

**Jihočeská univerzita v Českých Budějovicích
Přírodovědecká fakulta**

**Populačně-genomická analýza generalistického parazita –
tasemnice *Ligula intestinalis***

Diplomová práce

Bc. Pavlína Kočová

Školitel: doc. RNDr. Jan Štefka, Ph.D.

České Budějovice 2018

Kočová, P., 2018: Populačně-genomická analýza generalistického parazita – tasemnice *Ligula intestinalis* [Population-genomic analysis of adaptation in a parasite with a wide host range – tapeworm *Ligula intestinalis*. Mgr. Thesis, in Czech.] 52 p., Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic.

Annotation:

New insight to the population structure of the tapeworm *Ligula intestinalis* were obtained using genomic data generated by reduced represented genomic methods (ddRAD and high-throughput sequencing).

Prohlašuji, že svoji diplomovou práci jsem vypracovala samostatně pouze s použitím pramenů a literatury uvedených v seznamu citované literatury. Prohlašuji, že v souladu s § 47b zákona č. 111/1998 Sb. v platném znění souhlasím se zveřejněním své diplomové práce, a to v nezkrácené podobě elektronickou cestou ve veřejně přístupné části databáze STAG provozované Jihočeskou univerzitou v Českých Budějovicích na jejich internetových stránkách, a to se zachováním mého autorského práva k odevzdanému textu této kvalifikační práce. Souhlasím dále s tím, aby toutéž elektronickou cestou byly v souladu s uvedeným ustanovením zákona č. 111/1998 Sb. zveřejněny posudky školitele a oponentů práce i záznam o průběhu a výsledku obhajoby kvalifikační práce. Rovněž souhlasím s porovnáním textu mé kvalifikační práce s databází kvalifikačních prací Theses.cz provozovanou Národním registrem vysokoškolských kvalifikačních prací a systémem na odhalování plagiátů.

V Českých Budějovicích dne 17.4.2018

.....
Pavína Kočová

Poděkování

Touto cestou bych ráda poděkovala vedoucímu naší laboratoře Janu Štefkovi za příležitost pracovat ve skvělé laboratoři, za vždy projevenou trpělivost, ochotu poradit a pomoci. Velké díky patří Liborovi Piálkovi a Jakubovi Vlčkovi za nedocenitelnou pomoc při zpracování genomových dat. Díky patří také Marušce za podporu, dobré nápady, překonávání nástrah bioinformatiky a zásobování kanceláře sladkostmi. Dále bych chtěla poděkovat všem, co poskytli vzorky ke zpracování, tedy členům laboratoří Tomáše Scholze a Ivana Fialy, Alimu a členům Hydrobiologického ústavu za příležitost jezdit s jejich týmem na sběry vzorků. Nakonec bych chtěla poděkovat svým přátelům, kteří ve chvílích nejvyšší nouze vždy stáli při mně a i když já pochybovala, oni nikdy, své rodině za bezmeznou důvěru, finanční i psychickou podporu poskytovanou během mého studia. Ještě jednou Vám moc děkuji!

Obsah

1	Úvod.....	1
1.1	Ekologická adaptace a speciace.....	1
1.2	Parazité a adaptace.....	2
1.3	Molekulární markery.....	4
1.4	RADseq.....	4
1.5	Modelový organismus.....	5
1.5.1	Životní cyklus.....	5
1.5.2	<i>L. intestinalis</i> předmětem ekologických studií.....	7
2	Cíle práce.....	9
3	Materiál a metodika.....	10
3.1	Laboratorní část.....	10
3.1.1	Vzorky a DNA.....	10
3.1.2	Detekce mitochondriálních genů.....	10
3.1.3	Gelová elektroforéza.....	11
3.1.4	Příprava podknihojen pro sekvenaci na Illumina.....	11
3.2	Bioinformatická část.....	12
3.2.1	Fylogenetické analýzy mitochondriálních genů.....	12
3.2.2	Analýza genomových dat.....	13
4	Výsledky.....	16
4.1	Fylogeneze mitochondriálních genů.....	16
4.2	Kontrola kvality a filtrování sekvencí.....	18
4.3	Globální rozložení diverzity <i>L. intestinalis</i> pomocí SNP dat.....	21
4.4	Analýzy lokálních patternů.....	28
4.4.1	Analýzy populační struktury.....	28
4.4.2	Analýza „outliers“ u sympatrických lokalit Těrlicko a Žermanice.....	30
5	Diskuze.....	33
6	Závěr.....	40
7	Literatura.....	41
8	Přílohy.....	47

1 Úvod

1.1 Ekologická adaptace a speciace

Všudypřítomná prostorová a ekologická proměnlivost životního prostředí nutí populace různých druhů přizpůsobovat se místním abiotickým i biotickým podmínkám, čímž dochází ke vzniku lokálních adaptací. Jedním z hlavních kritérií lokální adaptace je vyšší fitness jejích nositelů než u původní populace (Conover *et al.* 2009). Jako příklad lze uvést geneticky podmíněnou variabilitu ve fenologických fázích podél zeměpisných šířkových gradientů u mnoha rostlin a živočichů. Ekologická specializace realizovaná formou lokálních adaptací pak může vést ke značné fenotypové a genetické diferenciaci, nebo až k samotné speciaci. A naopak, fenotypová a genetická diferenciaci podél ekologických gradientů nebo mezi kontrastními typy biotopů mohou být také indikátory pro hledání lokálních adaptací (Savolainen *et al.* 2013). Podle modelu ekologické speciace tak přírodní selekce přizpůsobuje jednotlivé populace danému prostředí různým způsobem, což vede nejprve k diferenciaci druhu do různých linií/klastrů, a nakonec do samostatných, oddělených druhů (Savolainen *et al.* 2013).

Druhým významným, a zároveň velmi intuitivním, mechanismem vzniku genetické vnitrodruhové variability je izolace prostorová. V případě, že se dvě populace stanou alopatrickými (tj. zcela geograficky oddělenými), mohou se začít geneticky odlišovat z důvodu nepřítomnosti toku genů a působení genetického driftu. Na obou stranách bariéry se populace vyvíjí odděleně a vlivem hromadění pomalých změn vznikají dva druhy, v konečné fázi dochází k reprodukční izolaci. Pokud však bariéra zmizí před vytvořením kompletní reprodukční izolace vzniká tzv. hybridní zóna (Rundle a Nosil 2005). Vznik nových biogeograficky oddělených druhů (např. vznik druhů na jednotlivých ostrovech souostroví) je v porovnání s divergencí za přítomnosti výměny genů (sympatrické speciace) poměrně jednoduchý proces (Sousa a Hey 2013).

Model sympatrické speciace, tedy evoluce reprodukční izolace za nepřítomnosti (bio)geografických bariér, byl mnohými považován za nepravděpodobný, protože výměna genů napříč areálem druhu byla považována za silnou homogenizační sílu, která působí proti odchylkám způsobeným přirozeným výběrem. Nedávno však genetická data (například sekvence mitochondriální DNA a mikrosatelitů) společně s biogeografickými okolnostmi poskytly přesvědčivé důkazy, že sympatrickou speciací lze nalézt v mnoha kontextech

(Bolnick a Fitzpatrick 2007). Například u rostlin může dojít k rychlé separaci díky chybě při meióze a vzniku polyploidie, čímž je dceřiná populace reproduktivně oddělená od mateřské. Další z příkladů sympatrické speciace můžeme nalézt v malých izolovaných prostředí jako jsou oceánské ostrovy, postglaciální jezera nebo kráterová jezera. Nemálo zastoupené jsou případy ekologické speciace např. u fytofágního hmyzu, kdy změna hostitele může přímo vést k sympatrické speciaci (Bolnick a Fitzpatrick 2007).

Základním mechanismem umožňujícím sympatrickou diferenciaci je tzv. „speciation with gene flow“, tedy situace, kdy je selekce pro daný znak (znaky) natolik silná, že v genomu vytvoří ostrov(y) diferenciaci (speciation islands), které postupně odlišují specializované populace, či linie. Výzkum v poslední době však ukazuje, že i některé vnitřní genomové procesy (např. míra rekombinace nebo „background selection“) mohou vytvářet podobný obraz genetické diferenciaci jako ekologická speciace (Nadeau *et al.* 2012, Cruickshank a Hahn 2014, Wolf a Ellegren 2016). Proto je třeba mechanismy ekologické speciace testovat na dalších (nových) modelech.

1.2 Parazité a adaptace

Paraziti představují velmi vhodný model pro studium ekologické diferenciaci a speciace. Vynikají schopností přizpůsobovat se neustále se měnícím podmínkám vně (volně se vyskytující stádia) i v hostiteli. Tímto způsobem může docházet k diferenciaci populací parazita v důsledku lokálních adaptací (Isolation By Adaptation, IBA). Vzniku a udržování rezistence na léčiva, osvojení si nového hostitele, nebo expanzi na exotická místa napomáhá jedna ze zásadních vlastností parazita, a tím je schopnost úniku před imunitou hostitele (Criscione *et al.* 2005). Vlastnosti, které nalézáme u parazitů, odrážejí jejich schopnost produkovat rychlé evoluční změny (Price 1980) a IBA je poměrně komplikovaný fenomén, který nelze vnímat zjednodušeně bez dalších faktorů. Samotný průběh speciace parazitů je značně ovlivňován tokem genů a genetickým driftem nejen uvnitř parazitických populací, ale i hostitelských populací, na nichž jsou parazité závislí (Hufbauer 2005). Koevoluční změny mezi hostiteli a jejich parazity jsou již dlouho uznávány jako základní hnací síla evolučních procesů. Stupeň hostitelské specifity přitom může snadno ovlivňovat vnitrodruhovou genetickou strukturu a vytvoření bariér genetickým tokem (Johnson *et al.* 2012). Parazité pravděpodobně často podstupují proces speciace spolu s hostitelem, který do značné míry ovlivňuje jejich genetický charakter (Poulin a Morand 2000).

Dostupnost hostitelů může být jedním z hlavních ovlivňujících faktorů toku genů u parazitů (je zde poměrně málo volně žijících stádií s omezenými schopnostmi šíření). Z toho vyplývá hypotéza, že tok genů a migrační schopnosti parazitů se složitým (komplexním) životním cyklem jsou limitovány nejmobilnějším hostitelem (Louhi *et al.* 2010). Tento fakt zmiňuje Louhi ve své práci, která byla zaměřena na genetickou strukturu populace *Diplostomum pseudospathaceum* (Trematoda). V této studii autoři představili hypotézu, ve které nesmí být opomíjen vliv geografické struktury oblastí na zkoumání fragmentace populací (Louhi *et al.* 2010). Pokud se v prostředí nevyskytuje žádná přírodní bariéra (moře, pohoří apod.), znamená to, že je disperze organismů mnohem jednodušší. Stejně jako existuje teorie šíření parazitů v závislosti na přírodním prostředí, existuje také hypotéza, která se soustředí na rozdílnou schopnost šíření parazitů vzhledem k ekologii jejich hostitelů. Parazité, kteří mají ve svém životním cyklu zapojeny pouze hostitele obývající sladkovodní habitaty, jsou více limitováni v disperzi než parazité, kteří mají do svého cyklu zapojeny i terestrické hostitele (Criscione a Blouin 2004). Hostitelské populace se mohou lišit v prostorovém i společenském uspořádání, tím pádem se různé genotypy parazitů setkávají s různorodým prostředím. Například dva kmeny parazitů (patogenů) mohou mít odlišný přístup ke zdrojům, tedy různou lokální hustotu vnímavých hostitelů, a tím pádem může docházet ke specializaci na více dostupný zdroj (Lion a Gandon 2015). Pokud všechna tato fakta shrneme, dostaneme se k závěru, že populační struktura parazitů je určována mnoha faktory jak na makroevoluční, tak na mikroevoluční úrovni. Schopnost rychlé adaptace, upevnění evolučních změn v populacích a často omezené možnosti šíření na větší geografické vzdálenosti vedou k častému vzniku kryptických druhů parazitů (Huyse *et al.* 2005).

U druhů s nízkou schopností disperze je očekávána, i při absenci zjevné migrační bariéry, vyšší genetická podobnost mezi sousedními populacemi než mezi vzdálenými – tento jev se nazývá genetická izolace podle vzdálenosti („*genetic isolation by distance* – IBD“) (Wright 1943). Analýza IBD má zásadní význam pro studium procesů zodpovědných za genetickou strukturu populací v prostoru. Lze říci, že evoluční potenciál druhu závisí právě na genetické struktuře daného druhu (Garnier *et al.* 2004). Díky IBD může být identifikována rovnováha mezi migrací a genetickým driftem (současné procesy) nebo může propojit disperzní schopnosti s genetickou diferenciací. U geograficky omezených populací, které se nacházejí v rovnováze, nebo se rovnovážnému stavu blíží, může být detekována IBD. Samozřejmě můžou být detekovány i různé druhy nerovnovážných systémů. Absence jakéhokoli modelu IBD u druhu, kde je očekávána, naznačuje, že systém je mimo

rovnováhu a je možné, že pronikl na dané území teprve nedávno, nebo představy o disperzních tendencích zkoumaného druhu neodpovídají skutečnému stavu (Slatkin 1993).

1.3 Molekulární markery

I když jsou mitochondriální geny, nukleární DNA i mikrosatelitní lokusy tradičním nástrojem populačních, ekologických, fylogenetických i fylogeografických studií, jsou často limitovány množstvím informací, které z nich lze získat. Obecně mají mitochondriální (respektive plastidové) lokusy nižší informativní hodnotu pro odlišení jemných populačních rozdílů než mikrosatelitní lokusy. Pro studium speciálních procesů na počátku diferenciace a zejména pro studium adaptací je zapotřebí adekvátního množství genetických dat, zpravidla s ještě hustějším pokrytím genomu. Řešením je celogenomový screening lokusů obsahujících „single nucleotide polymorphism – SNP“, tedy jedno nukleotidového polymorfismu, který umožňuje dostatečnou statistickou sílu pro odhalení zkoumaných procesů. Odhalování genetické variace díky na celogenomové úrovni se stává stále více dostupnější pomocí „next generation sequencing methods – NGS“ (Jeffries *et al.* 2016, Querejeta *et al.* 2016, Allendorf 2017).

Polygenní kvantitativní znaky jsou nejčastěji spojovány se vznikem lokálních adaptací a identifikace lokusů řídící jejich variace je často náročná, zejména pak s omezenými genomovými daty daného organismu (Savolainen *et al.* 2013).

Nově se rozšiřující metody NGS zajišťují více než uspokojivé množství SNP. Fakt, že studie pracující s celogenomovými daty jsou možné jen u modelových organismů, už neplatí. Stále se posouvající hranice práce s genomickými nástroji nyní umožňují práci na celogenomové úrovni i u nemodelových organismů (Savolainen *et al.* 2013).

Za poslední desetiletí došlo k výraznému rozvoji těchto metod a jsou považovány za jedny z nejdůležitějších vědeckých průlomů poslední doby. Využití obrovského informačního potenciálu NGS se odráží v ekologických, evolučních, fylogenetických i genomových studiích (Andrews *et al.* 2016, Davey *et al.* 2011).

1.4 RADseq

Sekvenování celých genomů je i při velkém pokroku NGS metod stále příliš nákladné pro dostatečně velké množství jedinců potřebných pro populčně genetický výzkum. Velice

populárními se stávají různé metody umožňující „navzorkování“ (subsampling) genomu tak, aby získaná sekvenční informace pokrývala náhodně, ale rovnoměrně celý genom. Nejvyužívanější jsou RAD (*Restriction site-Associated DNA*) metody, které pomocí restričního enzymu specificky naštipou genomovou DNA. Dochází tak k náhodné fragmentaci a tím pádem k širokému velikostnímu rozsahu získaných lokusů napříč celým genomem. Se zdokonalením této metody přišli Peterson *et al.* (2012), kdy ve své studii použili dvě restriční endonukleázy – double digest RAD. Kombinací více, či méně často štípajících enzymů lze poměrně přesně navrhnout sekvenační projekt tak, aby získaný počet lokusů odpovídal potřebám projektu. Tím lze dosáhnout snížení nákladů na sekvenaci a možnosti multiplexování velkého počtu jedinců do knihoven (Peterson *et al.* 2012).

Takto získaná genomová data jsou dobrým nástrojem k ověření, prohloubení a rozšíření stávajících poznatků (Andrew *et al.* 2014, Emerson *et al.* 2010, Querejeta *et al.* 2016). S velice zajímavým výzkumem přišel Nosil *et al.* (2012). V této práci se zaměřili na systém hostitel/herbivor, v tomto případě rostlina/herbivorní hmyz, kdy byl popsán zásadní fenotypový rozdíl mezi jedinci *Timema cristinae* žijících se rozdílných rostlinách. U zkoumaného hmyzu byla prokázána i reprodukční izolace, jako jeden ze závěrečných kroků ekologické speciace. Pozorované rozdíly mezi ekotypy byly potvrzeny pomocí genomových dat – které potvrdily domněnku ekologické speciace prostřednictvím adaptace na hostitele (Nosil *et al.* 2012).

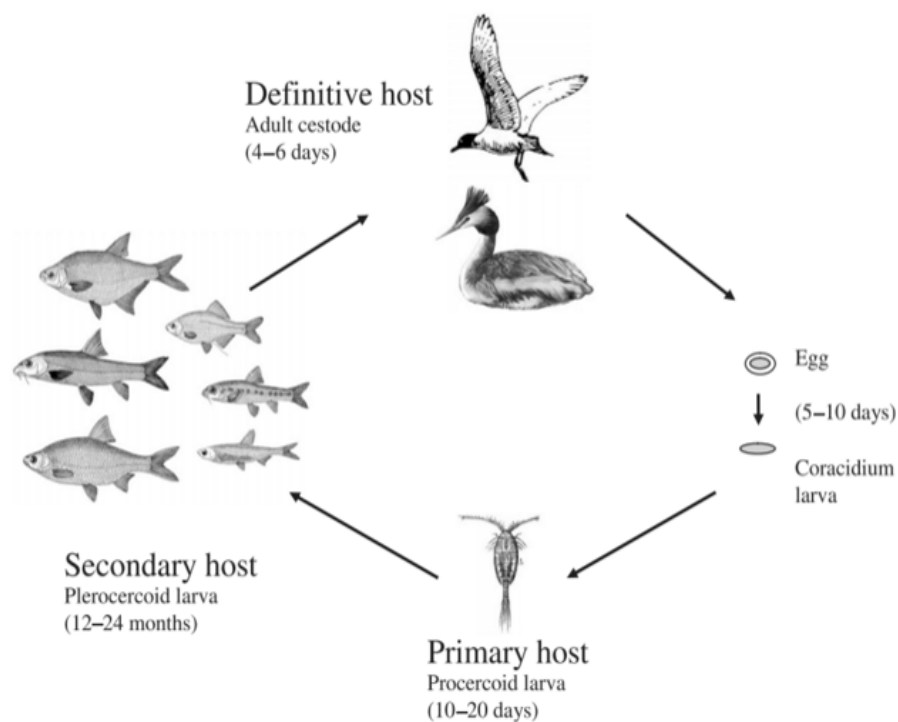
1.5 Modelový organismus

Jako modelový organismus pro studium procesu genetické diferenciaci prostřednictvím adaptací byla zvolena tasemnice *Ligula intestinalis: Cestoda – Diphyllbothriidea*, (Linnaeus 1758). Jedná se o parazita sladkovodních ryb s globální distribucí. *L. intestinalis* je zaznamenávána především v holarktických oblastech, ale její výskyt byl popsán například v Mexiku a Austrálii (Hoole 2009, Morgan 2003), nejnovější výskyt byl zaznamenán na Novém Zélandu, (Lagruet *et al.* 2018)

1.5.1 Životní cyklus

Tasemnice *L. intestinalis* se vyznačuje komplexním životním cyklem. Během svého vývoje infikuje celkem 3 odlišné hostitele. Životní cyklus je zahájen od vajíčka, které se nachází volně ve sladkovodním prostředí. Během několika dní se transformuje na volné

plovoucí larvální stádium koracidium, které je požřeno klanonožcem (Copepoda). Zpravidla doba strávená v prvním hostiteli nepřesahuje dobu delší než 20 dní (Dubinina 1980). Během této doby dochází k přeměně koracidia na následující stádium vývoje – procerkoid, které je infekční pro ryby. Z větší části jsou napadáni zástupci řádu kaprovitých, nejedná se však o striktní podmínku, infekce konkrétního druhu je závislá na geografických a ekologických podmínkách (Loot 2001). Tasemnice následně proniká stěnou střeva do abdominální oblasti, kde se vyvíjí v třetí stádium – plerocerkoid. Červ je schopen v rybím mezihostiteli přežít až dva roky (Dubinina 1980) a způsobuje hostiteli řadu patologických změn (zbrzděný růst, kastrace). Rybí hostitel tedy představuje nejdůležitější část hostitelského cyklu z hlediska vzniku možných adaptací a specializace na hostitelské druhy. Konečným hostitelem *L. intestinalis* je rybožravý pták, např. zástupci řádů potápek (Podicipediformes), brodivých (Ciconiiformes), dlouokřídých (Charadriiformes), vrubozobých (Anseriformes), nebo dravců (Falconiformes). Dospělec infikuje konečného hostitele po dobu maximálně 4-6 dní (viz Obr. 1).



Obr. 1: Životní cyklus tasemnice *Ligula intestinalis*, převzato Štefka *et al.* 2009.

1.5.2 *L. intestinalis* předmětem ekologických studií

L. intestinalis byla častým středem zájmu nejrůznějších studií zaměřených na ekologická a parazitologická témata. Dá se říci, že dnes má tento druh (nebo lépe seskupení kryptických druhů) poměrně dobře zmapovanou vnitro – až mezidruhovou fylogenezi (např.: Li a Lao 2003, Logan *et al.* 2004). Dále byly zkoumány i interakce hostitel/parazit (např.: Loot *et al.* 2002, Brown *et al.* 2002, Bouzid *et al.* 2008 b), populačně geneticky orientované studie (např.: Li a Lao 2000, Morgan 2002, Bouzid *et al.* 2008 a, Štefka *et al.* 2009) i epidemiologické výzkumy (Bouzid *et al.* 2013). *L. intestinalis* nevykazuje striktní hostitelskou specifitu na úrovni rybích mezihostitelů ani konečných hostitelů. Pro infekci konkrétních hostitelů však existuje závislost na prostředí, dostupnosti hostitele, geografických a ekologických podmínkách (Loot *et al.* 2001, Bouzid *et al.* 2008 a, b; Štefka *et al.* 2009, Bouzid *et al.* 2013).

Molekulárně orientované studie naznačují, že vnitrodruhová struktura *L. intestinalis* se jeví více fragmentovaná, než se původně předpokládalo na základě minimálních morfologických odlišností. Bylo objeveno několik statisticky dobře podpořených geneticky oddělených linií, z nichž některé se navzájem geograficky překrývají – euro-mediteránní klady A a B, které se liší i preferencí pro různé rybí hostitele (Bouzid *et al.* 2008 b). Vzhledem k množství sympatricky se vyskytujících rybích hostitelů v mnoha dalších oblastech výskytu *L. intestinalis* je zcela namístě se domnívat, že sympatrický výskyt více linií *L. intestinalis*, s odlišnou hostitelskou specifitou, je mnohem častější, než bylo dosud popsáno. Dále jsou známy populace monofyletických linií, které jsou geograficky izolované (přírodní bariéry, u *L. intestinalis* především moře a oceány). Jednalo se o skupiny izolované z oblastí Mexika, Kanady, Číny a Austrálie (Bouzid *et al.* 2008 a, b). Zmiňované práce se liší zejména v množství analyzovaných dat a v aplikaci různých statistických postupů ve fylogenetických analýzách. Jako první byly použity mitochondriální geny (COI a COB), ale ukázalo se, že nejsou dostatečně citlivé pro rozlišení vnitřní struktury sympatrických linií. V navazující studii se autoři rozhodli zmapovat populační strukturu *L. intestinalis* pomocí mikrosatelitních markerů (Štefka *et al.* 2009). Tento postup přinesl přesnější výsledky především v populační struktuře mezi mediteránními liniemi. Rozložení linií ovšem stále kopírovalo předchozí studie s těmi rozdíly, že v tomto případě byly rozpoznatelné jemnější rozdíly v genetické struktuře sympatrických populací, například v Africe (jezero Tana, Štefka *et al.* 2009).

Geograficky oddělené linie vznikly pravděpodobně díky omezenému toku genů mezi místy výskytu. Značné genetické rozdíly byly pozorovány u linií geograficky velmi

vzdálených. Vzhledem k životnímu cyklu *L. intestinalis* vázanému na sladkovodní organismy a krátkému trvání infekce v ptačích, tedy konečných, hostitelích zajišťujících disperzi funguje např. Středozevní moře jako účinná bariera genovému toku. To je přitom v kontrastu s výskytem linie A jak v Evropě, tak v severní Africe (Alžír a Tunisko). Výskyt linie A v severní Africe byl ale přisouzen lidskému faktoru. V 60. letech minulého století byly do Tuniska importovány některé evropské druhy kaprovitých ryb za ekonomickými účely (podpora diverzity místních akvakultur). Tímto způsobem zřejmě došlo i k introdukci evropské linie *L. intestinalis* spolu s infikovanými hostiteli. Směr této introdukce (z Evropy do severní Afriky) byl pak ověřen pomocí koalescenční analýzy multilokusových dat (Bouzid *et al.* 2013). Mikrosatelitní analýzou se však dosud nepodařilo vysvětlit některá jiná pozorování. Populace kladu A vykazovaly panmiktický charakter bez pozorovatelné vnitřní struktury napříč Evropou, a to i přes pozorované rozdíly ve schopnostech infikovat odlišné druhy ryb v různých částech areálu (Loot *et al.* 2001). Zůstalo otázkou, zda lze jemnější genetické rozdíly zachytit pomocí citlivějších analýz založených na celogenomovém screeningu.

2 Cíle práce

Cílem diplomové práce je prohloubení znalostí populačních trendů tasemnice *L. intestinalis* za použití genomických dat získaných pomocí metody ddRAD. Vedle lepšího obrazu globální populační struktury a genového toku budou oproti dřívějším pracím testovány zejména dvě následující otázky:

- Existuje patern izolace populací vzdáleností, mezi populacemi evropské linie *L. intestinalis* z různých částí areálu?
- Lze v genomu evropské linie *L. intestinalis* nalézt stopy lokálních adaptací na odlišné hostitele (např. nejběžnější druhy jako cejn, plotice a ouklej)?

3 Materiál a metodika

3.1 Laboratorní část

3.1.1 Vzorky a DNA

Do studie bylo zahrnuto celkem 218 jedinců tasemnice *Ligula intestinalis* a 3 jedinci *Ligula interrupta* z celkem 18 zemí světa a 13 různých hostitelů (viz Příloha I). Z větší části se jednalo o izoláty DNA, které byly analyzovány v dřívějších fylogenetických i ekologických studiích (Bouzid et al. 2008a, 2008b, Štefka et al. 2009). Jako outgroups byly zařazeny také dva vzorky *Dibothriocephalus latus* a dva vzorky *Schistocephalus solidus* poskytnuté z helmintologické laboratoře Parazitologického ústavu BC, AV ČR. Nové vzorky ve stádiu plerocerkoid byly získány od spolupracujících laboratoří v zahraničí (africké lokality, Oregon), poskytnutím rybích mezihostitelů k odběru vzorků a vlastním sběrem v terénu (české lokality). U nově zařazených vzorků byla nejdříve izolována DNA pomocí Blood & Tissue kit (Quiagen), bylo postupováno podle návodu výrobce. Izolovaná DNA byla skladována při teplotě -22°C . U nově izolovaných i starších vzorků byla provedena fluorescenční kvantifikace DNA pomocí přístroje Qubit 2.0 (Invitrogen, Broad range double – stranded DNA kit). Pro jednotlivá měření bylo použito 5 μl vstupní DNA každého vzorku. Tento krok byl nezbytný pro výpočet vstupního množství DNA pro přípravu ddRAD podknihočen. Následně byla provedena PCR analýza mitochondriálních genů COB a COI pro ověření mitochondriální linie nově zařazených tasemnic a získání základních fylogenetických informací.

3.1.2 Detekce mitochondriálních genů

Nově zařazené vzorky (Tab. 1) pocházející z různých oblastí Afriky a jedné lokality Severní Ameriky – Oregon (USA) byly analyzovány pomocí sekvenace dvou mitochondriálních genů (COB a COI). Složení reakcí o celkovém objemu 20 μl bylo pro oba sledované geny shodné, lišilo se pouze v použití párů primerů: 0,1 μl Taq polymeráza, 0,5 μl dNTP's (5 pmol), 2 μl 2x Taq Buffer, 1 μl forward primer, 1 μl reverse primer, 15,4 μl H₂O, 1 μl DNA. Získané PCR produkty byly enzymaticky přečištěny (enzymy Exonukláza I, alkalická fosfatáza; New England Biolabs) a odeslány k sekvenaci do společnosti SeqMe.

Tab. 1: Zeměpisný a hostitelský původ nově získaných vzorků.

Země	Hostitel	Počet
Jihoafrická Republika	<i>Barbus anoplus</i>	3
	<i>Barbus neefi</i>	1
	<i>Barbus trimaculatus</i>	1
DRK	<i>Barbus</i> spp.	1
Namibie	<i>Barbus paludinosus</i>	4
Keňa	neznámý hostitel	4
Oregon	<i>Rhinichthys osculus</i>	4

3.1.3 Gelová elektroforéza

Jako kontrolní metoda byla použita gelová elektroforéza, a to jak průběhu PCR amplifikace mitochondriálních genů, tak u závěrečného kroku přípravy ddRAD knihoven. Vzorky byly naneseny na 1 % (mitochondriální geny), nebo 2 % (produkty ddRAD) agarosový gel obarvený GelRed (BIOTIUM). Na gel bylo naneseno 2 μ l produktů PCR amplifikace mitochondriálních genů, respektive 5 μ l produktů v případě ddRAD. Jednotlivé vzorky byly naneseny pomocí 6x DNA Loading Dye (Fermentas) v daném poměru. Pro odečtení délky produktů byl použit Ladder 100 pb Plus (Fermentas). Vizuální kontrola fragmentů byla provedena pomocí UV transiluminátoru (Eppendorf).

3.1.4 Příprava podknihoven pro sekvenaci na Illumina

Postup pro přípravu ddRAD knihovny byl převzat z Peterson et al. (2012), zde je uvedeno pouze shrnutí nejdůležitějších částí protokolu. Pracovní postup se skládal ze čtyř základních kroků: enzymatická restrikce genomové DNA, ligace identifikačních barcodů, velikostní separace jedinců smíchaných do podknihoven (pooled DNA) a závěrečná PCR amplifikace při níž byly podknihovněm přiřazeny indexy a flow-cell adaptéry pro Illumina sekvenování. Vzorky byly rozděleny celkem do 5 podknihoven po maximálně 48 vzorcích (maximální počet adaptérů P1-flex). Pro restrikci genomové DNA byla zvolena následující kombinace restrikčních endonukleáz: SphI a MluCI (NEB), které byly vybrány jako nejvhodnější pár s ohledem na neoptimálnější počet získaných lokusů vzhledem k předpokládané velikosti genomu *L. intestinalis* (Peterson et al. 2012). Po dokončení restrikce byly vzorky jednotlivě purifikovány pomocí Agencourt AMPure XP Beads (Beckman Coulter) a následně byla opět stanovena koncentrace DNA pomocí fluorescenční metody na

přístroji Quibit 2.0. Při ligaci byly použity adaptéry P1-flex a P2-flex obsahující barcodes přiřazující specifickou identifikační sekvenci jednotlivým vzorkům. Po přidělení individuálních adaptérů ke každému jedinci mohly být všechny vzorky v dané podknihovně ekvimolárně smíšeny do jednotlivých podknihoven. Vzniklá směs byla následně opět purifikována pomocí magnetických kuliček Agencourt AMPure XP Beads (Beckman Coulter). Pro následující analýzy bylo nutné vyselektovat pouze frakci fragmentů o shodné délce, 300 bp – velikostní separace byla provedena na Pippin prep (Sage Science) pro každou podknihovnu zvlášť. Přiřazení “Illumina flowcell annealing sequences”, multiplexových indexů a “sequencing primer annealing regions” bylo provedeno pomocí PCR Multiplex PCR kitu (QIAGEN) nebo za použití polymerázy Herkuláza (Agilent Technologies). PCR reakce byly pro každou podknihovnu rozděleny do 8 reakcí, aby se zamezilo případné chybě při amplifikaci (amplification bias). V závěrečném kroku byly jednotlivé podknihovny ekvimolárně smíchány do finální knihovny, která byla odeslána k sekvenaci na Illumina HiSeq platformě do laboratoře EMBL, Heidelberg.

3.2 Bioinformatická část

3.2.1 Fylogenetické analýzy mitochondriálních genů

Obdržené sekvence mitochondriálních genů byly zpracovány v programu Geneious8.1. Pro fylogenetické analýzy byly oba geny konkatenovány, a to v pořadí COI, COB. Vzorky byly následně převedeny na haplotypy pomocí softwaru ALTER (Glez-Peña *et al.* 2010). Pro fylogenetické analýzy byly nově získané sekvence přiřazeny k datasetu z předchozích studií (Bouzid *et al.* 2013; Kočová, bakalářská práce 2015)

Pro správný výběr fylogenetických modelů byl použit program PartitionFinder2 (Lanfear *et al.* 2016), který hledá nejlepší evoluční model pro každou pozici v kodonu. Pro dataset v této studii byly nalezeny celkem 3 nejvhodnější modely – pro každou pozici jiný, a to: model1 – TIM+G, model2 – F81+G, model3 – TrN+I+G. Samotná fylogenetická analýza byla provedena v programu Garli (Bazineta *et al.* 2014). Jakou outgroup skupiny byly použity vzorky *D. latus* a *S. solidus*. Statistická podpora uzlů byla odhadnuta pomocí Bootstrap analýzy (1000 permutací) v programu Garli. Konsenzus bootstrapových stromů byl zpracován v programech PAUP a FigTree.

3.2.2 Analýza genomových dat

Data obdržena po sekvenaci byla již demultiplexovaná do podknižoven. Prvotní analýza kvality dat byla provedena pomocí nástroje FastQC (Andrews 2010) pro forward i reverse ready každé podknižovny. Následně byly výsledky porovnány pomocí programu MultiQC (Ewels *et al.* 2016).

S daty se nadále pracovalo v programu Stacks-1.45 (Catchen *et al.* 2011). Stacks byl vytvořen primárně pro práci s daty získaných na základě enzymatických restričních metod za účelem vytváření genetických map, populačně – genomických a fylogenetických analýz. Stacks nabízí kaskádu dílčích programů, které na sebe navzájem navazují a tím je možné data zpracovat od filtrace až po finální výstupy pro navazující analýzy.

Pomocí základního nástroje *process_radtags* byly ze sekvencí odfiltrovány adaptéry, vyřazeny sekvence s nízkou kvalitou a rozříděny dle barcodů na jedince.

Software Stacks nabízí několik možností, jak zpracovávat genomová data. Vzhledem k nedostupnosti referenčního genomu *L. intestinalis*, byl zvolen nástroj *denovo_map*. V tomto kroku byly z dostupných readů, získaných po prvním filtrovacím procesu, vytvořeny “balíčky” (stacks) shodných krátkých sekvencí, z nichž byly následně sestaveny lokusy a katalogy SNP.

Pro následující program *populations* bylo zapotřebí sestavit populační mapy, které programu poskytovaly informaci o příslušnosti jednotlivých vzorků do patřičných skupin. Tento program byl spuštěn několikrát pro různé nastavení parametrů týkající se striktnosti výběru lokusů jako hloubky prosekvenování lokusů (-m parametr), nutnosti zastoupení lokusů napříč skupinami definovanými v populačních mapách (-p parametr) i minimu požadovaných výskytů u jedinců v populacích (-r parametr) za účelem získání optimálního počtu lokusů a SNP pozic. U všech analýz byl nastaven parametr --min_maf 0,05 specifikující minimální frekvenci alel potřebnou pro zpracování nukleotidového místa v lokusu, parametr --write_random_snp zajistil, že z každého lokusu byla do výstupních souborů přenesena pouze jedna náhodně vypraná SNP pozice, čímž se zabránilo ovlivnění dalších analýz případnou genovou vazbou. Po zadání parametru --fstat *populations* porovnává párově všechny populace a vytváří základní F_{ST} statistiky. V programu *populations* lze nastavit požadované výstupní formáty, pro účely v této práci byly vygenerovány soubory .genepop, .vcf, .phylip, .plink.

Z kompletního datasetu vzorků bylo zapotřebí získat několik různých výstupních souborů podle typu následných analýz. Vygenerovaná data obsahovala buď všechny vzorky, nebo jen vzorky z určitých lokalit pro lepší náhled do vnitřní populační struktury daných oblastí. Bylo vytvořeno celkem 5 nastavení: 1) kompletní dataset, 2) všechny vzorky zbylé po

redukci jedinců kladu A (výběr 20 vzorků), 3) vzorků spadajících dle mitochondriálních fylogenetických analýz pouze do kladu A, 4) blízké české lokality, které obsahovaly vzorky z různých rybích meziphostitelů (Těrlicko a Žermanice).

Softwar PLINK (Purcell *et al.* 2007) umožňujícím práci s celogenomovými daty byl použit pro konverzi .plink výstupů z programu Stacks do binárního kódu a vygenerování formátu akceptovaného programem Admixture (Alexander *et al.* 2009).

Klastrovací analýza byla provedena v softwaru Admixture (Alexander *et al.* 2009), který používá podobný statistický model jako software Structure (Falush 2007). Oproti Structure umí zpracovávat daleko větší matice a pracuje o poznání rychleji. Tímto programem byly zpracovány matice, respektive výstupy z programu Stacks pro následující skupiny vzorků: kompletní dataset, světové populace s částí kladu A, samotný klad A, africké vzorky. Jednotlivé analýzy byly prováděny v 10, respektive 20 replikacích pro nalezení optimální hodnoty K, tedy nejnižší hodnoty CV (“cross validation”), reprezentující počet populací/klastrů nejlépe podpořený daty. Pro jednotlivé replikace byly spočítány průměrné hodnoty, nalezeny nejmenší a nejvyšší hodnoty K (viz Příloha II) a výsledné hodnoty byly graficky zpracovány pomocí LibreOffice.

Populační struktura byla také zkoumána pomocí nového programu fineRADstructure (Malinsky *et al.* 2018), který aplikuje metodu coancestry matic založených na celogenomových datech (programu fineSTRUCTURE) na data obsahující kratší sekvence, získané právě RAD sekvenováním. Tato metoda je založena na měření stupně společného původu dvou či více jedinců, tedy nejbližšího sousedního haplotypu (“coancestry matrix”) pomocí MCMC klastrovacího algoritmu. Tato analýza byla spuštěna pro matici všech vzorků, světové populace s částí kladu A a samotný klad A. Data byla konvertována pomocí python skriptu s množstvím „missing data“ maximálně 60 % (většina lokusů ale měla missing data <40%). Program byl spuštěn na 1 mil. MCMC generací, se 100 tis. generacemi burn-in. Parametry byly ukládány každých 1000 MCM kroků (thinning). Grafické zobrazení bylo vytvořeno v programu R (R Core Team 2013) pomocí upravených skriptů (Malinský *et al.* 2018).

Detekce genetických markerů souvisejících s biologickou adaptací tzv. “outliers” byla provedena pomocí softwaru Bayescan (Foll a Gaggiotti 2008) a PCAdapt (Duforet – Frebourg *et al.* 2014). Tato analýza byla provedena jen u vzorků z geograficky blízkých populací se sympatrickým výskytem *L. intestinalis* u různých hostitelů. Tyto podmínky splňovaly lokality Těrlicko a Žermanice v Moravskoslezském kraji. Vzhledem k tomu, že program BayeScan vyžaduje specifický formát vstupních dat, byly výstupy ze Stacks .genepop nejdříve

konvertovány do formátu Geste/BayeScan pomocí PGDSpider (Lischer *et al.* 2011). Program byl spouštěn ve výchozím nastavení, pouze parametr --pr_odds byl z nastavení 10 pozměněn na hodnotu 30. Výstupy byly zpracovány pomocí R skriptu dostupného pro BayeScan (Foll a Gaggiotti 2008).

PCAdapt je program schopný identifikovat “outliers” lokusy a zároveň je s jeho pomocí možné odvodit populační strukturu. Metoda využívá hierarchický Bayesiánský model (Duforet – Frebourg *et al.* 2014). PCAdapt analýzy byly opět provedeny pro kompletní dataset vzorků (odhalení populační struktury) a pro dílčí matici zahrnující vzorky z oblastí Těrlicko/Žermanice (zkoumání populační struktury a identifikace “outliers” lokusů). Pro výběr hlavních komponent (PCA) byla provedena analýza komponent s maximálním počtem $K = 20$. Pro jednotlivé SNP lokusy byla vypočítána testovací statistika podle prvních dvou komponent vyplývajících z předchozích výpočtů. Seznam “outliers” lokusů byl získán nastavením alfa = 0,1 a vygenerováním seznamu jen těch lokusů, které korelovaly s první komponentou. “Outliers” detekované oběma metodami (BayeScan a PCAdapt) byly porovnány s F_{ST} hodnotami vygenerovanými programem Stacks.

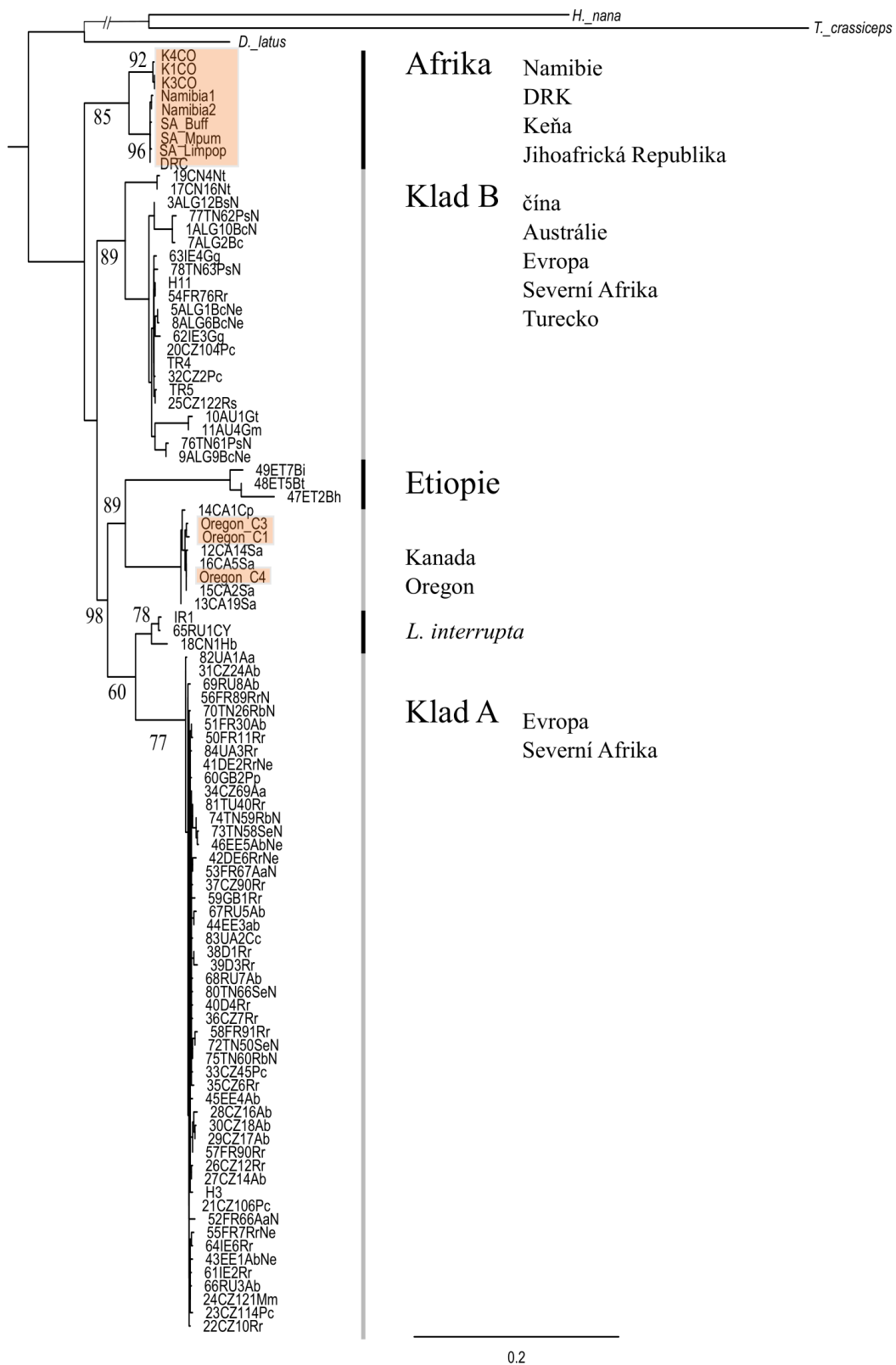
4 Výsledky

4.1 Fylogeneze mitochondriálních genů

Mezi konkatenovanými sekvencemi COI a COB bylo detekováno 12 nových haplotypů: 9 z afrických lokalit a 3 pocházející z Oregonu. Nově získaná data byla přiřazena k převzatému datasetu z předchozí studie (Bouzid *et al.* 2013) doplněné o vzorky zpracované v bakalářské práci (Kočová, 2015), čímž vznikl nový soubor dat pro fylogenetické analýzy. Tento dataset obsahoval 82 haplotypů z původní studie, 5 haplotypů z bakalářské práce a 12 nejnovějších haplotypů (viz Obr. 2).

Všechny nově detekované africké haplotypy vytvořily velmi dobře podpořenou větev odštěpující se na bázi stromu. Uvnitř této větve byly vzorky rozštěpeny na dvě větve. Vzorky z Keni vytvořily jednu skupinu a druhou větev tvořily zbývající vzorky z Namibie, Jihoafrické republiky a Demokratické republiky Kongo.

Vzorky pocházející z Oregonu (povodí řeky McKenzie) klastrovaly v topologii s kanadskými vzorky a tato větev byla velmi dobře podpořena. Toto klastrování odpovídalo biogeografii i ekologii daných populací, vzorky si byly geograficky nejbližší a zároveň pocházely z rybích hostitelů patřících do stejné podčeledi Leuciscinae (*Rhinichthys osculus* a *Semotilus atromaculatus*).



Obr. 2: Maximum Likelihood fylogeneze datasetu konkatenovaných sekvencí mitochondriálních genů COI a COB. Bootstrapové hodnoty nad 50 % jsou uvedeny u jednotlivých větví. Nové haplotypy oproti předchozím studiím (Bouزيد *et al.* 2008, Štefka *et al.* 2009, Kočová, bakalářská práce 2015) jsou barevně zvýrazněny.

4.2 Kontrola kvality a filtrování sekvencí

Pomocí programu FastQC byl u všech souborů spočítán počet readů na podknižovnu. Kvalita sekvencí byla kontrolována na několika různých úrovních, například kontrola délky sekvencí, kvalita čtení bází, obsah CG, „overrepresented“ sekvence a jejich analýza apod. I přes to, že žádná z podknižoven nevykazovala hodnoty znemožňující další práci s daty, byla pozorována celkově horší kvalita u podknižoven L2, L3. Oproti zbývajícím podknižovnám (L1, L3 – L5) obsahovaly L1 a L2 značně menší počty sekvencí a celkově horší kvalitu. Podobně tomu bylo i u reverse sekvencí oproti forward sekvencím. Nižší kvalita se projevila především zhoršením čtení bází v sekvencích a vyšším počtu „overrepresented“ sekvencí, které se v naprosté většině ukázaly jako sekvence adaptérů.

Tab. 2: Základní statistiky z FastQC.

Podknižovna	Počet sekvencí	Délka sekvencí	% CG
L1	39 342 973	125	44
L2	24 003 120	125	45
L3	22 269 233	125	49
L4	30 593 691	125	47
L5	28 204 688	125	44

Dále byly zpracovány pouze forward sekvence z důvodu celkově lepší kvality, vyvarování se případné vazbě variabilních pozic a lepší zpracovatelnosti programem Stacks.

Po prvním kroku v programu Stacks byly odfiltrovány sekvence obsahující adaptéry, sekvence s nízkou kvalitou čtení bází, sekvence, u kterých nebylo možné jednoznačně určit barcode a sekvence u kterých nebyl správně rozpoznán RAD-Tag (viz Tab. 3).

Tab. 3: Hodnoty pro každou podknihovnu forward readů zvlášť po *process_radtags*.

	Získaný počet lokusů	Illumina sekvence	Sekvence obsahující adaptér	Nízká kvalita	Nejednoznačně identifikovaný barcode	Nejednoznačný RAD-Tag	Celkem
L1	76862937	0	85227	538587	449198	749997	78685946
L2	46334580	0	55616	310458	475872	829714	48006240
L3	42447278	0	44458	285422	519958	1241350	44538466
L4	59554012	0	65404	427850	503164	636952	61187382
L5	55032543	0	58713	378833	412902	526385	56409376

Tab. 4: Tabulka všech filtrovaných dat – pro celou knihovnu (podknihovny L1-L5).

Celkový počet sekvencí	288827410
Illumina reads	0
Ready obsahující sekvence adaptérů	309418
Nejednoznačný barcode	2361094
Nízká kvalita	1941150
Nejednoznačný RAD-Tag	3984398
Počet obdržených sekvencí	280231350

Po filtraci dat a rozřídění všech readů po jednotlivcích byly zjištěny markantní rozdíly v počtu sekvencí na jednotlivce mezi podknihovny. Výrazně horších výsledků dosahovaly vzorky řazené v podknihovně L2 a L3. Rozdíly nebyly zanedbatelné, jednalo se až o stovky tisíc lokusů. Toto byl první silný signál, že pravděpodobně bude s těmito daty problém, což se potvrdilo v následujícím kroku, kdy program *populations* nebyl schopen nalézt téměř žádné homologní lokusy při hledání napříč všemi vzorky. Po vyřazení vzorků pocházejících z podknihoven L2, L3 byl tento problém odstraněn. Celkem bylo tedy z následujících analýz vyřazeno 96 jedinců a zbylo 125 (viz Tab. 5). Dalšími vzorky vyřazenými z následného zpracování byly všechny vzorky z tasemnic, které měly sloužit jako outgroup skupiny (*D. latus*, *S. solidus*). Program nebyl schopen nalézt žádné homologní lokusy společné s *L. intestinalis*, a proto nebylo možné vzorky nadále vzájemně porovnávat.

Tab. 5: Souhrnné počty vzorků vyřazených a zůstávajících ve finálních analýzách.

Vyřazené vzorky		Analyzované vzorky	
Země	Počet	Země	Počet
Velká Británie	2	Česká republika	80
Island	5	Estonsko	5
Mexiko	2	Austrálie	5
Rusko	8	Írán	3
Čína	10	Maďarsko	5
Tunis	14	Etiopie	17
Německo	4	Namibie	4
Kanada	10	DRK	1
Francie	38	Jihoafrická republika	5
Česká republika	3		
celkem	96	celkem	125

Pro zbývajících 125 vzorků byl znovu spuštěn program *denovo_map.pl* a následně několik různých běhů programu *populations*, při kterých byly specifikovány skupiny vzorků (1 až 5) náležící do stejných geografických oblastí a striktnější výběry vzorků potřebné pro detailnější náhled struktury sympatrických lokalit.

Tab. 6: Výběr vzorků pro program *populations*, počty skupin v jednotlivých bězích, získané lokusy a SNP.

Dataset	Počet vzorků	Porovnávané skupiny populací	Počet lokusů	Počet SNP
1) Všechny vzorky	125	6	1929	1897
2) Světové populace + výběr kladu A	65	6	1912	1882
3) Klad A	85	4	3037	3014
4) Těrlicko/Žermanice	48	2	4097	4064

4.3 Globální rozložení diverzity *L. intestinalis* pomocí SNP dat

Z výsledků primárních F_{ST} statistik (Stacks) vyplývá, že v globálním měřítku jsou populace *L. intestinalis* z různých oblastí jasně divergované. Vysoké hodnoty naznačují, že populace jsou od sebe prokazatelně geneticky vzdálené (viz Tab. 7 a 8). Po porovnání výsledků F_{ST} u datasetu obsahující všechny vzorky (1) a datasetu počítajícím pouze s výběrem vzorků kladu A (2) nebyly pozorovány zásadní změny v hodnotách porovnávacích jednotlivé páry lokalit. Toto však neplatilo ve všech případech, k výraznějšímu posunu hodnoty F_{ST} došlo v porovnávání skupin klad A a Írán. Po redukci vzorků v kladu A se výrazně zvýšila hodnota F_{ST} . Tento jev byl pravděpodobně způsoben nevyváženým počtem vzorků v jednotlivých skupinách (klad A = 85, Írán = 3).

Tab. 7: F_{ST} statistika všech vzorků (125) rozdělených dle geografického výskytu.

	Klad A	Jižní Afrika	Írán	Maďarsko	Etiopie	Austrálie
Klad A		0,457	0,149	0,259	0,561	0,350
Jižní Afrika			0,846	0,734	0,831	0,876
Írán				0,705	0,766	0,962
Maďarsko					0,738	0,818
Etiopie						0,838

Tab. 8: Hodnoty F_{ST} pro světové populace a výběr kladu A (20).

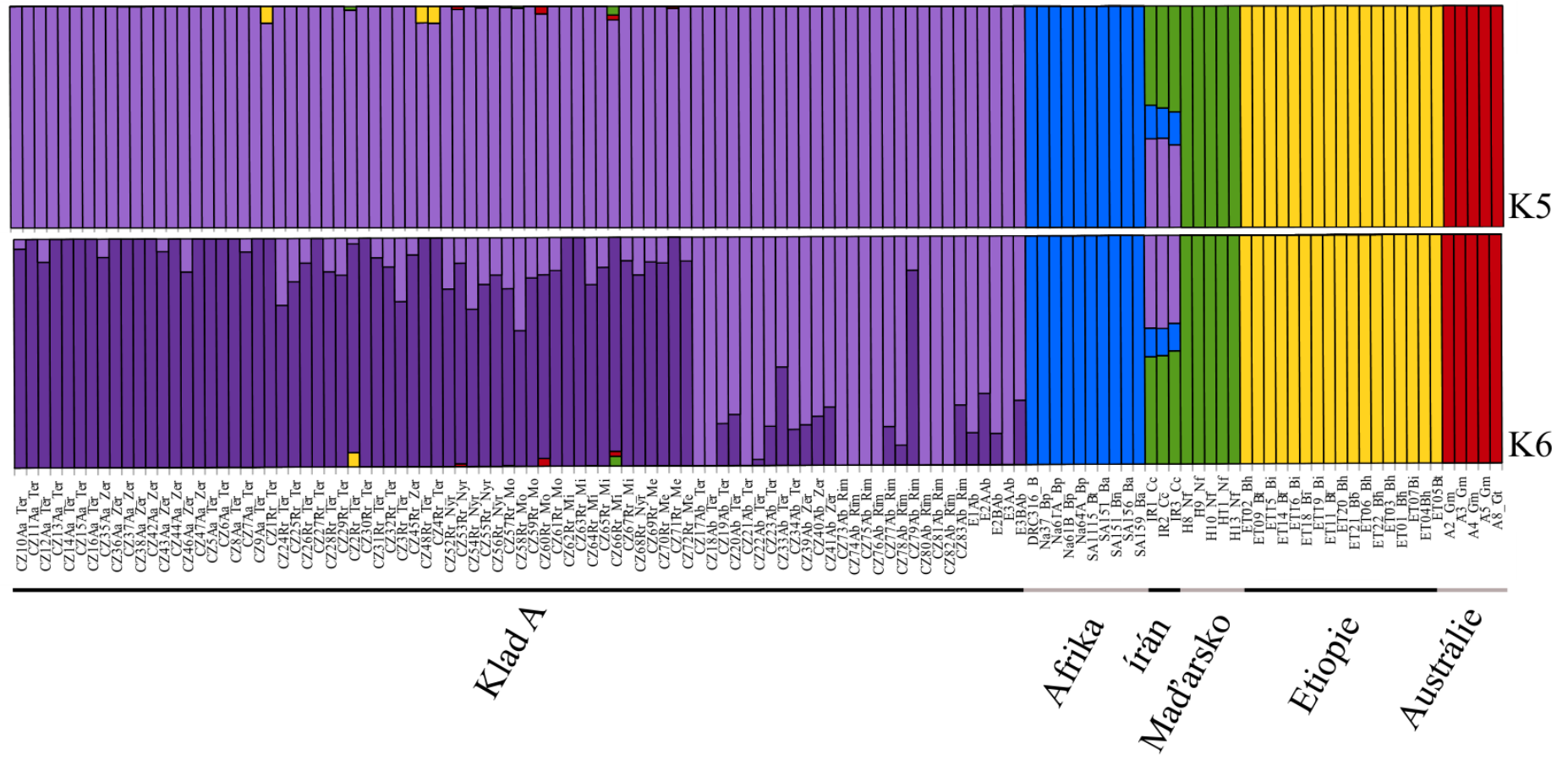
	Klad A	Jižní Afrika	Írán	Maďarsko	Etiopie	Austrálie
Klad A		0,556	0,455	0,472	0,627	0,537
Jižní Afrika			0,792	0,700	0,744	0,838
Írán				0,735	0,741	0,941
Maďarsko					0,707	0,786
Etiopie						0,807

Globální rozložení genetických klastrů *L. intestinalis* bylo testováno na celkovém datasetu (1) (viz Obr. 3, 4) a na datasetu u kterého bylo z kladu A vybráno pouze 20 vzorků (2) (viz Obr. 5, 6). Užší výběr měl zabránit případnému zkreslení výsledků důsledkem podstatně vyššího

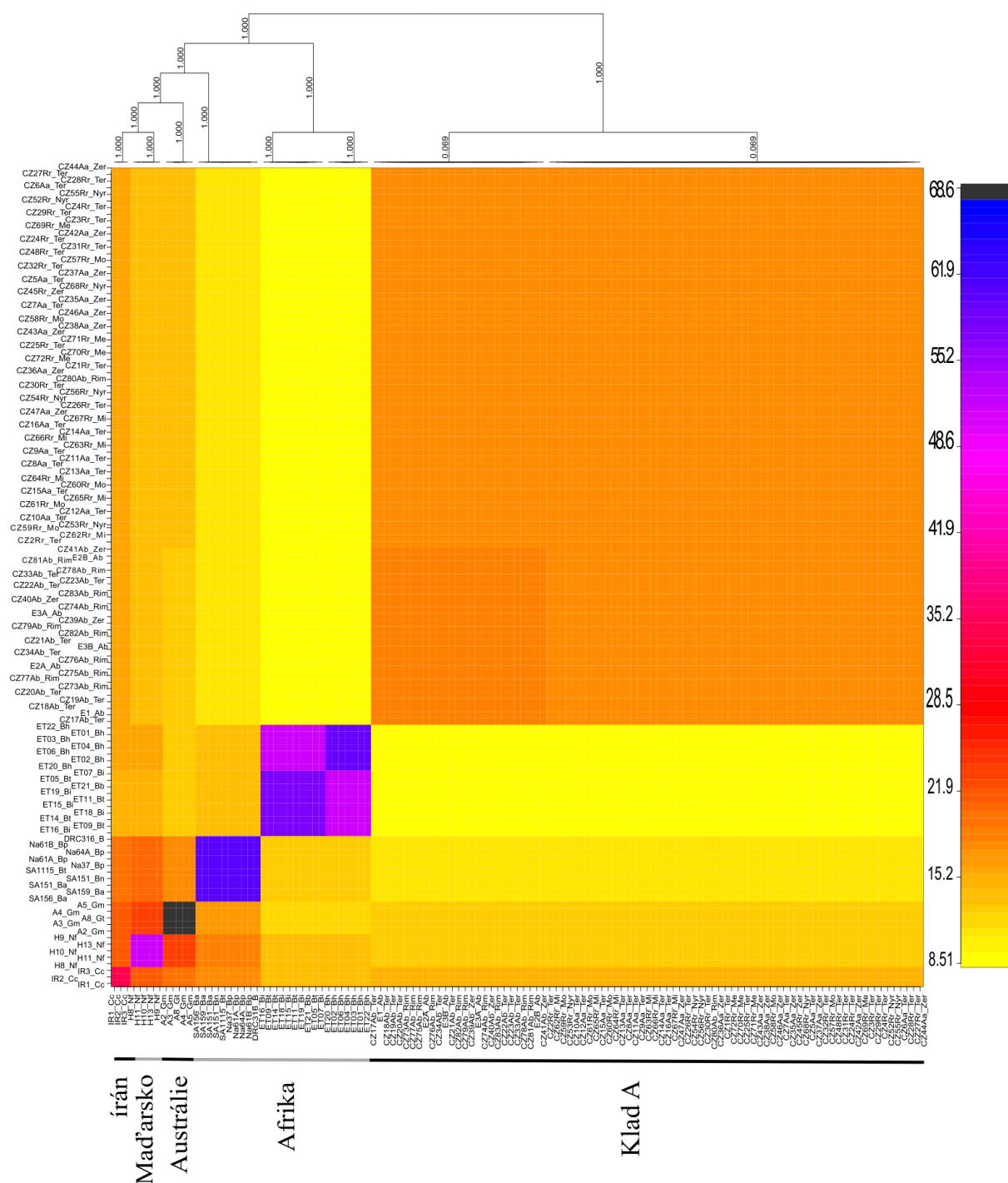
počtu vzorků v jednom kladu, tedy vzorků fylogeneticky spadajícím do kladu A. Správné zvolení hodnot K pro oba datasety (1, 2) bylo testováno pomocí „cross – validation error“ (viz Příloha II). Zvolení neoptimálnějších hodnoty K nebylo zcela jednoznačné, avšak nejnižší hodnoty oscilovaly kolem hodnoty $K = 5$ (1). Zde jsou graficky znázorněné výstupy pro hodnoty $K5$ a $K6$ pro kompletní datadet (1) (viz Obr. 3), respektive pro $K5 - K9$ pro dataset (2) obsahující pouze výběr vzorků kladu A (viz Obr. 5). Klastrovací analýza jasně odhaluje kompaktnost každé z geograficky vzdálených populací. V případě analyzování kompletního počtu vzorků 1 vykazovaly íránské vzorky afinitu k několika klastřům, což se jeví jako artefakt analýzy. (Neodpovídá výsledkům z předchozích mitochondriálních studií/nedává biologický smysl). V druhém případě (2), kdy byl optimalizován počet vzorků z evropských lokalit, se tento jev odstranil a Íránské vzorky utvořily stálý klastř při $K = 6$. Při vyšších hodnotách $K > 6$ docházelo ke štěpení skupin Klad A, Etiopie a klastřu zahrnující africké lokality. Toto štěpení nekopírovalo jemnější geografické rozložení, ale hostitelskou specifitu původu vzorků. Klad A se štěpí dle hostitelů *R. rutilus* a *A. alburnus* versus *A. brama*. Africké lokality a vzorky pocházející z Etiopie se štěpí podle dvou skupin druhů parem z rodu *Barbus*.

Stejné schéma, kopírující geografické rozložení, bylo odhaleno i pomocí „coancestry“ analýz (FineRADstructure). Z grafu je jasně patrná soudružnost vzorků pocházejících se stejných světových lokalit (viz Obr. 4, 6). U vzorků kladu A a Etiopie je možné pozorovat naznačení vnitřní struktury. Tato struktura opět odpovídá hostitelské specifitě stejně jakou u Admixture analýzy. Pokud se zaměříme na výběr vzorků, tedy optimalizovaný počet zástupců kladu A, je zde patrný silnější signál a vyšší počet sdílených lokusů pro jednotlivé skupiny (viz Obr. 6).

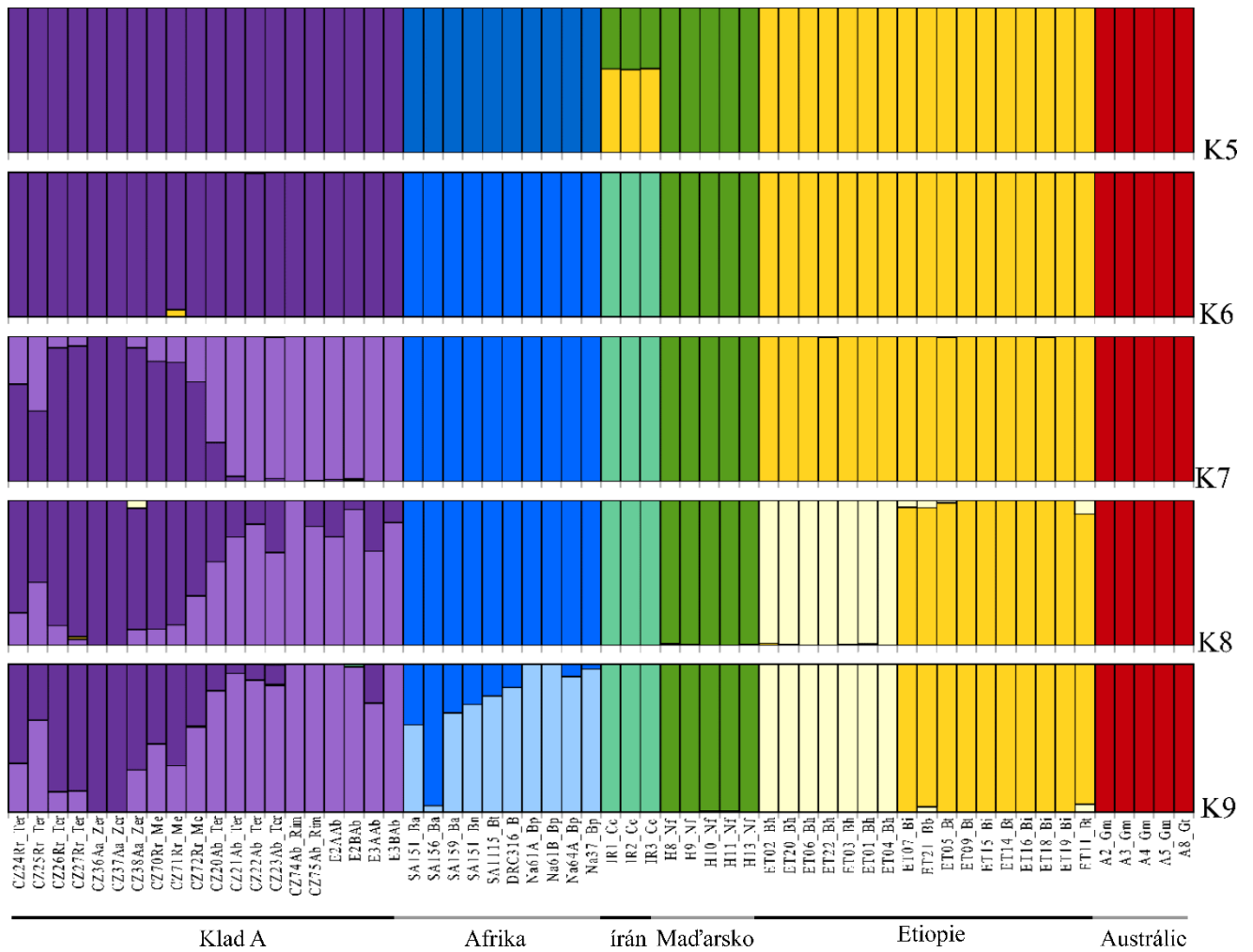
Populační struktura kompletního datasetu 1 byla také zkoumána pomocí PCA analýzy (PCAdapt) (viz Obr. 7). Nejvyšší procento odchylek vysvětlující populační strukturu bylo pozorováno při $K = 2$, tedy mezi první a druhou osou PC (viz Příloha III). Výsledky potvrzují trend, který můžeme nelézt v předchozích analýzách, tedy že geograficky vzdálené populace vytvářejí oddělené klastry.



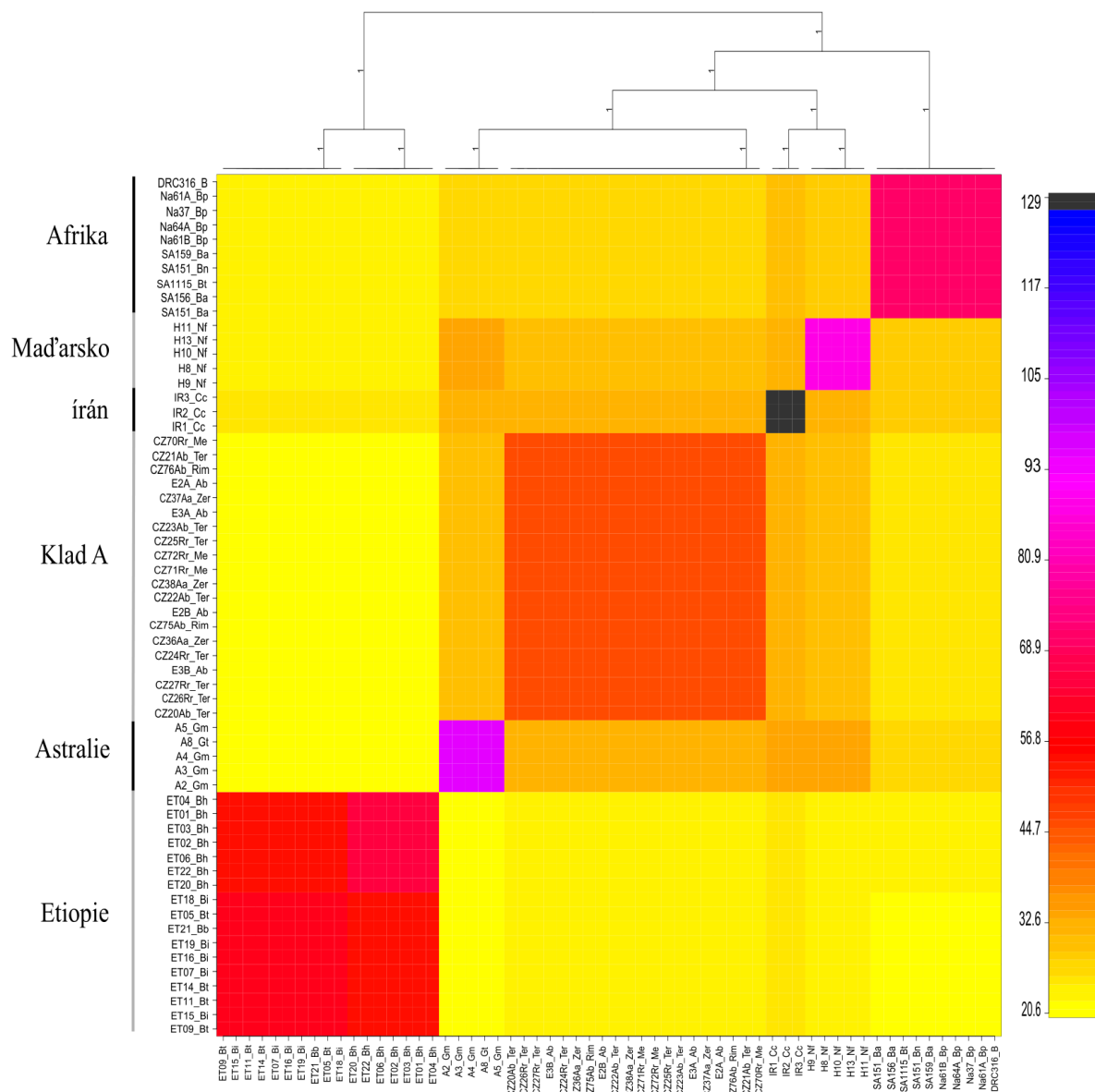
Obr. 3: Analýzy populační struktury všech jedinců (125 vzorků, 1929 lokusů). Výsledky klastrovací analýzy v programu Admixture pro K = 5 a 6.



Obr. 4: Analýzy populační struktury všech jedinců (125 vzorků, 1929 lokusů). Výsledky koancestrální shlukovací analýzy vygenerované programem FineRADstructure, max. 60 % missing data. Barevná škála vyznačuje míru sdílení lokusů mezi jednotlivci. Šedé a černé linie po levé straně označují geografický původ vzorků.

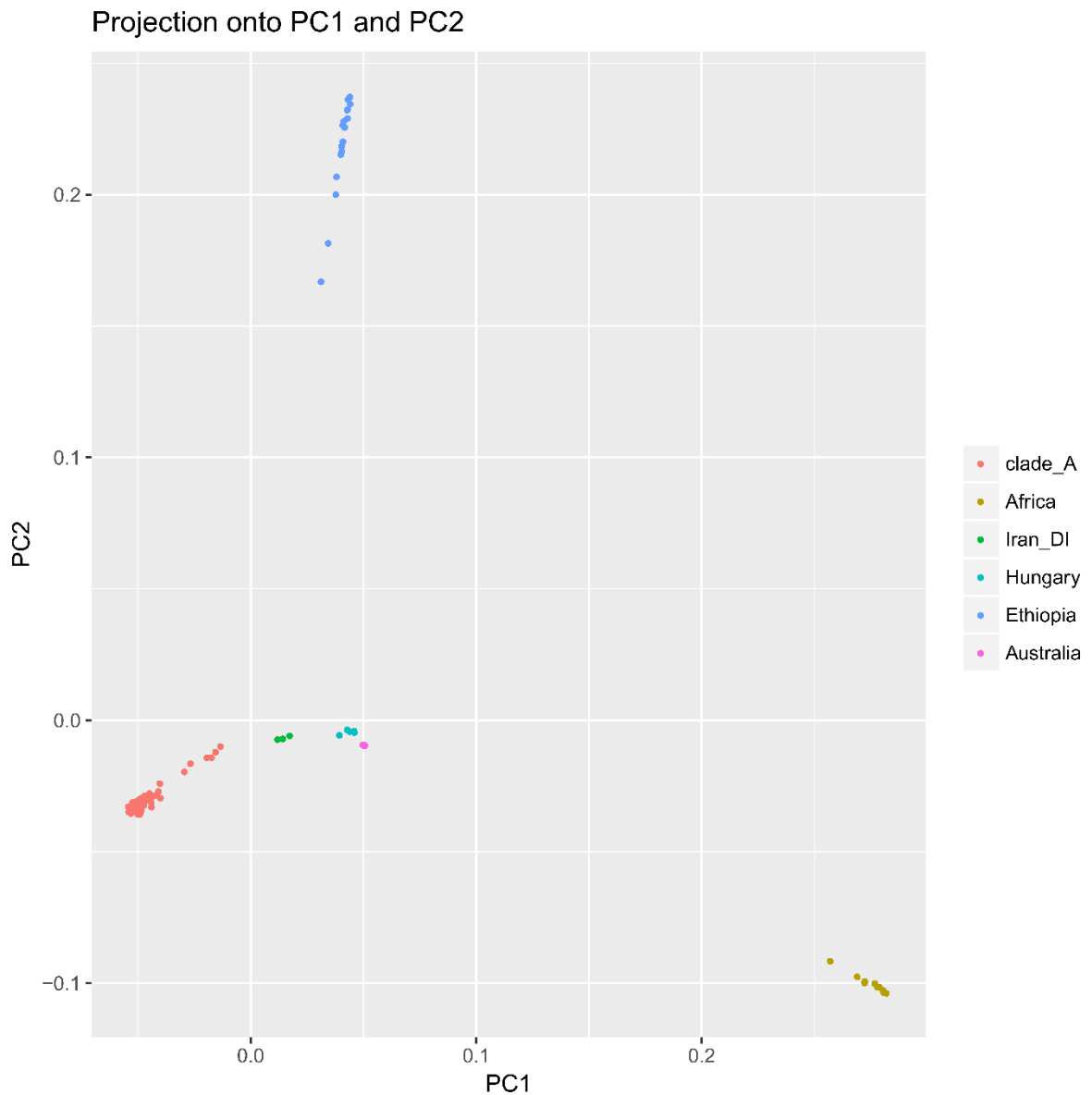


Obr. 5: Analýzy populační struktury světových lokalit a výběru 20 vzorků z evropského kladu A (65 vzorků, 1912 lokusů). Výsledky klastrovací analýzy vygenerované programem Admixture pro $K = 5-9$.



Obr. 6: Analýzy populační struktury světových lokalit a výběru 20 vzorků z evropského kladu A (65 vzorků, 1912 lokusů). Výsledky koancestrální shlukovací analýzy vygenerované programem FineRADstructure, max. 60 % missing data. Barevná škála vyznačuje míru shodných lokusů mezi jednotlivci. Šedé a černé linie po levé straně označují geografický původ vzorků.

Výsledky PCAdapt globální struktury populací *L. intestinalis* také odhalují soudružnost vzorků dle geografického původu. Výsledky PCA (viz Obr. 7) a výsledky F_{ST} statistiky se navzájem podporují, lokality s vyšší hodnotou F_{ST} jsou v grafu dále od sebe (např. Etiopie se zbytkem Afriky, 0,831), naopak lokality, které měly vzájemnou hodnotu F_{ST} nižší se shlukují blízko sebe (např. klad A a Maďarsko, 0,259).



Obr. 7: Populační struktura všech vzorků rozložená podle PC1 a PC2 os, geografické lokality vyznačeny barevně, červená = evropský klad A, žlutá = africké lokality, zelená = Írán, tyrkysová = Maďarsko, modrá = Etiopie, lila = Austrálie.

4.4 Analýzy lokálních patternů

Pro podrobnější náhled do vnitřní struktury lokálních adaptací byly vybrány skupiny vzorků v těsnější geografické vazbě, tedy pro vzorky kladu A (3).

4.4.1 Analýzy populační struktury

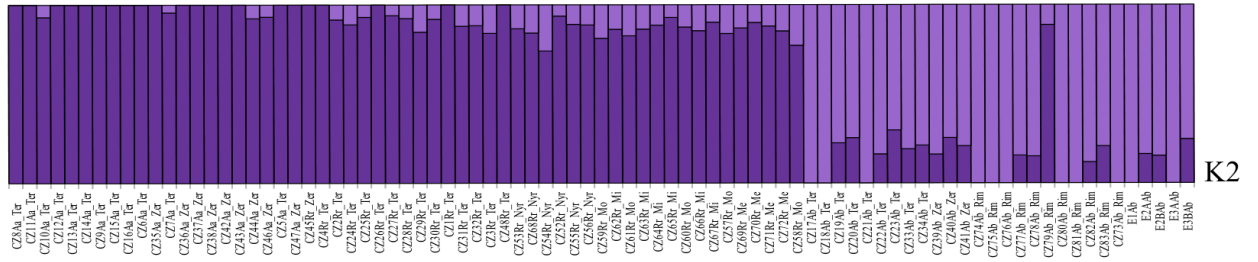
Pro výpočet F_{ST} statistiky byly vzorky kladu A rozděleny dle lokalit výskytu na oblast severu, jihu a východu Česka a Estonsko. Základní F_{ST} statistika neodhalila výrazný genetický rozdíl mezi geografickými skupinami (Tab. 9), ani při porovnávání hostitelských skupin ($Rr+Aa / Ab$) nebyla obdržena výrazná hodnota $F_{ST} = 0,036$. Přesto kastrovací analýza (bez vzorků rozdělených *a priori* do populací) odhalila velmi zřetelné signifikantní štěpení do několika skupin. Optimální počet klastrů byl nalezen pomocí nejnižší hodnoty „cross-validation error“, která jednoznačně určila nejlepší hodnotu $K = 2$ (viz přílohy). Nešlo však o rozdělení dle geografického výskytu, ale podle hostitelského spektra, což vzhledem k výsledkům globální struktury nebylo překvapivé. Stejně jako tomu bylo kompletního datasetu 1, jsou v grafu (viz Obr. 8 A) zobrazeny dvě skupiny – vzorky pocházející z *R. rutilus* a *A. alburnus* vykazují vzájemnou afinitu, proti nim stojí vzorky z hostitele *A. brama*. Jediný vzorek CZ79Ab_Rim (hostitel *A. brama*, lokalita Římov) se vymyká a klastruje k vzorkům pocházejícím z plotic a ouklejí.

Již ve výsledcích koancestrální analýz kompletního datasetu 1 byla naznačená vnitřní strukturovanost kladu A. Tato strukturovanost je výrazněji znát právě na výstupech koancestrální analýzy zahrnující pouze vzorky kladu A 3 (viz Obr. 8 B). Rozdělení opět kopíruje hostitelskou specifičnost, a stejně jako tomu bylo u Admixture analýz, vzorek CZ79Ab_Rim neprojevuje afinitu k ostatním vzorkům z *A. brama*.

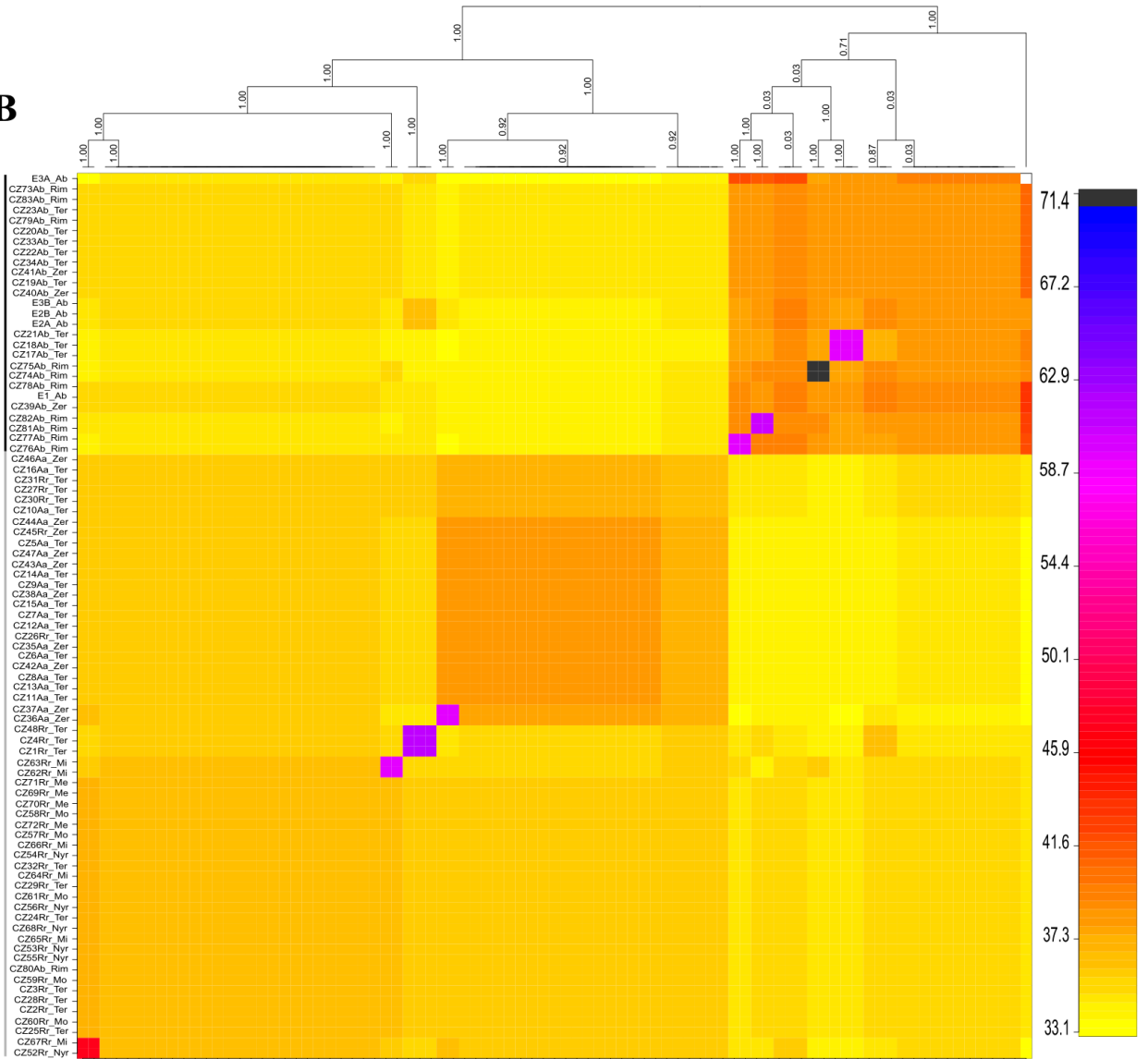
Tab. 9: F_{ST} hodnoty mezi geografickými lokalitami kladu A.

	Východní Čechy	Jižní Čechy	Severní Čechy	Estonsko
Východní Čechy		0.0210147	0.0168204	0.019502
Jižní Čechy			0.0361347	0.0320746
Severní Čechy				0.0598564

A



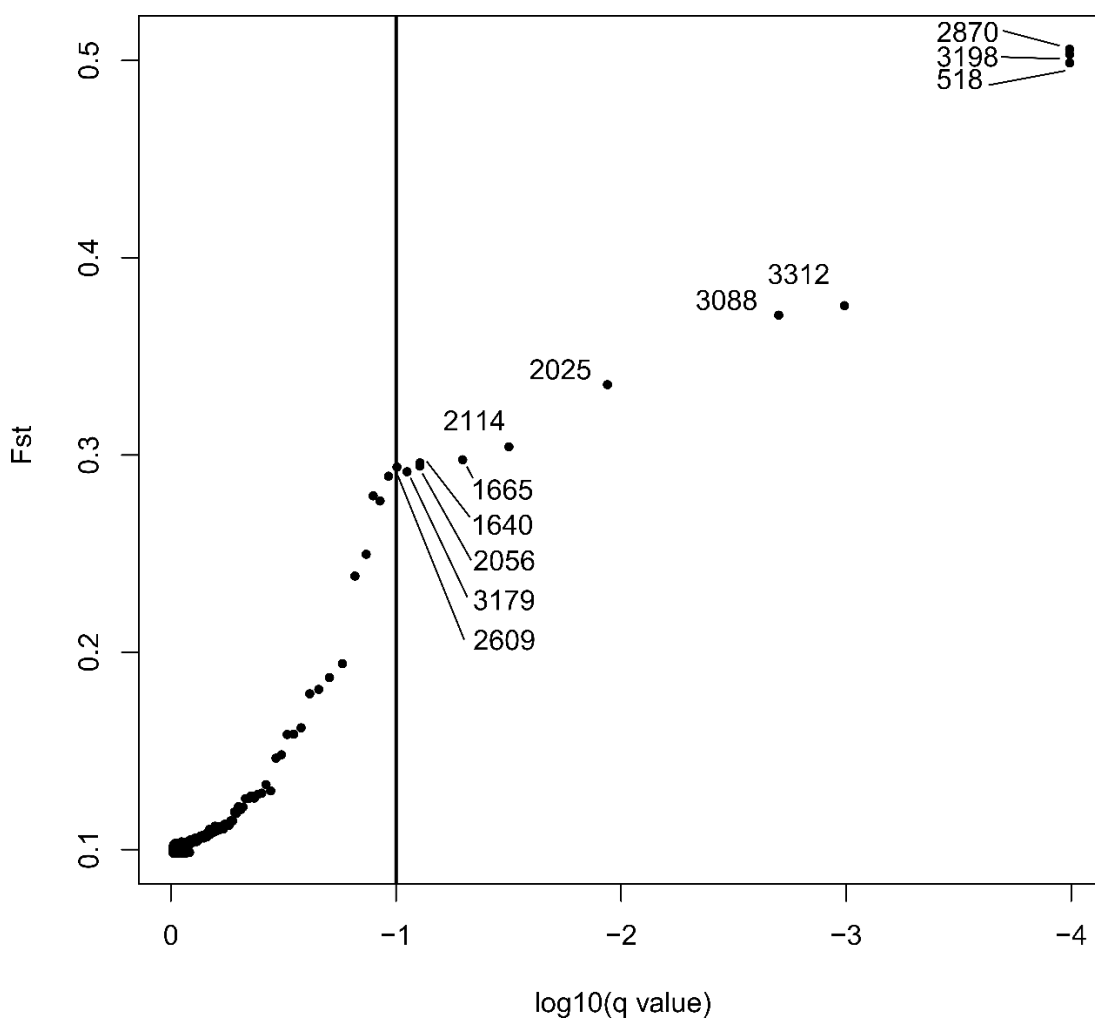
B



Obr. 8: Analýzy populační struktury kladu A (85 vzorků). **(A)** Výsledky klastrovaný analýzy vygenerované programem Admixture pro $K = 2$. **(B)** Výsledky koancestrální shlukovací analýzy vygenerované programem FineRADstructure, max. 60 % missing data. Černou linií označeny vzorky pocházející z hostitele druhu *A. brama*, šedá označuje vzorky z hostitelů *R. rutilus* a *A. alburnus*. Barevná škála vyznačuje míru shodných lokusů mezi jednotlivci.

4.4.2 Analýza „outliers“ u sympatrických lokalit Těrlicko a Žermanice

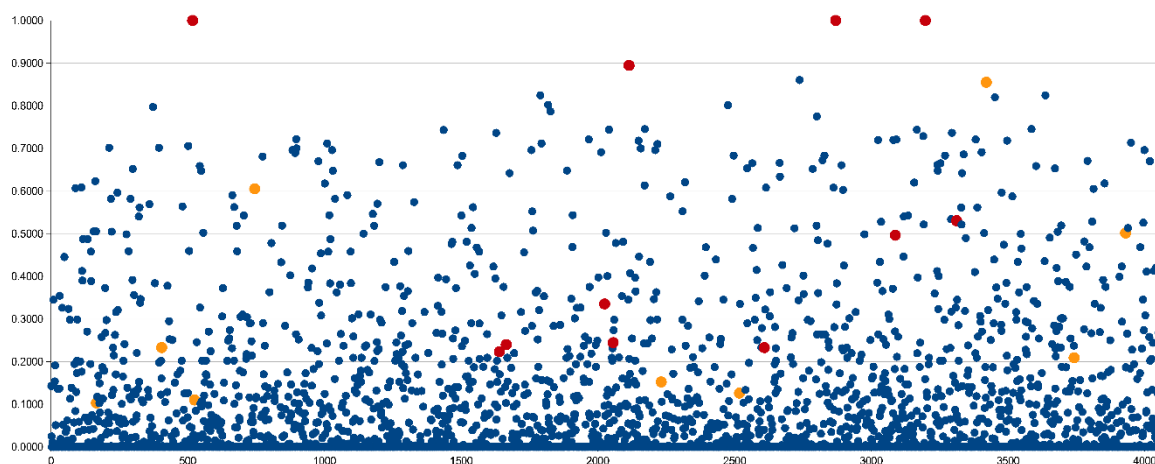
Pro vzorky pocházejících z blízkých lokalit Těrlicko a Žermanice byla vypočítána F_{ST} pro dvě různé situace. V prvním případě byly porovnávány populace z jednotlivých lokalit, tedy Těrlicko a Žermanice, výsledná hodnota F_{ST} byla 0,017. V druhém případě byly vzorky porovnávány na základě rozdělení dle hostitelů, tedy *R. rutilus* (dále jen Rr) + *A. alburnus* (dále jen Aa) vs. *A. brama* (dále jen Ab), v tomto případě byla hodnota F_{ST} 0,047. Dále bylo pracováno s výstupy *populations* (Stacks) s mapováním dle hostitele (Rr+Aa / Ab). V programu BayeScan bylo detekováno 12 „outlier“ lokusů přesahujících hodnotu $FDR \geq 0,1$ (false discovery rate) (viz Obr. 9).



Obr. 9: Detekce F_{ST} outliers lokusů použitím programu BayeScan, rozhodovací faktor pro určení výběru lokusů v logaritmickém měřítku; svislá čára označuje "rozhodující" faktor pro výběr „outliers“ lokusů ($FDR \geq 0,1$). Plné body označují F_{ST} hodnoty pro jednotlivé SNP pozice všech lokusů. Lokusy označené čísly (pořadí RAD lokusů ze Stacks) leží nad hranicí FDR.

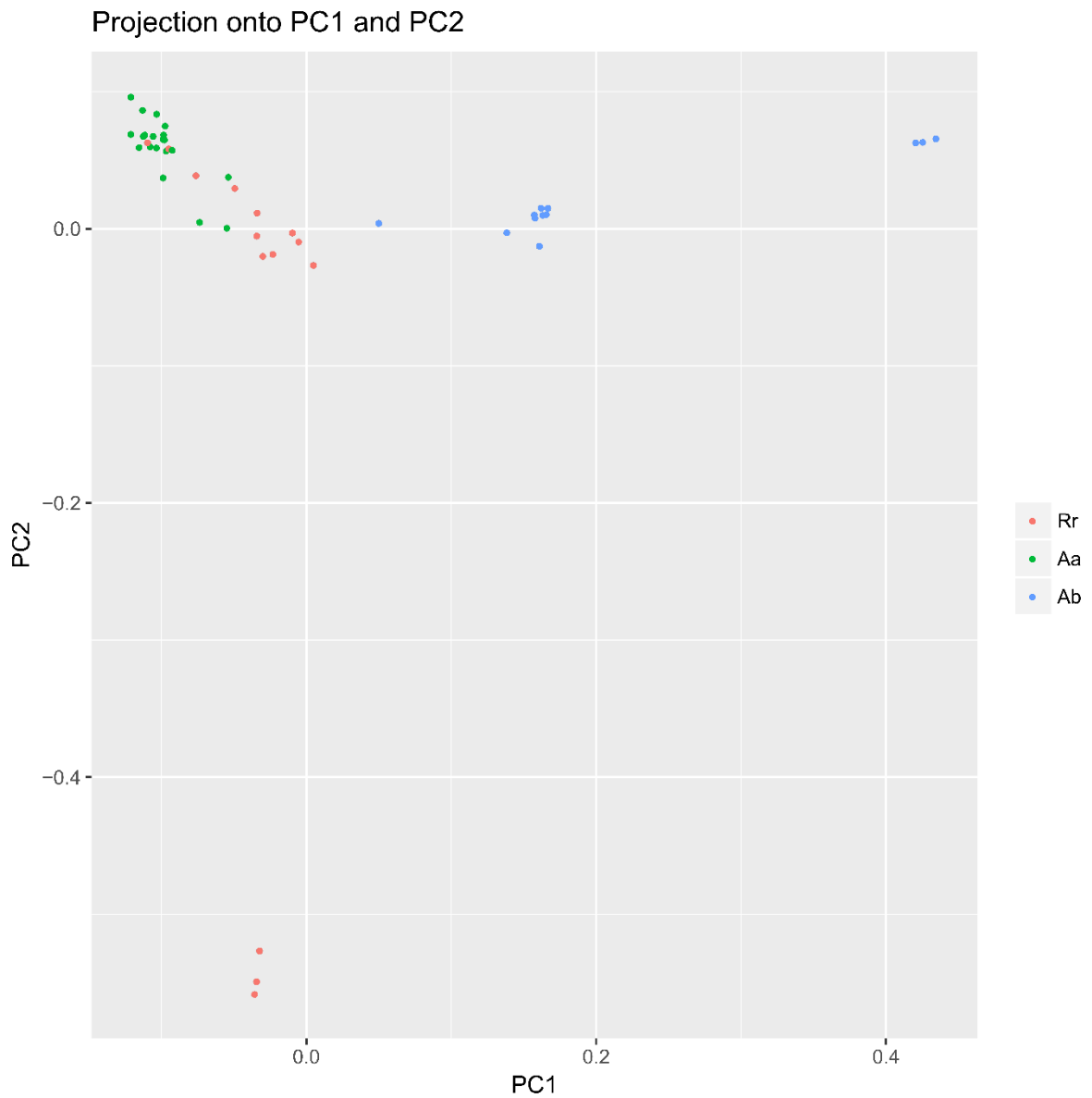
Dalším programem testujícím „outliers“ lokusy a zároveň populační strukturu, který byl v této práci použit byl PCAdapt. Lokusy byly vybírány na základě hodnot P-value při $\alpha = 0,1$. Celkem jich bylo detekováno 801, ale pro porovnání s výstupy dalších analýz byly vybrány pouze ty pozitivní lokusy, které zároveň v BayeScan statistice vykazovaly F_{ST} vyšší než 0,15.

Porovnání obou metod testování „outlier“ lokusů, tedy Bayescan a PCAdapt je znázorněno v grafu promítajícím výsledky obou metod přes sebe (viz Obr. 10). Graf je vyobrazením F_{ST} hodnot každé variabilní pozice získaných ve Stacks, červeně jsou zvýrazněny lokusy vyhodnocené jako „outliers“ pomocí obou metod, tedy BayeScan a PCAdapt, žlutě pak pouze lokusy vybrané programem PCAdapt, a to lokusy s nejvyšší hodnotou P (P – value) a zároveň s hodnotou $F_{ST} > 0,15$.



Obr. 10: Manhattan graf F_{ST} hodnot „outlier“ lokusů (Stacks). Červeně označené lokusy jsou shodně vybrané metodami BayeScan ($FDR \geq 0,1$) a PCAdapt ($P - value > F_{ST} 0,15$), žlutě označené lokusy nalezené pouze programem PCAdapt při stejných kritériích.

Pomocí PCAdapt bylo zkoumáno i populační rozložení jedinců mezi lokalitami Těrlicko a Žermanice. Nejvyšší procento odchylek vysvětlující populační strukturu bylo pozorováno při $K = 2$, tedy mezi první a druhou osou PC (viz přílohy). Z výsledku (viz Obr. 11) je zřejmé, že jedinci *L. intestinalis* pocházející z Rr+Aa vytváří jednu celkem kompaktní skupinu, zatímco jedinci pocházející z Ab se odlišují, zejména podél první osy.



Obr. 11: Populační struktura založená na PC analýze (PCAdapt). Rozložení podle prvních dvou os PC1 a PC2, barevně vyznačeny vzorky pocházející z různých mezihostitelů, Rr = *Rutilus rutilus*, Aa = *Alburnus alburnus*, Ab = *Abramis brama*.

5 Diskuze

Fylogenetická analýza mtDNA sekvencí

Tasemnice *Ligula intestinalis*, jakožto parazit s globálním rozšířením, se vyznačuje několika geneticky oddělenými liniemi, což vyplynulo z dřívějších fylogenetických a populačně strukturních studií založených na sekvenaci mitochondriálních genů, respektive na detekci mikrosatelitních lokusů (Bouzid *et al.* 2008 b, Štefka *et al.* 2009). Tyto studie odhalily jasný vzor v uspořádání linií vzhledem ke geografickému výskytu populací a zároveň byl zjištěn náznak hostitelské specifity u sympatrických linií (klad A a B, Euro-mediteránní oblast). Geografická vzdálenost se zdála být hlavním faktorem určujícím příslušnost do jednotlivých skupin. Toto vysvětlení se jevílo jako logické, obzvláště u populací, kde zjevně nedochází ke genetickému toku s ohledem na možnosti disperze dané životním cyklem parazita, tedy u geograficky velmi vzdálených populací, u kterých nedochází k přemísťování/migraci konečného hostitele (Štefka *et al.* 2009).

Tento trend se dá očekávat u dostatečně vzdálených lokalit. Na druhou stranu i u oblastí, kde je díky konečnému hostiteli (rybožravý pták) umožněn intenzivní genetický tok, se objevují geneticky izolované sympatricky se vyskytující linie. Jak již bylo řečeno, v mediteránní oblasti (Evropa a severní Afrika) se sympatricky vyskytují dva dobře fylogeneticky podpořené linie/klady A a B, což vyplývá z fylogenetických analýz mitochondriálních genů. Mezi klady A a B byla pozorována rozdílnost v preferenci rybiho mezhospitele, která může být odrazem adaptace parazita na rozdílné imunitní systémy odlišných druhů ryb (Arme, 1997). Zástupci kladu A infikují odvozené kaprovité ryby podčeledi Leuciscinae, zatímco jedinci spadající do kladu B infikují druhy, které se klastrují blíže ke kořeni čeledě kaprovitých ryb (Bouzid *et al.* 2008 b).

Z výsledků fylogenetických analýz nových vzorků zkoumaných v této práci vyplývá, že podobná situace, jako u kladů A a B nastává i u jedinců pocházejících z oblastí subsaharské Afriky. I přes nižší počet vzorků je zřejmá vnitřní populační struktura u lokalit, z nichž některé dělí poměrně velké geografické vzdálenosti (Etiopie, Keňa, Demokratická Republika Kongo, Jihoafrická Republika), avšak nejedná se o oblasti oddělené jasnou bariérou jako v případě mezikontinentálních izolací. Vzorky byly rozděleny do dvou zcela oddělených, dobře podpořených větví. Jedinci pocházející z Keni, DRK a JAR jsou na fylogenetickém stromě postaveny bazálně, zatímco etiopské vzorky okupují jednu z vnitřních větví. Pokud se zaměříme na hostitelské spektrum, všechny zkoumané vzorky byly nalezeny v různých

zástupcích palem rodu *Barbus* (Cyprinidae). Je velice pravděpodobné, že tasemnice *L. intestinalis* je v afrických lokalitách rozdělena do několika linií v závislosti na kombinaci hostitelské specifity spolu s omezeným genovým tokem. Každopádně je subsaharská Afrika zajímavou oblastí budoucích studií. V Africe lze nalézt více než 300 zástupců rodu *Barbus*. Taxonomie, založená na fenotypových znacích, těchto ryb není zatím ustálená a molekulární data nejsou k dispozici (Tsigenopoulos *et al.* 2002). Vzhledem k doposud nejednoznačné fylogenezi nelze porovnávat koevoluci parazit vs. hostitel ve smyslu kospeciální analýzy, ale lze hledat linie, které mají preference k určitému druhu (skupině) ryb. Nabízí se tedy prostor k dalším výzkumům, sběru vzorků a analýze dat k pochopení komplexu linií *L. intestinalis*, k čemuž mohou přispět zejména analýzy založené na SNP datech.

Metodické aspekty RADseq analýzy

Získávání genomových dat pomocí stále se rozvíjejících a dostupnějších RADseq metod se za poslední desetiletí stalo průlomovým objevem a stále více využívaným nástrojem pro ekologické, populační i fylogenetické studie (Andrews *et al.* 2016). Příprava genomových knihoven metodou ddRAD slibuje získání až desítek tisíc SNP pro desítky až stovky jedinců naráz (Peterson *et al.* 2012). V této práci se podařilo po filtraci dat získat nižší počet SNP, v závislosti na porovnávaných vzorcích, a to 1897–4064. Tento menší počet SNP může být způsoben více faktory, například samotnou biologickou podstatou *L. intestinalis* (předpokládaným malý genom, výrazná genetická vzdálenost zkoumaných populací) nebo metodologickým aspektem (možné snížení kvantity DNA v průběhu přípravy knihovny, metodickou chybou). Také vyřazení celých dvou podknihoven L2 a L3 z analýz mohlo být způsobeno už možnými nedostatky v přípravě podknihoven. V těchto dvou podknihovnách převažovaly starší izoláty DNA, u kterých nebylo možné provést nově extrakce DNA (vzorky z dřívějších studií, u kterých již nebyla tkáň pro izolaci k dispozici). Tyto vzorky mohly obsahovat již degradovanou DNA, a tím mohlo dojít k narušení průběhu přípravy knihoven. Dalším limitujícím faktorem mohl být fakt, že tyto dvě podknihovny byly separovány na přístroji Pippin Prep v jiném běhu, mohlo tak dojít k separaci fragmentů s mírně odchylenou šířkou rozpětí délek fragmentů. Z toho vyplývá, možný problém při hledání homologních sekvencí s ostatními podknihovny. Nedá se vyloučit ani chyba při PCR, kdy mohlo dojít k amplifikaci některých fragmentů více než ostatních (amplification bias), tím může nastat problém při sekvenaci, kdy může dojít k výraznějšímu prosekvenování jedné skupiny lokusů oproti jiné, což způsobuje rozdílnost v hloubce pokrytí přes všechny lokusy (Andrews *et al.* 2016). Pro následné analýzy tedy bylo nutné redukovat dataset jen na vzorky tří podknihoven

(125 jedinců), což však umožnilo vyprodukovat SNP datasety s daleko menší „děravostí“ pokryvu jednotlivými lokusy.

Globální diverzita linií pomocí SNP dat

Při analýze globální populační struktury (všechny vzorky, 125 jedinců, 1929 lokusů) byla odhalena silná hierarchie populací dle geografického původu. Toto rozdělení velmi dobře podporují výsledky *F_{st}* statistik (viz Tab. 7, 8). Vysoké hodnoty naznačují silnou diferenciaci mezi velmi vzdálenými populacemi (např. Austrálie vs. Írán, Austrálie vs. Etiopie, Etiopie vs. Maďarsko apod.). Při vizualizaci výsledku klastrovacích analýz (Admixture) byl tento trend potvrzen, většina geografických oblastí je jasně vymezena (dle barevného rozložení grafu) (viz Obr. 3). Jen u íránských vzorků je patrná vnitřní fragmentovanost jedinců, respektive afinita k více odlišným klastrům. Sdílený signál by se dal interpretovat jako možnost výskytu hybridních linií, spíše se však jedná o artefakt analýzy z důvodu menšího počtu íránských jedinců a nutnosti modelovat vyšší počet klastrů pro jejich oddělení do samostatné linie. Hybridizace také není dobrým vysvětlením z hlediska geografického rozšíření sdílených klastrů a výsledků ostatních analýz: a) podle výsledků Admixture je íránská populace fragmentovaná mezi populace Maďarska, kladu A a Jižní Afriku, čemuž neodpovídají výsledky *F_{st}* statistiky (viz Tab. 7), b) ve výsledku koancestrální shlukovací analýzy (fineRADstructure) stojí íránské vzorky samostatně a zcela oddělené od kladu A, i ostatních populací se kterými mají sdílený signál v klastrovací analýze (viz Obr. 4), c) vzorky z Íránu v předchozích fylogenetických analýzách mitochondriálních genů spadaly k ostatním jedincům tasemnice *L. interrupta* zařazených do studie, d) dřívější mikrosatelitní analýzy také přiřadily íránské vzorky k *L. interrupta* bez náznaku fragmentovanosti. Íránští jedinci byli tedy klasifikováni jako příbuzný druh tasemnice *L. intestinalis*, která je ve fylogenetickém smyslu parafyletickým druhem, nebo spíše komplexem kryptických druhů. Hybridizaci mezi jednotlivými liniemi nelze zcela vyloučit, ale pravděpodobnost, že by k ní došlo, není velká. Na druhou stranu mohla být analýza ovlivněna mnohonásobně vyšším počtem jedinců kladu A.

Po snížení tohoto počtu, vygenerováním nového datasetu zahrnujícího všechny světové lokality a výběr dvaceti vzorků kladu A (celkem 65 jedinců), byl tento problém, i přes lehce snížený počet lokusů oproti kompletnímu datasetu (1929/1912), eliminován. Z výsledků klastrovacích analýz provedených na redukovaném datasetu (viz Obr. 5) je pozorovatelná kompaktnost světových populací při nižších hodnotách *K* (5–6), při $K \geq 6$ bylo možně sledovat

samostatný klastř už i v případě íránské populace, který vykazoval soudružnost i při narůstajícím počtu K. Při hodnotách $K \geq 7$ docházelo ke štěpení celků zahrnujících více lokalit, tedy kladu A, jihoafrických populací a etiopského klastř. Zajímavým zjištěním bylo, že se klastř nerozdělovaly podle jemnější geografie, ale podle hostitelů. Klad A se rozdělil na dvě skupiny, a to na vzorky pocházející z *R. rutilus* společně s *A. alburnus* vs. *A. brama*. Etiopské vzorky se rozdělily také na dvě skupiny dle rybích mezihostitelů, *B. humilis* vs. *B. tsanensis*, *B. intermedius* a *B. brevicephalus*. Jihoafrické vzorky se také štěpily na dva klastř, ale nebylo možné jednoznačně pozorovat nějaký trend, ani v geografii ani v hostitelské specifitě, pravděpodobně kvůli nízkému počtu vzorků.

U dřívějších mikrosatelitních analýz bylo taktéž pozorováno rozdělení linií dle geografie, avšak na jemnější úrovni nebyla odhalena zjevná souvislost mezi genetickou strukturou *L. intestinalis* a odlišnými rybími mezihostiteli (Štefka *et al.* 2009). Náznak obdobné situace byl při mikrosatelitních analýzách pozorován pouze etiopských vzorků, kdy došlo k rozdělení na klastř podle hostitelů různých druhů palem rodu *Barbus* (Štefka *et al.* 2009).

Porovnáním mikrosatelitních dat a dat získaných metodami RAD sekvenováním nezjišťujeme zásadní rozdíl ve výsledcích globální populační struktury, ale výhody RAD dat přesto nelze ignorovat. Genomová data umožňují zřetelně hlubší náhled do jemnější struktury lokálních populací (např. Jeffries *et al.* 2016). Jeffries s kolegy (2016) zkoumali populační strukturu *Carassius carassius* napříč evropskými lokalitami a porovnávali výsledky získané sekvenací mitochondriální DNA, mikrosatelitních lokusů a RAD sekvenováním. Z jejich výsledků vyplývá, že první dvě metody nejsou dostatečně silné, aby do hloubky odhalily dělení linií (Jeffries *et al.* 2016). V této práci tak bylo dosaženo podobných rozdílů v citlivosti analýz podle typu dat. Za zmínku stojí i zřejmě lepší rozlišovací schopnost RADseq dat oproti AFLP analýzám. Zejména v některých koevolučních studiích, např. ve studii porovávající populační strukturu tasemnic druhu *Wenyonia virilis* jejich rybích hostitelů (Jirsová *et al.* 2017), měla AFLP data dokonce nižší rozlišovací schopnost než mtDNA sekvence, a to i přes poměrně vysoký počet analyzovaných lokusů (řádově stovky až tisíce).

Silnější signál pro rozlišení klastřů *L. intestinalis* po vyčlenění reprezentativního vzorku kladu A a vyšší počet sdílených lokusů jednotlivých lokalit byl pozorován i v koancestrální analýze (FineRADstructure) (viz Obr. 6). Jasně vymezené oblasti v populační struktuře jedinců jasně oddělují geografické oblasti.

Překvapivým zjištěním byla vysoká hodnota F_{st} mezi etiopskými vzorky a vzorky ze zbytku lokalit subsaharské Afriky – „Jižní Afrika“ (0,831 v případě kompletního datasetu, 0,744 při redukci počtu jedinců kladu A). Tyto lokality se od sebe signifikantně oddělují již při fylogenetických analýzách mitochondriálních genů a jejich genetická odlišnost je dalšími analýzami (PCA, Admixture, FineRADstructure) potvrzena. I při celkem nízkém počtu vzorků a nevelkém počtu zkoumaných lokalit je tento výsledek zajímavý. Je zároveň velmi pravděpodobné, že při hlubším zkoumání afrických lokalit by byla odhalena rozsáhlejší a jemnější populační struktura.

Lokální diverzita populací se sympatrickými hostiteli

Již v roce 2001 narazila Géraldine Loot se svými kolegy na situaci, kdy nebylo možné za stejných laboratorních podmínek infikovat tasemnicemi *L. intestinalis* dva rozdílné druhy kaprovitých ryb cejny (*Blicca bjoerkna*) a plotice (*Rutilus rutilus*). Loot předpokládala, že u *B. bjoerkna* (pocházející z lokality, kde přirozeně nedochází k infekci tohoto druhu ryby tasemnicí *L. intestinalis*), se mohou vyskytovat fyziologické a imunologické bariéry zabráňující vývoji tasemnice z procerkoidu do dalšího larválního stádia, tedy plerocerkoidu. V práci připouští hypotézy, že jiné linie *L. intestinalis* jsou schopny kolonizovat *B. bjoerkna* v odlišných lokalitách, nebo že *L. intestinalis* vytváří komplex druhů, které lokálně parazitují na různých rybích populacích (Loot *et al.* 2001). V předchozích populačních studiích zaměřených na analýzu mikrosatelitních lokusů nebyl objeven žádný zjevný trend v rozdělení kladu A dle hostitelské specifity ani geografického původu v rámci Evropy (Štefka *et al.* 2009). Díky analýzám genomových dat byla odhalena jasná genetická diverzita mezi jedinci kladu A pocházející z různých hostitelů. Klastrovací i koancestrální analýza jasně oddělila dvě skupiny, první zahrnovala jedince z *R. rutilus* a *A. alburnus*, druhá z *A. brama*. Na druhou stranu vzorky nevykazovaly žádnou strukturovanost dle geografického původu (viz Obr. 8). Zdá se tedy, že hlavním determinačním jevem geneticky rozdělujícím evropské populace *L. intestinalis* na více linií, je hostitelská specifita, nikoli geografická vzdálenost, nebo jiné environmentální bariéry.

Situace, kdy dochází ke speciaci u geograficky blízkých linií/ekotypů jednoho živočišného druhu, není ojedinělá. U křivek (*Loxia curvirosta* – komplex morfologicky a vokálně rozlišených ekotypů) nebyla potvrzena geografická izolace (IBD). Přesto došlo ke genetické izolovanosti jedné linie v závislosti na potravní preferenci, tedy v závislosti na specializovanosti vůči hostitelské rostlině, v tomto případě na borovicích *Pinus contorta*.

Zároveň dochází k reprodukční izolaci tohoto typu od ostatních. I přes mírnou celogenomovou divergenci mezi ekotypy byl objeven důkaz relativní genetické diferenciace (měřené F_{ST}) mezi jinak genetickou soudržností v ostatních ekotypech (Parchman *et al.* 2016). Obdobný vzorec byl nalezen i zde. Přes poměrně nízkou diferenciaci ve smyslu F_{ST} distance byly pozorované preference odlišných skupin rybích mezipřehoditelů téměř absolutní. Pouze v jednom případě došlo k „přeskoku“ genetické formy typické pro *R. rutilus* a *A. alburnus* na jiného hostitele – *A. brama* z nádrže Římov. To ale zároveň naznačuje, že nejde o zcela evolučně izolované linie, které tak mohou být zajímavé pro navazující výzkum vzniku a selekce adaptací v sympatrických populacích.

K vyhledávání lokusů, na které působí selekce, je osvědčeným (byť ne bezchybným) nástrojem tzv. „ F_{ST} outliers test“ (Holsinger a Weir 2009). Pro toto testování je dostupný nemalý počet různých programů, v této práci byly výstupy ze Stacks (SNP a hodnoty F_{ST}) testovány pomocí programů BayeScan a PCAdapt. Vhodnými adepty na vyhledávání lokusů podléhající selekci se staly vzorky pocházející z geograficky blízkých lokalit Těrlicko a Žermanice (severo-východní Morava). Obě tyto lokality se vyznačovaly přítomností *L. intestinalis* u tří rybích mezipřehoditelů: *R. rutilus*, *A. alburnus* a *A. brama*. Tím bylo zaručeno testování rozdílných mezipřehoditelů a zároveň zamezení případnému zkreslení výsledků při testování geograficky výrazně oddělených linií. Nejdříve byly porovnány F_{ST} hodnoty mezi dvěma různými nastaveními dat – rozdělení podle geografie (Těrlicko vs. Žermanice) a rozdělení podle hostitelské preference (také dvě skupiny a to *R. rutilus* a *A. alburnus* vs. *A. brama*). V prvním případě byla hodnota F_{ST} 0,017, v druhém případě bylo naměřeno 0,047, v obou případech se jedná o nízké hodnoty a divergence mezi geografickými, respektive hostitelskými, skupinami není výrazná. Vzhledem k tomu, že nás zajímalo rozložení lokusů mezi hostitelskými skupinami, bylo dále pracováno s daty vygenerovanými právě v tomto nastavení. Celkem bylo testováno 4064 lokusů získaných po filtraci dat programem Stacks. Lokusy označené jednotlivými programy jako „outliers“ byly vyneseny do Manhattan grafu F_{ST} hodnot vypočítaných programem Stacks. BayeScan vyhodnotil jako „outliers“ 12 lokusů (viz Obr. 9), všechny tyto lokusy byly shodně vyhodnoceny i PCAdapt programem, který nebyl tak striktní a jako „outliers“ označil dalších 9 lokusů. Byly zaznamenány lokusy s vysokými hodnotami F_{ST} , podle Stacks dokonce 3 lokusy dosahovaly hodnoty 1, což by znamenalo, že jsou tyto lokusy pod silným tlakem divergentní selekce. V porovnání všech hodnot F_{ST} „outliers“ lokusů (Stacks vs. BayeScan) je zřejmé, že výsledky BayeScan jsou poměrně konzervativní, a zdaleka ne všechny lokusy s vysokými hodnotami F_{ST} (Stacks) byly

BayeScanem identifikovány jako “outliers”, stejně jako lokusy PCAdapt. BayeScan je tedy nejstriktnější metodou výběru lokusů pod selekcí. Z výsledků analýz každopádně vyplývá, že hostitelsky specifické skupiny zkoumané tasemnice mají mezi sebou nezanedbatelný genetický rozdíl.

Z hlediska budoucího výzkumu bude rozhodující získat referenční genom *L. intestinalis*, aby bylo možné “outliers” lokusy namapovat na referenci, ověřit přítomnost kódujících lokusů v genetické vazbě s F_{ST} outliers a zjistit genetickou funkci pro dané lokusy. Bohužel genom *L. intestinalis* zatím osekvenován nebyl. Nejpříbuznější tasemnicí s osekvenovaným genomem je *S. solidus*. Ale vzhledem k nenalezení žádných homologních lokusů během analýzy ddRAD dat mezi zástupci *L. intestinalis* a zařazenými vzorky *S. solidus*, nebylo považováno za relevantní F_{ST} outlier lokusy mapovat na genom *S. solidus*.

6 Závěr

Díky genomovým datům získaným restriční metodou ddRAD a následným next-generation sekvenováním se i přes ztrátu části dat při filtraci podařila objasnit populační struktura *L. intestinalis* na mnohem jemnější úrovni, než bylo doposud možné pomocí klasických metod (analýza mitochondriálních genů, mikrosatelitní analýzy). Tyto výsledky byly potvrzeny několika různými statistickými přístupy.

Zkoumáním hostitelsky specifických, sympatrických populací uvnitř evropského kladu A byly odhaleny lokusy, které zjevně souvisí se selekcí způsobenou preferencí odlišných druhů rybích meziphostitelů. Tím byla potvrzena hypotéza, že *L. intestinalis* se separuje do různých linií nejen v závislosti na geografickém výskytu, ale i na hostitelské preferenci.

Africké lokality vykazující geograficky i hostitelsky podmíněnou genetickou diverzitu nabízejí obrovský potenciál pro další výzkumy zaměřené jak na ekologii druhu, tak na diferenciaci podmíněnou vznikem lokálních adaptací.

7 Literatura

- Allendorf FW (2017) Genetics and the conservation of natural populations: allozymes to genomes. *Molecular Ecology*, **26**, 420–430.
- Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, **19**, 1655–1664.
- Andrew HL, Eaton DAR, Cavender-Bares J, Fitzek E, Nipper R, Manos PS (2014) A framework phylogeny of the American oak clade based on sequenced RAD data. *PLoS ONE*, **9**, e93975.
- Andrews S (2010) FastQC: a quality control tool for high throughput sequence data. Dostupný na webu: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
- Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA (2016) Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, **17**, 81–92.
- Arme C (1997) Ligulosis in two cyprinid hosts: *Rutilus rutilus* and *Gobio gobio*. *Helmithologia*, **34**, 191–196.
- Bazinet AL, Zwickl DJ, Cummings MP (2014) A gateway for phylogenetic analysis powered by grid computing featuring GARLI 2.0. *Systematic Biology*, **63**, 812–818.
- Bouزيد W, Lek S, Mace M, Ben Hassine OK, Etienne R, Legal L, Loot G (2008 a) Genetic diversity of *Ligula intestinalis* (Cestoda: Diphylobothriidea) based on analysis of inter-simple sequence repeat markers. *Journal of Zoological Systematics and Evolutionary Research*, **46**, 289–296.
- Bouزيد W, Štefka J, Hypša V, Lek S, Scholz T, Legal L, Ben Hassine OK, Loot G (2008 b) Geography and host specificity: two forces behind the genetic structure of the freshwater fish parasite *Ligula intestinalis* (Cestoda: Diphylobothriidae). *International Journal of Parasitology*, **38**, 1465–1479.
- Bolnick DI, Fitzpatrick BM (2007) Sympatric speciation: models and empirical evidence. *Annual Review of Ecology, Evolution, and Systematics*, **38**, 459–487.
- Bouزيد W, Štefka J, Bahir-Sfar L, Beerli P, Loot G, Lek S, Haddaoui N, Hypša V, Scholz T, Dkhil-Abbes T, Meddour R, Ben Hassine OK (2013) Pathways of cryptic invasion in a

- fish parasite traced using coalescent analysis and epidemiological survey. *Biological Invasions*, **15**, 1907–1923.
- Brown SP, Loot G, Grenfell BT, Guégan JF (2002) Host manipulation by *Ligula intestinalis*: accident or adaptation? *Parasitology*, **123**, 519–529.
- Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011) Stacks: building and genotyping loci de novo from short-read sequences. *G3: Genes, Genome, Genetics*, **1**, 171–182.
- Conover DO, Duffy TA, Hice LA (2009) The covariance between genetic and environmental influences across ecological gradients: reassessing the evolutionary significance of countergradient and cogradient variation. *Annals of the New York Academy of Sciences*, **1168**, 100–129.
- Criscione CD, Blouin MS (2004) Life cycles shape parasite evolution: comparative population genetics of salmon trematodes. *Evolution*, **58**, 198–202.
- Criscione CD, Poulin R, Blouin MS (2005) Molecular ecology of parasites: elucidating ecological and microevolutionary processes. *Molecular Ecology*, **14**, 2247–2257.
- Cruickshank TE, Hahn MW (2014) Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, **23**, 3133–3157.
- Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, **12**, 499–510.
- Dubinina MN (1980) Tapeworms (Cestoda, Ligulidae) of the fauna of the USSR. *New Delhi, India, Amerind Publishing Co.*
- Duforet-Frebourg N, Bazin E, Blum MGB (2014) Genome scans for detecting footprints of local adaptation using a Bayesian factor model. *Molecular Biology and Evolution*, **31**, 2483–2495.
- Emerson KJ, Clayton RM, Catchen JM, Hohenlohe PA, Cresko WA, Bradshaw WE, Holzapfel CM (2010) Resolving postglacial phylogeography using high-throughput sequencing. *Proceedings of the National Academy of Sciences*, **107**, 16196–16200.

- Ewels P, Magnusson M, Lundin S, Kaller M (2016) MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, **32**, 3047–3048.
- Falush D, Stephens M, Pritchard JK (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular Ecology Notes*, **7**, 574–578.
- Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*, **180**, 977–993.
- Garnier S, Alibert P, Audiot P, Prieur B, Rasplus J-Y (2004) Isolation by distance and sharp discontinuities in gene frequencies: implications for the phylogeography of an alpine insect species, *Carabus solieri*. *Molecular Ecology*, **13**, 1883–1897.
- Glez-Peña D, Gómez-Blanco D, Reboiro-Jato M, Fdez-Riverola F, Posada (2010) ALTER: program-oriented conversion of DNA and protein alignments. *Nucleic Acids Research*, **38**.
- Holsinger KE, Weir BS (2009) Genetics in geographically structured populations: defining, estimating and interpreting F_{ST} . *Nature Reviews Genetics*, **10**, 639–650.
- Hoole D, Carter V, Dufour S (2010) *Ligula intestinalis* (Cestoda: Pseudophyllidea): an ideal fish-metazoan parasite model? *Parasitology*, **137**, 425–438.
- Hufbauer RA, Roderick GK (2005) Microevolution in biological control: mechanisms, patterns, and processes. *Biological Control*, **35**, 227–239.
- Huyse T, Poulin R, Théron A (2005) Speciation in parasites: a population genetics approach. *Trends in Parasitology*, **21**, 469–475.
- Jeffries DL, Copp GH, Handley LL, Olsén KH, Sayer CD, Hänfling B (2016) Comparing RADseq and microsatellites to infer complex phylogeographic patterns, an empirical perspective in the Crucian carp, *Carassius carassius*, L. *Molecular Ecology*, **25**, 2997–3018.
- Jirsová D, Štefka J, Jirků M (2017) Discordant population histories of host and its parasite: a role for ecological permeability of extreme environment? *PLoS ONE*, **12**, e0175286. <https://doi.org/10.1371/journal.pone.0175286>

- Johnson KP, Williams BL, Drown DM, Adams RJ, Clayton DH (2002) The population genetics of host specificity: genetic differentiation in dove lice. *Molecular Ecology*, **11**, 25–38.
- Kočová P. (2015) Genetická struktura populací tasemnice *Ligula intestinalis*. *Bakalářská práce, České Budějovice, Přírodovědecká fakulta Jihočeské Univerzity*.
- Lanfear R, Frandsen P B, Wright AM, Senfeld T, Calcott B (2016) PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution*, **34**, 772–773.
- Lagrange C, Presswell B, Dunckley N, Poulin R (2018) The invasive cestode parasite *Ligula* from salmonids and bullies on the South Island, New Zealand. *Parasitology Research*, **117**, 151–156.
- Li J, Liao X, Yang H (2000) Molecular characterization of a parasitic tapeworm (*Ligula*) based on DNA sequences from formalin-fixed specimens. *Biochemical Genetics*, **38**, 309–322.
- Li J, Liao X (2003) The taxonomic status of *Digramma* (Pseudophyllidae: Ligulidae) inferred from DNA sequences. *Journal of Parasitology*, **89**, 792–799.
- Lion S, Gandon S (2015) Evolution of spatially structured host-parasite interactions. *Journal of Evolutionary Biology*, **28**, 10–28.
- Lischer HEL, Excoffier L (2012) PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, **28**, 298–299.
- Logan FJ, Horák A, Štefka J, Aydogdu A, Scholz T (2004). The phylogeny of diphyllbothriid tapeworms (Cestoda: Pseudophyllidea) based on ITS-2 rDNA sequences. *Parasitology Research*, **94**, 10–15.
- Loot G, Francisco P, Santoul F, Lek S, Guegan JF (2001). The three hosts of the *Ligula intestinalis* (Cestoda) life cycle in Lavernose Lacasse gravel pit, France. *Archiv für Hydrobiologie-Hauptbände*. **152**, 511–525.
- Loot G, Poulin R, Lek S, Guegan JF (2002) The differential effects of *Ligula intestinalis* (L.) plerocercoids on host growth in three natural populations of roach, *Rutilus rutilus* (L.). *Ecology of Freshwater Fish*, **11**, 168–177.

- Louhi KR, Karvonen A, Rellstab C, Jokela J (2010) Is the population genetic structure of complex life cycle parasites determined by the geographic range of the most motile host? *Infection, Genetics and Evolution*, **10**, 1271–1277.
- Malinsky M, Trucchi E, Lawson DJ, Falush D (2018) RADpainter and fineRADstructure: population inference from RADseq data. *Molecular Biology and Evolution*, OpenAccess.
- Morgan DL (2003). Distribution and biology of *Galaxias truttaceus* (Galaxiidae) in southwestern Australia, including first evidence of parasitism of fishes in Western Australia by *Ligula intestinalis* (Cestoda). *Environmental Biology of Fish*, **66**, 55–167.
- Nadeau NJ, Whibley A, Jones RT, Davey JW, Dasmahapatra KK, Baxter SW, Quail MA, Joron M, French-Constant RH, Blaxter ML, Mallet J, Jiggins CD (2012) Genomic islands of divergence in hybridizing *Heliconius* butterflies identified by large-scale targeted sequencing. *Philos Trans R Soc Lond B*, **367**, 343–53.
- Nosil P, Gompert Z, Farkas TE, Comeault AA, Feder JL, Buerkle CA, Parchman TL (2012) Genomic consequences of multiple speciation processes in a stick insect. *Proceedings of the Royal Society B-Biological Sciences*, **279**, 5058–5065.
- Parchman TL, Buerkle AC, Soria-Carrasco V, Benkman CW (2016) Genome divergence and diversification within a geographic mosaic of coevolution. *Molecular Ecology*, **25**, 5705–5718.
- Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-Model species. *Plos One*, **7**, e37135.
- Poulin R, Morand S (2004) Parasite Biodiversity. *Smithsonian Institution Press*.
- Price PW (1980) Evolutionary Biology of Parasites. *Princeton, New Jersey, Princeton University Press*.
- Purcell S, Chang S (2015) <https://www.cog-genomics.org/plink2>
- Querejeta M, González-Esteban J, Gómez A, Fernández-González A, Aymerich P, Gosálbez J, Escoda L, Igea J, Castresana J (2016) Genomic diversity and geographical structure of the Pyrenean desman. *Conservation Genetics*, **17**, 1333–1344.

- R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Rundle HD, Nosil P (2005) Ecological speciation. *Ecology Letters*, **8**, 336–352.
- Savolainen O, Lascoux M, Merilä J (2013) Ecological genomics of local adaptation. *Nature Reviews Genetics*, **14**, 807–820.
- Slatkin M (1993) Isolation by distance in equilibrium and non-equilibrium populations. *Evolution*, **47**, 264–279.
- Sousa V, Hey J (2013) Understanding the origin of species with genome-scale data: modelling gene flow. *Nature Reviews Genetics*, **14**, 404–414.
- Štefka J, Hypša V, Scholz T (2009) Interplay of host specificity and biogeography in the population structure of a cosmopolitan endoparasite: microsatellite study of *Ligula intestinalis* (Cestoda). *Molecular Ecology*, **18**, 1187–1206.
- Tsigenopoulos CS, Ráb P, Naran D, Berrebi P (2002) Multiple origins of polyploidy in the phylogeny of southern African barbs (Cyprinidae) as inferred from mtDNA markers. *Heredity*, **88**, 466–473.
- Wolf JBW, Ellegren H (2017) Making sense of genomic islands of differentiation in light of speciation. *Nature Review Genetics*, **18**, 87–100.
- Wright S (1943) Isolation by distance. *Genetics*, **28**, 114–138.

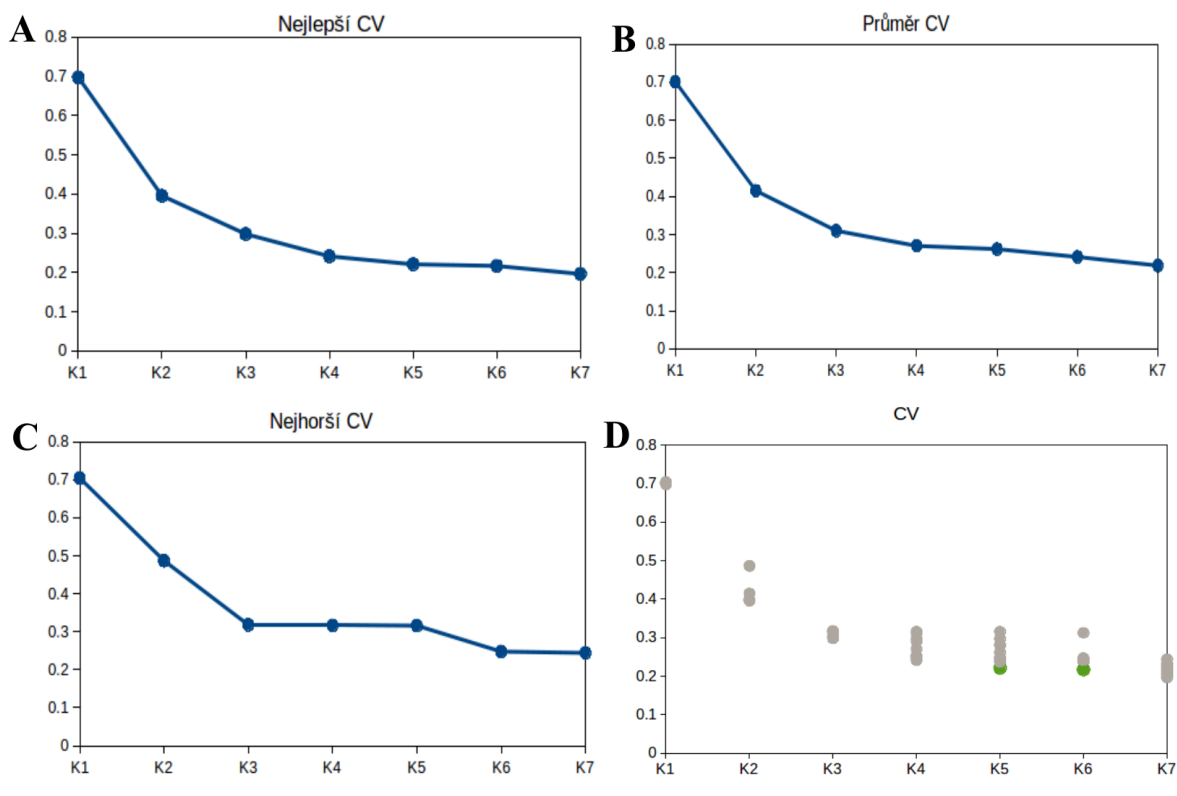
8 Přílohy

Příloha I Tabulka vzorků zařazených do genomové studie, geografický původ a hostitelské spektrum.

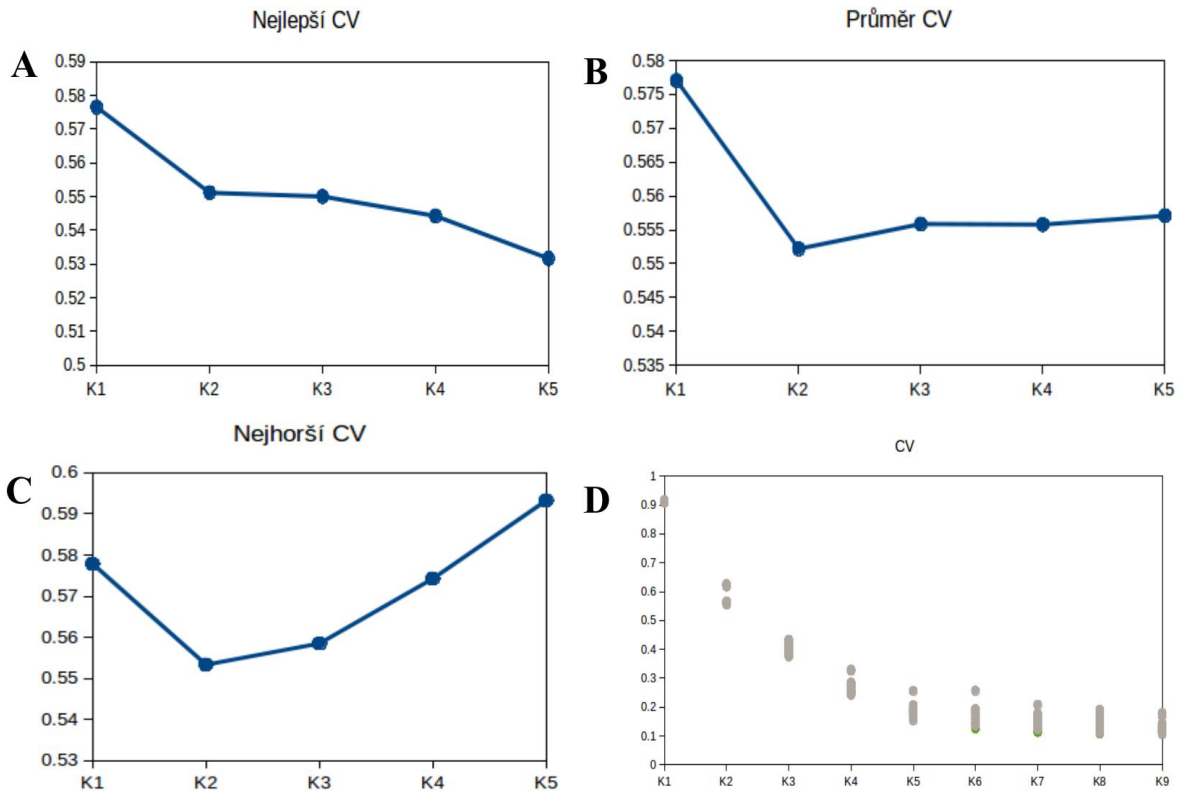
Tab. 1: Geografický původ, hostitel a počty vzorků *L. intestinalis* zařazených do genomové studie.

Kontinent	Země	Hostitel	Počet vzorků v RADseq/po filtrování	mt sekvence
Evropa	Česká republika	<i>Scardinius erythrophthalmus</i>	3/0	
		<i>Rutilus rutilus</i>	36/36	•
		<i>Alburnus alburnus</i>	21/21	•
		<i>Abramis brama</i>	23/23	•
	Francie	<i>Rutilus rutilus</i>	36/0	•
		<i>Blicca bjoerkna</i>	1/0	•
		<i>Alburnus alburnus</i>	1/0	•
	Island	<i>Rutilus rutilus</i>	3/0	•
		<i>Gobio gobio</i>	2/0	•
	Velká Británie	<i>Phoxinus phoxinus</i>	2/0	•
	Německo	<i>Rutilus rutilus</i>	4/0	•
	Estonsko	<i>Abramis brama</i>	5/5	•
	Maďarsko	<i>Neogobius fluviatilis</i>	5/5	•
Austrálie	<i>Galaxias truttaceus</i>	1/1	•	
	<i>Galaxias maculatus</i>	4/4	•	
Amerika	Kanada	<i>Semotilus atromaculatus</i>	10/0	•
	Mexiko	<i>Girardinichthys multiradiatus</i>	2/0	•
Asie	Rusko	<i>Abramis brama</i>	8/0	•
	Írán	<i>Cyprinus carpio</i>	3/3	•
	Čína	<i>Neosalanx taihuensis</i>	10/0	•
Afrika	Tunis	<i>Rutilus rutilus</i>	10/0	•
		<i>Scardinius erythrophthalmus</i>	4/0	•
	Etiopie	<i>Barbus brevicephalus</i>	1/1	•
		<i>Barbus humilis</i>	7/7	•
		<i>Barbus intermedius</i>	5/5	•
		<i>Barbus tsanensis</i>	4/4	•
	DRK	<i>Barbus sp.</i>	1/1	•
	Namibie	<i>Barbus paludinosus</i>	4/4	•
	JAR	<i>Barbus anoplus</i>	3/3	•
		<i>Barbus neefi</i>	1/1	•
<i>Barbus trimaculatus</i>		1/1	•	

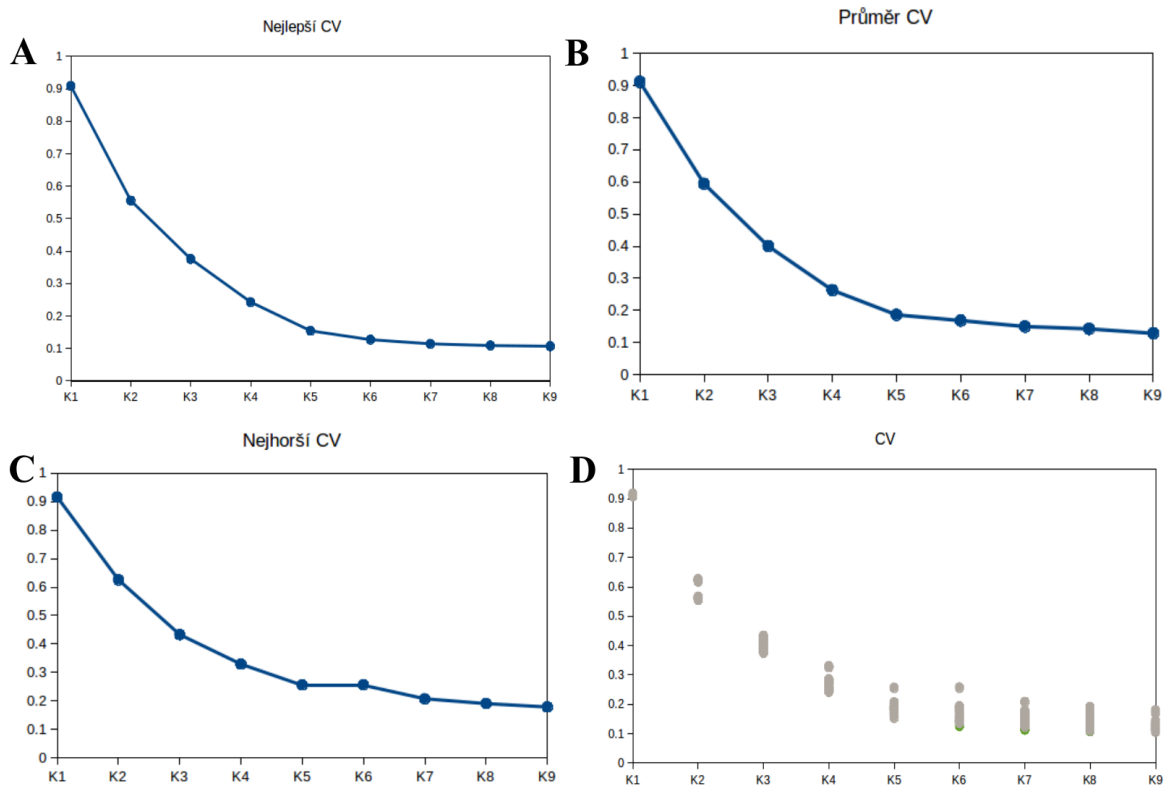
Příloha II Admixture, výběr správné hodnoty K.



Obr. 1: Výsledky replikačních běhů Admixture cross-validation (CV) pro kompletní dataset (1). (A) Vyneseny nejlepší hodnoty přes všechny běhy, (B) průměrné hodnoty CV, (C) nejhorší hodnoty přes všechny běhy, (D) všechny hodnoty CV pro dané K, zeleně zvýrazněny graficky zpracované výstupy pro dané K

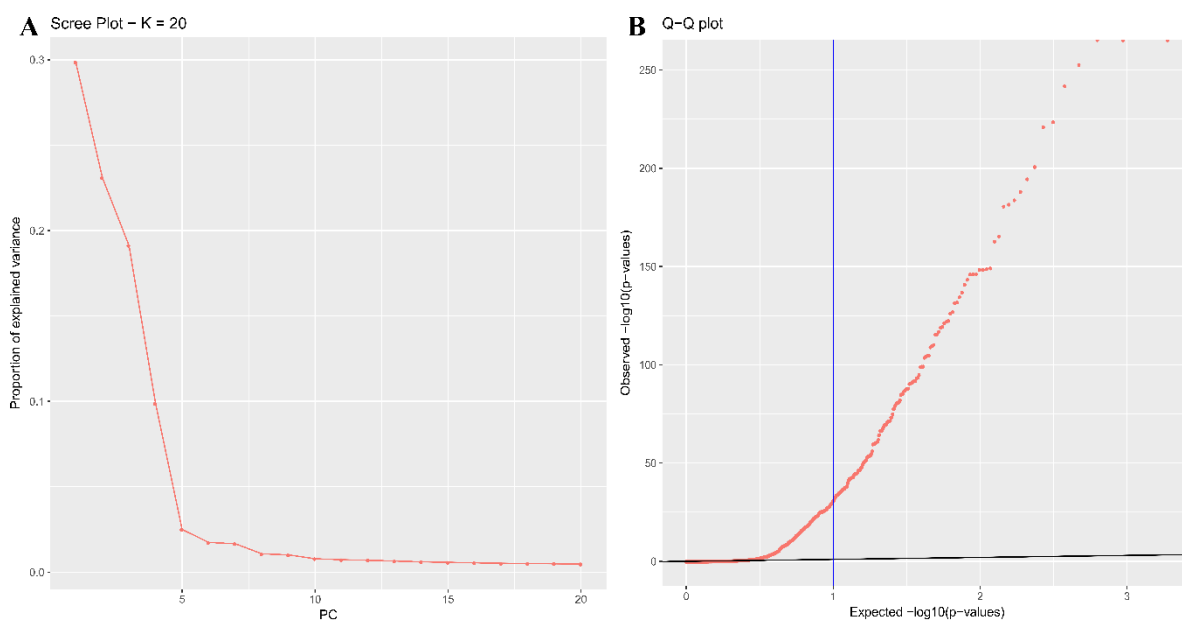


Obr. 2: Výsledky replikačních běhu Admixture cross – validation (CV) pro světové populace a výběr vzorků z kladu A (2). (A) Vyneseny nejlepší hodnoty přes všechny běhy, (B) průměrné hodnoty CV, (C) nejhorší hodnoty přes všechny běhy, (D) všechny hodnoty CV pro dané K, zeleně zvýrazněny graficky zpracované výstupy pro dané K.

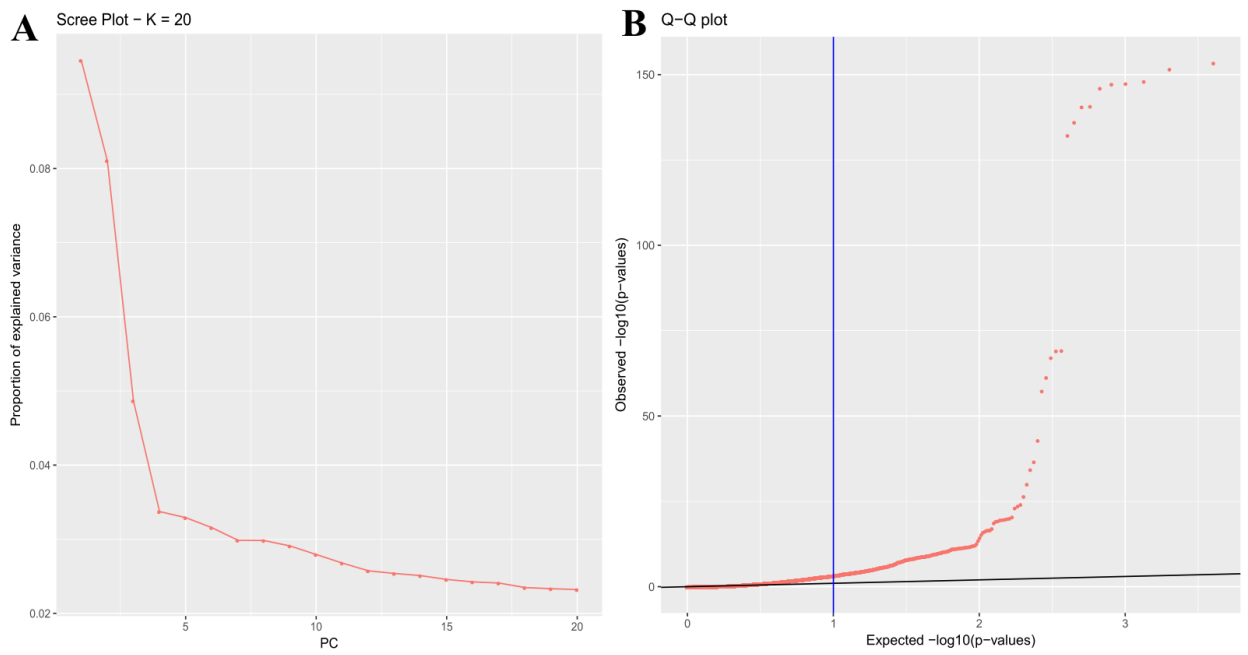


Obr. 3: Výsledky replikačních běhu Admixture cross – validation (CV) pro samostatný klad A (3). (A) Vyneseny nejlepší hodnoty přes všechny běhy, (B) průměrné hodnoty CV, (C) nejhorší hodnoty přes všechny běhy, (D) všechny hodnoty CV pro dané K, zeleně zvýrazněny graficky zpracované výstupy pro dané K.

Příloha III Úvodní testy PCAdapt pro správné určení K hodnot



Obr. 4: Úvodní testy PCAdapt pro kompletní dataset (1): (A) Zobrazení procent variance vysvětlených každou z hlavních komponent (PC) při $K = 20$. (B) Distribuce P – hodnot (P – value).



Obr. 5: Úvodní testy PCAadapt pro dataset Těrlicko/ Žermanice (4): (A) Zobrazení procent variance vysvětlených každou z hlavních komponent (PC) při $K = 20$. (B) Distribuce P – hodnot (P – value).