

Jihočeská univerzita v Českých Budějovicích
Přírodovědecká fakulta

**Bioinformatická analýza pohlavních chromozomů *Lycorea
halia* (Lepidoptera, Danainae)**

Bakalářská práce

Pavel Hronek

Vedoucí práce: RNDr. Petr Nguyen, Ph.D.

České Budějovice, 2022

Hronek P., (2022): Bioinformatická analýza pohlavních chromozomů *Lycorea halia* (Lepidoptera, Danainae) [Bioinformatic analysis of sex chromosomes of *Lycorea halia* (Lepidoptera, Danainae), Bc. Thesis, in Czech] - 30 pp., Faculty of Science, University of South Bohemia in České Budějovice, Czech Republic.

Annotation:

The thesis presents *de novo* assembly of the male *Lycorea halia* genome and identification of its sex-linked sequences by means of coverage analysis.

Prohlášení

Prohlašuji, že jsem autorem této kvalifikační práce a že jsem ji vypracoval pouze s použitím pramenů a literatury uvedených v seznamu použitých zdrojů.

V Českých Budějovicích dne

.....7.12.2022.....

Pavel Hronek

Poděkování:

Rád bych na začátek poděkoval svému školiteli Petru Nguyenovi a jeho kolegům Pablu Mora Ruizovi a Aničce Voleníkové za podporu, pomoc a trpělivost při této náročné cestě. Dále bych chtěl poděkovat i své rodině, která mi byla vždy oporou a pomohla mi při každém kroku, obzvláště můj bratr. A nakonec bych rád poděkoval i své přítelkyni, která mi poskytovala neuvěřitelnou morální podporu. Jsem si jistý, že bez této pomoci bych nikdy nebyl schopen zvládnout tuto práci, a proto bych chtěl všem z hloubky mého srdce ještě jednou poděkovat.

1.	ÚVOD	1
1.1.	LEPIDOPTERA	1
1.2.	<i>DANAUS PLEXIPPUS</i>	2
1.3.	CHROMOZOMÁLNÍ FÚZE	5
1.4.	NEO-POHLAVNÍ CHROMOZOMY MOTÝLŮ	7
2.	CÍLE PRÁCE:	9
3.	MATERIÁL A METODY	10
3.1.	MATERIÁL	10
3.2.	METODY	10
3.2.1	<i>Odhadování velikosti genomu</i>	10
3.2.2	<i>Úprava a vizualizace kvalitních krátkých Illumina čtení</i>	10
3.2.3	<i>Úprava a vizualizace dlouhých čtení</i>	11
3.2.4	<i>Složení genomu</i>	12
3.2.5	<i>Začištění genomové sekvence pomocí krátkých i dlouhých čtení</i>	12
3.2.6	<i>Finální vyčištění genomu od kontaminace a repetitivních úseků</i>	13
3.2.7	<i>Analýza hloubky čtení</i>	13
3.2.8	<i>Přiřazení sekvencí <i>Lycorea halia</i> k chromozomům pomocí ortologů.</i>	14
4.	VÝSLEDKY	15
4.1.	SKLÁDÁNÍ GENOMU	15
4.2.	ANALÝZA HLOUBKY ČTENÍ A IDENTIFIKACE Z-VÁZANÝCH SEKVENCÍ	18
4.3.	PŘIŘAZENÍ SEKVENCÍ <i>LYCOREA HALIA</i> K CHROMOZOMŮM <i>MELITAEA CINXIA</i> POMOCÍ ORTOLOGŮ	19
5.	DISKUZE	21
6.	ZÁVĚR	23
7.	LITERATURA	24

1. Úvod

1.1. Lepidoptera

Motýli (Lepidoptera) představují ekonomicky i pro výzkum důležitý hmyzí řád čítající přibližně 160 000 (Mitter et al., 2017) druhů. Motýli se stali velice důležitým modelem pro řadu molekulárních, ekologických a genetických výzkumů. Kvůli dlouhé tradici produkce hedvábí a s tím spojený zájem o genetiku byl genom bource morušového *Bombyx mori* vybrán k osekvenování jakožto první zástupce motýlů (Mita et al., 2004) již v roce 2004. Mezi největší problémy, se kterými se vědci při sekvenování hmyzího genomu setkávají, patří jejich vysoký polymorfismus (Li et al., 2019). Ideální pro sekvenování je použití čistých homozygotních linií. Bohužel získání takových vzorků u hmyzu je velmi obtížné až nerealistické. Většina hmyzích druhů se nedá chovat a křížit v laboratorních podmínkách za vzniku čistých homozygotních linií a jedinci pro výzkum se musí získat v terénu, což činí získání čistě homozygotních linií nemožný. A jako poslední problém, se kterým se musíme potýkat, je fyzické množství DNA získané z jednoho jedince (Li et al., 2019). Vzhledem k malé velikosti většiny druhů hmyzu, je pro získání dostatečného vzorku potřeba použít více jedinců, kde se opět vracíme k problému jejich polymorfismu a neschopnosti vytvořit homozygotní populaci (Li et al., 2019). Snažit se poskládat kvalitní knihovnu DNA určitého druhu je velice obtížné, pokud se jako vzorek používá směs několika polymorfních jedinců (Richards & Murali, 2015). Dalšími problémy, se kterými se setkáme během sekvenování genomu Lepidopter je repetitivní DNA způsobená transpozony a holocentrickými chromozomy. Tyto chromozomy nemají centrální centromeru jako metacentrické chromozomy a vyznačují se koncentrací centromery podél celého chromozomu, což má potenciálně vliv na evoluci druhu, kdy se v buňkách mohou udržet fragmenty chromozomu i skrze několik buněčných dělení. Chromozomální centromery jsou charakteristické obsahem velkého množství krátkých repetitiv, což představuje další problém při sekvenování jedinců.

Z celého hmyzího řádu předběhli v osekvenování genomu bource morušového pouze octomilka *Drosophila melanogaster* (Myers et al., 2000) a komár *Anopheles gambiae* (Holt et al., 2002.), jenž je přenašečem malárie. Bourec morušový *Bombyx mori* představuje největšího producenta hedvábí a spolu se včelami patří mezi jediné domestikované druhy hmyzu. Protein fibroin, který tvoří základ tohoto hedvábí, má díky svému složení a svým vlastnostem celou řadu uplatnění v biomedicině. Kromě své unikátní vlastnosti produkovat hedvábí se bourec morušový používá pro výzkum chromozomů, kdy se u nich objevil gen Fem (Traut et al., 2008),

což je pohlaví determinující faktor nacházející se na pohlavním chromozomu W a též se používá i pro výzkum kompenzace genové dávky WZ/ZZ (Zha et al., 2009; Walters & Hardcastle, 2011). Pro přiblížení důležitosti těchto organismů z hmyzího řádu jsem si vybral monarcha stěhovavého *Danaus plexippus*, jehož genom byl osekvenovaný po bourci jako druhý z řádu Lepidoptera. Monarchové jsou zajímaví svou unikátní migrací, kterou každoročně podnikají přes Severní Ameriku (Nail et al., 2019).

1.2. *Danaus plexippus*

Monarcha stěhovavý se původně ze Severní Ameriky (Nail et al., 2019) rozšířil po celém světě a osídlil země Jižní a Centrální Ameriky, Španělska, Maroka, Austrálie, Nového Zélandu a Havajských ostrovů (Pierce et al., 2014; Nail et al., 2019). Nejvíce se proslavil již zmíněnou migrací, kterou každoročně podniká severoamerická populace skrze celou severní Ameriku až do zimoviště v Mexiku. Ke zvládnutí celé podzimní migrace mají monarchové řadu adaptací nutných k přežití, jako například větší abdominální zásoby tuku, delší životnost z několika týdnů až na několik měsíců a zvýšenou rezistenci vůči zimě (Reppert, 2006). Tyto změny jsou zapříčiněny sníženou hladinou juvenilních hormonů v těle jedinců a s tím spojenou diapauzou. Díky tomu se *Danaus plexippus* rychle stal důležitým modelem pro studium migrací a předmětem řady výzkumů ve snaze odhalit molekulární a genetickou podstatu adaptací nutných k podniknutí a přežití této 4500 km dlouhé a vyčerpávající cesty. Vědci se během svého bádání zaměřili i na roli epigenetiky a smyslového vnímání během této migrace (Reppert et al., 2004; Merlin et al., 2009; Heinze & Reppert, 2011; Guerra & Reppert, 2013; Merlin & Liedvogel, 2019).

Kromě východní populace, která migruje do Mexika na přezimování, se v Severní Americe ještě nachází západní populace, jenž je od východní oddělená Skalnatými horami a migruje pouhých 500 km k pobřeží Kalifornie (Reppert & de Roode, 2018).

Svoji migraci začíná východní populace na podzim, kdy dochází ke snižování teplot a ke zkracování délky dne, což pro monarchy signalizuje začátek zimy. Díky adaptacím, které jim poskytují větší vytrvalosti a délku života, jsou schopni doletět až do Mexika, kde přecházejí v diapauze do jara a v další etapě se stávají pohlavně aktivními, migrují zpátky na sever a cestou se rozmnožují. Ke konci migrace se vrací zpět do svého severního teritoria, odkud původně před několika generacemi zahájili svoji výpravu jejich pra-prarodiče. Z tohoto místa opět na

podzim při klesajících teplotách začíná nová migrace a tím se celý cyklus uzavírá. Tímto se stal *Danaus plexippus* i skvělým modelem pro sledování epigenetických změn, kdy se v průběhu migrace monarchy stěhovavého vystřídá několik generací.

Z epigenetického pohledu je zvláštní, jak může daná generace vědět, kam a kdy má letět, když mají stejné genetické pozadí. Zde hrají důležitou roli epigenetické značky. Tyto značky ovlivňují, jaká část genomu bude aktivní v daný konkrétní okamžik a umožní jedinci správně reagovat na přítomné podmínky (Lyons et al., 2012; Reppert et al., 2016). Kvůli své komplexnosti se migrace monarchů stěhovavých stala i velmi důležitým modelem pro studování orientace během migrace. Byly provedeny pokusy se snahou zjistit, jak je tato cesta vlastně spuštěna a jaké jsou molekulární, genetické a fyziologické procesy doprovázející jejich migraci, s důrazem na fungování orientačních smyslů.

Celý tento migrační cyklus, zahrnuje dlouhou migraci na jih i na sever a několik generací, kdy každá musí na okolní podmínky správně zareagovat. Čili generace musí vědět, kdy mají letět na sever, kdy mají letět na jih a kdy zůstat na místě a přezimovat. S příchodem zimy a tím souvisejícími faktory, jako zkracování dne a snižování teplot, monarcha započne svojí migraci na jih do Mexika za teplejšími a příznivějšími podmínkami (Guerra, Reppert, 2013). Samotný spouštěč pro začátek migrace u monarchů je tedy změna délka fotoperiody (Guerra, Reppert, 2013, 2015). Při zkracování dne dochází u monarchů stěhovavých v Severní Americe k důležitým fyziologickým změnám, které je připraví na dlouho cestu na jih (Reppert, 2006). Ale samotný směr cesty, už není ovlivněn délkou fotoperiody, nýbrž teplotami (Guerra, Reppert, 2013). Vystavováním pokusných jedinců různým teplotám dokázali Guerra a Reppert se svým týmem ovlivnit směr jejich migrace, zatímco změna délky fotoperiody u směru nehrála žádnou roli a zjistili, že centrem pro vnímání a zpracování teplot není mozek, nýbrž tykadla (Guerra, Reppert, 2013).

Obecně existuje mnoho způsobů, podle kterých se živočichové orientují během své migrace. Někteří využívají magnetické pole, jiní polohu slunce či směr polarizovaného světla anebo si pamatují cestu podle specifických struktur, ať už to jsou hory, budovy či jiné. *Danaus plexippus* využívá několik těchto systémů najednou (Heinze & Reppert, 2011; Guerra & Reppert, 2013, Guerra et al., 2014; 2015; Reppert et al., 2016). Pro udržení správného směru je monarcha stěhovavý schopný využít polohu slunce díky speciálnímu slunečnímu kompasu, jenž

je uložen v tykadlech (Merlin et al., 2009; Reppert et al., 2016). S pomocí tohoto kompasu je monarcha stěhovavý schopný rozpoznat v jaké části dne se nachází a podle těchto informací i správně zkorigovat svůj směr. Pokud je však slunce zastíněné mraky, může se dále orientovat pomocí polarizovaného světla (Heinze, Reppert, 2011) či magnetického pole (Guerra et al., 2014), stejně jako mnoho jiných migrujících živočichů. Stále však není vyřešeno mnoho otázek, týkající se jejich migrace. Doposud se nezjistilo, zda má *Danaus plexippus* kompletní představu o směru svého letu a přesného umístění jeho destinace (Guerra et al., 2014; Guerra, Reppert, 2015; Reppert et al., 2016) a do jaké míry využívají při migraci magnetické pole. Prozatím se vnímání magnetického pole pro orientaci považuje za záložní plán (Guerra et al., 2014) a hlavním mechanismem stále zůstává sluneční kompas (Nguyen et al., 2021).

Je zde tedy stále mnoho otázek týkající se principu a fungování různých systémů podílejících se na migraci, na jejichž odpovědi budeme potřebovat ještě podrobnější výzkum. Tyto odpovědi by nám poskytly cenné informace pro pochopení základních principů používaných při migraci jinými živočichy. Donedávna ještě bylo záhadou, podle jakých podnětů *Danaus plexippus* ví, jestli migrovat na sever, či na jih (Guerra, Reppert, 2013) a jakou roli při migraci hraje mozek a smysly lokalizované v tykadlech (Merlin et al., 2009), ale jsou prováděny pokusy, které se snaží na tyto otázky odpovědět (Nguyen et al., 2021). Kdybychom dokázali pochopit a vysvětlit mechanismus procesů umožňující monarchům podstoupit takto unikátní cestu, mohli bychom naše znalosti aplikovat i na jiné migrující organismy v živočišné říši.

Během své evoluce, si monarchové získali výhodu nad ostatními živočichy tím, že se u jejich larev objevila adaptace na konzumaci jedovaté, jinak nepřístupné rostliny. Tímto se stal monarcha stěhovavý i velmi zajímavým modelem pro studium koevoluce, sekvestrace jedovatých látek a evoluční adaptaci na novou potravu.

Jedinou potravou pro larvy *Danaus plexippus* jsou klejichy *Asclepias*, které jsou známé svou toxicitou, kvůli vyšší koncentraci kardenolidů (Dobler et al., 2012; Petschenka et al., 2013). Kardenolidy inaktivují sodno-draselné pumpy, jinak též Na^+/K^+ ATPázy, které jsou esenciální pro správné fungování organismů (Malcolm, Brower, 1989; Dobler et al., 2012; Petschenka et al., 2013; Karageorgi et al., 2019). Bez správného fungování sodno-draselných pump nefunguje například přenos nervových vzruchů, či dochází k narušení rovnováhy látek uvnitř buňky a mimo ni.

Díky rezistenci vůči kardenolidům a jejich následnému zabudování do svých těl, získal monarcha stěhovavý užitečnou obranu vůči parazitům (Gowler et al., 2015; Petschenka & Agrawal, 2016) a predátorům (Pierson, Fink, 1985; Gowler et al., 2015), kteří poznají toxicitu monarchů kvůli jejich specifickému zbarvení. Rezistence vůči kardenolidům je u monarchů způsobená třemi mutacemi ovlivňující sodno-draselnou pumpu. Hypotetizuje se, že z těchto tří mutací, právě dvě poskytují největší rezistenci vůči kardenolidům, ale zároveň způsobují neurologické problémy. Monarchové nejspíše prvně získali mutaci způsobující slabou rezistenci proti kardenolidům a s tím spojené malé neurologické problémy. Následně se u nich objevila druhá mutace, která tyto problémy znegovala a jako poslední získali třetí mutaci, která jim ze všech tří poskytla největší rezistenci. Nebýt druhé mutace, tak by sama o sobě byla třetí mutace s největší pravděpodobností letální (Karageorgi et al., 2019). Kombinace těchto mutací jim tedy přinesla značnou výhodu oproti ostatním. Naproti tomu sesterská skupina *Danaus chrysippus* nezískala všechny tyto tři mutace, a proto nejsou vůči kardenolidům tak rezistentní (Karageorgi et al., 2019). Proto se dá na *Danaus plexippus* studovat, jak probíhala evoluce na potravu, která je pro ostatní druhy jedovatá, společně s koevolucí mezi monarchou a klejichami, kdy jejich migrace následuje kvetení klejich po celém území Ameriky (Guerra, Reppert, 2015).

1.3. Chromozomální fúze

Karyotyp u organismů bývá často konzervovaný (Talavera et al., 2013) a napříč řádem Lepidoptera je to ve většině případů pravda. Existují však výjimky, např: rod *Lysanda* se vyznačuje velkou diverzitou karyotypu od $n=24$ až do $n=93$ (Talavera et al., 2013) či jejich příbuzní *Agrodiaetus* z čeledi modráskovitých *Lycaeniade* s karyotypem od $n=10$ do $n=224-226$ (Lukhtanov et al., 2005; Lukhtanov, 2015). Změny v počtu chromozomů jsou u motýlů často způsobeny chromozomálními fúzemi (Talavera et al., 2013; Potter et al., 2017; de Vos et al., 2020) či rozpady, které se u motýlů objevují častěji kvůli holocentrickým chromozomům.

Podle novějších modelů jsou chromozomální fúze jedny z příčin vzniku reprodukčních bariér mezi populacemi a tím podporují speciaci některých druhů. Chromozomální fúze může fyzicky spojit geny, které se dříve nacházely na jiných chromozomech a tím změni frekvenci jejich rekombinace, jenž může ovlivnit jejich evoluci. Pokud se propojí lokusy obsahující geny s lokální adaptivní výhodou, které mají sníženou rekombinaci, může dojít vlivem selekce či driftu k zafixování tohoto chromozomálního přeskupení (Guerrero, Kirkpatrick, 2014; de Vos et al., 2020) a eventuálně i k speciaci.

Stejně tak může dojít i k fúzi autozomu s pohlavním chromozomem Z či W za vzniku neo-Z/neo-W pohlavního systému (Dalíková et al., 2017). Jeden další možný vznik pohlavního chromozomu je, když autozom získá pohlaví determinující gen. Kolem tohoto genu dojde pomocí selekce k akumulaci antagonistických alel, které jsou výhodné pro jedno pohlaví ale nežádoucí pro to druhé. Následně zde dochází ke chromozomálním přestavbám jako např. k inverzím, které ukončí rekombinaci mezi tímto chromozomem s jeho párem a selekce zapříčiní akumulaci dalších genů důležitých pro jedno pohlaví a následnou degeneraci (Bachtrog, 2006, 2013; Dalíková et al., 2017). Jak vývoj nově vzniklých chromozomů pokračuje, stále více se hromadí antagonistické alely, pseudogeny a chromozomální přestavby. Takto vzniklý chromozom je z velké části tvořen heterochromatinem a na většině tohoto chromozomu pak nedochází k vůbec žádné rekombinaci. Tímto může chromozom časem zdegenerovat a úplně se ztratit. Ztráta pohlavního chromozomu však není letální a postupem času může fúzí opět vzniknout chromozom, který by převzal funkci ztraceného pohlavního chromozomu a celý cyklus se může opakovat (Wright et al., 2016).

Původní konstituce pohlavních chromozomů u motýlů byla Z0/ZZ (Traut et al., 2008, Nguyen, Carabajal Paladino, 2016) vzhledem k výskytu tohoto systému u bazálních skupin motýlů a sesterského řádu chrostíků (Traut, Marec, 1996, Traut, Marec, 1997). Napříč řádem Lepidoptera se vyskytují nově vzniklé chromozomy W a Z jež vznikly nezávisle či fúzí, s různými stádii vývoje (Traut et al., 2008, Mongue et al., 2017). Jak přesně a kdy u Lepidopter vznikl chromozom W není zcela jasné. Existuje však hypotéza tvrdící, že chromozom W vznikl u společného předka *Tischeriidae* a *Dytrisia* (Lukhtanov 2000) a v důsledku vznikly studie s detailními analýzami s jedinci z těchto skupin (Dalíková et al., 2017; Fraïsse et al., 2017)

1.4. Neo-pohlavní chromozomy motýlů

Motýli tvoří potenciálně nejlepšího kandidáta na výzkum neo pohlavních chromozomů u systému se samičím heterogametickým pohlavím. Existuje celá řada výzkumů zabývajících se fúzí pohlavních chromozomů s autozomy mezi Lepidoptera s cílem odhalit historii a podstatu vzniku neo pohlavních chromozomů například u *Tortricidae* (Nguyen et al., 2013), *Nymphaliade* (Smith et al., 2016, Mongue et al., 2017) *Saturniidae* (Yoshido et al., 2011), *Pyrilidae* (Bachtrog, 2013), *Yponomeutida* (Traut et al., 2008). Stejně tak můžeme naleznout různé nezávislé fúze u drosophil za vzniku neo-pohlavních chromozomů. Stáří těchto neo-pohlavních chromozomů se odhaduje na 0,1My u *D. albomicans*, 1My u *D. Miranda* a 15My u *D. pseudoobscura* a jejich výzkum nám tedy umožní sledovat, jak se může vyvíjet chromozom Y u velmi lehce dostupného modelu (Bachtrog, 2013). Tyto různé úrovně vývoje, nám umožňují studovat a sledovat, jak vůbec pohlavní chromozomy vznikají a zanikají a jak se mohou postupně vyvíjet a následně degenerovat. Práci na těchto modelech můžeme nahlédnout na možnou budoucnost našeho vlastního chromozomu a umožní nám odhalit řadu procesů a principů, které probíhají u vývoje chromozomů.

Ve snaze pochopit podstatu těchto procesů a mechanismů společně s mnoha dalšími, se klade stále větší důraz na vytvoření kvalitnějších osekvenovaných genomů studovaných organismů. S přibývajícím dostupností sekvenování se databáze obsahující nukleotidové sekvence organismů stále více rozšiřují. Je však zajímavé, že i přes sekvenovaný genom (Zhan et al., 2011) takto významného druhu monarchy stěhovavého, jsme se dozvěděli o vzniku neo-Z pohlavního chromozomu až při opětovné analýze o několik let později (Mongue et al., 2017). V této analýze bylo též potvrzeno, že se tento neo-Z chromozom vyskytuje i u blízkých příbuzných *Danaus plexippus* a to u *D. gilippus*, *D. erippus*, and *D. eresimus*. Z toho vyplývá, že samotný vznik tohoto neo-Z předchází diverzifikaci rodu *Danaus*.

K zajímavé fúzi mezi pohlavním samičím chromozomem W s autozomem ovládající zbarvení jedince a náchylnost k endosymbiotické bakterii *Spiroplasma ixodeti*, za vzniku neo-W chromozomu, došlo u africké *Danaus chrysippus* (Smith et al., 2016). *Spiroplasma ixodeti* je maternálně přenášena bakterie, která zabíjí samčí populaci již ve vajíčku a tím značně ovlivňuje poměr pohlaví v samičí prospěch. Ve východní Africe se vyskytují poddruhy *Danaus c. chrysippus* (cc), *Danaus c. dorippus* (CC) a hybridní *Danaus c. transiens* (Cc), u nichž se jednotlivě vyskytuje specifický pattern pro každý poddruh řízený autozomálním lokusem C.

Existuje zde však hybridní zóna, kde se často střetává *Danaus chrysippus* a *dorippus* za vzniku již výše zmíněného *Danaus transiens*. Za zmínění rozhodně stojí, že všechny samice *Danaus chrysippus* (cc) a *Danaus transiens* (Cc), které jsou infikované bakterií *Spiroplasma*, neprodukují žádné samčí jedince (Smith et al., 2016). Na druhou stranu, potomci *Danaus dorippus* (CC) nejsou nijak ovlivněni a produkují jak samce, tak i samice. Tato autozomální fúze je tedy zodpovědná za nenáhodnou segregaci pohlaví jedinců, a tudíž i jejich barevných vzorů. Dochází zde k velké nerovnováze napříč všemi poddruhy mezi samčím a samičím pohlavím a tato fúze podporuje segregaci příbuzných poddruhů v hybridní zóně. Toto narušuje Haldenovo pravidlo, jenž tvrdí: „pokud se mezi hybridy vyskytuje pohlaví, které je vzácné, sterilní, či chybí, bude to pohlaví heterozygotické“, což v našem případě není pravda, protože je zde znevýhodněno právě homozygotní pohlaví, kdy umírají všichni samci a přežívají pouze heterogametické samice

2. Cíle práce:

Cílem této práce je sestavit genom motýla *Lycorea halia*, zástupce skupiny Euploinea sesterské zbytku tribu Danaini, a identifikovat jeho pohlavně vázané sekvence na základě analýzy hloubky čtení. Dále pak na základě srovnání s referenčním genomem *Melitaea cinxia* otestovat hypotézu o společném původu neo-Z chromozomu dříve identifikovaného u zástupců rodu *Danaus* v rámci tribu Danaini.

3. Materiál a metody

3.1. Materiál

Kukly *Lycorea halia* byly zakoupeny od Stratford butterfly farm (Stratford-upon-Avon, Anglie). Druh byl molekulárně potvrzen pomocí amplifikace části genu pro cytochrom oxidázu 1 (Voleníková A., nepublikovaná data).

Dlouhá čtení byla získána ze samce *Lycorea halia* pomocí sekvenovací technologie Oxford Nanopore sequencing (ONT). Tímto se získalo 58 Gb dat s N50 23,053 kb. K dispozici jsem též měl krátká Illumina čtení od tří samců a tří samic, která měla dohromady 76,9 Gb.

3.2. Metody

3.2.1 Odhadování velikosti genomu

Před samotným složením sekvence genomu je třeba data filtrovat, abychom se zbavili nekvalitních a nežádoucích sekvencí, které by mohly negativně ovlivnit výsledné složení genomu.

Pro úvodní analýzu a odhadnutí velikosti genomu motýla *Lycorea halia* jsem použil krátká kvalitní Illumina čtení a program GenomeScope-2.0 (Vurture et al., 2017) s Jellyfish-2.3.0 (Marçais, Kingsford, 2011). K tomuto odhadu používá Jellyfish a GenomeScope k-mery, tj. čtení jsou rozložena na sekvence délky K. Např. pokud si vezmeme sekvenci AATGC, dostaneme 2-mery AA, AT, TG, GC, či 3-mery AAT, ATG, TGC. Kromě odhadu velikosti genomu GenomeScope poskytuje i předpokládanou heterozygotnost či množství unikátních a duplikovaných sekvencí. Pro co nejpresnější výsledek jsem použil Jellyfish count -m 21 -s 32g -t 16 opakovaně s narůstajícími hodnotami pro velikost k-merů od 21 po 50 s nárůstem po 5 reprezentováno parametrem -m a paměť 32 Gb reprezentováno parametrem -s. Následně použité -t bylo pouze pro určení množství vláken. V pokračování mé práce jsem se rozhodl použít výsledky s k-mery o velikosti 40 bp.

3.2.2 Úprava a vizualizace kvalitních krátkých Illumina čtení

Pro krátká čtení jsem použil program FastQC-0.11.5 (Andrews, S. 2010) k vizualizaci kvality dat a následně Trimmomatic-0.36 (Bolger et al., 2014) na jejich úpravu. Abych

vyhodnotil vliv filtrace, použil jsem opět FastQC-0.11.5 na výsledná data. Pomocí FastQC je možné zobrazit statistiky týkající se délky našich jednotlivých čtení a jejich kvality. Skóre kvality je reprezentováno hodnotou Q. Skóre kvality 30Q znamená, že pravděpodobnost špatného přiřazení báze je 1 z 1000 a tudíž bude přesnost 99,9 %. Zatímco 20Q představuje přesnost pouze 99 %, kdy se špatně přiřadí 1 báze ze 100. Kromě toho poskytuje tento program informace např. i o předpokládaném zastoupení GC bází či přítomnost různých technických sekvencí použitých při tvorbě sekvenačních Illumina knihoven.

I s velkou přesností Illumina čtení je nutné tato data filtrovat. V našem případě byl použit zmíněný program Trimmomatic-0.36. Parametr ILLUMINACLIP:/PATH/TruSeq3-PE-2.fa:2:30:10 byl použit pro odstranění jakýkoliv Illumina adaptérů, jenž by v genomu mohly zůstat, abych nepracoval s genomem obsahující nesmyslné sekvence. Další parametr MINLEN:100 po filtrování odřízl všechna čtení kratší než 100 bp a HEADCROP:10 odřízl prvních 10 sekvencí z těchto čtení. DNA polymeráza často na začátku chybí, a proto jsem se rozhodl tyto sekvence odstranit. Poslední parametr byl SLIDINGWINDOW:4:20 se záměrem odstranit nekvalitní data pod Q20.

3.2.3 Úprava a vizualizace dlouhých čtení

Pro vizualizaci dlouhých čtení získaných z ONT či PacBio se používá program NanoPlot (de Coster et al., 2018). NanoPlot byl použit téměř po každém následujícím kroku, abych ověřil efektivitu provedených úprav.

Jeden z největších problémů pro ONT či PacBio je právě jejich vysoká chybovost, která dosahuje hodnot až 15 % (Laver et al., 2015; Weirather et al., 2017). Byl použit program Ratatosk-0.1 (Holley et al., 2021) provádějící hybridní opravu dlouhých čtení pomocí krátkých s využitím de Bruijn grafů. Ratatosk dokáže snížit chybovost až 6-násobně na výslednou přesnost 99 %. Pro práci s Ratatoskem jsem použil parametry Ratatosk correct -v -c 64 -i 450. Tyto parametry představují požadovaných 64 procesorů a velikost vkládaných čtení o délce 450 bp.

Pro následující úpravu jsem použil NanoFilt (de Coster et al., 2018), ve kterém jsem odřízl všechna čtení kratší než 10 kbp a méně kvalitní než Q10. Samotné filtrování lze provést v závislosti na délce či průměrné délce našich čtení anebo na průměrném zastoupení GC bází. Lze též specifikovat oříznutí přesného počtu bází z obou konců (de Coster et al., 2018).

Parametry, které jsem použil pro svůj skript byly Nanofilt -l 10000 -q 10 zajišťující žádanou délku a kvalitu čtení.

3.2.4 Složení genomu

Opravená dlouhá čtení jsem následně složil s použitím Flye-2.8 (Kolmogorov et al., 2019). Flye využívá algoritmus, který zrekonstruuje cílový genom pomocí překrývajících se úseků jednotlivých čtení a vypíše důležité informace ke složeným genomům, jako např. průměrnou hloubku čtení, délku genomu či počet scaffoldů. Pro moje účely jsem ve svém skriptu použil parametr s očekávanou velikostí genomu –genome-size 300m, jenž jsem získal přes již zmíněný Genomescope.

Pro ověření kvality našich výsledných dat jsem použil program BUSCO-5.2.2 (Simão et al., 2015, Manni et al., 2021), jenž porovnává evolučně konzervované unikátní („single copy“) ortology přítomné v sekvenci genomu s předem zvoleným datasetem. Ortology jsou geny, které si u jedinců zachovávají podobné funkce napříč evolucí, neboť se vyvíjely ze společných ancestrálních genů pomocí speciace. Program tak zkontroluje kvalitu a kompletnost našeho vzorku a poskytne nám biologicky smysluplné parametry v závislosti na očekávaném genetickém obsahu (Manni et al., 2021), vzhledem k tomu, že parametry jako počet „scaffoldů“ a „contigů“ či jejich délka a hodnota N50 nenabízí zhodnocení genetického obsahu. „Contig“ je sekvence po sobě jdoucích bází A, T, G, C bez mezer. „Scaffold“ vzniká přiložením jednotlivých „contigů“ k sobě po získání detailnější informace ohledně jejich relativní pozice. Tyto scaffoldy však obsahují mezery, které vznikly nedokonalým poskládáním genomu. Hodnota N50 představuje délku nejkratšího čtení, kdy suma předchozích čtení pokryje právě polovinu celého genomu. Pro svoji práci jsem ke srovnání využil dataset Lepidoptera a následovně jsem BUSCO použil pro genomy jiných druhů ke srovnání.

3.2.5 Začištění genomové sekvence pomocí krátkých i dlouhých čtení

Vzhledem k relativně velkému výskytu chyb u dlouhých čtení jsem nadále pokračoval se začištěním získané sekvence genomu („polishing“). Byl použit program Medaka 1.6.0, využívající pro začištění dlouhá čtení se základním nastavením. Dále byl použit program Nextpolish-1.3.0 (Hu et al., 2020) spolu s bwa-0.6.2 (Li & Durbin, 2009), který k opravě využívá vysoce kvalitní krátké čtení. Proto jsem použil filtrovaná krátká čtení z předchozích kroků. Opět bylo použito pouze základní nastavení.

3.2.6 Finální vyčištění genomu od kontaminace a repetitivních úseků

Jedním z posledních kroků bylo zjistit, zda náš genom neobsahuje kontaminace. Pro odlišení a vizualizaci taxonomického složení našeho genomu nám posloužil program Blobtools (Laetsch & Blaxter, 2017) a samtools-1.9 (Danecek et al., 2021). Blobtools je program vytvořený ke kontrole kvality a vizualizaci taxonomického rozdělení genomových datasetů. Ve výsledném grafu od Blobtools můžeme vidět jednotlivé zastoupení taxonomických skupin, vyskytujících se v našem vzorku s jejich procentuálním obsahem GC v závislosti na hloubce čtení. Nalezené sekvence, které patřily proteobakteriím, či nebyly nikam zařazeny, byly následně stejným programem odstraněny pomocí parametru „seqfilter -v“ a pro kontrolu byl vygenerován nový graf s taxonomickým složením našeho genomu.

Posledním krokem bylo zamaskování repetitivních sekvencí jako např: transpozony a retrotranspozony. K zamaskování byly použity programy RepeatModeler-2.0.2 (Flynn et al., 2020) a následně RepeatMasker-4.0.7 (<http://repeatmasker.org/>) se základním nastavením. Tyto repetitivní sekvence by zkreslily sekvence následujících analýz.

3.2.7 Analýza hloubky čtení

Následně jsem kromě původních krátkých čtení použil resekvenační data obsahující krátká čtení od tří dalších samic a dvou samců, která jsem filtroval pomocí programu Trimmomatic-0.36 se stejnými parametry s následnou kontrolou pomocí FastQC-0.11.5 jak je uvedeno v kap. 3.2.2. Takto filtrovaných šest vzorků krátkých čtení jsem použil k mapování na můj vyčištěný genom pomocí programu Bowtie2-2.3.5.1 (Langmead & Salzberg, 2012) a k indexaci pomocí samtools-1.9. Pro citlivější mapování byly použity parametry bowtie2 –very-sensitive-local –no-mixed –no-discordant.

Výsledná data byla analyzována Bedtools-2.25.0 (Quinlan & Hall, 2010) a Samtools-1.11-intel-19.0.4-wzth4e4, které pomocí parametru „genomecov“ spočítaly hloubku čtení na bázi a pomocí „groupby“ seřadily mediány hloubky čtení pro každou bázi pro jednotlivé „scaffoldy“.

Pro finální analýzu porovnání hloubky čtení samčích a samicích vzorků byl použit R skript poskytnutý Dr. Pablem Ruiz Morou, který odstranil sekvence kratší než 100 kbp. Medián hloubky čtení byl pro všechny vzorky normalizován celkovým pokrytím. Pro každé pohlaví byl

spočten průměr normalizované hloubky čtení. Pro každý „scaffold“ byla spočtena hodnota $\log_2$ poměru průměrné hloubky čtení u samců a samic. Očekává se, že hloubka čtení autozomálních sekvencí bude u obou pohlaví stejná, a proto by měl být poměr $\log_2 = 0$. Hloubka čtení sekvencí vázaných na pohlavní chromozom Z by měla být u samců oproti samicím dvojnásobná vzhledem k jejich ZZ/ZW systému. Proto by měly hodnoty pro Z vázané sekvence u $\log_2 = 1$. Výsledné hodnoty byly vizualizovány histogramem

3.2.8 Přirazení sekvencí *Lycorea halia* k chromozomům pomocí ortologů.

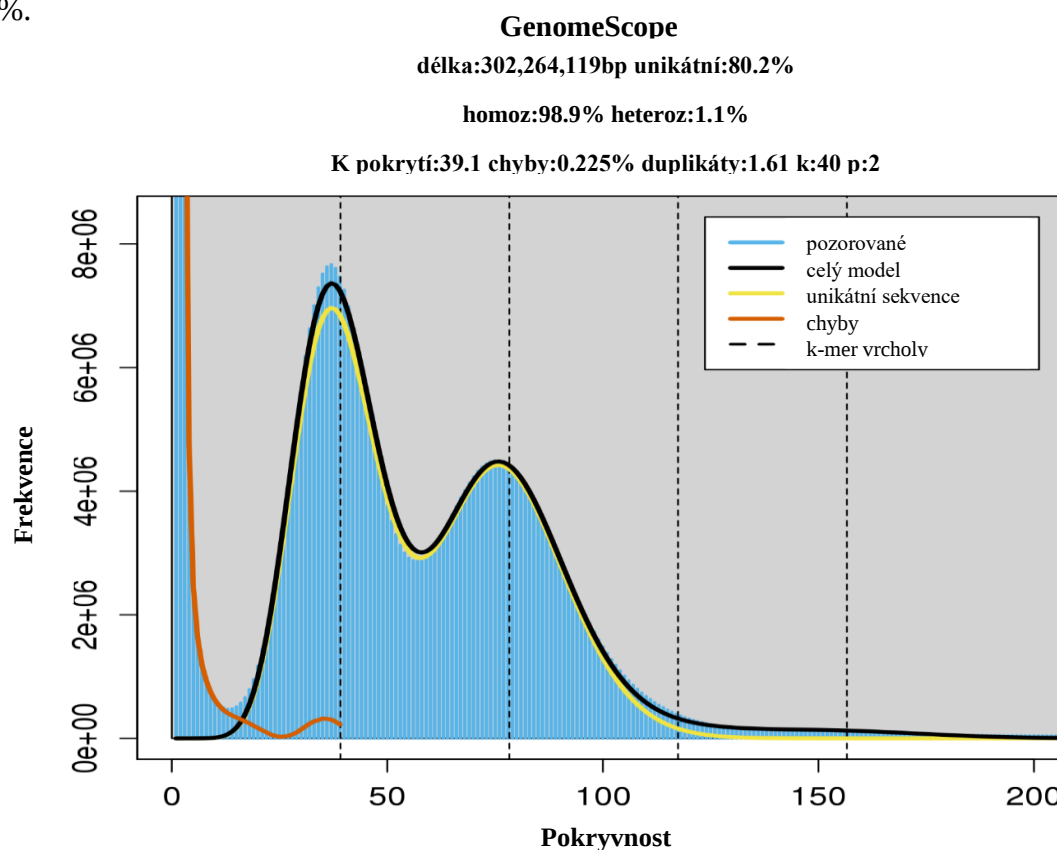
Na základě námi provedené BUSCO analýzy je dále možné přiřadit sekvence *L. halia* k ancestrálním syntením blokům pomocí referenčního genomu *Melitaea cinxia*.

A nyní můžeme jejich geny přiřadit k náležitým chromozomům od *Melitaea cinxia* na základě velké syntenie mezi motýly, skrze kterou je možné identifikovat jednotlivé ortology. Na základě tohoto srovnání byly sekvence přiřazeny k chromozomům *Melitaea cinxia* s největším počtem ortologů. Podle vyplývající statistiky můžeme určit původ neo-Z chromozomu *Lycorea halia* a určit, jestli došlo ke stejné fúzi jako u monarchů.

4. Výsledky

4.1. Skládání genomu

Jako výsledek z Jellyfish-2.3.0 a GenomeScope-2.0 jsem dostal odhadovanou velikost genomu 302,26 mbp. Na grafu (Graf 1) je dále vidět, že heterozygotnost našeho motýlího vzorku je kolem 1,1%. Pro srovnání u lidského genomu heterozygotnost dosahuje přibližně 0,1%.



Graf 1: Vizualizace k-merového spektra krátkých Illumina čtení *Lycoreia halia*.

Celkově bylo sekvenováním DNA získáno 58 Gb Oxford Nanopore (ONT) dlouhých čtení s N50 23,053 kb a 76,9 Gb krátkých Illumina čtení. Po úvodním filtrování krátkých Illumina čtení pomocí programu Trimmomatic-0.36 byla odstraněna nekvalitní či příliš krátká čtení a z původních 76,9 Gb nám zůstalo 55.6 Gb opravených dat. Před složením genomu bylo filtrováním dlouhých čtení pomocí programů NanoFilt a Ratatosk-0.1 získáno 57 Gb opravených čtení s N50=23,346 kb.

V následující tabulce (Tabulka 1) je shrnutí efektů jednotlivých filtrací na náš genom. Největší rozdíl po použití programu Nanofilt je vidět u hodnot týkající se průměrné délky a

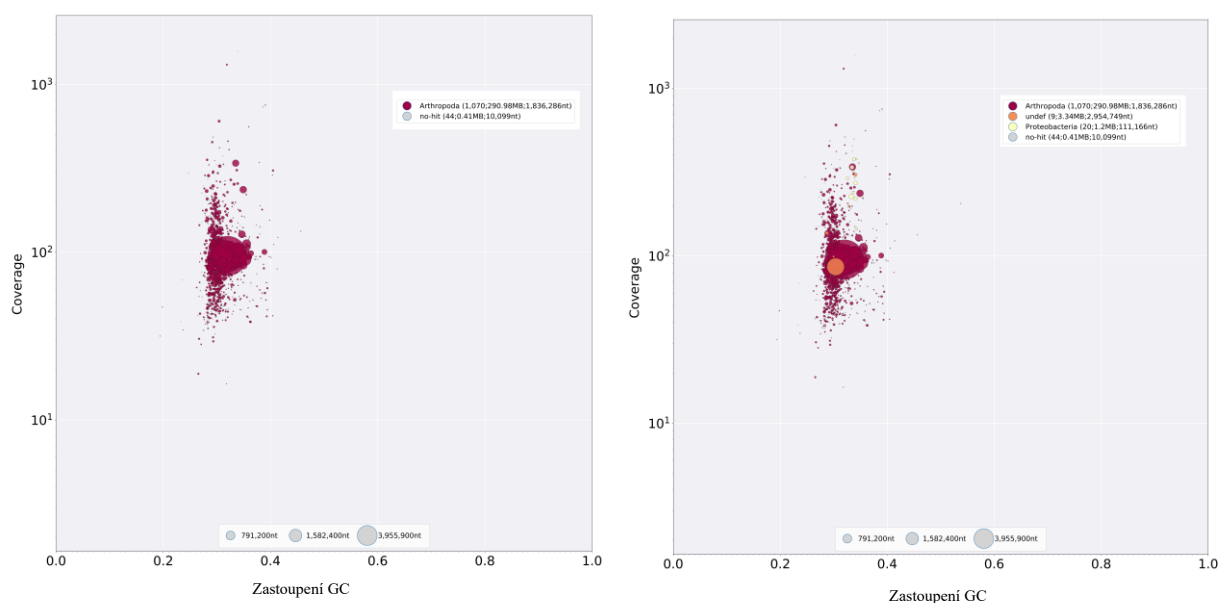
počtu čtení, kdy průměrná délka vzrostla o necelých 28%. Ratatosk-0.1 neměl žádný vliv na délky čtení, zato měl značný efekt na vylepšení jejich kvality.

Tabulka 1: Porovnání upravovaných dat pomocí programu Nanoplot.				
Kontrolované hodnoty	Bez úprav	Ratatosk	NanoFilt	Ratatosk+NanoFilt
Průměrná délka čtení (bp)	17,245.7	17,245.7	21,486.6	22,050.6
Počet čtení	1,740,138.0	1,740,138.0	995,532.0	946,257.0
N50	23,053.0	23,369.0	22,673.0	23,346.0
Celkový počet bází (gbp)	30,01	30,40	21,39	20,87

V tabulce 2 je vypsáno srovnání po složení genomu pomocí programu Flye-2.8, kde je vidět největší rozdíl v hodnotě N50, počtu fragmentů a průměrné hloubce čtení. Po odstranění a filtraci nekvalitních sekvencí došlo ke značnému zkrácení délky z 312,9 mbp na 294,6 mbp. Na základě pozorování efektů jednotlivých filtrací, jsem se rozhodl použít genom opravený programy NanoFilt a Ratatosk pro finální složení, který byl nejméně fragmentovaný, a i přes filtraci přibližně 20 mbp kratších nekvalitních dat si udržel relativně vysokou hloubku čtení na bázi.

Tabulka 2: Srovnání jednotlivých úprav při závěrečném poskládání genomu pomocí programu Flye.				
Kontrolované hodnoty	Bez úprav	Ratatosk	NanoFilt	Nanofilt+Ratatosk
Celková velikost (bp)	312,970,799	310,617,161	312,043,220	294,622,219
Fragmenty	2270	2014	2010	1469
N50 fragmentů	1,000,624	1,111,200	1,282,458	1,838,061
Největší fragment (bp)	15,224,657	8,534,491	8,546,222	21,158,623
Scaffold	43	44	50	36
Průměrná hloubka čtení	94	95	67	71

S využitím programu Blobtools byla identifikována a následně odstraněna kontaminace našeho genomu (Graf 2). Kromě 91,73 % našeho vzorku odpovídající členovcům (Arthropoda) bylo 1,13 % přiřazeno k proteobakteriím a 1,09 % (undef) nebylo přiřazeno k jakémukoliv druhu. Po odstranění 29 sekvencí, které nepatřily k členovcům, nám v genomu zbylo 1114 sekvencí.



Graf 2: Porovnání genomu po odstranění kontaminace pomocí programu Blobtools

Pro zamaskování repetitivních elementů v našem genomu byly použity programy RepeatModeler-2.0.2 spolu s RepeatMasker-4.0.7. Celkově bylo v našem genomu zamaskováno 19,39 % repetitivních (56,5 mbp). Nejvíce zastoupeny byly repetice LINE (long interspersed nuclear elements) 6,37 %, dále pak jednoduché repetice 2,01 % a DNA elementy 1,69 %.

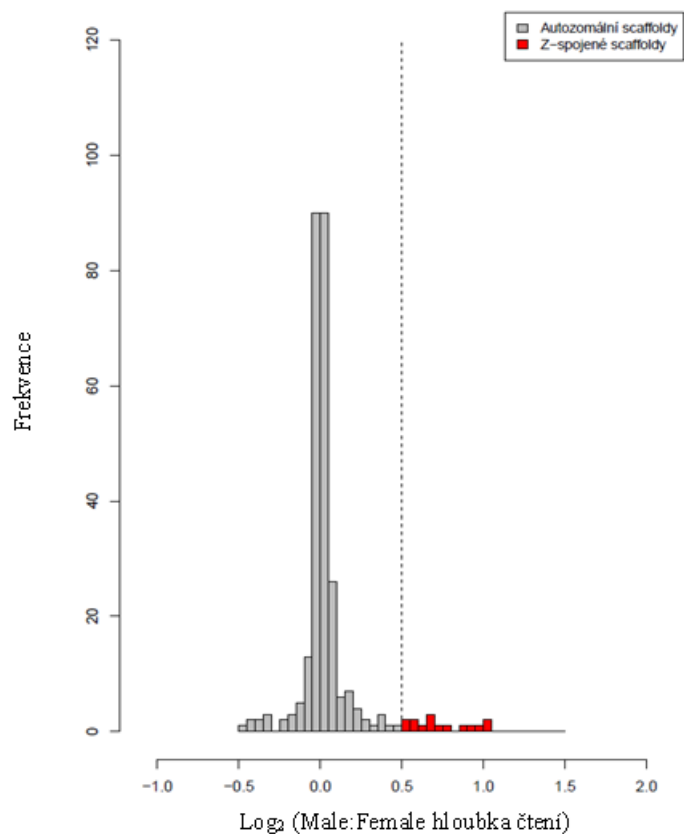
Po všech filtračních krocích měl finální genom 1114 sekvencí o délce 291,38 mbp, jenž byl z 31,76 % zastoupen bázemi GC a hodnota N50 dosahovala 1,86 mbp. Komplettnost našeho genomu byla posouzena pomocí ortologů BUSCO-5.2.2 s genovým datasetem Lepidoptera, kdy 89,4% (4729) ortologů bylo kompletních. Z toho 88,7% (4690) se vyskytovalo jako „single-copy“, 0,7% (39) bylo duplikovaných, 3,2% (168) fragmentovaných a 7,4% (389) chybělo. Následně byl náš poskládaný genom porovnán s dalšími motýlími genomy (Tabulka 3).

Tabulka 3: Statistika složeného genomu <i>Lyocrea halia</i> spolu s vybranými zástupci z řádu Lepidoptera.							
Druh (čeleď)	Velikost genomu (Mbp)	Délka složeného genomu (Mbp)	Počet fragmentů	N50 (Mpb)	Heter. (%)	Repetice (%)	BUSCO (Lepidoptera)
<i>Lyocrea halia</i> (Nymphalidae)	303,3 ^c	291,38	1469	1,86	1,1	19,4	C:89,4%[S:88,7%,D:0,7%] F:3,2%,M:7,4% ^a
<i>Bombyx mori</i> (Bombycidae)	508,6 ^d	460,3	696	16,8	N/A	46,8	C:98,9%[S:97,9%,D:1,0%] F:0,2%,M:0,9%
<i>Danaus chrysippus</i> (Nymphalidae)	N/A	354	4,115	11,45	N/A	35,5	C:92,7%[S:92,2%,D:0,5%] F:2,2%,M:5,1%
<i>Drepana arcuata</i> (Drepanidae)	303,2 ^d	270,5	11,493	0,1	1,0	8,3	C:79,4%[S:77,6%,D:1,8%] F:5,4%,M:15,2% ^a
<i>Ephestia elutella</i> (Pyralidae)	N/A	576,9	804	19	N/A	58,1	C:82,1%[S:81,0%,D:1,1%] F:5,3%,M:12,6% ^a
<i>Helicoverpa armigera</i> (Noctuidae)	N/A	337,07	997	1,0	N/A	14,6	C:88,5%[S:81,0%,D:7,5%] F:4,5%,M:7,0% ^a
<i>Manduca sexta</i> (Sphingidae)	420,5 ^d	470,0	4,057	14,3	N/A	33,9	C:89,4%[S:86,1%,D:3,3%] F:3,7%,M:6,9%
<i>Melitaea cinxia</i> (Nymphalidae)	390 ^d	484,0	31	17,3	N/A	42,0	C:88,4%[S:87,7%,D:0,7%] F:4,4%,M:7,2% ^a
Heter. heterozygotnost, N/A not available							
^a Busco z této práce							
^b Databáze velikostí zvířecích genomů (https://www.genomesize.com/). Nezávisle zjišťována							
^c Odhad velikosti genomu podle k-merů							
^d Odhad velikosti podle cytometrie							

4.2. Analýza hloubky čtení a identifikace Z-vázaných sekvencí

S využitím R skriptu byly porovnány mediány hloubky čtení tří samic a tří samců. Sekvence kratší než 100kbp byly z tohoto srovnávání vyřazeny. Tímto nám z původních 6684 sekvencí zbylo 4956.

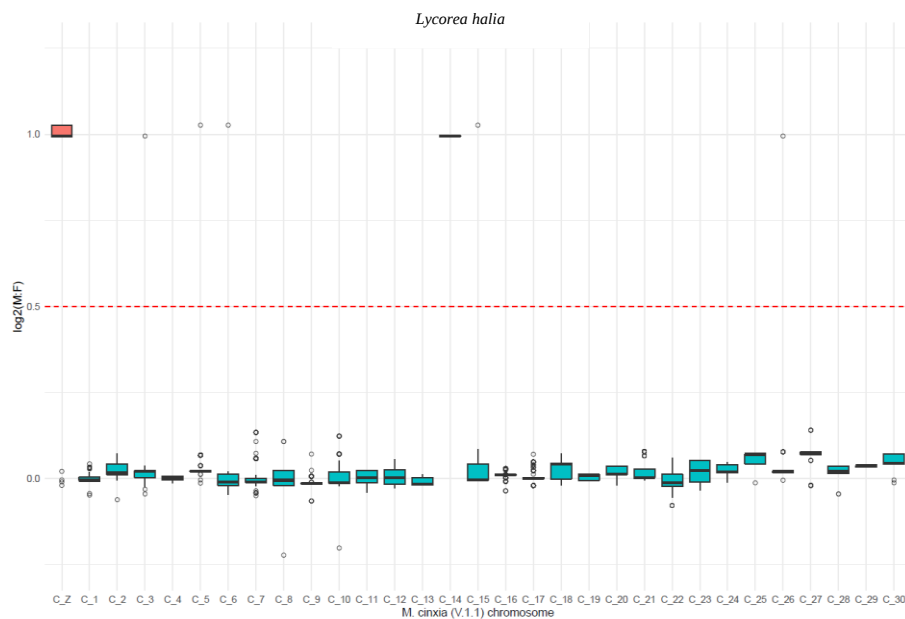
Autozomální sekvence by měly být zastoupeny ve stejné míře u obou pohlaví. Samci však mají oproti samicím dvojnásobný počet pohlavně vázaných sekvencí na chromozom Z kvůli jejich ZZ/ZW systému. Vynesním histogramu $\log_2$ hodnot poměru průměrné samčí (M) a samičí (F) normalizované hloubky čtení (Graf 3) ukazuje hlavní pík s $\log_2$ M:F = 0, který reprezentuje autozomální sekvence. Čtyři sekvence o celkové délce 21,57 mbp mají $\log_2$ M:F = 1 a představují tak pohlavně vázané sekvence. Pohlavní sekvence s $\log_2$ M:F $\approx$ 0,7 mohou být chiméry autozomálních a pohlavně vázaných sekvencí vzniklé vlivem chyby při sestavování genomu.



Graf 3: Zobrazení autozomálních a pohlavně vázaných sekvencí *Lycorea halia*.

4.3. Přiřazení sekvencí *Lycorea halia* k chromozomům *Melitaea cinxia* pomocí ortologů

Vzhledem k velké úrovni syntenie pozorované mezi motýly (Pringle et al., 2007; Ahola et al., 2014; Kanost et al., 2016), můžeme přiřadit naše sekvence skrze jejich ortologní geny k chromozomům jiných referenčních genomů. Porovnáním těchto ortologních genů můžeme identifikovat, na jakém chromozomu jiných druhů se nacházejí a tím i určit k jakým genetickým přestavbám či genetickému toku mezi druhy došlo. Pro porovnání byly použity BUSCO ortology našeho genomu *Lycorea halia* (n=30) a referenční genom *Melitaea cinxia* reprezentující ancestrální karyotyp (n=31) (Ahola et al., 2014).



Graf 4: Porovnání sekvencí *Lycorea halia* s genomem *Melitaea cinxia* a jejich přiřazení ke chromozomům.

Z výsledného grafu porovnávajícího ortologní geny je patrné, že pohlavně vázané sekvence na chromozomu Z v našem genomu jsou homologní s chromozomy Z a 14 referenčního genomu *Melitaea cinxia* ([Graf 4](#)).

5. Diskuze

K otestování hypotézy o možném společném původu neo-pohlavního chromozomu v rámci tribu *Danaini* byl osekvenován a následně složen genom *L. halia*. Pomocí k-merů byla jeho velikost odhadnuta na 303,3 mbp. Tímto by se velikostně řadil mezi menší genomy známé u zástupců skupiny *Danainae*, které se pohybují mezi 249-445 mbp (Liu et al., 2020). Finální délka výsledné sekvence genomu byla 291,38 mbp s kompletními 89,4% BUSCO ortology. V našem genomu bylo celkem zamaskováno 19,39% repetice, jejichž podíl u danainů koreluje s velikostí genomu. U příbuzného monarchy *D. plexippus* s velikostí genomu odpovídající přibližně 245-248 mbp bylo nalezeno mezi 11,2-14,3% repetice, zatímco u sesterského druhu *D. chrysippus* s mnohem větším genomem 354 mbp bylo identifikováno 35,5% repetice (Singh et al., 2022). Velikost genomu v poměru s procentuálním zastoupením repetitivních elementů se však u motýlů znatelně liší (Talla et al., 2017).

Srovnávacími analýzami složených genomů je taky možné sledovat změny karyotypu příbuzných druhů a studovat různé chromozomální fúze spolu s jejich evolucí. Právě jedna taková fúze byla studována v práci Mongue et al., (2017). Autoři objevili u *D. plexippus* a blízce příbuzných druhů *D. gilippus*, *D. erippus* a *D. erismus* neo-Z chromozom vzniklý fúzí chromozomu Z s autozomem odpovídajícím chromozomu 20 v referenčním genomu *M. cinxia*. Tento objevený neo-Z chromozom je tak pravděpodobně společným znakem rodu *Danaus* a je tedy minimálně 5 milionů let starý (Lushai et al., 2003). Je však možné, že je tato fúze společná všem členům tribu *Danaini*, kteří mají často karyotyp $n=30$ (Traut et al., 2018). Autoři navrhli, že se stejným autozomem sfúzoval také pohlavní chromozom W za vzniku neo-W chromozomu. K další fúzi pohlavního chromozomu s autozomem došlo u afrického *D. chrysippus* (Smith et al., 2016). Mezi poddruhy *Danaus chrysippus* můžeme najít u hybridů fúzi pohlavního chromozomu W s autozomem, která ovlivňuje jejich barevné vzory a náchylnost k bakterii *Spiroplasma*, jenž zabíjí samce a tím ovlivňuje poměr pohlaví. V této hybridní zóně došlo k zamezení genetického toku mezi poddruhy, neboť výsledný hybrid s neo-W chromozomem neplodí živé samce. Takto vzniklý genetický „sink“ od sebe odděluje oba poddruhy a přispívá k jejich speciaci napříč hybridní zónou. Časté fúze činí motýly rodu *Danaus* zajímavým modelovým systémem pro studium evoluce pohlavních chromozomů.

V této práci jsme identifikovali pohlavně vázané sekvence, které jsme přiřadili k chromozomům referenčního genomu *M. cinxia* (Graf 4). Byla nalezena homologie mezi pohlavně vázanými sekvencemi *Lycorea halia* a chromozomy Z a 14 *Melitaea cinxia*. Druh

Melitaea cinxia představuje první složený genom motýlů, který si zachoval ancestrální karyotyp $n=31$, u kterého se předpokládá, že je alespoň 140 My starý (Ahola et al., 2014). Proto je možné přes komparativní analýzy právě s tímto druhem sledovat změny karyotypů příbuzných druhů. Tomuto přispívá i fakt, že je napříč motýly zachována relativně vysoká úroveň syntenie (Dasmahapatra et al., 2012; Ahola et al., 2014; Kanost et al., 2016).

Naše analýza ukázala, že u *L. halia* došlo k fúzi mezi ancestrálním chromozomem Z a autozomem 14, který tak nesdílí fúzi identifikovanou u monarchy (Mongue et al., 2017). Námi studovaný druh *L. halia* je součástí skupiny Euploinea, která se v rámci tribu Danaini oddělila jako první. Jeho neo-pohlavní chromozom tak může být starý až 33 My (Chazot et al., 2019). Značnému stáří tohoto systému neo-pohlavních chromozomů odpovídá i jejich vysoká diferenciace, kdy se hodnoty $\log_2$ M:F pro chromozom 14 pohybují kolem jedné (Graf 3).

Dalším zajímavým výsledkem, který přineslo studium neo-pohlavních chromozomů *D. plexippus* je dichotomie v kompenzaci genové dávky. Na pohlavním chromozomu Z objevili dva způsoby této kompenzace, kdy se ancestrální forma chromozomu Z vyznačovala genovou kompenzací, kterou najdeme např. u hlístic *Nematoda*. Dochází zde ke snížení genové exprese u samců (ZZ) a jejich ancestrálního Z o polovinu (u hlístic dochází ke snížení regulace chromozomu X). Tento výsledek koreluje i s ostatními motýlími druhy nesoucí ancestrální Z, kde se vyskytuje stejný způsob kompenzace genové dávky. U neo-Z chromozomu, který vznikl fúzí s autozomem však dochází k jinému způsobu kompenzace, kterou nalezneme např. u octomilek *Drosophila*. Při tomto typu kompenzace dochází ke dvojnásobnému navýšení exprese samičího Z (ZW). Tímto byl získán velmi unikátní poznatek o existenci dvou různých mechanismů kompenzace genové dávky na jediném chromozomu (Gu et al., 2019).

Neo-pohlavní chromozomy jsou u motýlů velmi časté, například u *Tortricidae* (Nguyen et al., 2013), *Nymphaliade* (Smith et al., 2016, Mongue et al., 2017) *Saturniidae* (Yoshido et al., 2011). Proto nám neo-pohlavní chromozomy paralelně v rámci tribu Danaini mohou pomoci objasnit různé aspekty evoluce neo-pohlavních chromozomů motýlů.

6. Závěr

Složil jsem genom druhu *Lycorea halia*, který jsem použil pro analýzu hloubky čtení a srovnání s referenčním genomem *M. cinxia*. Byla objevena nezávislá fúze autozomu 14 s chromozomem Z, čímž byla vyvrácena naše původní hypotéza o sdíleném neo-Z chromozomu *D. plexippus* v rámci tribu Danaini.

7. Literatura

- Ahola, V., Lehtonen, R., Somervuo, P. et al. (2014). The Glanville fritillary genome retains an ancient karyotype and reveals selective chromosomal fusions in Lepidoptera. *Nature Communications*, 5, 4737.
- Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data.
- Bachtrog, D. (2006). A dynamic view of sex chromosome evolution. *Current Opinion in Genetics and Development*, 16(6), 578-585.
- Bachtrog, D. (2013). Y-chromosome evolution: Emerging insights into processes of Y-chromosome degeneration. *Nature Reviews Genetics*, 14(2), 113-124.
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114–2120.
- Dalíková, M., Zrzavá, M., Hladová, I., Nguyen, P., Šonský, I., Flegrová, M., Kubičková, S., Voleníková, A., Kawahara, A. Y., Peters, R. S., & Marec, F. (2017). New insights into the evolution of the W chromosome in Lepidoptera. 30;108(7), 709-719.
- Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H. (2021). Twelve years of SAMtools and BCFtools. *Gigascience*. 10(2).
- Dasmahapatra, K., Kanchon, Walters, R., James, Briscoe, D., Adriana, (2012). The Heliconius Genome Consortium. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature* 487, 94–98.
- de Coster, W., D’Hert, S., Schultz, D. T., Cruts, M., & van Broeckhoven, C. (2018). NanoPack: Visualizing and processing long-read sequencing data. *Bioinformatics*, 34(15), 2666–2669.
- de Vos Jurriaan M., Augustijnen Hannah, Bäscher Livio and Lucek Kay (2020). Speciation through chromosomal fusion and fission in Lepidoptera. *Phil. Trans. R. Soc.* 375.
- Dobler, S., Dalla, S., Wagschal, V., & Agrawal, A. A. (2012). Community-wide convergent evolution in insect adaptation to toxic cardenolides by substitutions in the Na,K-ATPase. *Proceedings of the National Academy of Sciences of the United States of America*, 109(32), 13040–13045.

- Flynn, Jullien M., Hubley, R., Goubert, C., Rosen, J., Clark, G., A., Feschotte, C., Smit, F., A. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences*, 117(17), 9451-9457.
- Fraïsse, C., Picard, M.A.L. & Vicoso, B. The deep conservation of the Lepidoptera Z chromosome suggests a non-canonical origin of the W (2017). *Nature Communications* 8,
- Gowler, C. D., Leon, K. E., Hunter, M. D., & de Roode, J. C. (2015). Secondary Defense Chemicals in Milkweed Reduce Parasite Infection in Monarch Butterflies, *Danaus plexippus*. *Journal of Chemical Ecology*, 41(6), 520–523.
- Gu L., Reilly P., F., Lewis J., J., Reed R., D., Andolfatto P., Walters J., R. (2019). Dichotomy of Dosage Compensation along the Neo Z Chromosome of the Monarch Butterfly. *Current Biology*. 29(23), 4071-4077.
- Guerra, P. A., Gegear, R. J., & Reppert, S. M. (2014). A magnetic compass aids monarch butterfly migration. *Nature Communications*, 5, 4164.
- Guerra, P. A., & Reppert, S. M. (2013). Coldness triggers northward flight in remigrant monarch butterflies. *Current Biology*, 23(5), 419–423.
- Guerra, P. A., & Reppert, S. M. (2015). Sensory basis of lepidopteran migration: Focus on the monarch butterfly. *Current Opinion in Neurobiology*, 34, 20–28.
- Guerrero, R.F. and Kirkpatrick, M. (2014), Local adaptation and the evolution of chromosome fusions. *Evolution*, 68: 2747-2756.
- Heinze, S., & Reppert, S. M. (2011). Sun compass integration of skylight cues in migratory monarch butterflies. *Neuron*, 69(2), 345–358.
- Holley, G., Beyter, D., Ingimundardottir, H., Møller, P. L., Kristmundsdottir, S., Eggertsson, H. P., & Halldorsson, B. v. (2021). Ratatosk: hybrid error correction of long reads enables accurate variant calling and assembly. *Genome Biology*, 22(1).
- Holt, R. A., Subramanian, † G Mani, Halpern, A., Sutton, G. G., Charlab, R., Nusskern, D. R., Wincker, P., Clark, A. G., Ribeiro, J. M. C., Wides, R., Salzberg, S. L., Loftus, B., Yandell, M., Majoros, W. H., Rusch, D. B., Lai, Z., Kraft, C. L., Abril, J. F., Anthouard, V., ... Hoffman, L. (2002.). The Genome Sequence of the Malaria Mosquito *Anopheles gambiae*. *Science*. 4, 298(5591), 129-149.

- Hu, J., Fan, J., Sun, Z., Liu, S. (2020) NextPolish: a fast and efficient genome polishing tool for long-read assembly, *Bioinformatics*, 36(7), 2253–2255.
- Chazot N., Wahlberg N., Freitas AVL, Mitter C., Labandeira C., Sohn JC., Sahoo RK., Seraphim N., de Jong R., Heikkilä M. (2019). Priors and Posteriors in Bayesian Timing of Divergence Analyses: The Age of Butterflies Revisited. *Systematic Biology*, 68(5), 797–813.
- Kanost, M., R., Arrese, L., E., Cao, X., et al. (2016). Multifaceted biological insights from a draft genome sequence of the tobacco hornworm moth, *Manduca sexta*. *Insect Biochemistry and Molecular Biology*. 76, 118-147.
- Karageorgi, M., Groen, S. C., Sumbul, F., Pelaez, J. N., Verster, K. I., Aguilar, J. M., Hastings, A. P., Bernstein, S. L., Matsunaga, T., Astourian, M., Guerra, G., Rico, F., Dobler, S., Agrawal, A. A., & Whiteman, N. K. (2019). Genome editing retraces the evolution of toxin resistance in the monarch butterfly. *Nature*, 574(7778), 409–412.
- Kolmogorov, M., Yuan, J., Lin, Y., & Pevzner, P. A. (2019). Assembly of long, error-prone reads using repeat graphs. *Nature Biotechnology*, 37(5), 540–546.
- Krejčí, A., Müller, P., Vojtěšek, B. (2015) Bioinformatics and Next- generation Sequencing. *Klinická onkologie* [online]. 28(Suppl 2), 2S91-2S96.
- Laetsch DR and Blaxter ML. (2017). BlobTools: Interrogation of genome assemblies. *F1000Research*, 6, 1287.
- Langmead, B., Salzberg, S. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9, 357–359 (2012).
- Laver, T., Harrison, J., O'Neill, P. A., Moore, K., Farbos, A., Paszkiewicz, K., & Studholme, D. J. (2015). Assessing the performance of the Oxford Nanopore Technologies MinION. *Biomolecular Detection and Quantification*, 3, 1–8.
- Li H, Durbin R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 25(14), 1754-1760
- Li, F., Zhao, X., Li, M., He, K., Huang, C., Zhou, Y., Li, Z., & Walters, J. R. (2019). Insect genomes: progress and challenges. *Insect Molecular Biology*, 28(6), 739–758.
- Liu, G., Zhou Ch., Lei Ch., et al. (2020). Genome size variation in butterflies (Insecta, Lepidoptera, Papilionoidea): a thorough phylogenetic comparison. *Systematic Entomology*, 45(3), 571-582.

- Lukhtanov, V. A. (2000). Sex chromatin and sex chromosome systems in nonditrysian Lepidoptera (Insecta). *Journal of Zoological Systematics and Evolutionary Research*, 38, 73-79.
- Lushai, G., Smith, D.A.S., Goulson, D. et al. (2003). Mitochondrial DNA Clocks and the Phylogeny of *Danaus* Butterflies. *International Journal of Tropical Insect Science*. 23(04), 309–315.
- Lyons, J. I., Pierce, A. A., Barribeau, S. M., Sternberg, E. D., Mongue, A. J., & de Roode, J. C. (2012). Lack of genetic differentiation between monarch butterflies with divergent migration destinations. *Molecular Ecology*, 21(14), 3433–3444.
- Malcolm, S.B., Brower, L.P. (1989). Evolutionary and ecological implications of cardenolide sequestration in the monarch butterfly. *Experientia* 45, 284–295.
- Manni, M., Berkeley, M. R., Seppey, M., & Zdobnov, E. M. (2021). BUSCO: Assessing genomic data quality and beyond. *Current Protocols*, 1, e323.
- Merlin, C., & Liedvogel, M. (2019). The genetics and epigenetics of animal migration and orientation: Birds, butterflies and beyond. *Journal of Experimental Biology*, 222.
- Merlin, C., Reppert, S. M., & Gegear, R. J. (2009). Antennal Circadian Clocks Coordinate Sun Compass Orientation in Migratory Monarch Butterflies. *Science*, 325(5948), 1700–1704.
- Mita, K., Kasahara, M., Sasaki, S., Nagayasu, Y., Yamada, T., Kanamori, H., Namiki, N., Kitagawa, M., Yamashita, H., Yasukochi, Y., Kadono-Okuda, K., Yamamoto, K., Ajimura, M., Ravikumar, G., Shimomura, M., Nagamura, Y., Shin, T., Abe, H., Shimada, T., ... Sasaki, T. (2004). The Genome Sequence of Silkworm, *Bombyx mori*. *DNA Research*, 11(1), 27-35.
- Mitter, C., Davis, D., Cummings, M. (2017). Phylogeny and evolution of Lepidoptera. *Annual review of entomology*, 62, 265-283.
- Mongue, A. J., Nguyen, P., Voleníková, A., & Walters, J. R. (2017). Neo-sex chromosomes in the monarch butterfly, *Danaus plexippus*. *G3: Genes, Genomes, Genetics*, 7(10), 3281–3294.
- Myers, E. W., Sutton, G. G., Delcher, A. L., Dew, I. M., Fasulo, D. P., Flanigan, M. J., Kravitz, S. A., Mobarry, C. M., Reinert, K. H. J., Remington, K. A., Anson, E. L., Bolanos, R. A., Chou, H. H., Jordan, C. M., Halpern, A. L., Lonardi, S., Beasley, E. M., Brandon, R. C., et al. (2000). A whole-genome assembly of *Drosophila*. *Science*, 287(5461), 2196–2204.

- Nail, K. R., Drizd, L., & Voorhies, K. J. (2019). Butterflies across the globe: A synthesis of the current status and characteristics of monarch (*Danaus plexippus*) populations worldwide. *Frontiers in Ecology and Evolution*, 7.
- Nguyen, P., & Carabajal Paladino, L. (2016). On the neo-sex chromosomes of lepidoptera. *Evolutionary Biology: Convergent Evolution, Evolution of Complex Traits, Concepts and Methods*, Springer International Publishing 171–185.
- Nguyen, P., Sýkorová, M., Šíchová, J., Kůta, V., Dalíková, M., Frydrychová, R. Č., Neven, L. G., Sahara, K., & Marec, F. (2013). Neo-sex chromosomes and adaptive potential in tortricid pests. *Proceedings of the National Academy of Sciences of the United States of America*, 110(17), 6931–6936.
- Nguyen, P., Tu Anh Thi, M. Jerome Beetz, Merlin, Ch., Jundi, El, B. (2021). Sun compass neurons are tuned to migratory orientation in monarch butterflies. *Proceedings of the Royal Society B: Biological Sciences*, 288(1945)
- Petschenka, G., & Agrawal, A. A. (2016). How herbivores coopt plant defenses: Natural selection, specialization, and sequestration. *Current Opinion in Insect Science*, 14, 17–24.
- Petschenka, G., Fandrich, S., Sander, N., Wagschal, V., Boppré, M., & Dobler, S. (2013). Stepwise evolution of resistance to toxic cardenolides via genetic substitutions in the Na^+/K^+ -ATPase of milkweed butterflies (Lepidoptera: Danaini). *Evolution*, 67(9), 2753–2761.
- Pierce, A. A., Zalucki, M. P., Bangura, M., Udawatta, M., Kronforst, M. R., Altizer, S., Haeger, J. F., & de Roode, J. C. (2014). Serial founder effects and genetic differentiation during worldwide range expansion of monarch butterflies. *Proceedings of the Royal Society B: Biological Sciences*, 281(1797).
- Pierson, L., & Fink, L. S. (1985). A Natural Toxic Defense System: Cardenolides in Butterflies versus Birds. *Annals of the New York Academy of Sciences*, 443, 171–188.
- Potter S, Bragg JG, Blom MP, Deakin JE, Kirkpatrick M, Eldridge MD, Moritz C. (2017). Chromosomal Speciation in the Genomics Era: Disentangling Phylogenetic Evolution of Rock-wallabies. *Front Genet*. 10(8), 10.
- Pringle, G., E., Baxter. W., S., Webster, L., C., Papanicolaou, A., Lee, F., S., Jiggins, D., Ch. (2007). Synteny and Chromosome Evolution in the Lepidoptera: Evidence From Mapping in *Heliconius melpomene*, *Genetics*, 177(1), 417-426.

- Quinlan AR, Hall IM. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 26(6), 841-842.
- Reppert, S. M. (2006). A colorful model of the circadian clock. *Cell*, 124(2), 233–236.
- Reppert, S. M., & de Roode, J. C. (2018). Demystifying Monarch Butterfly Migration. *Current Biology*, 28(17), 1009–1022.
- Reppert, S. M., Guerra, P. A., & Merlin, C. (2016). Neurobiology of Monarch Butterfly Migration. *Annual Review of Entomology*, 61, 25–42.
- Reppert, S. M., Zhu, H., & White, R. H. (2004). Polarized Light Helps Monarch Butterflies Navigate. *Current Biology*, 14(2), 155–158.
- Richards, S., & Murali, S. C. (2015). Best practices in insect genome sequencing: What works and what doesn't. *Current Opinion in Insect Science*, 7, 1–7.
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. v, & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation complete-ness with single-copy orthologs. 31(19), 3210-3212.
- Singh KS, De-Kayne R, Omufwoko KS, Martins DJ, Bass C, Ffrench-Constant R, Martin SH. (2022). Genome assembly of *Danaus chrysippus* and comparison with the Monarch *Danaus plexippus*. *G3 Genes|Genomes|Genetics*. 12(3).
- Smith, D. A. S., Gordon, I. J., Traut, W., Herren, J., Collins, S., Martins, D. J., Saitoti, K., Ireri, P., & Ffrench-Constant, R. (2016). A neo-W chromosome in a tropical butterfly links colour pattern, male-killing, and speciation. *Proceedings of the Royal Society B: Biological Sciences*, 283(1835).
- Talla, V., Suh, A., Kalsoom, F., Dinca, V., Vila, R., Fribeg, M., Wiklund, Ch., Backström, N., (2017). Rapid Increase in Genome Size as a Consequence of Transposable Element Hyperactivity in Wood-White (Leptidea) Butterflies. *Genome Biology and Evolution*, 9(10), 2491-2505.
- Traut, W., & Marec, F. (1996). Sex chromatin in lepidoptera. *Quarterly Review of Biology*, 71(2), 239–256.
- Traut, W., & Marec, F. (1997). Sex chromosome differentiation in some species of Lepidoptera (Insecta). *Chromosome Res.*, 5(5), 283–291.

- Traut, W., Sahara, K., & Marec, F. (2008). Sex chromosomes and sex determination in Lepidoptera. *Sexual Development*, 1 (6), 332–346.
- Traut W., Ahola V., Smith DAS, Gordon IJ, Ffrench-Constant RH. (2018). Karyotypes versus Genomes: The Nymphalid Butterflies *Melitaea cinxia*, *Danaus plexippus*, and *D. chrysippus*. *Cytogenetic and Genome Research*, 153(1), 46-53.
- Vurtture, G. W., Sedlazeck, F. J., Nattestad, M., Underwood, C. J., Fang, H., Gurtowski, J., & Schatz, M. C. (2017). GenomeScope: Fast reference-free genome profiling from short reads. *Bioinformatics*, 33(14), 2202–2204.
- Walters, J. R., & Hardcastle, T. J. (2011). Getting a full dose? reconsidering sex chromosome dosage compensation in the silkworm, *Bombyx mori*. *Genome Biology and Evolution*, 3(1), 491–504.
- Weirather JL, de Cesare M, Wang Y, Piazza P, Sebastiano V, Wang XJ, Buck D, Au KF. (2017) Comprehensive comparison of Pacific Biosciences and Oxford Nanopore Technologies and their applications to transcriptome analysis. *F1000Res*. 3(6), 100.
- Wright, A. E., Dean, R., Zimmer, F., & Mank, J. E. (2016). How to make a sex chromosome. *Nature Communications*, 7.
- Yoshido, A., Yasukochi, Y., & Sahara, K. (2011). *Samia cynthia* versus *Bombyx mori*: Comparative gene mapping between a species with a low-number karyotype and the model species of Lepidoptera. *Insect Biochemistry and Molecular Biology*, 41(6), 370–377.
- Zha, X., Xia, Q., Duan, J., Wang, C., He, N., & Xiang, Z. (2009). Dosage analysis of Z chromosome genes using microarray in silkworm, *Bombyx mori*. *Insect Biochemistry and Molecular Biology*, 39(5–6), 315–321.
- Zhan S, Merlin C, Boore JL, Reppert SM (2011). The monarch butterfly genome yields insights into long-distance migration. *Cell*. 147(5), 1171-85.