description of an XX/XY_1Y_2 sex chromosome system in catfish *Harttia carvalhoi* (Siluriformes, Loricariidae). Cytogenet Genome Res. 2006;112:320–4.
85. de Oliveira RR, Feldberg E, dos Anjos MB, Zuanon J. Occurrence of multiple sexual chromosomes (XX/XY_1Y_2 and $Z_1Z_1Z_2Z_2/Z_1Z_2W_1W_2$) in catfishes of the genus *Ancistrus* (Siluriformes: Loricariidae) from the Amazon basin. Genetica. 2008;134:243–9.
86. Traut W, Sahara K, Otto TD, Marec F. Molecular differentiation of sex chromosomes probed by comparative genomic hybridization. Chromosoma. 1999;108:173–80.
87. Meyne J, Baker JR, Hobart HH, Hsu TC, Ryder OA, Ward OG, et al. Distribution of non-telomeric sites of the $(TTAGGG)_n$ telomeric sequence in vertebrate chromosomes. Chromosoma. 1990;99:3–10.
88. Ruiz-Herrera A, Nergadze SG, Santagostino M, Giulotto E. Telomeric repeats far from the ends: mechanisms of origin and role in evolution. Cytogenet Genome Res. 2008;122:219–28.
89. Ocalewicz K. Telomeres in fishes. Cytogenet Genome Res. 2013;141:114–25.
90. Pellegrino KCM, Rodrigues MT, Yonenaga-Yassuda Y. Chromosomal evolution in the Brazilian lizards of genus *Leposoma* (Squamata, Gymnopthalmidae) from Amazon and Atlantic rain forests: banding patterns and FISH of telomeric sequences. Hereditas. 1999;131:15–21.
91. Ocalewicz K, Furgala-Selezniow G, Szmyt M, Lisboa R, Kucinski M, Lejk AM, et al. Pericentromeric location of the telomeric DNA sequences on the European grayling chromosomes. Genetica. 2013;141:409–16.
92. Slijepcevic P. Telomeres and mechanisms of Robertsonian fusion. Chromosoma. 1998;107:136–40.
93. Garrido-Ramos MA, de la Herrán R, Rejón CR, Rejón CR. A satellite DNA of the Sparidae family (Pisces, Perciformes) associated with telomeric sequences. Cytogenet Cell Genet. 1998;83:3–9.
94. Cioffi MB, Martins C, Bertollo LAC. Chromosome spreading of associated transposable elements and ribosomal DNA in the fish *Erythrinus erythrinus*. Implications for genome change and karyoevolution in fish. BMC Evol Biol. 2010;10:271.
95. Guillén AKZ, Hirai Y, Tanoue T, Hirai H. Transcriptional repression mechanisms of nucleolus organizer regions (NORs) in humans and chimpanzees. Chromosome Res. 2004;12:225–37.
96. Amemiya CT, Gold JR. Chromosomal NORs as taxonomic and systematic characters in North American cyprinid fishes. Genetica. 1988;76:81–90.
97. Rábová M, Ráb P, Ozouf-Costaz C. Extensive polymorphism and chromosomal characteristics of ribosomal DNA in a loach fish, *Cobitis vardarensis* (Ostariophysi, Cobitidae) detected by different banding techniques and fluorescence *in situ* hybridization (FISH). Genetica. 2001;111:413–22.
98. Boron A, Ozouf-Costaz C, Coutanceau J-P, Woroniecka K. Gene mapping of 28S and 5S rDNA sites in the spined loach *Cobitis taenia* (Pisces, Cobitidae) from a diploid population and a diploid-tetraploid population. Genetica. 2006;128:71–9.
99. Gromicho M, Coutanceau J-P, Ozouf-Costaz C, Collares-Pereira MJ. Contrast between extensive variation of 28S rDNA and stability of 5S rDNA and telomeric repeats in the diploid-polyploid *Squalius alburnoides* complex and in its maternal ancestor *Squalius pyrenaicus* (Teleostei, Cyprinidae). Chromosome Res. 2006;14:297–306.
100. Pereira CSA, Aboim MA, Ráb P, Collares-Pereira MJ. Introgressive hybridization as a promoter of genome reshuffling in natural homoploid fish hybrids (Cyprinidae, Leuciscinae). Heredity. 2014;112:343–50.
101. Cazaux B, Catalan J, Veyrunes F, Douzery EJP, Britton-Davidian J. Are ribosomal DNA clusters rearrangement hotspots? A case study in the genus *Mus* (Rodentia, Muridae). BMC Evol Biol. 2011;11:124.
102. Pedrosa-Harand A, de Almeida CCS, Mosiolek M, Blair MW, Schweizer D, Guerra M. Extensive ribosomal DNA amplification during Andean common bean (*Phaseolus vulgaris* L.) evolution. Theor Appl Genet. 2006;112:924–33.
103. Milhomem SSR, Scacchetti PC, Pieczarka JC, Ferguson-Smith MA, Pansonato-Alves JC, O'Brien PCM, et al. Are NORs always located on homeologous chromosomes? A FISH investigation with rDNA and whole chromosome probes in *Gymnotus* fishes (Gymnotiformes). PLoS One. 2013;8:e55608.
104. Miller DA, Dev VG, Tantravahi R, Miller O. Supression of human nucleolar organizer activity in mouse-human somatic hybrid cells. Exp Cell Res. 1976;101:235–43.
105. Martins C, Wasko AP. Organization and evolution of 5S ribosomal DNA in the fish genome. In: Williams CR, editor. Focus on genome research. Hauppauge: Nova; 2004. p. 335–63.
106. Mantovani M, Abel LDS, Moreira-Filho O. Conserved 5S and variable 45S rDNA chromosomal localisation revealed by FISH in *Astyanax scabripinnis* (Pisces, Characidae). Genetica. 2005;123:211–6.

107. Martins C, Galetti Jr PM. Two 5S rDNA arrays in Neotropical fish species: is it a general rule for fishes? Genetica. 2001;111:439–46.
108. Martins C, Galetti Jr PM. Chromosomal localization of 5S rDNA genes in *Leporinus* fish (Anostomidae, Characiformes). Chromosome Res. 1999;7:363–7.
109. Drouin G, de Sá MM. The concerted evolution of 5S ribosomal genes linked to the repeat units of other multigene families. Mol Biol Evol. 1995;12:481–93.
110. Rebordinos L, Cross I, Merlo A. High evolutionary dynamism in 5S rDNA of fish: state of the art. Cytogenet Genome Res. 2013;141:103–13.
111. Pisano E, Ghigliotti L. Ribosomal genes in notothenioid fishes: focus on the chromosomal organisation. Mar Genomics. 2009;2:75–80.
112. Collares-Pereira MJ, Ráb P. NOR polymorphism in the Iberian species *Chondrostoma lusitanicum* (Pisces: Cyprinidae) – re-examination by FISH. Genetica. 1999;105:301–3.
113. Tigano C, Rocco L, Ferrito V, Costagliola D, Pappalardo AM, Stingo V. Chromosomal mapping and molecular characterization of ribosomal RNA genes in *Lebias fasciata* (Teleostei, Cyprinodontidae). Genetica. 2004;121:95–100.
114. Eickbush TH, Eickbush DG. Finely orchestrated movements: evolution of the ribosomal RNA genes. Genetics. 2007;175:477–85.
115. Raskina O, Barber JC, Nevo E, Belyayev A. Repetitive DNA and chromosomal rearrangements: speciation-related events in plant genomes. Cytogenet Genome Res. 2008;120:351–7.
116. Raskina O, Belyayev A, Nevo E. Quantum speciation in *Aegilops*: molecular cytogenetic evidence from rDNA cluster variability in natural populations. Proc Natl Acad Sci U S A. 2004;101:14818–23.
117. Zhang X, Eickbush MT, Eickbush TH. Role of recombination in the long-term retention of transposable elements in rRNA gene loci. Genetics. 2008;180:1617–26.
118. da Silva M, Matoso DA, Vicari MR, de Almeida MC, Margarido VP, Artoni RF. Physical mapping of 5S rDNA in two species of knifefishes: *Gymnotus pantanal* and *Gymnotus paraguensis* (Gymnotiformes). Cytogenet Genome Res. 2011;134:303–7.
119. Nakajima RT, Cabral-de-Mello DC, Valente GT, Venere PC, Martins C. Evolutionary dynamics of rRNA gene clusters in cichlid fish. BMC Evol Biol. 2012;12:198.
120. Costa GWWF, Cioffi MB, Bertollo LAC, Molina WF. Transposable elements in fish chromosomes: a study in the marine *Cobia* species. Cytogenet Genome Res. 2013;141:126–32.
121. Silva DMZA, Pansonato-Alves JC, Utsunomia R, Daniel SN, Hashimoto DT, Oliveira C, et al. Chromosomal organization of repetitive DNA sequences in *Astyanax bockmanni* (Teleostei, Characiformes): dispersive location, association and co-localization in the genome. Genetica. 2013;141:329–36.
122. Symonová R, Majtánová Z, Sember A, Staaks GBO, Bohlen J, Freyhof J, et al. Genome differentiation in a species pair of coregonine fishes: an extremely rapid speciation driven by stress-activated retrotransposons mediating extensive ribosomal DNA multiplications. BMC Evol Biol. 2013;13:42.
123. Ichiyanagi K, Nishihara H, Duvernell DD, Okada N. Acquisition of endonuclease specificity during evolution of L1 retrotransposon. Mol Biol Evol. 2007;24:2009–15.
124. Gibbons JG, Branco AT, Yu S, Lemos B. Ribosomal DNA copy number is coupled with gene expression variation and mitochondrial abundance in humans. Nat Commun. 2014;5:4850.
125. Gibbons JG, Branco AT, Godinho SA, Yu S, Lemos B. Concerted copy number variation balances ribosomal DNA dosage in human and mouse genomes. Proc Natl Acad Sci U S A. 2015;112:2485–90.
126. Chénais B, Caruso A, Hiard S, Casse N. The impact of transposable elements on eukaryotic genomes: from genome size increase to genetic adaptation to stressful environments. Gene. 2012;509:7–15.
127. Casacuberta E, González J. The impact of transposable elements in environmental adaptation. Mol Ecol. 2013;22:1503–17.
128. Ozouf-Costaz C, Brandt J, Körting C, Pisano E, Bonillo C, Coutanceau J-P, et al. Genome dynamics and chromosomal localization of the non-LTR retrotransposons *Rex1* and *Rex3* in Antarctic fish. Antarct Sci. 2004;16:51–7.

PLOS | ONE

RESEARCH ARTICLE

Asexual Reproduction Does Not Apparently Increase the Rate of Chromosomal Evolution: Karyotype Stability in Diploid and Triploid Clonal Hybrid Fish (Cobitis, Cypriniformes, Teleostei)

Zuzana Majtánová[1,2]*, Lukáš Choleva[1,3], Radka Symonová[1,4], Petr Ráb[1], Jan Kotusz[5], Ladislav Pekárik[6,7], Karel Janko[1,3]

1 Laboratory of Fish Genetics, Institute of Animal Physiology and Genetics, CAS, v.v.i, Liběchov, Czech Republic, **2** Department of Zoology, Faculty of Science, Charles University in Prague, Prague, Czech Republic, **3** Department of Biology and Ecology, Faculty of Science, University of Ostrava, Ostrava, Czech Republic, **4** Research Institute for Limnology, University of Innsbruck, Mondsee, Austria, **5** Museum of Natural History, University of Wrocław, Wrocław, Poland, **6** Institute of Botany, SAS, Bratislava, Slovakia, **7** Department of Biology, Faculty of Education, Trnava University, Trnava, Slovakia

* majtanova@iapg.cas.cz

Citation: Majtánová Z, Choleva L, Symonová R, Ráb P, Kotusz J, Pekárik L, et al. (2016) Asexual Reproduction Does Not Apparently Increase the Rate of Chromosomal Evolution: Karyotype Stability in Diploid and Triploid Clonal Hybrid Fish (Cobitis, Cypriniformes, Teleostei). PLoS ONE 11(1): e0146872. doi:10.1371/journal.pone.0146872

Editor: Vincent Laudet, Ecole normale superieure de Lyon, FRANCE

Received: December 16, 2014

Accepted: December 25, 2015

Published: January 25, 2016

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This study was supported by the Grant Agency of the Charles University in Prague (www.cuni.cz; project number 158110), Czech Science Foundation (www.gacr.cz; project number GAP506/10/1155 to KJ, PR and LC; GPP506/12/P857 to LC; 14-02940S PR and RS; GA13-12580S to KJ and LC), by the Academy of Sciences of the Czech Republic (www.cas.cz; project number RVO 67985904) and by the National Science Centre in Poland (http://www.

Abstract

Interspecific hybridization, polyploidization and transitions from sexuality to asexuality considerably affect organismal genomes. Especially the last mentioned process has been assumed to play a significant role in the initiation of chromosomal rearrangements, causing increased rates of karyotype evolution. We used cytogenetic analysis and molecular dating of cladogenetic events to compare the rate of changes of chromosome morphology and karyotype in asexually and sexually reproducing counterparts in European spined loach fish (*Cobitis*). We studied metaphases of three sexually reproducing species and their diploid and polyploid hybrid clones of different age of origin. The material includes artificial F1 hybrid strains, representatives of lineage originated in Holocene epoch, and also individuals of an oldest known age to date (roughly 0.37 MYA). Thereafter we applied GISH technique as a marker to differentiate parental chromosomal sets in hybrids. Although the sexual species accumulated remarkable chromosomal rearrangements after their speciation, we observed no differences in chromosome numbers and/or morphology among karyotypes of asexual hybrids. These hybrids possess chromosome sets originating from respective parental species with no cytogenetically detectable recombinations, suggesting their integrity even in a long term. The switch to asexual reproduction thus did not provoke any significant acceleration of the rate of chromosomal evolution in *Cobitis*. Asexual animals described in other case studies reproduce ameiotically, while *Cobitis* hybrids described here produce eggs likely through modified meiosis. Therefore, our findings indicate that the effect of asexuality on the rate of chromosomal change may be context-dependent rather than universal and related to particular type of asexual reproduction.

ncn.gov.pl; project number DEC-2011/03/B/NZ8/02095) to JK. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Introduction

Most species differ by their karyotype, which are routinely defined by the variation in number, size and morphology of chromosomes. Differences originating during karyotype evolution play an important role in speciation [1]. Simultaneously, the rate at which chromosome rearrangements accumulate varies greatly from one species to another. For example, changes in chromosome number have occurred roughly 20 times faster in mammals than in frogs [2] (although even there one can find lineages with exceptionally fast rate of chromosomal evolution, e.g. [3]). With more than 27.000 taxonomically known species [4], teleost fishes show rather conserved karyotypes with a uniform haploid chromosome number n = 24–25 or close to it [5,6], suggesting a slow rate of karyotype change, mostly by intrachromosomal rearrangements [7]. However, a number of exceptions exist, especially in lineages of polyploid origin (see [8]).

The reasons why karyotypes remain conservative in some groups while they readily mutate and rearrange in others continue to be the focus of ongoing and lively debate (e.g. [9]). At least some factors have been proposed to modify rates of karyotype evolution. One of them, polyploidization, has been shown to instantaneously increase the chromosome numbers [10] but also to directly increase rates of chromosome breakages and aneuploidy incidence [11]. Asexuality (defined in a broader sense by the lack of efficient recombination and segregation; e.g. [12]) is another factor that is supposed to significantly influence the rate of chromosomal changes [13]. Relaxed constraints on the pairing of homologous chromosomes during gamete formation in clones may allow faster accumulation of chromosomal mutations than in sexual species (e.g. [14]). An absence of effective recombination ultimately leads to an accumulation of independent mutations on any locus including whole chromosomes (Meselson effect; [15]). Chromosomal complements of asexual organisms, where most of them are also polyploids, may therefore acquire strongly aneuploid or structurally changed characteristics relatively easily. The hypothesis of asexuality-linked increase in the rate of karyotype evolution is supported by studies of apomictic arthropods [13,14,16–18] and root knot nematodes [19] which suggest that once sex is lost, chromosomal structure can become irregular extremely quickly; e.g. chromosomal fissions in *Sitobion miscanthi* aphids occurred over several dozens of generations during seven years in laboratory culture [13].

However, number of uncertainties remains. Especially, it is unclear whether the aforementioned indications of increased rates of karyotype changes could be generally attributed to asexual reproduction *per se* because other features of those organisms, especially the holokinetic nature of their chromosomes, may also influence resulting patterns [20]. Moreover, since abovementioned studies were focused on apomictic parthenogens (i.e. with eggs formed by mitosis) it is unknown whether other types of asexual reproduction have the same effect on the rate of morphological change of karyotype. Actually, asexuality can be achieved by many cytological mechanisms, which have profoundly different predictions about the heterozygosity of resulting individuals [21,22]. This may affect the rate of chromosomal evolution differently [14]. Certain types of automixis may be exactly equivalent to apomixis [23], e.g. premeiotic endoduplication. In that case, the genetic material is duplicated before entering meiosis while segregation and recombination take place on bivalents that form between sister chromatids, the genetically identical copies derived from endoduplication event. The progeny is therefore genetically identical to its mother [24]. Other types of automixis may lead to intragenomic recombination or exclusions of large genomic parts (e.g. [25]). Therefore, it is important to test the general applicability of the hypothesis of asexuality-driven increase of karyotype mutation rate.

Spined loaches from the *Cobitis taenia* hybrid complex (*sensu* Janko *et al.* [26]) are fish model with included automictic reproduction, allowing us to study karyotype changes in

vertebrate automicts as an alternative to published cases from invertebrate apomicts. Primary hybridization between *C. elongatoides* and either *C. taenia* or *C. tanaitica* (two sister species distantly related to *C. elongatoides*; [27]) leads to diploid hybrid clonal lineages reproducing via gynogenesis. In this reproductive mode, the sperm-dependent parthenogenetic females produce unreduced eggs that are only activated but not truly fertilized by sperm from males of one of the parental species. Therefore, males serve as hosts for these so called clonal sexual parasites [28–30]. Clonal spined loaches have retained meiosis [31] but unreduced egg production is ensured by premeiotic genome doubling [32], a similar mechanism to the clonal pond loach *Misgurnus anguillicaudatus* [33]. Interestingly, having originated as diploid hybrids, these lineages expanded over large parts of the European continent (for distribution of particular *Cobitis* hybrid population types in Europe see [34,35]). They also largely tend to increase the ploidy level by accidental incorporation of sperm nucleus into unreduced clonal eggs at fertilization via genome addition, resulting in polyploid (mostly 3n) clonal lineages [26,28,29].

Considering that interspecific hybridization with subsequent clonal formation has been proceeding continuously at least since the last interglacial [26], contemporary clonal assemblages represent a mixture of very recent as well as relatively old clonal hybrid lineages. The oldest asexual lineage (referred to as the 'Hybrid clade I') apparently predates the last glacial maximum and its age was estimated to be roughly 0.3 MYA [26].

In the present study we compare rates of morphological evolution of karyotypes of three sexual species (*C. elongatoides*, *C. taenia* and *C. tanaitica*) and their wild-caught diploid and triploid hybrid biotypes with asexual reproduction. Hybrid biotypes differ by age represented by three classes: F1 laboratory-produced clonal hybrids; individuals derived from recent (Holocene) hybridizations; individuals derived from ancient (interglacial) hybridizations. Using conventional (Giemsa staining) and molecular (genomic *in situ* hybridization; GISH) cytogenetic methods we test following tasks: i) whether visible changes in chromosome number and/or morphology occurs in karyotypes of hybrid biotypes; ii) whether the original parental chromosomal sets can be unambiguously identified in hybrid karyotypes of clones of different age and iii) whether large-scale recombinations and/or rearrangements can be detected. Karyological observations were put into phylogenetic and time-based context with a newly dated phylogeny of parental species as well as hybrids by using available calibration point. This point was studied both on a relative and absolute time scales by visualizing the changes in karyotypes of parental species onto dated phylogeny. The dating allowed testing of the hypothesis whether asexual reproduction and/or hybridity have notably increased the rate of chromosomal changes in comparison with karyologically diversified sexual lineages. We particularly focused on whether karyotypes of older clonal lineages differ from contemporary parental species more than those recently derived.

Materials and Methods

Background knowledge about karyotypes of parental species

Sexual parental species included in this study are *C. elongatoides*, *C. taenia* and *C. tanaitica* and hereafter we use the acronyms E, T and N for their haploid genomes, respectively. These are used to designate species and also hybrids with various genomic combinations (e.g. triploid hybrid biotype containing two haploid genomes of *C. elongatoides* and one haploid genome of *C. tanaitica* is encoded EEN). To evaluate rates of karyotype evolution in sexual species and hybrid clones, we take an advantage of previous karyological studies of European spined loaches. While no intraspecific karyotype differences has been discovered so far, it has been shown that karyotypes of parental species have diverged from each other by various chromosomal rearrangements involving one probable fusion and numerous pericentric inversions

[34,36]. In contrast with the substantial karyotype differentiation among parental species, several studies demonstrated consistency of karyotypes within species across their distribution ranges [34,36–38]. Multilocus phylogenetic study [27] suggested the sister relationship of sexual *C. taenia* and *C. tanaitica* species with *C. elongatoides* being their distant relative. In concordance with this hypothesis, *C. elongatoides* (2n = 50) possesses the most differentiated karyotype dominated by meta- and submetacentric chromosomes, while karyotypes of *C. tanaitica* and *C. taenia* are more similar to each other and dominated by subtelo- and acrocentric chromosomes. They differ, however, in number of chromosomes and their centromere positions with *C. taenia* (2n = 48) having more acrocentrics than *C. tanaitica* (2n = 50). A detailed karyotype characteristic of sexual species is given in Table 1.

Animals used in this study

Altogether 57 *Cobitis* individuals of nine biotypes were used in the study (Table 2). The examined animals included di- and triploid clonal forms of various genomic constitutions, i.e. EN, ET, EEN, ENN, EET and ETT. Since the genomic composition of species and hybrid forms may not be identified by external morphology because of their cryptic nature, the genome of each individual was determined by species-specific allozyme markers (*Aat*, *Mdh*, *Pgm*, *Gpi-A* and *Ldh*) according to Janko *et al.* [28]. Interspecific hybrids mostly came from native populations in Europe from Bulgaria (2 localities, 4 individuals), Germany (3 localities, 4 individuals), Poland (2 localities, 10 individuals), Romania (4 localities, 18 individuals), Slovakia (1 locality, 8 individuals) and Czech Republic (1 locality, 1 individual). In addition, one laboratory-born first-generation (F1) ET and three backcrossed (B1) ETT hybrids were analysed as control specimens to test the karyotype stability and to standardize the GISH approach for correct classification of chromosomes between hybrid and parental complements. The maternal ancestry of a subset of studied hybrid individuals and their divergence from parental species were determined using mitochondrial cytochrome *b* gene (cyt *b*) according to Janko *et al.* [30].

Ethics statement

All specimens were collected in accordance with the national legislation of the countries concerned. The experimental procedures involving fish were approved by the Institutional Animal Care and Use Committee of the Institute of Animal Physiology and Genetics, CAS, v.v.i, according with directives from State Veterinary Administration of the Czech Republic, permit number 217/2010, and by the permit number CZ 00221 issued by Ministry of Agriculture of the Czech Republic. LC has Certificate of competency according to §17 of the Czech Republic Act No. 246/1992 coll. on the Protection of Animals against Cruelty (Registration number: CZU 955/06), provided by Central Commission for Animal Welfare, which authorizes animal experiments in the Czech Republic.

Chromosome preparation

Metaphase chromosomes were prepared according to Ráb and Roth [39] with slight modifications. Briefly, fish were injected with 0.1% colchicine solution (1 ml/100 g of body weight) 45 min before being sacrificed using an overdose of 2-phenoxyethanol anaesthetics agent (Sigma). The kidneys were removed, dissected in 0.075 M KCl at room temperature and the cell suspension free of tissue fragments was hypotonized for 8 min in 0.075 M KCl, fixed in fresh prepared fixative (methanol: acetic acid 3:1, v/v), washed twice in fixative, and finally spread onto slides. We also used a preparation of chromosomes from regenerated caudal fins according to Völker and Ráb [40].

Table 1. Karyotype characteristics of *Cobitis* genomes with various ploidy levels and genomic compositions of biotypes involved in this study.

ploidy level	(n)			(2n)					(3n)			
genome composition	E	T	N	EE	TT	NN	ET	EN	EET	ETT	EEN	ENN
metacentric chromosomes	11	5	5	22	10	10	16	16	27	21	27	21
submetacentric chromosomes	13	9	13	26	18	26	22	26	35	31	39	39
subtelocentric chromosomes	1	1	3	2	2	6	2	4	3	3	5	7
acrocentric chromosomes	-	9	4	-	18	8	9	4	9	18	4	8
total number of chromosomes	25	24	25	50	48	50	49	50	74	73	75	75

Abbreviations: Capital letters represent sets of haploid genomes: E, *Cobitis elongatoides*; T, *C. taenia*; N, *C. tanaitica*.

doi:10.1371/journal.pone.0146872.t001

Table 2. Individuals used in this study.

Biotype	Country	Locality	Lat	Long	AOC	NOI	NOMKA	NOGE	NOAGM
EE	Hungary	Szodrakosz Cr.	47°43'58.8"N	19°07'58.8"E	-	1	5	-	-
	Poland	Budkowiczanka R.	50°50'50.1"N	18°11'07.1"E	-	1*	-	-	-
TT	Germany	Haaren Cr.	53°04'58.8"N	7°49'58.8"E	-	2* + 1	4	-	-
NN	Romania	Danube R.	44°04'47.9"N	26°43'51.2"E	-	2*	-	-	-
	Romania	Sinoe at Histria	44°37'58.8"N	28°52'58.8"E	-	1	4	-	-
ET	Czech Rep.	Laboratory F1	50°24'37.6"N	14°27'16.9"E	F1 generation	1	9	1	4
	Poland	Dolna Barycz R.	51°36'59.1"N	16°30'49.1"E	Holocene	4	27	2	4
total						5	36	3	8
EN	Bulgaria	Jantra R.	43°09'58.0"N	25°55'53.8"E	hybrid clade I	2	22	1	1
	Romania	Danube R.	44°04'47.9"N	26°43'51.2"E	hybrid clade I	3	15	2	9
total						5	37	3	10
EET	Germany	Neisse R.	51°51'28.8"N	14°36'34.9"E	Holocene	1	12	1	8
	Poland	Polska Woda R.	51°31'17.0"N	17°30'07.0"E	Holocene	3	16	-	-
	Poland	Dolna Barycz R.	51°36'59.1"N	16°30'49.1"E	Holocene	2	5	1	2
	Czech Rep.	Pšovka Cr.	50°22'11.8"N	14°33'06.8"E	Holocene	1	2	-	-
total						7	35	2	10
ETT	Germany	Issel R.	51°51'00.0"N	06°15'00.0"E	Holocene	2	13	2	3
	Germany	Ilmenau R.	53°22'34.0"N	10°14'36.5"E	Holocene	1	2	-	-
	Czech Rep.	Laboratory B1	50°24'37.6"N	14°27'16.9"E	B1 generation	3	17	2	10
total						6	32	4	13
EEN	Bulgaria	Vit R.	43°15'47.0"N	24°19'30.1"E	hybrid clade I	2	56	2	6
	Poland	Polska Woda R.	51°31'17.0"N	17°30'07.0"E	hybrid clade I	1	7	1	1
	Romania	Barlat R.	46°40'23.8"N	27°40'07.5"E	Holocene	3	6	-	-
	Romania	Comana R.	44°10'07.1"N	26°08'54.9"E	hybrid clade I	1	2	-	-
	Romania	Adjud R.	46°04'21.0"N	27°12'32.2"E	Holocene	1	2	-	-
	Slovakia	Cierna voda R.	48°36'27.0"N	21°59'34.1"E	hybrid clade I	8	51	5	17
total						16	124	8	24
ENN	Romania	Danube R.	44°04'47.9"N	26°43'51.2"E	Holocene	10	35	3	12
total						10	35	3	12

Abbreviations: Capital letters represent sets of haploid genomes: E, *Cobitis elongatoides*; T, *C. taenia*; N, *C. tanaitica*. Shortcuts in the column captions: AOC, age of clones; NOI, number of individuals; NOMKA, number of metaphases karyologically analysed; NOGE, number of GISH experiments; NOAGM, number of analysed GISH metaphases.

In order to prepare probes for GISH experiments only gDNA was used from individuals marked with *.

doi:10.1371/journal.pone.0146872.t002

Karyotype analysis

Metaphases of each individual were inspected by Giemsa or DAPI staining in order to confirm number and morphology of their chromosomes. In total 299 metaphases of 49 hybrids were inspected (Table 2). Metaphases were classified using the nomenclature proposed by Levan *et al.* [41] and classified into four categories: metacentrics, submetacentrics, subtelocentrics and acrocentrics. Subset of metaphases from 23 hybrid individuals was arranged in a decreasing size order to demonstrate consistency of karyotypes (Fig 1A–1F and S1A–S1R Fig, S1 and S2 Tables).

Probe preparation, GISH

A subset of 23 hybrid individuals was used for further GISH analyses to determine whether the original parental chromosomal sets are distinguishable in hybrid karyotypes. At least two GISH experiments with different individuals were successfully performed *per* biotype (S1 and S3 Tables). Probes used in GISH experiments were prepared from whole genomic DNA (gDNA) of pure parental species *C. elongatoides*, *C. taenia* and *C. tanaitica*. gDNA was extracted from muscles or fins using the DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. gDNA samples were labelled with biotin-16-dUTP (Roche, Mannheim, Germany) and digoxigenin-11-dUTP (Roche) using a Nick Translation Mix (Roche) following the protocol supplied by the manufacturer. The best results were obtained after 1.5–5 h (time strongly depends on the starting gDNA quality and fragment length) of nick translation until labelled DNA fragments were approximately 200–500 bp long. Species-specific hybridization probes combined gDNA of sexual species (*C. elongatoides* and *C. taenia*, or *C. elongatoides* and *C. tanaitica*) to perform GISH experiments on chromosomes of hybrids (Table 2). Salmon sperm was used as blocking reagent for repetitive DNA. The hybridization and detection procedure were carried out under conditions described by Symonová *et al.* [42]. The biotin-dUTP labelled probes were detected using streptavidin-FITC (Invitrogen, San Diego, Calif., USA). The digoxigenin-dUTP labelled probes were detected using anti-digoxigenin-rhodamin (Roche). The chromosomes were counterstained with Vectashield/DAPI (1.5 mg/ml) (Vector, Burlingame, Calif., USA).

Image processing

Chromosomal preparations were examined by an Olympus Provis AX 70 epifluorescence microscope. Images of metaphase chromosomes were recorded with a cooled Olympus DP30BW CCD camera. The IKAROS and ISIS imaging programs (Metasystems, Altlussheim, Germany) were used to analyse grey-scale images. The captured digital images from GISH experiments were pseudocoloured (blue for DAPI, red for anti-digoxigenin-rhodamin, green for streptavidin-FITC) and superimposed using Microimage and Adobe Photoshop software, version CS5, respectively.

Estimates of clonal ages and speciation times in mutational-time units using the coalescent methods

We have applied molecular clocks in order to evaluate the absolute time scale over which observed karyotype changes have accumulated among parental sexual species and their hybrids. For estimates of ages of clonal lineages we used cytoplasmic markers (mtDNA) because it is haploid and traces the clonal evolution to its original formation event. We note that estimation clonal ages from nuclear markers is much more complicated since nucleus of clones is of hybrid nature and subject of polyploidization through subsequent genome

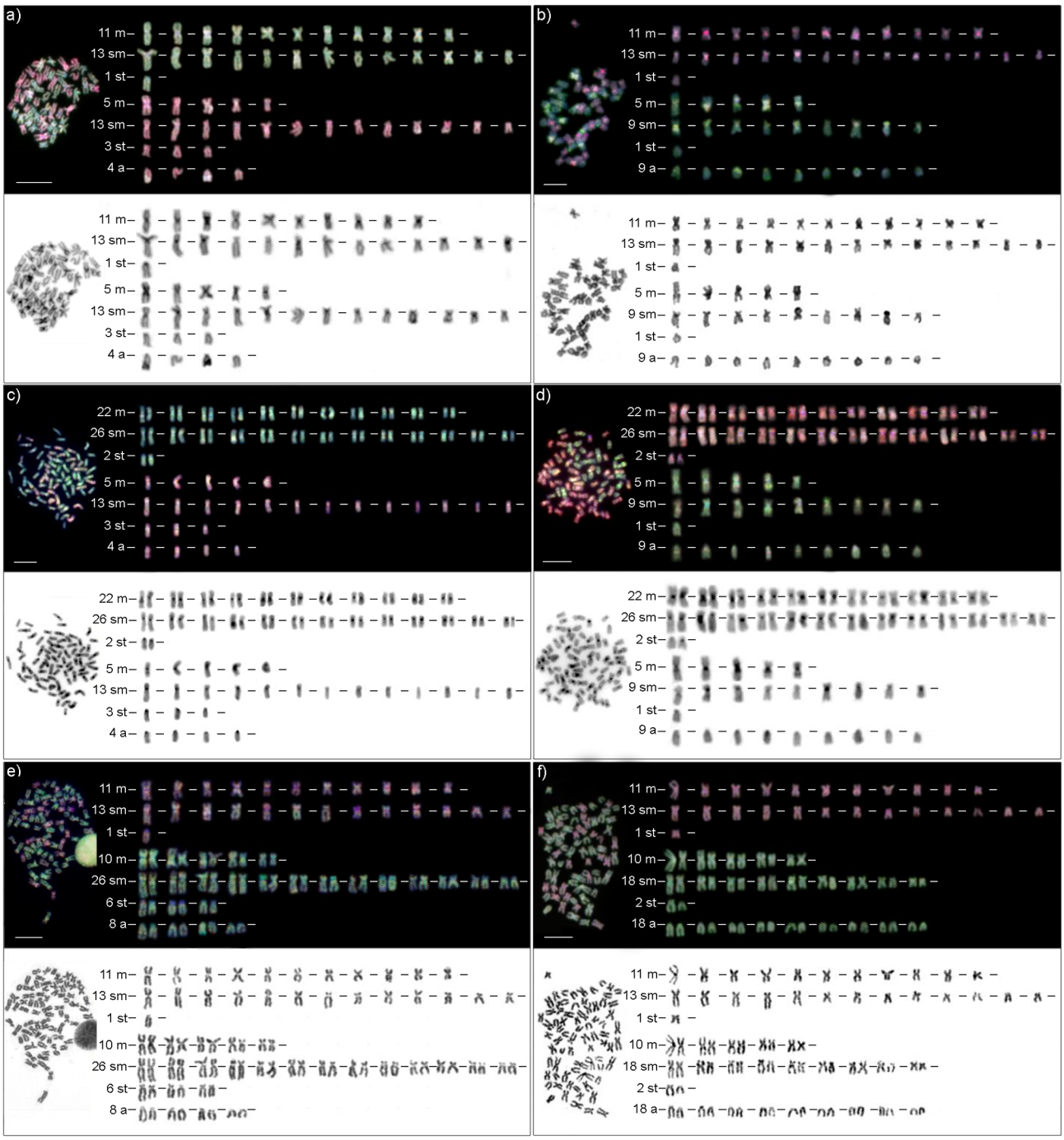


Fig 1. Representative karyotypes of hybrid biotypes after GISH and/or DAPI/Giemsa staining. (A) EN hybrid metaphase with hybridization pattern of *Cobitis elongatoides* gDNA in green, *C*. *tanaitica* gDNA in red. (B) ET hybrid with *C*. *elongatoides* in red, *C*. *taenia* in green. (C) EEN hybrid with *C*. *elongatoides* in green, *C*. *tanaitica* in red. (D) EET hybrid with *C*. *elongatoides* in red, *C*. *taenia* in green. (E) ENN hybrid with *C*. *elongatoides* in red, *C*. *tanaitica* in green. (F) ETT hybrid with *C*. *elongatoides* in red, *C*. *taenia* in green. Capital letters represent haploid genome sets: E, *C*. *elongatoides*; N, *C*. *tanaitica*; T, *C*. *taenia*. Chromosomes were arranged in a decreasing size order and classified in four morphological groups: metacentric (m), submetacentric (sm), subtelocentric (st) and acrocentric (a). Probes labelled with biotin-16-dUTP were detected with FITC-streptavidin (green signals on chromosomes); probes labelled with digoxigenin-11-dUTP were detected with anti-digoxigenin-rhodamin (red signals on chromosomes). To visualize the morphology of chromosomes DAPI (A, B, C, D) or Giemsa (E, F) stained karyotype was used. Captured DAPI stained karyotypes were inverted. Bars equal 5 μm. Detail information about individuals used is provided in S1 Table.

doi:10.1371/journal.pone.0146872.g001

incorporations ('leaky gynogenesis' *sensu* Janko *et al.* [28]). On the other hand, for the estimation of parental species speciation time, nuclear markers (nDNA) had to be used since Choleva *et al.* [27] demonstrated massive mtDNA introgression from *C. elongatoides* to *C. tanaitica* suggesting its mtDNA diversity reflects secondary hybrid introgression rather than its speciation from sister species *C. taenia*.

Given the aforementioned necessity to use distinct sets of markers for divergence time estimates of clones and sexual species, respectively we performed two-step analysis to obtain, comparable time estimates from both types of markers. First (1), we applied coalescent method to independently estimate the clonal ages and speciation times on the relative time scale (expressed in mutational-time units that are scaled by locus-specific mutation rates). In a second step (2), we translated such estimates into absolute-time units by consistently calibrating both markers (mtDNA and nDNA) with the same outgroup species.

1) Coalescence analysis: The age of the oldest known *Cobitis* clonal lineage was estimated as the time to the most recent common ancestor (TMRCA) for all cyt *b* sequences of asexual individuals sampled by Janko *et al.* [26] that cluster in the so called 'Hybrid clade I'. TMRCA was estimated by coalescent method of Griffiths and Tavare [43] and it was implemented in the Genetree software by R. C. Griffiths, which is particularly suitable because the observed pedigree of the Hybrid clade I conformed to the infinite-allele model (we observed no evidence for multiple mutational hits on sequence positions within this clade). We first determined the Maximum Likelihood (ML) value of θ ($\theta = 2N_f\mu$ for haploid maternally inherited mitochondrial locus; N_f represents the effective population size of females, and μ represents the mutation rate per locus per generation). Using the ML estimate of θ (1.156) we performed one million simulations of the coalescent process providing us with an estimate of TMRCA and 95% confidence intervals in coalescent units.

Speciation times in mutation units of *C. elongatoides*, *C. taenia* and *C. tanaitica* as well as their contemporary and ancestral population sizes, θ_{TN}, θ_N, θ_E, θ_{TNE} and θ_{TN} ($\theta = 4N_e\mu$) were estimated from sequences of nine nuclear loci from Choleva *et al.* [27]. The calculation was performed with the Bayesian Markov chain Monte Carlo (MCMC) algorithms implemented in BPP Version 2.0b [44], which allow for the incorporation of relative mutation rates among loci and multiple haplotypes per species. The analysis was performed under the assumption of (TT, NN) monophyly in the true species tree (see Choleva *et al.* [27]). As recommended in Yang [45], the mutation rate scalars (i.e. the information about the relative substitution rates of individual loci) were obtained by among-locus comparison of average distances between ingroup and outgroup species. As an outgroup we used the sexual species *C. paludica*, which is a phylogenetically distant taxon suggesting that the coalescent error (i.e. the deviations among loci caused by polymorphism within the common ancestor of in-group and outgroup taxa) would not be of a large magnitude [45]. With this outgroup the MCMC was run under conditions described in Choleva *et al.* [27].

2) In the second step, we converted abovementioned estimates into absolute-time units, which require a calibration point on *Cobitis* phylogeny that is suitable for employed nuclear markers. Since *Cobitis* phylogeny has been mostly addressed by mtDNA loci we first constructed an ultrametric tree of published cyt *b* sequences covering all major *Cobitis* lineages (dataset from Tang *et al.* [46]) using previously published calibration point, i.e. the opening of the Gibraltar strait (~5.33 MYA), which separated two sister species, i.e. *C. maroccana* and *C. paludica* [47]. Although the usage of this calibration point may have limitations due to its recency, obtained ultrametric tree provided two relevant facts: i) it allowed estimating the branch specific mutation rate of mitochondrial cyt *b*, which was used to convert relative estimates of clonal age into absolute time units, and ii) it provided an estimate of absolute divergence time between ingroup species and *C. paludica* outgroup. This time estimate was

subsequently used to convert relative estimates of speciation times from nuclear markers into absolute time units and hence, both mtDNA and nDNA markers were calibrated consistently.

The data of Tang *et al.* [46] were used for ultrametric tree construction with two modifications. Firstly, we excluded sequences of taxonomically problematic species belonging to *Misgurnus* and *Paramisgurnus* genera (see Šlechtová *et al.* [48] discussion of this topic). Secondly, we replaced the *C. paludica* sequence from Tang *et al.* [46] with the *C. paludica* specimen from the Valle River, Spain (GenBank accession No: AY860180.1), as recommended by Doadrio and Perdices [47] because this population represents a phylogenetically closer lineage to the African *C. maroccana* and thus minimizes the coalescent error. Using such a modified dataset, we reconstructed the ML tree of cyt *b* sequences with RAxML [49] and ultrameterised it with the penalized likelihood method (PL) [50] implemented in R package Ape ([51]; S2 Fig). PL is a computationally inexpensive method that is nonetheless more robust against violations of model assumptions than strict-clock and relaxed-clock models [52]. ML estimates of branching times and branch-specific substitution rates depend on the lambda parameter which controls the variation of substitution rate change among branches (lambda may oscillate between zero -saturated model with one distinct rate for each branch—and infinity—model converging to a strict clock). Cross validation criterion removing one tip each time was used to select the optimal lambda value to construct the final ultrametric tree and obtain the ML estimates of branching times and substitution rates for each branch.

Having obtained the lineage-specific mutation rates, the TMRCA of the Hybrid clade I was translated into the time in years (t) as $t = \mathrm{TMRCA}^{*} N_f^{*} g$ using the generation time (g) of two years (N_f was estimated from θ using the ML estimate of substitution rates provided by PL). In order to accommodate the uncertainty around mutation rate of clonal hybrids, which combine genetic material from two or more species, clonal age was estimated using an average rate over the branches leading to the species that participate in the origin of asexual *Cobitis* hybrids. Speciation times were translated into absolute time units by taking the ML estimate of the divergence time of *C. paludica* and ingroup species from the PL analysis.

Results

Karyotype analysis of hybrid metaphases

Giemsa stained metaphases of two diploid (EN, ET) and four triploid (EEN, ENN, EET and ETT) hybrid biotypes were analysed in order to evaluate the morphological stability of their karyotypes. Chromosomes were sorted out into one of four categories (meta-, submeta-, subtelo- and acrocentric chromosomes) based on their morphology. Without exception, all complete metaphases of all hybrid biotypes contained chromosomal numbers corresponding to appropriate combinations of haploid parental genomes involved, namely EN hybrid had 2n = 50, ET had 2n = 49, ENN and EEN had 3n = 75, ETT had 3n = 73, and EET had 3n = 74 (Fig 1, S1 Fig, Table 1). Also, numbers of chromosomes in particular morphological categories of all hybrids matched those expected by combining the karyotypes of parental species (e.g. in diploid hybrid of EN genomic constitution we observed 16 metacentrics– 11 metacentrics derived from E genome and 5 from the N genome; other categories could be derived similarly). These data are summarized in Table 1 and we further provide Giemsa or DAPI stained karyotypes of 23 individuals from all biotypes in Fig 1 and S1 Fig to document the consistency of such observations. Information about individuals used for karyotyping is given in S1 and S2 Tables.

We further focused on the differences among hybrids along the time gradient of their historical origin. We observed no detectable changes in karyotypes of hybrid clones irrespective of their age. In particular, Fig 1A and 1C, S1A–S1E Fig display representatives of ancient Hybrid

clade I, Fig 1B and 1E, and S1F–S1O, S1Q and S1R Figs show representatives of lineage originated in the Holocene epoch; Fig 1D and 1F and S1P Fig shows the first generation hybrid obtained from recent laboratory crosses of parental species (S1 and S2 Tables).

Identification of parental chromosomes in karyotypes of hybrid individuals based on GISH

Cytogenetic analyses were performed on metaphases of all 2n and 3n hybrid biotypes in order to identify parental chromosomal sets, and to detect potential large-scale recombinations and/or rearrangements. Reliability of conventional approach was validated by GISH analysis of first generation hybrids in which we clearly sorted karyotypes into haploid sets categorized after parental species (Fig 1D and 1F, S3H and S3L Fig). After having applied GISH to wild caught hybrids, we got the same result and clearly assigned chromosomes to haploid chromosomal sets of sexual species involved in the primary hybridizations (i.e., to *C. elongatoides*, *C. taenia*, or *C. tanaitica*; Fig 1 and S3 Fig). For example, in GISH experiment to EN hybrid metaphases, the *C. elongatoides* signals were green and *C. tanaitica* signals were red (Fig 1A). According to GISH signal ratio we observed 11 meta-, 13 submeta- and 1 subtelocentric chromosome in green colour. Number and morphology of these chromosomes correspond to haploid set of chromosomes of pure *C. elongatoides* (Table 1). The rest of 50 chromosomes from EN hybrid metaphase were coloured in red and contained 5 meta-, 13 submeta-, 3 subtelo- and 4 acrocentric chromosomes which exactly matched to the haploid chromosome number of pure *C. tanaitica* (Fig 1A, Table 1). Analogously, after GISH experiments with all other hybrid biotypes (Fig 1B–1F and S3 Fig), the chromosomes matched by numbers and by morphology to combination of haploid and diploid chromosome set of respective sexual species. The summary of these observations is provided in Table 1.

In concordance with observation from Giemsa stained metaphases no detectable differences in karyotypes were observed among hybrids of different age of origin (Fig 1, S3 Fig, S1 and S3 Tables). No visible evidence for large-scale intergenomic exchanges was observed in the diploid and triploid hybrid metaphases of hybrids under this study. In some metaphases we found parts of individual chromosomes with balanced signals from both parental species but such observation resulted from non-specific hybridization rather than from recombination events and was never confirmed in other metaphases from the same individual including the same slide. GISH hybridization pattern was not uniform over the whole chromosomal length. The hybridization signals were stronger in the pericentromeric regions, particularly in metacentric chromosomes and in the telomeric regions of many other chromosomes (Fig 1 and S3 Fig).

Age estimates of studied clones and speciation times of parental species

The cyt *b* tree topology reconstructed with RAxML matched Tang *et al.* [46]. The cross-validation selected a relatively high lambda value (10^2), which is consistent with Bohlen's *et al.* [53] finding of no significant deviation from strict molecular clock in *Cobitis* cyt *b*. Using such a lambda value for PL method, we found relatively older cladogenetic events compared to Tang *et al.* [46] since these authors used a more distant *C. paludica* population as a putative sister lineage to *C. maroccana* for calibration, while we used a more closely related sequence from the Valle River, Spain (see [47]). The average cyt *b* substitution rate found by the PL method in the clade of hybridizing *Cobitis* species equalled 0.0065 per site per MYA. Having applied this rate to the ML estimate of the TMRCA of the Hybrid clade I provided a clonal age equal to 0.371 MYA (95% C I, 0.172–0.670 MYA) (Fig 2).

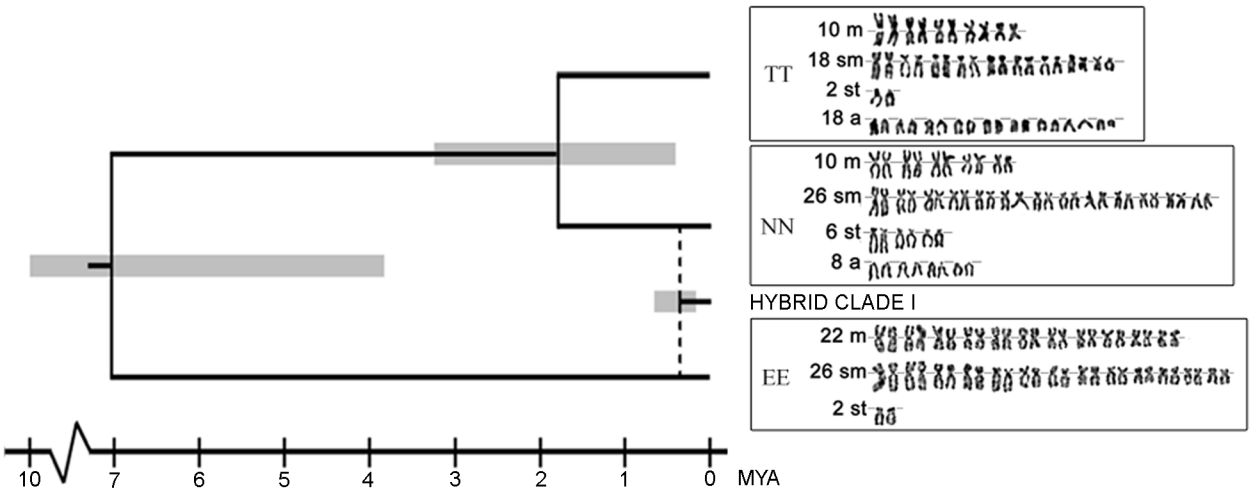


Fig 2. Ultrametric phylogenetic tree demonstrating estimated speciation times of parental species and the Hybrid clade I. Species-specific karyotypes arranged from Giemsa stained chromosomes are shown along the right side of cladogram. Confidence intervals of nodes of interest are in grey colour. TT, *C. taenia*; NN, *C. tanaitica*; EE, *Cobitis elongatoides*. Chromosomes were arranged in a decreasing size order and classified in four morphological groups: metacentric (m), submetacentric (sm), subtelocentric (st) and acrocentric (a).

doi:10.1371/journal.pone.0146872.g002

The ML estimate of a divergence time between *C. paludica* and the studied species within the *C. taenia* hybrid complex equalled 16.17 MYA. Using this divergence time as a calibration point for each nuclear locus from [27], the BPP software suggested that *C. elongatoides* diverged from the *C. taenia*–*C. tanaitica* ancestor during Miocene about 7 MYA (3.83–10.28 MYA), while both latter species diverged from each other for the period of Pleistocene-Pliocene about 1.8 MYA (0.42–3.25 MYA) (Fig 2).

Altogether, the results represent the hybrid biotypes under this study as representatives of a wide gradient of evolutionary ages (Table 2). The youngest hybrids are clearly represented by the first-generation individuals arisen by laboratory crosses. mtDNA analysis further revealed that natural hybrids formed two types with respect to their age. First type of hybrids shared haplotypes with the maternal species, demonstrating their recent origin during the Holocene epoch (see [26,30] for details). Second type of natural hybrids possessed haplotype E29 or E30 (*sensu* Janko *et al.* [26]) representing the oldest known *Cobitis* clonal lineage, the Hybrid clade I (Table 2).

Discussion

Stability of parental chromosomes in hybrid metaphases and their identification based on GISH

Analyses of Giemsa stained metaphases of hybrid clones enabled sorting of all chromosomes into morphological categories whose counts corresponded to expected numbers upon combining haploid complements of parental species. These observations were confirmed by GISH because chromosomal staining by species specific probes allowed unambiguous distinguishing of parental components in genomes of all studied hybrids. The observed pattern shows that chromosomes of hybrids did not undergo large-scale restructuring otherwise one would expect a failure to assign them to the morphological categories corresponding to karyotypes of respective parental species. Karyotypes of particular biotypes were uniform, regardless of whether these were of recent or ancient origin (including the oldest known *Cobitis* clonal lineage—the Hybrid clade I). This indicates the long-term stability of karyotypes of these asexuals. Our

findings are not trivial from several reasons. Application of GISH to identify parental components in hybrids of cold-blooded vertebrates often faced technical hurdles stemmed from their small sized chromosomes. Hence, only a few previous studies reported successful results allowing further interpretations ([25] and subsequent references on *Ambystoma* salamanders, [54–62]). Moreover, our findings contrast other asexual organisms, e.g. *Sitobion miscanthi* where extra chromosomal element (originated by fission) was observed after seven years or less in laboratory culture [13].

Although various cytological mechanisms are supposed to maintain the long-term integrity of clonally inherited chromosomes, asexual organisms may have so called 'minimal sex' (i.e. occasional incorporation of low amount of alien genetic material into ova [63]) or sometimes switch between phases of sexual and asexual reproduction [64]. Such 'sexual periods' are expected to produce recombinant patterns of hybrid asexuals lineages, which were not observed in *Cobitis* hybrids studied here. Moreover, asexual transmission of genomes is seldom perfect, especially in automictic asexual lineages where limited recombinations between co-inherited genomes have been found (e.g. [25,63]). This may lead to a notable introgression of one parental genome onto the genomic background of the other one. Such intragenomic recombinations have the consequence in less heterozygous progeny. Arai and Mukaino [65] found that genetically variable progeny may also be formed by automictic organisms with premeiotic endoduplication and attributed such an observation to occasional formation of bivalents between the chromosomes from different parental taxa. Mitotic intraclonal gene conversion or recombination was also considered as one of possible explanations for exceptionally high rate of homozygosity in species-diagnostic SNP markers of hybrid clonal fish *Poecilia formosa* [66].

In contrast to aforementioned studies, this study showed that asexually transmitted *Cobitis* genomes maintain their integrity even in relatively old clonal lineages. We found no evidence of recombination between homeologous chromosomes from different parental species despite the fact that gametogenesis in hybrid *Cobitis* involves quasi normal meiosis [31,32]. Data from this and previous studies of *Cobitis* (see above) suggest that intraclonal recombinations among chromosomal complements originating from hybridizing parental species seem to be either absent or rare as they have not accumulated to a notable extent in hybrid clonal lineages whose ages span over hundreds of thousands of generations (Hybrid clade I). Although cytogenetic approach has a limitation in resolution power, there is possibility of detection of large-scale recombination as observed in unisexual salamanders [25]. On the other hand there is a case of *Pelophylax* hybrid complex where ongoing and site-specific interspecific genetic transfer mediated by hybridogenetic hybrids has been evidenced [63,67]. GISH experiments on metaphases of diploid hybrid *P. esculentus* which originate through hybridization between *P. ridibundus* and *P. lessonae* have detected two clear groups of chromosomal constituents, each originating from one of the sexual parental species but no intergenomic exchanges in metaphase chromosomes [54]. Similarly in *Cobitis* hybrid complex, if potential recombination occurred, these are small-scale changes and undetectable by conventional karyotype analyses or by GISH method. Our finding of integrity of karyotypes in *Cobitis* hybrids is well in line with previous cytological studies [31,32], experimental crossing experiments (e.g. [29]), and multilocus genotyping [28,30]. The result congruently indicates that *Cobitis* clones maintain a fixed heterozygosity in essayed species-specific loci. The only source of intra-lineage variation known so far stems from post formational mutations and incorporations of entire paternal genomes leading to increased ploidy of the clonally reproducing progeny. The integrity of clonal *Cobitis* genomes may theoretically be explained by high sequence divergence between parental species which prevents large-scale chromosomal recombinations. However, Bi and Bogart [25] described intergenomic recombination blocks on chromosomes of *Ambystoma* hybrids composed of *A.*

laterale and *A*. *jeffersonianum* parental species. Despite authors did not provide divergence time of these species, Robertson *et al*. [68] showed that their divergence in cyt *b* (~ 10%) is more than twice as large than between *A*. *laterale* and *A*. *texanum* (~ 4%), which diversified about 10 MYA [69]. Recombination can therefore occur between genomes of species that are even more diverged than *Cobitis* under study (~ 7 MYA). Stability of Cobitid clonal genomes may, therefore, have another explanation.

Is there an evidence for increased rates of chromosomal evolution?

The genetic variability of extant asexuals is derived from the variability possessed by their sexual progenitors at the time of primary hybridization. It means the variability of sexual genomes became frozen in forming clonal lineages. The clonal progeny observed nowadays thus bears the combination of genomes derived from the first generation of an interspecific hybrid. Therefore, our finding that parental chromosomal complements in clones exactly match morphological categories of karyotypes in contemporary sexual relatives suggests that neither the karyotypes of clones, nor the karyotypes of sexual species have accumulated morphological changes since the origin of hybrid lineages sampled in natural populations.

According to the molecular dating, parental species diverged from each other during Miocene and Pleistocene, respectively. Since then, their karyotypes differentiated in both number and categories of chromosomes [34,36] (Fig 2, Table 1) including the large changes in the positions of centromeres (particularly dynamic regions of chromosomal evolution in most eukaryotes [70]), as well as centric fusion. We cannot distinguish whether these changes have accumulated gradually or have been acquired during or right after speciation events. However, taking into account the level of karyotype differentiation among sexual species and the fast mutation rates documented in some clonal organisms [13,14,16–19], one would expect to observe at least some morphologically visible mutations accumulated in karyotypes of the much older *Cobitis* Hybrid clade I (ca 0.37 MYA which represents over one fifth of the time since the separation of *C*. *taenia* and *C*. *tanaitica*). Instead of increased rates of karyotype evolution, our current data suggest that asexual karyotypes lacked any notable reorganization despite they have been clonally transmitted more than several hundreds of thousands of generations since the initial clonal formation.

Abovementioned molecular dating should be considered with a large grain of salt due to large confidence intervals around the age estimates and other well-known problems inherent in molecular clock dating. However, another point that strongly supports the conclusion that clonally transmitted karyotypes evolve conservatively over long periods comes from biogeographic data. These imply that the clonal Hybrid clade I has survived at least one and probably more interglacials in multiple separated refugee [26]. We compared Hybrid clade I individuals that were sampled from areas that belong to different Danubian refugial areas as identified by Janko *et al*. [26]. Here, samples from Bulgaria and Romania represented the Lower Danubian refugium, samples from Eastern Slovakia represented the Pannonian refugium, and samples from Poland aimed to cover Central European/upper Danubian refugium (Table 2). Yet, we did not find karyotype differentiation among Hybrid clade I individuals originating from those distinct populations, which most likely evolved in isolation [26] suggesting that their identical karyotypes cannot be explained by homogenizing gene flow. Rather, we prefer the explanation that karyotypes of different refugial populations did not accumulate any morphological changes at least since the last interglacial.

We are aware that our conclusions are valid only for the resolution power of applied methods and sample number. Also, we may not exclude the accumulation of small-scale chromosomal mutational changes. For example, we observed indications of the accumulation of

heterochromatin in the pericentromeric region in hybrid karyotypes. The phenomenon of the stronger signals in the centromeric regions is similar to the one described in *Ambystoma* salamanders [25], *Pelophylax* water frogs [54], and *Brassica* plants [71]. This may be explained by a higher degree of homeology of repetitive sequences on chromosome arms than at the centromeres with consequent weak blocking by the competitor DNA and a higher concentration of species specific repetitive sequences at centromeres [71]. Further investigation of the nature of karyotypic variation and genome evolution in *Cobitis* fishes will require the isolation and characterization of the major repetitive DNA sequences in the heterochromatin of each species to determine how homogeneous they are, and to map the sites of particular sequences on chromosomes by *in situ* hybridization. The same applies for mobile genetic elements such as retrotransposons, which are expected to degenerate in asexuals with the absence of meiosis [72,73], but there is nothing known about asexuals with modified meiosis, as for example in *Cobitis* hybrids.

Conclusion

Conventional karyological studies of apomictic parthenogens show that asexuality increases the rate of chromosomal evolution [13,14,16,18] and could lead to accumulation of visible chromosome changes on short evolutionary time scale [13]. We applied conventional as well as molecular methods on karyotypes of automictic *Cobitis* hybrids and observed no morphological structural changes even in longer time scales. This indicates an interesting possibility that the effect of asexuality on the rate of chromosomal evolution is rather context-dependent than universal, i.e. it may depend e.g. on particular type of asexual reproduction. Future studies should take this possibility explicitly into account.

Supporting Information

S1 Fig. Representative karyotypes of hybrid biotypes after DAPI/Giemsa staining. (A-C) EN hybrids. (D-F) EEN hybrids. (G-I) ENN hybrids. (J-L) ET hybrids. (M-O) EET hybrids. (P-R) ETT hybrids. Chromosomes were arranged in a decreasing size order and classified in four morphological groups: metacentric (m), submetacentric (sm), subtelocentric (st) and acrocentric (a). To visualize the morphology of chromosomes DAPI (K and Q) or Giemsa (A-J, L-P and R) stained karyotype was used. Captured DAPI stained karyotypes were inverted. Bars equal 5 μm. Detail information about individuals used is provided in S2 Table.
(TIF)

S2 Fig. Maximum likelihood phylogenetic tree of *Cobitis* and *Sabanejewia* species. ML phylogenetic tree constructed from cytochrome *b* gene sequences from Tang *et al.* [46] and from Doadrio and Perdices [47]. Calibration point follows Doadrio and Perdices [47]. Widths of branches are proportional to ML estimate of substitution rate. *Cobitis* species are highlighted (grey colour), species mentioned in this study are highlighted.
(TIF)

S3 Fig. Representative metaphases of hybrid biotypes after GISH experiments. (A, B) EN hybrids. (C, D) EEN hybrids. (E, F) ENN hybrids. (G, H) ET hybrids. (I, J) EET hybrids. (K, L) ETT hybrids. Probes labelled with biotin-16-dUTP were detected with streptavidin-FITC (green signals on chromosomes); probes labelled with digoxigenin-11-dUTP were detected with anti-digoxigenin-rhodamin (red signals on chromosomes). Bars equal 5 μm. Detail information about individuals and hybridization patterns is provided in S3 Table.
(TIF)

S1 Table. Hybrid individuals used for GISH experiments presented in Fig 1.
(DOCX)

S2 Table. Hybrid individuals used for karyotyping presented in S1 Fig.
(DOCX)

S3 Table. List of hybrid individuals used for GISH experiments presented in S3 Fig.
(DOCX)

Acknowledgments

The authors would like to express their gratitude to Jörg Bohlen for his help with obtaining samples, Věra Šlechtová and Jana Kopecká for their help with the laboratory work. We are very grateful to František Marec for his useful comments and corrections and to Christopher Murray Johnson and for language revision. We also thank the Axios Review service for valuable comments on an earlier version of the manuscript.

Author Contributions

Conceived and designed the experiments: RS. Performed the experiments: ZM. Analyzed the data: LC KJ ZM. Contributed reagents/materials/analysis tools: LC KJ PR ZM. Wrote the paper: ZM LC RS PR JK LP KJ. Designed the whole study: PR KJ LC. Collected and raised material: LC KJ ZM JK LP.

References

1. Rieseberg LH. Chromosomal rearrangements and speciation. Trends Ecol Evol. 2001; 16: 351–358. doi: 10.1016/S0169-5347(01)02187-5 PMID: 11403867
2. Wilson AC, Maxson LR, Sarich VM. Two types of molecular evolution. Evidence from studies of interspecific hybridization. Proc Natl Acad Sci U S A. 1974; 71: 2843–2847. PMID: 4212492
3. Schmid M, Steinlein C, Bogart JP, Feichtinger W, León P, La Marca E, et al. The chromosomes of terraranan frogs. Insights into vertebrate cytogenetics. Cytogenet Genome Res. 2010; 130–131: 1–568. doi: 10.1159/000301339 PMID: 21063086
4. Nelson JS. Fishes of the world. Fourth edition. Nelson, Joseph S.; Department of Biological Sciences, University of Alberta, Edmonton, Alberta T6G 2E9, Canada, Canada.: John Wiley & Sons, Inc.; 2006.
5. Kohn M, Högel J, Vogel W, Minich P, Kehrer-Sawatzki H, Graves JAM, et al. Reconstruction of a 450-My-old ancestral vertebrate protokaryotype. Trends Genet TIG. 2006; 22: 203–210. doi: 10.1016/j.tig.2006.02.008 PMID: 16517001
6. Mank JE, Avise JC. Phylogenetic conservation of chromosome numbers in Actinopterygiian fishes. Genetica. 2006; 127: 321–327. doi: 10.1007/s10709-005-5248-0 PMID: 16850236
7. Arai R. Fish karyotypes: a check list. 2011 edition. Tokyo ; New York: Springer; 2011.
8. Ravi V, Venkatesh B. Rapidly evolving fish genomes and teleost diversity. Curr Opin Genet Dev. 2008; 18: 544–550. doi: 10.1016/j.gde.2008.11.001 PMID: 19095434
9. Phillips R, Ráb P. Chromosome evolution in the Salmonidae (Pisces): an update. Biol Rev Camb Philos Soc. 2001; 76: 1–25. PMID: 11325050
10. Otto SP, Whitton J. Polyploid incidence and evolution. Annu Rev Genet. 2000; 34: 401–437. doi: 10.1146/annurev.genet.34.1.401 PMID: 11092833
11. Mayer VW, Aguilera A. High levels of chromosome instability in polyploids of *Saccharomyces cerevisiae*. Mutat Res Mol Mech Mutagen. 1990; 231: 177–186. doi: 10.1016/0027-5107(90)90024-X
12. Martens K, Loxdale HD, Schön I. The elusive clone—in search of its true nature and identity. In: Schön I, Martens K, Dijk P, editors. Lost Sex. Springer Netherlands; 2009. pp. 187–200. Available: http://link.springer.com/chapter/10.1007/978-90-481-2770-2_9
13. Sunnucks P, England PR, Taylor AC, Hales DF. Microsatellite and chromosome evolution of parthenogenetic *Sitobion aphids* in Australia. Genetics. 1996; 144: 747–756. PMID: 8889535

14. Schneider MC, Cella DM. Karyotype conservation in 2 populations of the parthenogenetic scorpion *Tityus serrulatus* (Buthidae): rDNA and its associated heterochromatin are concentrated on only one chromosome. J Hered. 2010; 101: 491–496. doi: 10.1093/jhered/esq004 PMID: 20231264

15. Welch MD, Meselson M. Evidence for the evolution of bdelloid rotifers without sexual reproduction or genetic exchange. Science. 2000; 288: 1211–1215. PMID: 10817991

16. Normark BB. Evolution in a putatively ancient asexual aphid lineage: recombination and rapid karyotype change. Evolution. 1999; 53: 1458. doi: 10.2307/2640892

17. Spence JM, Blackman RL. Inheritance and meiotic behaviour of a de novo chromosome fusion in the aphid *Myzus persicae* (Sulzer). Chromosoma. 2000; 109: 490–497. PMID: 11151679

18. Welch JLM, Welch DBM, Meselson M. Cytogenetic evidence for asexual evolution of bdelloid rotifers. Proc Natl Acad Sci U S A. 2004; 101: 1618–1621. doi: 10.1073/pnas.0307677100 PMID: 14747655

19. Triantaphyllou AC. Oogenesis and the chromosomes of the parthenogenic root-knot nematode *Meloidogyne incognita*. J Nematol. 1981; 13: 95–104. PMID: 19300730

20. Blackman RL, Spence JM, Normark BB. High diversity of structurally heterozygous karyotypes and rDNA arrays in parthenogenetic aphids of the genus *Trama* (Aphididae: Lachninae). Heredity. 2000; 84 (Pt 2): 254–260. PMID: 10762396

21. Suomalainen E, Saura A, Lokki J. Cytology in evolution of parthenogenesis. 1 edition. Boca Raton, Fla: CRC Press; 1987.

22. Stenberg P, Saura A. Cytology of asexual animals. In: Schön I, Martens K, Dijk P, editors. Lost Sex. Springer Netherlands; 2009. pp. 63–74. Available: http://link.springer.com/chapter/10.1007/978-90-481-2770-2_4

23. Normark BB, Judson OP, Moran NA. Genomic signatures of ancient asexual lineages. Biol J Linn Soc. 2003; 79: 69–84. doi: 10.1046/j.1095-8312.2003.00182.x

24. Lutes AA, Neaves WB, Baumann DP, Wiegraebe W, Baumann P. Sister chromosome pairing maintains heterozygosity in parthenogenetic lizards. Nature. 2010; 464: 283–286. doi: 10.1038/nature08818 PMID: 20173738

25. Bi K, Bogart JP. Identification of intergenomic recombinations in unisexual salamanders of the genus *Ambystoma* by genomic *in situ* hybridization (GISH). Cytogenet Genome Res. 2006; 112: 307–312. doi: 10.1159/000089885 PMID: 16484787

26. Janko K, Culling MA, Ráb P, Kotlík P. Ice age cloning—comparison of the Quaternary evolutionary histories of sexual and clonal forms of spiny loaches (*Cobitis*; Teleostei) using the analysis of mitochondrial DNA variation. Mol Ecol. 2005; 14: 2991–3004. doi: 10.1111/j.1365-294X.2005.02583.x PMID: 16101769

27. Choleva L, Musilova Z, Kohoutova-Sediva A, Paces J, Rab P, Janko K. Distinguishing between incomplete lineage sorting and genomic introgressions: complete fixation of allospecific mitochondrial DNA in a sexually reproducing fish (*Cobitis*; Teleostei), despite clonal reproduction of hybrids. PLoS ONE. 2014; 9: e80641. doi: 10.1371/journal.pone.0080641 PMID: 24971792

28. Janko K, Bohlen J, Lamatsch D, Flajshans M, Epplen JT, Ráb P, et al. The gynogenetic reproduction of diploid and triploid hybrid spined loaches (*Cobitis*: Teleostei), and their ability to establish successful clonal lineages on the evolution of polyploidy in asexual vertebrates. Genetica. 2007; 131: 185–194. doi: 10.1007/s10709-006-9130-5 PMID: 17216551

29. Choleva L, Janko K, De Gelas K, Bohlen J, Šlechtová V, Rábová M, et al. Synthesis of clonality and polyploidy in vertebrate animals by hybridization between two sexual species. Evol Int J Org Evol. 2012; 66: 2191–2203. doi: 10.1111/j.1558-5646.2012.01589.x

30. Janko K, Kotusz J, De Gelas K, Šlechtová V, Opoldusová Z, Drozd P, et al. Dynamic formation of asexual diploid and polyploid lineages: multilocus analysis of *Cobitis* reveals the mechanisms maintaining the diversity of clones. PLoS ONE. 2012; 7: e45384. doi: 10.1371/journal.pone.0045384 PMID: 23028977

31. Saat TV. Reproduction of the diploid and polyploid spinous loach (*Cobitis*, Teleostei), oocyte maturation and fertilization in the triploid form. Russ J Dev Biol. 1991; 22: 332–338.

32. Juchno D, Boroń A, Spóz A, Kujawa R, Kolczyńska J. Diplotene chromosomes of oocytes of polyploid hybrid *Cobitis* (Pisces, Cobitidae). Konferencja Embriologiczna, Acta Biologica Cracoviensia,series Botanica. Poznań; 2014. p. 64.

33. Itono M, Morishima K, Fujimoto T, Bando E, Yamaha E, Arai K. Premeiotic endomitosis produces diploid eggs in the natural clone loach, *Misgurnus anguillicaudatus* (Teleostei: Cobitidae). J Exp Zoolog A Comp Exp Biol. 2006; 305: 513–523. doi: 10.1002/jez.a.283

34. Janko K, Flajšhans M, Choleva L, Bohlen J, Šlechtová V, Rábová M, et al. Diversity of European spined loaches (genus *Cobitis* L.): an update of the geographic distribution of the *Cobitis taenia* hybrid complex

with a description of new molecular tools for species and hybrid determination. J Fish Biol. 2007; 71: 387–408. doi: 10.1111/j.1095-8649.2007.01663.x

35. Choleva L, Apostolou A, Ráb P, Janko K. Making it on their own: sperm-dependent hybrid fishes (*Cobitis*) switch the sexual hosts and expand beyond the ranges of their original sperm donors. Philos Trans R Soc B Biol Sci. 2008; 363: 2911–2919. doi: 10.1098/rstb.2008.0059

36. Rábová M, Pelikánová Š, Choleva L, Ráb P. Cytogenetics of bisexual species and their asexual hybrid clones in European spined loaches, genus *Cobitis*. II. Mapping of telomeric (TTAGGG)n sequences and DAPI-positive heterochromatins in four parental species. ECI XII European Congress of Ichthyology, Book of Abstracts. Cavtat, Croatia; 2007. p. 59.

37. Boroń A. Replication banding patterns in the spined loach, *Cobitis taenia* L. (Pisces, Cobitidae). Genetica. 2003; 119: 51–55. PMID: 12903746

38. Ráb P, Bohlen J, Rábová M, Flajšhans M, Kalous L. Cytogenetics as a tool in fish conservation: the present situation in Europe. In: Pisano E, Ozouf-Costaz C, Foresti F, Kapoor BG, editors. Fish Cytogenetics. Enfield, NH 03748, USA: Science Publishers; 2007. pp. 215–241.

39. Ráb P, Roth P. Cold-blooded vertebrates. In: Balíček P, Forejt J, Rubeš J, editors. Methods of chromosome analysis. Brno: Cytogenet Sect Cs Biol Soc Publishers; 1988. pp. 115–124.

40. Völker M, Ráb P. Direct chromosome preparation from regenerating fin tissue. In: Ozouf-Costaz C, Pisano E, Foresti F, Foresti de Almeida-Toledo L, editors. Fish Cytogenetic Techniques. Enfield, NH 03748, USA: CRC Press, Inc.; 2015. pp. 37–41.

41. Levan A, Fredga K, Sandberg AA. Nomenclature for centromeric position on chromosomes. Hereditas. 1964; 52: 201–220. doi: 10.1111/j.1601-5223.1964.tb01953.x

42. Symonová R, Sember A, Majtánová Z, Ráb P. Characterization of fish genomes by GISH and CGH. In: Ozouf-Costaz C, Pisano E, Foresti F, de Almeida L, editors. Fish Cytogenetic Techniques. CRC Press; 2015. pp. 118–131. Available: http://www.crcnetbase.com/doi/10.1201/b18534-17

43. Griffiths RC, Tavare S. Simulating probability distributions in the coalescent. Theor Popul Biol. 1994; 46: 131–159. doi: 10.1006/tpbi.1994.1023

44. Yang Z, Rannala B. Bayesian species delimitation using multilocus sequence data. Proc Natl Acad Sci. 2010; 200913022. doi: 10.1073/pnas.0913022107

45. Yang Z. Likelihood and Bayes estimation of ancestral population sizes in hominoids using data from multiple loci. Genetics. 2002; 162: 1811–1823. PMID: 12524351

46. Tang Q, Freyhof J, Xiong B, Liu H. Multiple invasions of Europe by East Asian cobitid loaches (Teleostei: Cobitidae). Hydrobiologia. 2008; 605: 17–28. doi: 10.1007/s10750-008-9296-1

47. Doadrio I, Perdices A. Phylogenetic relationships among the Ibero-African cobitids (*Cobitis*, cobitidae) based on cytochrome b sequence data. Mol Phylogenet Evol. 2005; 37: 484–493. PMID: 16150615

48. Šlechtová V, Bohlen J, Perdices A. Molecular phylogeny of the freshwater fish family Cobitidae (Cypriniformes: Teleostei): delimitation of genera, mitochondrial introgression and evolution of sexual dimorphism. Mol Phylogenet Evol. 2008; 47: 812–831. doi: 10.1016/j.ympev.2007.12.018 PMID: 18255319

49. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics. 2006; 22: 2688–2690. doi: 10.1093/bioinformatics/btl446 PMID: 16928733

50. Sanderson MJ. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. Mol Biol Evol. 2002; 19: 101–109. PMID: 11752195

51. Paradis E. Analysis of phylogenetics and evolution with R. 1 edition. New York: Springer; 2006.

52. Paradis E. Molecular dating of phylogenies by likelihood methods: a comparison of models and a new information criterion. Mol Phylogenet Evol. 2013; 67: 436–444. doi: 10.1016/j.ympev.2013.02.008 PMID: 23454091

53. Bohlen J, Perdices A, Doadrio I, Economidis PS. Vicariance, colonisation, and fast local speciation in Asia Minor and the Balkans as revealed from the phylogeny of spined loaches (Osteichthyes; Cobitidae). Mol Phylogenet Evol. 2006; 39: 552–561. PMID: 16439160

54. Zaleśna A, Choleva L, Ogielska M, Rábová M, Marec F, Ráb P. Evidence for integrity of parental genomes in the diploid hybridogenetic water frog *Pelophylax esculentus* by genomic *in situ* hybridization. Cytogenet Genome Res. 2011; 134: 206–212. doi: 10.1159/000327716 PMID: 21555873

55. Rampin M, Bi K, Bogart JP, Collares-Pereira MJ. Identifying parental chromosomes and genomic rearrangements in animal hybrid complexes of species with small genome size using Genomic *In Situ* Hybridization (GISH). Comp Cytogenet. 2012; 6: 287–300. doi: 10.3897/CompCytogen.v6i3.3543 PMID: 24260669

56. Pereira CSA, Rab P, Collares-Pereira MJ. Chromosomes of Iberian Leuciscinae (Cyprinidae) revisited: evidence of genome restructuring in homoploid hybrids using dual-color FISH and CGH. Cytogenet Genome Res. 2013; 141: 143–152. doi: 10.1159/000354582 PMID: 24107574

57. Pereira CSA, Aboim MA, Ráb P, Collares-Pereira MJ. Introgressive hybridization as a promoter of genome reshuffling in natural homoploid fish hybrids (Cyprinidae, Leuciscinae). Heredity. 2014; 112: 343–350. doi: 10.1038/hdy.2013.110 PMID: 24220087

58. Knytl M, Kalous L, Symonová R, Rylková K, Ráb P. Chromosome studies of European cyprinid fishes: cross-species painting reveals natural allotetraploid origin of a *Carassius* female with 206 chromosomes. Cytogenet Genome Res. 2013; 139: 276–283. doi: 10.1159/000350689 PMID: 23652770

59. Symonová R, Flajšhans M, Sember A, Havelka M, Gela D, Kořínková T, et al. Molecular cytogenetics in artificial hybrid and highly polyploid sturgeons: An evolutionary story narrated by repetitive sequences. Cytogenet Genome Res. 2013; 141: 153–162. doi: 10.1159/000354882 PMID: 24051427

60. Pokorná M, Rens W, Rovatsos M, Kratochvíl L. A ZZ/ZW sex chromosome system in the thick-tailed Gecko (*Underwoodisaurus milii*; Squamata: Gekkota: Carphodactylidae), a member of the ancient gecko lineage. Cytogenet Genome Res. 2014; 142: 190–196. doi: 10.1159/000358847 PMID: 24603160

61. Zhu H-P, Gui J-F. Identification of genome organization in the unusual allotetraploid form of *Carassius auratus gibelio*. Aquaculture. 2007; 265: 109–117. doi: 10.1016/j.aquaculture.2006.10.026

62. Valente TG, Schneider HC, Gross CM, Feldberg E, Martins C. Comparative cytogenetics of cichlid fishes through genomic *in-situ* hybridization (GISH) with emphasis on *Oreochromis niloticus*. Chromosome Res Int J Mol Supramol Evol Asp Chromosome Biol. 2009; 17: 791–799. doi: 10.1007/s10577-009-9067-5

63. Uzzell T, Gűnther R, Berger L. *Rana ridibunda* and *Rana esculenta*: a leaky hybridogenetic system (Amphibia, Salientia). Proc Acad Nat Sci Phila. 1977; 128: 147–171.

64. Alves MJ, Coelho MM, Collares-Pereira MJ. Evolution in action through hybridisation and polyploidy in an Iberian freshwater fish: a genetic review. Genetica. 2001; 111: 375–385. PMID: 11841181

65. Arai K, Mukaino M. Clonal nature of gynogenetically induced progeny of triploid (diploid × tetraploid) loach, *Misgurnus anguillicaudatus* (Pisces: Cobitididae). J Exp Zool—J EXP ZOOL. 1997; 278: 412–421.

66. da Barbiano LA, Gompert Z, Aspbury AS, Gabor CR, Nice CC. Population genomics reveals a possible history of backcrossing and recombination in the gynogenetic fish *Poecilia formosa*. Proc Natl Acad Sci. 2013; 201303730. doi: 10.1073/pnas.1303730110

67. Mikulíček P, Kautman M, Demovic B, Janko K. When a clonal genome finds its way back to a sexual species: evidence from ongoing but rare introgression in the hybridogenetic water frog complex. J Evol Biol. 2014; 27: 628–642. doi: 10.1111/jeb.12332 PMID: 26227900

68. Robertson AV, Ramsden C, Niedzwiecki J, Fu J, Bogart JP. An unexpected recent ancestor of unisexual *Ambystoma*. Mol Ecol. 2006; 15: 3339–3351. doi: 10.1111/j.1365-294X.2006.03005.x PMID: 16968274

69. Bi K, Bogart JP. Time and time again: unisexual salamanders (genus *Ambystoma*) are the oldest unisexual vertebrates. BMC Evol Biol. 2010; 10: 238. doi: 10.1186/1471-2148-10-238 PMID: 20682056

70. Eichler EE, Sankoff D. Structural dynamics of eukaryotic chromosome evolution. Science. 2003; 301: 793–797. doi: 10.1126/science.1086132 PMID: 12907789

71. Snowdon RJ, Köhler W, Friedt W, Köhler A. Genomic *in situ* hybridization in *Brassica* amphidiploids and interspecific hybrids. Theor Appl Genet. 1997; 95: 1320–1324. doi: 10.1007/s001220050699

72. Hickey DA. Selfish DNA: a sexually-transmitted nuclear parasite. Genetics. 1982; 101: 519–531. PMID: 6293914

73. Arkhipova I, Meselson M. Transposable elements in sexual and ancient asexual taxa. Proc Natl Acad Sci. 2000; 97: 14473–14477. doi: 10.1073/pnas.97.26.14473 PMID: 11121049