

Jihočeská univerzita v Českých Budějovicích
Přírodovědecká fakulta

**Vývoj molekulárních markerů pro určení
pohlaví u předivky brslenové
(*Yponomeuta cagnagella*)**

Diplomová práce

Mgr. Aneta Pilíková

Školitel: RNDr. Petr Nguyen, Ph.D.

České Budějovice 2023

Pilíková, A., 2023: Vývoj molekulárních markerů pro určení pohlaví u předivky brslenové (*Yponomeuta cagnagella*). [Development of molecular markers for sex determination of the spindle ermine moth (*Yponomeuta cagnagella*). Mgr. Thesis, in Czech] – 53 p., Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic.

Annotation

The thesis focuses on developing a female-specific marker for the spindle ermine moth *Yponomeuta cagnagella* (Lepidoptera). The genome of a female specimen was sequenced using Oxford Nanopore technology, yielding a highly complex assembly. Female-specific k-mers were extracted, and candidate sequences corresponding to the W chromosome were identified using the YGS k-mer based pipeline. In silico verification confirmed the specificity of these sequences and primers were designed for the four predicted female markers. However, PCR experiments revealed that these sequences were not sex-specific and therefore could not be used for molecular sexing.

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

V Českých Budějovicích 13. 4. 2023

.....

Mgr. Aneta Pilíková

Poděkování

Srdečné poděkování patří mému školiteli Petru Nguyenovi za udržování přátelského pracovního prostředí, za jeho humor, podporu a pomoc, které mi poskytl v průběhu celého mého studia, nejen při práci na tomto projektu.

Děkuji také celému týmu naší laboratoře, zejména Aniče za lekce bioinformatiky, a také Martině a Monče za rady ohledně k-merové analýzy. Za příjemnou pracovní atmosféru děkuji Xeně, Majdě, Adri, Monice, Lukášovi, Jamiemu, Pablovi a všem ostatním, kteří naší pracovní skupinou v průběhu mých studií prošli. Jsem vděčná za zázemí, které nám poskytuje Katedra molekulární biologie a genetiky na PřF JU a díky patří taktéž kolektivu, který katedru tvoří.

Děkuji své rodině, která mi vždy poskytla podporu. Díky také svým spolužákům za pocit sounáležitosti na naší převážně radostné cestě vysokoškolským vzděláním. Děkuji i ostatním kamarádům, kteří mě povzbuzovali, a v neposlední řadě mému příteli Davidovi, za všechny teplé večere, které mi během tůkání do počítače přistály pod nosem.

Obsah

1. Úvod	1
1.1. Pohlavní chromozomy motýlů.....	1
1.2. Identifikace pohlavně specifických sekvencí DNA	6
1.3. Novodobá sekvenační data	8
1.4. K-merová analýza.....	9
1.5. Předivka brslenová jako modelový druh	11
2. Cíle práce.....	12
3. Metody.....	13
3.1. Materiál a data	13
3.2. Sestavení samičího referenčního genomu	14
3.2.1. Izolace vysokomolekulární genomové DNA	14
3.2.2. Sekvence DNA s využitím technologie Oxford Nanopore.....	14
3.2.3. Skládání samičího genomu.....	14
3.3. K-merová analýza s využitím YGS	16
3.3.1. Analýza krátkých sekvenačních čtení	16
3.3.2. Rozdělení sekvencí na k-mery.....	16
3.3.3. YGS	16
3.3.4. Zpracování výsledků YGS	18
3.4. Navržení PCR primerů pro ověření kandidátních sekvencí	20
3.5. PCR ověření kandidátních markerů.....	22
4. Výsledky.....	23
4.1. Assembly genomu	23
4.2. Výsledky YGS.....	25
4.3. Identifikace kandidátních samičích sekvencí	28
4.4. Ověření PCR markerů	31
5. Diskuze	38
5.1. Sestavení genomu	38
5.2. K-merová analýza YGS.....	40
5.3. Další cesty k identifikaci pohlavně vázaných sekvencí.....	44
6. Závěr.....	46
7. Seznam použité literatury	47

1. Úvod

1.1. Pohlavní chromozomy motýlů

Pohlavní rozmnožování je mezi eukaryoty široce rozšířeno, protože ze své podstaty umožňuje promíchání genetického materiálu, a zvyšuje tak šanci potomků na přežití (Zimmer 2009). Systémy **určení pohlaví** lze rozdělit do dvou skupin. Buďto je pohlaví určeno environmentálními faktory, a není tak závislé na genetické výbavě, nebo přímo genetickými faktory, kdy je pohlaví jedince určováno chromozomy nesoucími geny pro vývoj daného pohlaví. Druhý jmenovaný způsob determinace pohlaví je mnohem častější. Existují v zásadě dva chromozomální systémy: samčí heterogamie ($\text{♀XX}/\text{♂XY}$), která je typická pro savce a většinu skupin hmyzu, a samičí heterogamie ($\text{♀WZ}/\text{♂ZZ}$), kterou nalezneme u ptáků, některých hadů a motýlů. Pohlavní chromozomy se u různých skupin organismů vyvinuly opakovaně a nezávisle (Bachtrog 2006), přesto vykazují vysokou podobnost co do funkce a složení. Tato pozoruhodná konvergentní evoluce, diferenciací a role pohlavních chromozomů ve vzniku nových druhů představují významná a fascinující témata v oblasti evoluční genetiky.

Motýli (Lepidoptera) představují druhově nejrozmanitější skupinu živočichů s heterogametickými samicemi. Nejčastější a ancestrální **karyotyp** tvoří 31 chromozomů v haploidní sadě (Lukhtanov 2000), ale mezi diploidními organismy najdeme právě u motýlů největší variabilitu v počtu chromozomů, který se může velmi lišit i u blízce příbuzných druhů (Traut et al. 2007). Nacházíme karyotypy v rozmezí $n = 5$, například u jihoamerického motýla *Hypothyris thea* (Nymphalidae) (Brown et al. 2004) až $n = 226$ u modráška *Polyommatus atlanticus* (Lycaenidae) (Lukhtanov 2015). Chromozomy motýlů jsou uniformní, malé a **holocentrické**, tedy postrádající primární konstrikci v oblasti centromery. Při dělení buňky se vlákna dělicího vřeténka připojují na 30-70 % délky chromozomu (Senaratne et al. 2021). Variabilita v počtu chromozomů je přičítána právě holocentrické povaze motýlích chromozomů, které při meióze snadno podlehnou fragmentacím nebo fúzím a dochází tak k přestavbám karyotypů (Bureš a Zedek 2014; Zedek a Bureš 2018; Hill et al. 2019). Jednotlivé motýlí chromozomy není možné rozlišit běžnými technikami pruhování, a jejich studium je umožněno až díky molekulárně-cytogenetickým metodám, například fluorescenční in situ hybridizaci. Variace této metody genomová in situ hybridizace (GISH) nebo komparativní genomová hybridizace (CGH) umožňují identifikovat pohlavní chromozomy a analyzovat chromozomální přestavby (Traut et al. 2001; Yoshido et al. 2006).

Chromozom Z u motýlů je podobný autozomům, často je v karyotypu největší a nese řadu pohlavně vázaných genů. Sekvenčně se od chromozomu W značně liší (Traut et al. 2007). Sekvence DNA chromozomu Z je mezidruhově konzervativní (Sahara et al. 2012), ačkoli výsledky studie Van't Hof et al. (2013) potvrzují, že na chromozomu Z dochází k dynamickým chromozomálním přestavbám v podobě inverzí, které mohou přispět k adaptivní evoluci a procesu speciace.

Chromozom W se nachází u skupin Tischeriidae a Ditrysia zahrnujících 98 % známých druhů motýlů (Traut et al. 2007) a přesto o jeho molekulárním složení a roli v určení pohlaví máme velmi málo informací. Tento chromozom je obvykle tvořen převážně konstitutivním heterochromatinem, který obsahuje hojné množství repetitivní DNA, zejména mobilní elementy (Hejníčková et al. 2023). To způsobuje těžkosti při skládání genomu („assembly“) (Carey et al. 2022). Doposud bylo na chromozomu W nalezeno jen několik málo protein-kódujících genů, které jsou obvykle kopií genů z chromozomu Z, například *period (per)* u martináče *Antheraea pernyi*, Saturniidae) (Gotter et al. 1999) nebo *laminin a* u drsnokřídlece březového (*Biston betularia*, Geometridae) (Van't Hof et al. 2013). Jediný protein-kódující gen specifický pro chromozom W, který byl dosud objeven, je *GUWI* u černopásky bavlníkové (*Helicoverpa armigera*, Noctuidae) a jeho funkce není prozkoumána (Deng et al. 2020).

V karyotypu bazálních skupin motýlů (Lepidoptera), stejně jako u sesterského řádu chrostíků (Trichoptera), chybí chromozom W a samičí pohlaví je charakterizováno přítomností pouze jednoho chromozomu Z (Sahara et al. 2012). Avšak i mezi bazálními skupinami byl nalezen druh disponující chromozomem W, *Phymatopus californicus* (Hepialidae) (Voleníková 2015). U samic evolučně mladších skupin motýlů se chromozom W obvykle nalézá, ale existují i druhy, u kterých došlo k sekundární ztrátě, např. *Samia cynthia ricini* (Saturniidae) (Yoshido et al. 2005). Některé druhy mají větší počet chromozomů W, například *Bactra lacteana* (Tortricidae) s konstitucí ♀W₁W₂Z/♂Z₁Z₂ (Suomalainen 1969), jiní zástupci mají násobný počet Z chromozomů, například *Trabala vishnou* ♀WZ₁Z₂/♂Z₁Z₁Z₂Z₂ (Lasiocampidae) (Rishi et al. 1999). Větší počet pohlavních chromozomů nalézáme u karyotypových ras a počínajících nebo kryptických druhů, například u rodu *Leptidea* byly detekovány multivalenty tvořené větším počtem W i Z chromozomů (Šíchová et al. 2015; Höök et al. 2023).

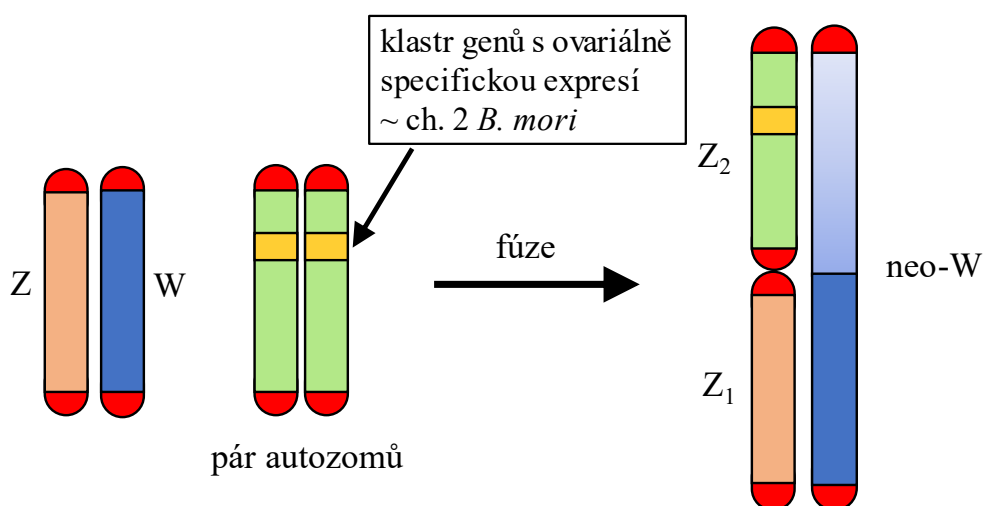
Vznik pohlavních chromozomů u eukaryot je tradičně vysvětlován univerzálním **modelem degenerace** páru autozomů. Předpokládaným začátkem procesu je mutace, která dá vzniknout pohlavní determinantě neboli faktoru určujícímu pohlaví na jednom

z homologů autozomálního páru. V blízkosti tohoto faktoru se začnou hromadit pohlavně antagonistické mutace, tedy mutace výhodné pro dané pohlaví případně škodlivé pro pohlaví opačné. Evolučně úspěšná kombinace pohlavně antagonistických mutací je udržována omezením rekombinace. Nerekombinující oblast postupně geneticky eroduje, tj. akumuluje mobilní elementy, tandemové repetice a pseudogeny a její struktura je tvořena heterochromatinem. Vzniká tak degenerovaný pohlavně specifický chromozom, označovaný Y v případě samců, respektive W v případě samičí heterogamie (Charlesworth et al. 2005). Alternativním způsobem vzniku pohlavních chromozomů je **adopce B chromozomu**, což je nadpočetný jaderný element DNA, jehož výskyt je v rámci populací a druhů variabilní. B chromozomy jsou z velké části tvořeny repetitivní DNA, která usnadňuje ektopickou rekombinaci a díky tomu může B chromozom získat pohlaví determinující faktor, případně se může začít párovat s existujícím univalentem chromozomu X nebo Z (Clark a Kocher 2019; Lewis et al. 2021).

Evoluční původ chromozomu W u motýlů doposud nebyl objasněn, roste však počet důkazů o jeho **nekanonickém původu** u různých skupin řádu Lepidoptera z B chromozomu (Dalíková et al. 2017; Dai et al. 2022). Konzervativní sekvence chromozomu Z a nedostatek jeho homologie s chromozomem W hypotézy o původu chromozomu W z B chromozomu podporují (Fraïsse et al. 2017). Navíc nedostatek podobnosti mezi chromozomy W různých druhů motýlů poukazuje na možnost opakované adopce chromozomu B (Lewis et al. 2021).

Naše znalosti o diferenciaci pohlavních chromozomů u živočichů vychází ze studia evolučně starých systémů, kde jsou chromozomy již diferencované a degenerované. Možnost **pozorovat proces diferenciaci** pohlavních chromozomů v různých stádiích nám přináší evolučně mladé systémy (s rychlou evolucí pohlavních chromozomů, „sex chromosome turn over“), jako jsou například právě motýli, u kterých se nezřídka vyskytují neo-pohlavní chromozomy (Bachtrog 2006; Bergero a Charlesworth 2008; Kaiser a Bachtrog 2010; Abbott et al. 2017). **Neo-pohlavní chromozomy** vznikají spojením pohlavních chromozomů s autozomy, což vede ke vzniku různých konstitucí, např. neo-WZ₁Z₂ nebo W₁W₂neo-Z. Neo-pohlavní chromozomy se poměrně často vyskytují u organismů s chromozomálním systémem XX/XY, a to jak u bezobratlých (např. u některých octomilek), tak u obratlovců (např. u ryb nebo šupinatých plazů). U organismů s pohlavními chromozomy ZZ/ZW jsou ale neo-pohlavní chromozomy vzácné (Pennell et al. 2015), s výjimkou motýlů, u nichž jsou fúze pohlavních chromozomů a autozomů běžné. Gonozomy se u takových druhů skládají z evolučně různě starých částí, což nám poskytuje příležitost zkoumat jejich vznik, ranou evoluci a diferenciaci a také faktory, které chromozomální fúze způsobují. Neo-pohlavní

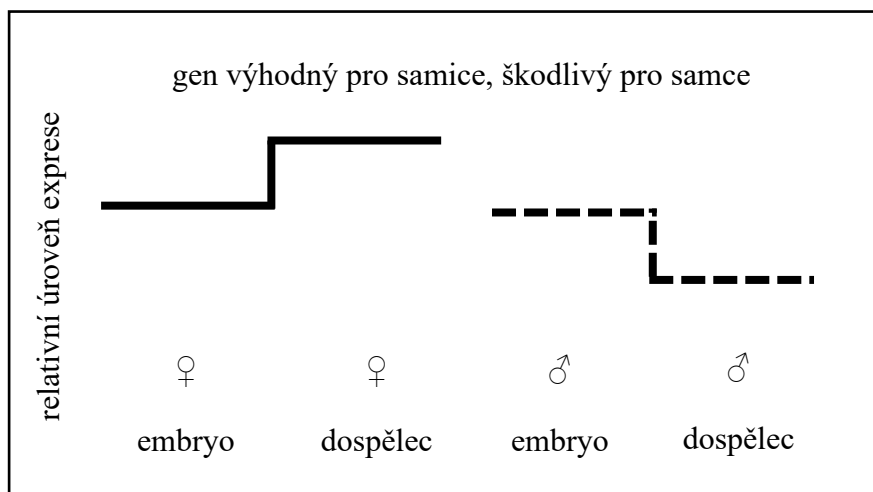
chromozomy byly objeveny například u obaleče jablečného *Cydia pomonella* a dalších zástupců čeledi obalečovitých (Tortricidae). Byla u nich potvrzena fúze chromozomu Z s autozomem, který u bource morušového (*Bombyx mori*, Bombycidae) odpovídá chromozomu 15 (Nguyen et al. 2013). U monarchy stěhovavého *Danaus plexippus* (Mongue et al. 2017) je neo-Z chromozom tvořen ancestrálním Z a autozomem, který je homologem chromozomu 16 u *B. mori*. Přítomnost mnohačetného pohlavního chromozomálního systému WZ_1Z_2 byla prokázána také u Yponomeutidae, např. u *Teinoptila gutella* a všech druhů *Yponomeuta* kromě *Y. tokyonella* (Provazníková 2022). U této čeledi vznikla tato chromozomální konstituce fúzí mezi chromozomem W a homologem chromozomu 2 u *B. mori*, na kterém se nachází klastr genů se expresí specifickou pro ovaria, viz **Obr. 1**.



Obrázek 1: Vznik neo-W u předivek rodu *Yponomeuta*. Chromozom W podstoupil fúzi s autozomem, na němž se nacházel klastr nesoucí geny s tkáňově specifickou expresí v samičích ováriích. Neo-W se nadále páruje s původním chromozomem Z, ale také s homologem autozomu, ze kterého se stal chromozom Z_2 . Původní autozomální pár odpovídá chromozomu 2 u *Bombyx mori*.

Doposud není jasné, čím jsou časté fúze pohlavních chromozomů s autozomy způsobeny. Bližší studium genového obsahu autozomů tvořících neo-pohlavní chromozomy u výše popsaných příkladů (čeledi Tortricidae, rodu *Danaus* a *Yponomeuta*) odhalilo, že tyto autozomy nesou klastry genů, které jsou u *B. mori* exprimovány v ováriích (Suetsugu et al. 2013). Nabízí se tak možnost, že vznik neo-pohlavních chromozomů je poháněn **pohlavně antagonistickou selekcí**. Geny prospěšné pro samice a škodlivé pro samce vedou ke konfliktu, který může být selekcí vyřešen pohlavně specifickou expresí pouze u samic, jak

je tomu u *B. mori*, kde jsou chorionové geny potřebné pro vytvoření obalu vajíčka na chromozomu 2 a exprimují se pouze v ovarích (Suetsugu et al. 2013). Druhou možností je vznik selekčního tlaku na těsnější vazbu těchto genů na samičí pohlaví, tj. na chromozom W. Pro zodpovězení otázky, zda jsou fúze pohlavních chromozomů s autozomy v populacích motýlů zvýhodněny pohlavně antagonistickou selekcí, je potřeba identifikovat pohlavně antagonistické geny. Mank a Ellegren (2009) **navrhli srovnání exprese** jako nepřímý způsob identifikace pohlavně antagonistických genů u kuřete (*Gallus gallus*). Ve své práci porovnávali expresi genů s různými účinky u různých vývojových stádií za předpokladu, že exprese pohlavně antagonistických genů je u raných embryí, kdy ještě není realizován rozdíl mezi pohlavími, u samce i samice stejná, ale liší se mezi dospělci, viz **Obr. 2**. Tento přístup není ovlivněn kompenzací genové dávky ani meiotickou inaktivací pohlavních chromozomů, které se vyvinuly u některých taxonů.



Obrázek 2: Příklad exprese pohlavně antagonistického genu. Úroveň exprese genu, který je výhodný pro samice ale nevýhodný pro samce je u embryí stejná, zatímco u dospělých samic je exprese vyšší a u samčích dospělců je potlačena (Mank a Ellegren 2009, upraveno).

Využití porovnání exprese u motýlů za účelem identifikace pohlavně antagonistických genů brání absenci metod pro **spolehlivé určení pohlaví** v časných stádiích vývoje, tj. u embryí a mladých larev, u kterých se obvykle neprojevuje pohlavní dimorfismus a samce a samice tedy nelze morfologicky odlišit (Fuková et al. 2009).

1.2. Identifikace pohlavně specifických sekvencí DNA

Ověření pohlaví u raných vývojových stádií hmyzu je zapotřebí nejen pro testování vlivu pohlavně antagonistické selekce zmíněné výše, ale také v celé plejádě biologických disciplín, od imunologie a parazitologie přes metabolomiku až po studium migrace a populační dynamiky (Belousova et al. 2019). Identifikace pohlaví je také nezbytná v biologických metodách **regulace šíření hmyzích škůdců**, kde se využívá sterilizace samců či samic (Koukidou a Alphey 2014; Marec a Vreysen 2019). Rozlišení pohlaví na základě morfologických znaků je v časných fázích vývoje u většiny hmyzu obtížné. Zároveň se však většina prací prováděných na motýlech (Lepidoptera) soustředí právě na **studium larev**, neboť zvířata v subadultních stádiích jsou největší hrozbou pro současné intenzivní zemědělství a lesnictví (Zalucki et al. 2012; Jumaev et al. 2021).

Spolehlivé určení pohlaví („sexing“) je umožněno, pokud jsou pro daný druh dostupné pohlavně specifické markery. Tradičním markerem použitelným u mnoha druhů motýlů je heterochromatinové tělísko, tzv. „**sex chromatin**“ nebo Smithovo tělísko, které se nachází v interfázních jádrech samic (Smith 1945; Traut a Marec 1996). Pohlavně specifický chromatin je tvořen chromozomem W a indikuje tak jeho přítomnost. Metoda však funguje pouze u druhů s jasně diferencovaným chromozomem W, zatímco u druhů s neo-pohlavními chromozomy není technika spolehlivá. U některých druhů Smithovo tělísko chybí, ačkoli chromozomem W disponují, například u jasoně dymnivkového (*Parnassius mnemosyne*, Papilionidae) (Vlašánek et al. 2017). Zároveň je tato metoda z praktického hlediska použitelná až u středních a pozdních larválních instarů. Raná vývojová stádia jsou na přípravu chromozomálních preparátů zpravidla příliš malá, ačkoli například v práci Fukové (2009) bylo tímto způsobem ověřováno pohlaví již u čerstvě vylíhlých larev.

Pohlaví lze ověřit také metodou **průtokové cytometrie**, která detekuje odlišné množství DNA v buňkách samic a samců na základě různé velikosti chromozomů Z a W, eventuelně absenci chromozomu W. Genom samic u druhů, které postrádají chromozom W je menší, než genom samců a zároveň lze touto metodou určit velikost chromozomu Z (Hejníčková et al. 2019). Tento způsob je náročný jak na potřebný materiál, tak na samotný postup.

Řada molekulárních testů pro určení pohlaví u zvířat je založena na **PCR amplifikaci známých pohlavně vázaných úseků DNA**, tedy na porovnávání alel či počtu kopií stejného genu u různých pohlaví. Rozšířeným způsobem genetického ověření pohlaví u ptáků je využití různé velikosti intronů v genu CHD („chromo helicase DNA-binding gene“). Poloha CHD genu v nerekombinantní části chromozomů Z a W umožnila rozdílnou evoluci těchto tzv.

gametologů, které tak mají na obou chromozomech různou délku, což je spolehlivý marker (Griffiths et al. 1998; Vucicevic et al. 2013). U vrtule velkohlavé (*Ceratitis capitata*, Diptera, Tephritidae) byl popsán postup molekulárního sexingu na základě homologických oblastí kódující ITS1 („internal transcribed spacer“), pohlavně vázané ribozomální DNA, přičemž samci a samice jsou rozlišeni na základě vzoru restrikčních fragmentů vzniklých po štěpení produktů PCR restrikčním enzymem (Douglas et al. 2004). V téže práci byla také jako marker pro samce použita repetitivní DNA pocházející z Y chromozomu.

Metody **molekulárního určení pohlaví** použitelné i pro vajíčka a časné instary byly vyvinuty pro několik druhů motýlů. U nejlépe prostudovaného motýla bource morušového *Bombyx mori* je znám gen *feminizer* (*fem*), který je na chromozomu W, kóduje krátké regulační RNA a zodpovídá za determinaci pohlaví (Abe et al. 2008). U zavíječe kukuřičného (*Ostrinia nubilalis*) byly detekovány dva mikrosatelity specifické pro samice (Coates a Hellmich 2003). Na obaleče *C. pomonella* byla nalezena sekvence specifická pro chromozom W (*CpW2-EcoRI*), která byla společně s genem *period* (*Cpper*) lokalizovaným na Z chromozomu použita jako marker pro metodu multiplex PCR (Fuková et al. 2009). Na základě této práce byl molekulární sexing proveden také u obaleče mramorovaného (*Lobesia botrana*, Tortricidae), u nějž byly nalezeny homologní sekvence *CpW2* a byly ověřeny metodou duplex PCR společně s částí genu *EF1-a* (*elongation factor 1 alpha*) jako vnitřní kontrolou (Aguirre et al. 2020). Na chromozomu W zavíječe moučného (*Ephestia kuehniella*, Pyralidae) bylo dokázáno, že obsahuje inserty mitochondriální DNA, jejichž fragmenty bylo možné použít jako PCR markery pro ověření samičího pohlaví (Traut et al. 2013). Gen *GUWI* byl použit jako marker pro identifikaci samic u černopásky *H. armigera*. Gen byl odhalen mapováním transkriptomu získaného z vajíček na samčí genom a jedná se o první protein-kódující gen specifický pro chromozom W, který byl u motýlů objeven (Deng et al. 2020), jeho funkce však není objasněna.

Překážkou pro širší uplatnění výše uvedených metod je skutečnost, že jsou založeny na detekci známých sekvencí vázaných na chromozom W. Tento chromozom však u žádného zástupce motýlů nebylo dosud technologicky možné kompletně osekvenovat kvůli jeho repetitivnímu obsahu (Sahara et al. 2003; Traut et al. 2007; Kawamoto et al. 2022). K rozšíření našich znalostí o molekulárním složení chromozomu W však mohou přispět neustále přibývající sekvenční data.

1.3. Novodobá sekvenační data

První rozšířenou sekvenační metodou bylo **Sangerovo sekvenování** (Sanger et al. 1977) založené na modifikované replikaci DNA. Sangerovo sekvenování fungující na principu inkorporace dideoxynukleotidů DNA polymerázou se používá dodnes, avšak rozšířenější je sekvenování tzv. **druhé generace**, které vzniklo v 90. letech. Přineslo rychlejší a méně pracné způsoby čtení molekul DNA. Dominantní metodou této generace je Illumina sekvenování, jehož princip spočívá v syntéze řetězce DNA na PCR amplikonech generovaných pomocí oligonukleotidových technických sekvencí přichycených k destičce tzv. „flow cell“, přičemž se měří emise světla, které je uvolňováno po každém přidání nukleotidu. Tato metoda je velmi účinná, maximální délka jednotných sekvencí sice dosahuje jen několik stovek párů bází (obvykle 100-250 bp), je však velmi přesná a chybovost se uvádí maximálně 1 % (McCombie et al. 2019).

Na počátku 21. století vznikly sekvenační metody **třetí generace**, které čtou v reálném čase pouze jednu molekulu DNA. Detekce složení molekul je rychlá, umožňuje získat až miliony bází dlouhé sekvence, které ale mají oproti Illumina sekvenování vyšší chybovost (v rozmezí zhruba 5-15 %). Opakované sekvenování však chybovost snižuje (Lin et al. 2021; Wang et al. 2021). V současné době na trhu převládají dvě sekvenační technologie třetí generace – Pacific Biosciences (PacBio, USA) a Oxford Nanopore Technologie (ONT, Velká Británie) (Udaondo et al. 2021).

Principem sekvenování **metodou ONT** je průchod DNA molekuly nanopórem ukotveným na membráně, inspirovaným bakteriálními transmembránovými proteiny. DNA molekula je díky motorovému proteinu protahována nanopórem, který je opatřen elektrodami zaznamenávajícími specifické změny iontového proudu při průchodu jednotlivých nukleotidů a v podobě křivky je tak identifikována celá procházející molekula DNA. Technologie umožňuje detekovat i epigenetické modifikace jednotlivých bází (Schatz 2017).

Průlom v dostupnosti sekvenování nastal v roce 2014 uvedením přenosného sekvenátoru MinION na trh. MinION je malé a cenově dostupné zařízení, které se připojuje k počítači standardním USB portem. Nalézá uplatnění v celé škále vědních disciplín, protože umožňuje v krátkém čase určit sekvenci jak DNA, tak RNA. Jádrem sekvenátoru tvoří průtoková komůrka („flow cell“), která má membránu s 512 měřicími kanály z nichž každý je napojen na čtyři nanopóry. V reálném čase je ale možné sekvenovat pouze na jednom z nich. DNA molekula prochází nanopórem rychlostí 400-450 bází za sekundu a výrobce uvádí zisk

až 100 Gbp během tří dní, přičemž délka sekvencí dosahuje stovky tisíc, ojediněle i miliony bází (nanoporetech.com).

Díky **rychlému vývoji sekvenačních technologií** je dnes daleko dostupnější i studium pohlavních chromozomů W a Y, které byly dříve ze sekvenačních projektů navzdory svému biologickému významu obvykle vyloučeny. Identifikace sekvencí vázaných na pohlaví se setkává především se dvěma obtížemi. Zaprvé jsou sekvence chromozomů W a Y vysoce repetitivní a zadruhé tyto chromozomy obecně získávají pouze poloviční sekvenační pokrytí v porovnání s autozomy (Tomaszkiewicz et al. 2017; Carey et al. 2022). Byly však popsány **komparativní genomické metody**, které své principy staví právě na výše uvedených faktorech a využívají **porovnávání pohlavně specifických k-merů**.

1.4. K-merová analýza

„K-mery“ jsou stejně dlouhé překrývající se úseky sekvence DNA o stanovené délce „k“, viz **Obr. 3**.

k=2:	k=4:	k=6:
AAGTCCATGG	AAGTCCATGG	AAGTCCATGG
AA	AAGT	AAGTCC
AG	AGTC	AGTCCA
GT	GTCC	GTCCAT
TC	TCCA	TCCATG
CC	CCAT	CCATGG
CA	CATG	
AT	ATGG	
TG		
GG		

Obrázek 3: Znárodnění principu rozkladu sekvence na k-mery. Sekvence AAGTCCATGG je rozložena na k-mery pro hodnoty $k = 2, 4$ a 6 .

K-mery se uplatňují jako velmi užitečný bioinformatický nástroj. Používají se při sekvenování DNA ke zjištění překryvů mezi náhodnými fragmenty, což je klíčové pro složení krátkých sekvencí do celého genomu (Manekar a Sathe 2018). K-merová analýza se také uplatňuje při studiích různých aspektů genomů, včetně jejich struktury a funkce, a také umožňuje detekovat sekvenační chyby a jednonukleotidové polymorfismy (SNP) (Leggett a MacLean 2014; Miyao et al. 2019).

Porovnání k-merů získaných ze sekvenačních dat umožňuje **detekovat oblasti**, které se mezi vstupními **sekvenačními daty liší**. Předpokladem pro využití k-merové analýzy k **určení pohlaví** je, že se na pohlavním chromozomu W (nebo Y) nachází unikátní sekvence DNA, které nejsou repetitivní, a odlišují se od autozomů i svých homologů na Z (nebo X) chromozomu. Jinými slovy při porovnání genomických dat od obou pohlaví mohou k-mery detekovat oblasti, které se vyskytují pouze v genomu heterogametického pohlaví (Carvalho a Clark 2013; Li et al. 2018; Rangavittal et al. 2019; Wang et al. 2020; Zhang et al. 2021).

Jedním z příkladů nástroje, který využívá k-merové analýzy k identifikaci pohlavně specifických sekvencí, je **Y chromosome Genome Scan („YGS“)** (Carvalho a Clark 2013). Jde o bioinformatický postup („pipeline“) sloužící k detekci sekvencí vázaných na chromozom Y u člověka a octomilky (*Drosophila virilis*, Diptera, Drosophilidae). V práci bylo využito krátkých sekvenačních čtení samice a samčího složeného genomu. Data byla rozdělena na k-mery a po odstranění těch, které odpovídali repetitívám a sekvenačním chybám, byly detekovány unikátní samčí k-mery. Ty byly následně namapovány na kompletní samčí sekvenci genomu, a tak autoři dosáhli zřetelného odlišení fragmentů náležících autozomům a chromozomu X. Na fragmenty DNA přítomné u jedinců obou pohlaví bylo namapováno méně než 5 % unikátních samčích k-merů, zatímco fragmenty specifické pro chromozom Y měli 95-100% shodu se samčími k-mery. Pomocí této metody bylo u *D. virilis* nalezeno 800 nových scaffoldů vázaných na chromozom Y o celkové velikosti 8,5 Mbp a čtyři nové geny. V lidském genomu, který byl použit jako pozitivní kontrola, byly pomocí YGS odhaleny všechny dříve známé geny nebo genové rodiny (Carvalho a Clark 2013).

Postup YGS byl úspěšně ověřen **i u organismů s chromozomálním určením pohlaví WZ/ZZ**, konkrétně na sekvenačních datech emu (*Dromaius novaehollandiae*) (Zhang et al. 2017) a na bourci morušovém (*B. mori*) (Li et al. 2018).

YGS vyžaduje vstupní data od jedné samice a jednoho samce, data z dalšího samce je možné použít v případě využití validačního kroku. Z toho plyne, že v případě aplikace YGS na outbreďní populace mohou být jako pohlavně specifické označeny i k-mery obsahující jednonukleotidové polymorfismy specifické pro daného jedince heterogametického pohlaví. Tento aspekt lze eliminovat použitím více biologických replik, a při vyhodnocení brát v potaz pouze k-mery, které se vyskytují ve všech replikách daného pohlaví.

Při k-merových analýzách je pro hmyzí genomy (přibližně o velikosti 100-500 Mbp) typicky využívána délka k-meru $k = 15$ (Carvalho a Clark 2013; Li et al. 2018), ale lze použít i delší k-mery, například 17-mery v práci Liu et al. (2019).

1.5. Předivka brslenová jako modelový druh

Studovaným modelem této práce je předivka brslenová *Yponomeuta cagnagella* (Yponomeutidae, Hübner 1813). Tento druh byl studován kvůli odvozenému karyotypu, který se liší mezi pohlavími. Samice mají 29 párů autozomů a tři pohlavní chromozomy: neo-WZ1Z2, které v meióze tvoří trivalent, viz **Obrázek 1**. Samci mají 31 chromozomálních párů (Nilsson et al. 1988; Veselá 2012).

Předivka brslenová je běžným škůdcem s holoarktickým rozšířením. Vyskytuje se lokálně, ale hojně a má jednu generaci za rok. Žije v řídkých lesích, křovinatých stráních, v ruderálních biotopech podél silnic a parcích s brslenovými výsadbami (Menken et al. 1992). Od poloviny do konce července kladou samičky vajíčka na větve hostitelské rostliny. Larvy se líhnou zhruba po dvou týdnech, ale přezimují pod ztvrdlými vaječnými obaly. První instary housenek se objevují koncem dubna nebo začátkem května a živí se čerstvými listy. Larvy předivky brslenové žijí ve společném zápředku a při přemnožení mohou opráskat svými vlákny celé keře a způsobit holožír. Kolem poloviny června se kuklí do kokonů, které jsou volně připevněny k hedvábné síti na rostlině. Dospělci vylétají z kokonů asi po dvou týdnech a brzy poté se páří.

Předivky jsou významným modelem pro výzkum speciace a evoluce s ohledem na změny hostitelských rostlin. Nadčeleď Yponomeutoidea patří mezi jednu z nejstarších skupin živočichů, u které se vyvinul externí příjem potravy a která již ve spodní křídě kolonizovala rostliny. Do rodu *Yponomeuta* patří 76 známých druhů (Ulenberg 2009), z nichž velká část způsobuje člověku ekonomické škody zejména na rostlinách čeledi jesencovitých (Celastraceae), růžovitých (Rosaceae) a vrbovitých (Salicaceae). *Y. cagnagella* spadá do komplexu devíti úzce příbuzných druhů (*Y. evonymella*, *Y. gigas*, *Y. grisea*, *Y. irrorela*, *Y. mahalebella*, *Y. malinella*, *Y. padella* a *Y. rorrella*), jejichž fylogenetické postavení není doposud zcela vyjasněno. Tyto druhy mohou žít sympatricky, jsou si morfologicky podobné a v laboratorních podmínkách některé z nich hybridizují (Turner et al. 2010; Hora et al. 2019). Společný předek druhů rodu *Yponomeuta* se pravděpodobně živil brslenem evropským (*Euonymus europaeus*, Celastraceae). U části druhů včetně výše zmíněného komplexu předivek došlo během geografické disperze ke změně hostitelské rostliny na některé zástupce růžovitých. Některé druhy včetně *Y. cagnagella* se však opět vrátili k původní živné rostlině (Turner et al. 2010; Ulenberg 2009).

2. Cíle práce

U většiny druhů motýlů je možné zjistit pohlaví jedince díky pohlavnímu dimorfismu dospělců, případně pitvou larev v 3.-5. instaru, chybí však možnost ověřit pohlaví u raných vývojových stádií. Pro studium role pohlavně antagonistické selekce v evoluci vzniku pohlavních chromozomů je nezbytně nutné vyvinout nástroje pro rychlé a spolehlivé určení pohlaví embryí a larev prvního instaru.

Předkládaná diplomová práce má za cíl identifikovat sekvence z pohlavního chromozomu W pomocí k-merové analýzy genomových sekvenačních dat obou pohlaví předivky brslenové, *Yponomeuta cagnagella*.

Dílčími cíli bylo:

- osekvenovat genom samice předivky *Y. cagnagella*,
- ze získaných sekvenačních dat poskládat samičí sekvenci genomu,
- identifikovat k-mery specifické pro samice,
- pomocí nástroje YGS nalézt fragmenty chromozomu W,
- navrhnout na tyto sekvence primery a ověřit jejich specificitu,
- ověřené PCR markery otestovat na jedincích raných vývojových stádií.

3. Metody

3.1. Materiál a data

Housenky a kukly zkoumaného organismu, tedy předivek *Yponomeuta cagnagella*, pocházely ze dvou různých divokých populací. První část byla získána z Amsterdamu (Nizozemsko) od Petera Roessingha (Amsterdamská universita). Další materiál byl sbírán v jihočeské obci Levín Petrem Nguyenem. Sebrané housenky byly do požadovaného instaru chovány na listech brsleny evropského (*Euonymus europaeus*).

Hrubá sekvenační data pro první část projektu (YGS) byla obdržena od vedoucího práce Petra Nguyena. Šlo o sekvenační čtení získané technologií Illumina o délce 150 bp a bylo použito párové čtení (pair end). V rámci této práce byl z DNA další samice poskládán nový genom a jako referenční byl použit samčí genom, získaný z práce (Volenikova et al. 2022). Přehled použitých sekvenačních dat je uveden v **Tabulce I**.

Tabulka I: Přehled vstupních dat *Yponomeuta cagnagella*. Místo původu je uvedeno zkratkou: CZ = Česká republika; NL = Nizozemsko. Sloupec „Rok“ uvádí rok odchyty a rok sekvenování daného jedince. „F_ONT“ je genom získaný v této práci, „M_ONT“ je poskládaný genom publikovaný v práci Volenikova et al. (2022).

Označení jedince	Pohlaví	Místo původu	Rok	Získáno	Typ dat
F1	♀	NL	2010/2016	P. Nguyen	Illumina
F2	♀	CZ	2015/2016	P. Nguyen	Illumina
M1	♂	NL	2016/2016	P. Nguyen	Illumina
M2	♂	NL	2016/2016	P. Nguyen	Illumina
M3	♂	NL	2010/2016	P. Nguyen	Illumina
F_ONT	♀	NL	2016/2021	A. Pilíková	ONT
M_ONT	♂	NL	2016/2021	A. Voleníková	gen. assembly

K bioinformatickým analýzám bylo využíváno prostředí operačního systému Linux a výpočetní zdroje a úložné kapacity virtuální organizace Metacentrum (metavo.metacentrum.cz). Skripty, které byly v této práci použity, jsou dostupné na webové platformě github: github.com/aneta-pilikova/w-chromosome-markers. Na tyto skripty používané v dílčích krocích výzkumu je v průběhu této práce odkazováno pomocí zkratk, např: (git/11_Jellyfish).

3.2. Sestavení samičího referenčního genomu

3.2.1. Izolace vysokomolekulární genomové DNA

Genomová DNA pro sekvenaci dlouhých čtení byla izolována z kukly samice předivky *Yponomeuta cagnagella* do té doby uchované v $-80\text{ }^{\circ}\text{C}$ (F_ONT, viz **Tab. I**). Samotná izolace byla provedena kitem Circulomics Nanobind HMW, podle protokolu Nanobind Insect Big DNA Kit, High Molecular Weight Insect DNA Extraction Protocol (Pacific Biosciences, USA) optimalizovaného pro motýly. Postup byl proveden s těmito změnami: ve 2. kroku bylo přidáno 400 μl pufru CT, v 8. kroku bylo použito 300 μl pufru IL1, v 16. kroku bylo přidáno 60 μl pufru BL3, v navazujícím kroku bylo použito 500 μl isopropanolu, ve 29. kroku bylo DNA rozpuštěno ve 100 μl pufru EB a v kroku 33 byl vzorek místo pokojové teploty uchován v lednici. Čistota získané DNA byla ověřena přístrojem NanoDrop One (ThermoFisher Scientific, USA), koncentrace DNA byla stanovena zařízením Qubit 4 Fluorometer s kitem Qubit dsDNA BR Assay Kit (ThermoFisher Scientific, USA) a délka fragmentů byla ověřena standardní elektroforetickou separací (1,5% agarózový gel v TAE, 5V/cm, barveno ethidium bromidem).

3.2.2. Sekvence DNA s využitím technologie Oxford Nanopore

Z genomové DNA byla připravena sekvenační knihovna podle protokolu Genomic DNA by ligation SQK-LSK110 pomocí kitů Ligation Sequencing Kit SQK-LSK109, SQK-LSK110 (Oxford Nanopore Technologies, Velká Británie) a NEBNext Companion Module for Oxford Nanopore Technologies Ligation Sequencing (New England BioLabs, USA). Cílem tohoto kroku bylo pročištění směsi, oprava konců molekul DNA a následné navázání adaptérů. Následně byla DNA kvantifikována pomocí Qubit 1x dsDNA HS Assay Kit na zařízení Qubit 4 Fluorometer (ThermoFisher Scientific, USA). Sekvence DNA byla provedena podle Genomic DNA by ligation SQK-LSK110 protokolu, na komůrce FlowCell R10.3 (1070 pórů) přístrojem MinION Mk1B (ONT). Na FlowCell bylo naneseno 22 fmol DNA. Sekvence trvala 52 hodin při -180 mV . Výsledkem sekvenování byla data ve formátu FAST5.

3.2.3. Skládání samičího genomu

Primární data získána ze sekvenační platformy MinION byla převedena na sekvenci příslušných nukleotidů procesem „basecallingu“. Jednotlivým signálům elektrického proudu zaznamenaným při průchodu nanopórem tak byly přiřazeny odpovídající dusíkaté báze

a hodnoty kvality, tedy pravděpodobnost chyby překladu. Byl použit basecaller Guppy-4.4.1 (ONT), ([git/01_Basecalling](#)). V tomto kroku bylo zároveň provedeno filtrování sekvencí na kvalitu a čtení s menším skórem než 7 byla odstraněna, protože u nich byla nízká pravděpodobnost přiřazení správného nukleotidu. Výsledné sekvence byly uloženy ve formátu FASTQ. Následně byla ověřena kvalita sekvenačních dat pomocí nástroje NanoPlot (De Coster et al. 2018), ([git/02_NanoPlot](#)). Na základě zjištěné délky sekvenačních čtení nebyla data dále filtrována a k samotnému poskládání byl použit program Flye-2.8 (Kolmogorov et al. 2020), ([git/03_Flye1](#)). Zároveň byly v tomto kroku sekvence opraveny sebou samými na základě jejich překryvů („polishing“) pomocí samotného nástroje Flye. Výsledný soubor je tvořen „contigy“, tedy souvislými genomovými sekvencemi, u nichž je pořadí DNA bází známo s vysokou mírou spolehlivosti a ty jsou v některých případech uspořádány do delších „scaffoldů“, obsahujících řetězce contigů a mezery. V této práci je pro contigy i scaffoldy sestaveného genomu využíván termín fragmenty. Kvalita získaného genomu byla vyhodnocena nástrojem BUSCO (Simão et al. 2015), ([git/04_BUSCO1](#)), který ověřuje informační obsah a komplexitu assembly, a také Quast (Gurevich et al. 2013), ([git/05_Quast](#)). Tímto způsobem byl získán soubor, který je v průběhu práce označován jako F_ONT, viz **Tab. I**, a který byl použit pro k-merovou analýzu (YGS). Při vyhodnocení kvality assembly (viz **kapitolu 4.1**) je tento soubor považován za první verzi (v1).

Genom byl posléze skládán znovu, a to postupem, který ve své práci použila Hrubá (2022). Účelem byla snaha o zvýšení kvality a následné použitelnosti genomu pro další analýzy. Ze sekvenačních dat byla před sestavením genomu odstraněna krátká čtení („ready“) s využitím nástroje NanoFilt (De Coster et al. 2018) a to se dvěma variantami nastavení. Prvně byla odstraněna sekvenační čtení menší než 5000 bp (v2) a v druhém případě čtení kratší než 10000 bp (v3). Vznikly tedy dvě sady dat („datasety“), s nimiž bylo dále pracováno zvlášť, a byly nejprve opět vyhodnoceny pomocí nástroje NanoPlot ([git/06_NanoFilt](#)). Pro zvýšení kvality sekvencí byl pro polishing použit také program Canu-2.1.1. (Koren et al. 2017), ([git/07_Canu](#)). Genom byl následně skládán assemblerem Flye-2.8 ([git/08_Flye2](#)), což opět zahrnovalo i polishing. Pro vyhodnocení kvality obou genomů sestavených z datasetů s různou minimální délkou sekvenčních čtení byly použity nástroje BUSCO, s modulem Augustus (Manni et al. 2021), ([git/09_BUSCO2](#)), a Quast ([git/05_Quast](#)). Tento způsob assembly zahrnoval filtrování a více polishing procedur.

3.3. K-merová analýza s využitím YGS

3.3.1. Analýza krátkých sekvenačních čtení

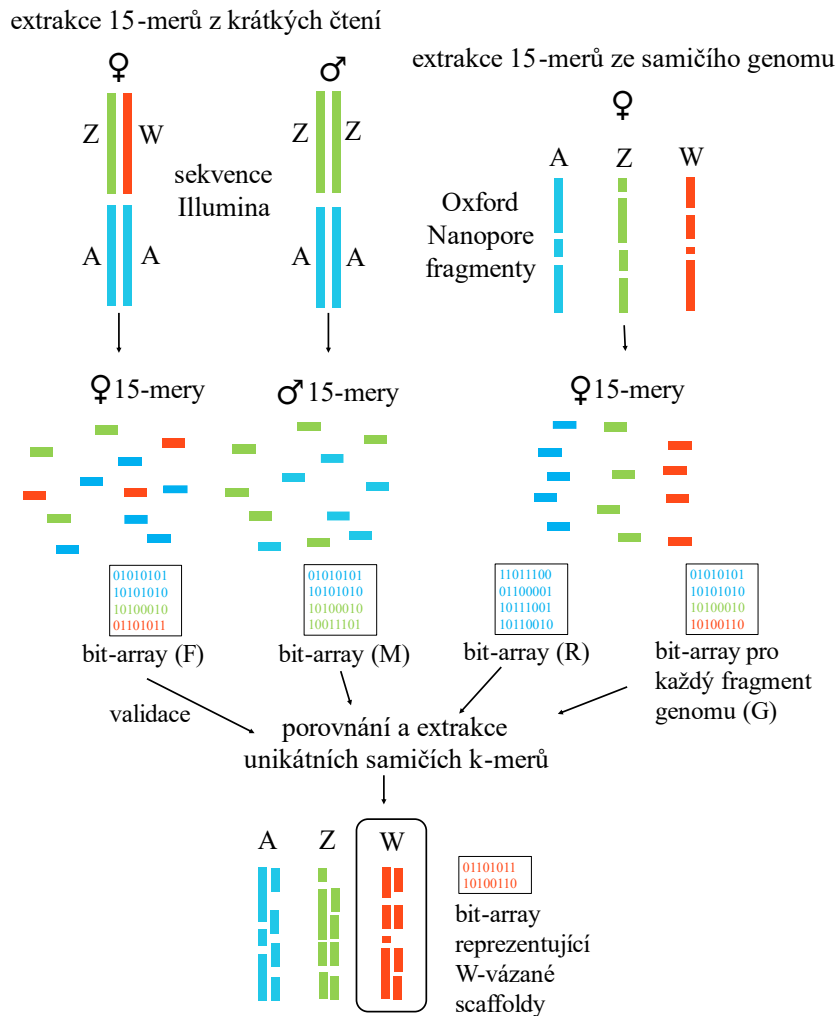
Soubory se sekvenačními daty získanými technologií Illumina (F1, F2, M1, M2, M3, viz **Tab. I**) ve formátu FASTQ byly extrahovány a všechny soubory s datovými soubory pro stejného jedince sloučeny příkazem „cat“. Kvalita sekvenačních čtení byla ověřena nástrojem FastQC-0.11.5 (Andrews 2010), (**git/10_FastQC**).

3.3.2. Rozdělení sekvencí na k-mery

V souladu s postupem v publikované pipeline YGS (Carvalho a Clark 2013) byl k filtrování krátkých čtení a jejich rozložení na k-mery použit program Jellyfish-1.1.11 (Marçais a Kingsford 2011), (**git/11_Jellyfish**). Byl použit příkaz „jellyfish dump“, jehož výstupem je seznam všech k-merů a jejich počet, s doporučenou délkou k-meru pro hmyz $k = 15$. Minimální kvalita sekvencí byla stanovena na 20 a minimální frekvence k-meru byla 3. Výstupem tohoto kroku byly soubory ve formátu FASTA obsahující všechny 15 mery nalezené v daném jedinci a jejich počet.

3.3.3. YGS

Primárním přístupem k naplnění cílů bylo využití publikovaného postupu Y chromosome Genome Scan (YGS) pipeline (Carvalho a Clark 2013), viz **kapitulu 1.4**. Bioinformatický postup („pipeline“) YGS slouží k identifikaci sekvencí specifických pro Y chromozom, tedy nacházející se u samců. Pro ověření tohoto přístupu u motýlů, kde je heterogametickým pohlavím samice, byla **pohlaví v publikovaných skriptech zaměněna**. YGS je založena na přímém porovnávání krátkých slov DNA (k-merů) genomu heterogametického pohlaví s krátkými ready z opačného pohlaví, přičemž odstraňuje všechny typy identických opakování. YGS pracuje s bitovými poli (bit-arrays), která mohou kompaktně uchovávat přítomnost (kódovanou jako „1“) nebo nepřítomnost („0“) každého k-meru. Každý k-mer a jeho reverzní komplement považuje za shodný a ukládá ten, který je abecedně první. Jednotlivé kroky pipeline YGS jsou znázorněny na **Obrázku 4**.



Obrázek 4: Hlavní kroky YGS pipeline. Postup zahrnuje extrakci k-merů z krátkých čtení, extrakci repetitivních k-merů z genomických fragmentů a porovnání mezi genomickými k-mery a k-mery z krátkých čtení. Zkratky použité v bitových polích (bit-array): F = samice, M = samec, R = repetice, G = genom. Sekvence pro validační krok je z jiné samice, než ze které byl poskládán genom. Nejprve bylo vytvořeno bitové pole pro 15 mery přítomné v krátkých čteních **samců („M“)** a bitové pole reprezentující 15 mery v krátkých čteních **samic („F“)**, která sloužila pro validaci. Další bitové pole bylo vytvořena pro 15 mery, které se v sestaveném samičím genomu vyskytovaly více než jednou, **opakované k-mery**, uložené v bitovém poli („R“). Čtvrté bitové pole bylo získáno ve finálním kroku pro každý fragment sestaveného **genomu („G“)**. Ta byla porovnána s bit-array F, čímž byly získány validované samičí 15 mery, které byly následně porovnány s bit-arrays M a R, čímž byl pro každý fragment genomu získán podíl unikátních samičích k-merů, které nemají shodu se samčími čteními (Zhang et al. 2017, upraveno).

Skript YGS.pl (Carvalho a Clark 2013) je napsán v programovacím jazyce Perl. Postup zahrnuje tři hlavní kroky, které běží samostatně.

1. YGS.pl mode trace. V prvním kroku byla vytvořena bitová pole (bit-arrays), reprezentující 15 mery přítomné v krátkých čteních samců „M“ a také samic „F“ (**git/12_YGS_trace**).

2. YGS.pl mode contig. V této části byly extrahovány 15 mery z fragmentů reprezentujících samičí genom, získány složením dlouhých sekvenčních čtení (viz **kapitulu 3.2.3**). Tento krok slouží k získání těch k-merů, které se v sestaveném samčím genomu vyskytovaly více než jednou. Tyto repetitivní k-mery byly uloženy v bitovém poli „R“, (**git/13_YGS_contig**).

3. YGS.pl mode final run. V tomto kroku bylo získáno třetí bitové pole pro každý fragment sestaveného genomu „G“. Logická operace $G \text{ NOT } R$, poskytuje bitové pole s unikátními 15 mery pro fragmenty genomu („GSC“) a s použitím bit-array „F“ je díky operaci „GSC AND F“ získán soubor unikátních validovaných 15 merů vyskytujících se v jedné kopii („VSC“) v krátkých čteních samic. Následně je pomocí operace „VSC NOT M“ dosaženo porovnání sestaveného genomu s krátkými čteními samců. Tento validační krok je podle autorů volitelný, ale poskytuje jasnější oddělení mezi fragmenty DNA vázanými na chromozom W a těmi ostatními. Vedle tohoto validačního kroku je součástí pipeline logická operace „GSC NOT M“ poskytující porovnání sestaveného samičího genomu a krátkých čtení sameců (**git/14_YGS_finalrun**). Výstupem kroku final run je tabulka s hodnotami uvádějícími pro každý genomový fragment například počet různých 15 merů, počet chybějících u homogametického pohlaví, počet unikátních, a také % validovaných pohlavně specifických 15 merů, tedy chybějících v druhém jedinci heterogametického pohlaví.

YGS pipeline byla šestkrát zopakována pro všechny dostupné kombinace dat, aby byla zvýšená pravděpodobnost, že odhalené W-specifické sekvence nebudou falešně pozitivní.

3.3.4. Zpracování výsledků YGS

Výsledkem výše uvedeného postupu bylo šest tabulek, z nichž každá uváděla podíl unikátních samičích 15 merů (z F1 nebo F2) pro každý fragment samičího genomu (F_ONT), které nemají shodu s k-mery získanými ze samčích čtení (M1, M2, M3). Následně byl z každé tabulky exportován sloupec s názvem scaffoldu či contigu a příslušné procento validovaných unikátních 15 merů (v originále P_VSC_UK: percentage of valid single-copy k-mers unmatched by the male reads), (**git/15_proceeding**). Pro další analýzu byly vybrány pouze ty fragmenty DNA chybějící v samčích vzorcích ve všech provedených kombinacích. Jako průnik všech provedených analýz byl vypočítán průměr hodnot P_VSC_UK pro každý

fragment. Výsledkem těchto kroků byla finální tabulka zobrazující výsledky porovnání genomických sekvencí všech samců a se všemi samicemi.

Z této tabulky byl na základě **průměrů hodnot P_VSC_UK** vytvořen histogram v programu RStudio (Boston, USA) s využitím balíku pro vizualizaci dat ggplot2 (Wickham 2016). Histogram zobrazil míru mapování samčích čtení na samičí fragmenty z YGS (**git/16_histogram**).

Podle Zhang et al. (2017) lze za potenciálně pohlavně specifické považovat ty scaffoldy, které měly proporcí nenamapovaných samčích k-merů větší než 70 % a je doporučeno jejich ověření metodou PCR. V této práci byly hodnoty P_VSC_UK filtrovány s hranicí 75 % napříč všemi analýzami, čímž byly identifikovány fragmenty, které obsahovaly sekvence potenciálně specifické pro chromozom W. Filtrování bylo provedeno v tabulkovém procesoru Microsoft Excel (Office 16, Microsoft® Excel® pro Microsoft 365 MSO, Version 2108, Build 14332.20145).

YGS postupem byl získán seznam identifikátorů (fragmentů assembly F_ONT), které mají vysoké procento samičích 15 merů (téměř se nevyskytují v krátkých sekvenčních datech M1, M2 a M3), a jsou tedy potenciálně specifické pro samice. Sekvence těchto fragmentů byla získána z původního assembly (F_ONT) a následně namapována na samčí genom z dalšího jedince, který v dosavadní analýze nebyl zahrnut (M_ONT; Volenikova et al. 2022) s využitím programu **BLAST** (Altschul et al. 1990). Byla nastavena poměrně relaxovaná hodnota e-value (10^{-20}), tedy počet očekávaných shod podobné kvality, které by mohly být nalezeny pouhou náhodou. Ty fragmenty, jež měly mezi pohlavími shodu větší než 90 % byly z dalších analýz vyloučeny, čímž byl získán seznam sekvencí, které nemají shodu ani v samčím genomu dalšího jedince, a tudíž lze očekávat, že jsou vázané na chromozom W. Byl vytvořen soubor s kandidátními samičími fragmenty ve formátu FASTA, (**git/17_BLAST_proceeding**).

Následně byly v těchto kandidátních sekvencích zamaskovány repetice s využitím RepeatMasker (repeatmasker.org). Tento nástroj byl použit s výchozím nastavením a jako referenční knihovna DNA byla zvolena *Drosophila melanogaster*.

Protože YGS pracuje s bitovými poli, jsou ztraceny informace o konkrétních použitých k-merech. Jako další krok ověření byly z výstupů Jellyfish získány k-mery unikátní pouze pro samice (**git/18_F_kmers**), které byly namapovány na samičí assembly (F_ONT) s využitím nástroje **bowtie2-2.2.9** (Langmead a Salzberg 2012) a **samtools-1.3.1** (Li et al. 2009) (**git/19_Bowtie2**). Byl tak získán soubor ve formátu BAM, na který byly v programu **Geneious Prime 2022.0.2** (Biomatters, Nový Zéland) nástrojem „Map to reference“

namapovány také kandidátní W-specifické fragmenty získané postupem uvedeným výše. Takto byly identifikovány oblasti s hojným výskytem samičích 15 merů.

Kandidátní fragmenty po zamaskování RepeatMasker, které byly identifikovány jako samičí postupem YGS, a které byly bohaté na samičí k-mery, byly namapovány na referenční samčí genom M_ONT pomocí nástroje Map to reference v programu Geneious. Byl použit standardní assembler Geneious s výchozím nastavením. Sekvence, u kterých byla nalezena shoda byly vyřazeny, a výběr kandidátních W-specifických oblastí tak byl tímto krokem ještě zúžen.

3.4. Navržení PCR primerů pro ověření kandidátních sekvencí

Na kandidátní oblast, dle předpokladů specifickou pro chromozom W, byly navrženy PCR primery s využitím programu Geneious. Kritéria pro navržení primerů byla nastavená na délku 20 bází, 0 offtargets, 50-55% CG, Hairpin T_m a Selfdimer T_m pod 40 °C a velikost produktu 400-600 bp. Primery, které byly softwarem navrženy byly in silico ověřeny mapováním jak na samičí, tak na samčí genom, skrze nástroj Map to reference v Geneious.

Navrženy byly také primery na **geny pro vnitřní kontrolu**. Jako referenční autozomální geny, které se v genomu vyskytují u obou pohlaví pouze v jediné kopii, byly použity *acetylcholinesteráza-1 (Ace1)* (ncbi.nlm.nih.gov/gene/692783) a *acetylcholinesteráza-2 (Ace2)* (ncbi.nlm.nih.gov/gene/692776). Sekvence těchto primerů byly staženy z databáze NCBI (National Center for Biotechnology Information, USA) pro druh *Bombyx mori* a jejich přítomnost u samice i samce *Yponomeuta cagnagella* byla ověřena v programu Geneious. Gen *Ace1* byl jako vnitřní kontrola použit u studovaného druhu již v bakalářské práci autorky (Pílíková 2019) a to jako vnitřní standard pro qPCR. Jako další referenční gen byl použit *elongační faktor 1 alpha, (EF1-a)* získaný od Anny Voleníkové, který byl jako vhodný autozomální referenční gen ověřen i u jiných druhů (Cho et al. 1995; Carabajal Paladino et al. 2019). Primery navržené pro vnitřní kontrolu byly in silico otestovány společně s ostatními navrženými primery, aby byla ověřena jejich komplementarita a vyloučena tvorba dimerů pomocí softwaru Multiple Primer Analyzer (ThermoFisher Scientific, USA, dostupné z: thermofisher.com/cz/en/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library/thermo-scientific-web-tools/multiple-primer-analyzer). Teplota nasedání primerů byla kromě programu Geneious ověřena také nástrojem T_m Calculator (ThermoFisher Scientific, USA, dostupné z: [20](https://thermofisher.com/cz/en/home/brands/thermo-scientific/molecular-biology/molecular-</p></div><div data-bbox=)

biology-learning-center/molecular-biology-resource-library/thermo-scientific-web-tools/tm-calculator).

Primery byly syntetizovány firmou Generi Biotech s.r.o. (Hradec Králové, ČR). Přehled primerů navržených na oblasti W chromozomu je uveden v **Tabulce II** a přehled referenčních primerů v **Tabulce III**.

Tabulka II: Primery navržené pro amplifikaci části chromozomu W.

Název primeru	Sekvence primeru	Amplifikovaný fragment	Velikost produktu (bp)
Ycag-1-F	TATGTCTGCCGAATACGCCA	contig_24514	643
Ycag-1-R	TTCAAGGCGGGATTTCGGATT		
Ycag-2-F	TCCGAATCCCGCCTTGAAAA	contig_24514	600
Ycag-2-R	GCGGGTTACATGTTTGGCAG		
Ycag-3-F	GAGTGTCGACAGCTGCAGTA	contig_24514	672
Ycag-3-R	CCCACATGCGCTTTCTACCT		
Ycag-4-F	GCTGACGCATACATTCGACG	contig_24514	433
Ycag-4-R	ACTAGTCAAAGCGGGCTTGA		
Ycag-5-F	TCAAGCCCGCTTTGACTAGT	contig_24514	413
Ycag-5-R	GGGCCGACTGGGTATGTTTT		
Ycag-6-F	GTGTTTGTGTGAGTGTCCGC	contig_16238	311
Ycag-6-R	TAGAGAGTGCAGGTGGTCA		
Ycag-7-F	AGAGGTAGGCAGAGACCAGG	contig_24139	444
Ycag-7-R	ATGTGATATCGGGCGGCATT		
Ycag-8-F	GAGGTGAAACTGGCTGACGA	contig_3883	340
Ycag-8-R	AAACATTGTTGGTGAGCGGC		

Tabulka III: Přehled primerů navržených na referenční autozomální geny.

Název primeru	Sekvence primeru	Velikost produktu (bp)
Ace1-F	CCACCTCCTCTGGCGAATAAT	121
Ace1-R	GATACGCCGAGACTGGCAAC	
Ace2-F	TTTGCTCGATCTTGGGCCAA	542
Ace2-R	ACGTCGCTTGTAGATGCCAT	
EF1-a-F	CACATYAACATTGTCGTSATYGG	541
EF1-a-R	CATRTTGTCKCCGTGCCAKCC	

3.5. PCR ověření kandidátních markerů

Všechny primery byly otestovány na vhodné nasedací teploty teplotním gradientem 55-65 °C, neboť se doporučené teploty velmi lišily mezi dodavatelem Generi Biotech sro., analyzátořem ThermoFisher Scientific a softwarem Geneious, kde byly primery navrženy. Ze dvou testovaných polymeráz - ExTaq (TaKaRa, Japonsko) a OneTaq (New England BioLabs, USA) se jako optimální ukázala být druhá jmenovaná. Byla testována i různá koncentrace DNA (10-90 ng/μl), přičemž se nejlépe osvědčilo množství 50 ng na reakci.

Na PCR ověření primerů byla použita reakční směs o objemu 20 μl se složením: 50 ng DNA, 1x OneTaq Quick-Load Reaction Buffer, 1 U OneTaq Quick-Load DNA Polymerase (New England BioLabs, USA); 10 μM forward a 10 μM reverse pro každý testovaný primer a 2,5 mM dNTPs (New England BioLabs, USA).

Pro PCR byl použit termocycler Biometra TAdvanced (Analytik Jena, Německo) a reakce měla následující teplotní profil: počáteční denaturace při 94 °C (3 minuty), 25 cyklů: denaturace při 94 °C (30 sekund), nasedání primerů při odpovídající teplotě (30 sekund) a elongace při 68 °C (1 minuta na 1000 bází produktu). Následovala finální elongace při 68 °C (3 minuty) a zchlazení na 4°C.

DNA použitá pro PCR ověření byla získána z desítek jedinců obou pohlaví z české i nizozemské populace, a to buďto izolací pomocí kitu NucleoSpin DNA Insect (Macherey-Nagel, Německo) dle protokolu výrobce, nebo pomocí cetyltrimetylamoniumbromidu (CTAB), podle postupu uvedeného v diplomové práci Kreklové (2020). Část vzorků byla získána během bakalářské práce autorky (Píliková 2019), další vzorky byly izolovány stejnými metodami v letech 2021-2022. Páry primerů byly vždy testovány na dvou různých jedincích od obou pohlaví a obou geografických populací.

Výsledek PCR byl ověřován gelovou elektroforézou. Vzorky a velikostní standard PCR BIO Ladder i (PCR Biosystems, Velká Británie), v některých případech Gene Ruler 1 kb DNA Ladder (ThermoFisher Scientific, USA), byly nanoseny na 1% agarózový gel v 1x TAE pufru (Tris-Acetate-EDTA: 40 mM Tris, 20 mM kyseliny octové, 1 mM EDTA; pH 8). Fragmenty DNA byly následně barveny v 0,05% roztoku ethidium bromidu po dobu 15 minut a zobrazeny v UV světle dokumentačním zařazením AZURE 200, nebo AZURE 280 (Azure Biosystems, USA).

4. Výsledky

4.1. Assembly genomu

Kukla samice předivky brslenové, ze které byla izolována DNA, vážila 49,5 mg a koncentrace vyizolované DNA byla podle přístroje Qubit 26,7 ng/μl a dle stanovení na spektrofotometru Nanodrop byla DNA čistá bez proteinů nebo dalších kontaminantů ($260/280 = 1,87$; $260/230 = 1,82$). Šlo o dlouhé molekuly DNA ($> 10\,000$ bp), což bylo ověřeno elektroforeticky. Po přípravě sekvenační knihovny byla koncentrace DNA dle přístroje Qubit 44,8 ng/μl. Na flow cell bylo nanášeno 12 μl, tedy 537,6 ng = 22 fmol. Sekvence probíhala celkem 52 hodin. Bylo získáno 12,5 Gb dat, z toho čtení delší než 10 kbp tvořila 7,2 Gb. Délka nejkratší sekvence, která pokrývá alespoň 50 % celkové délky genomu readů (N50) odpovídala 16 635 bp, nejdelší sekvence dosahovala 430 859 bp. Velikost samčího genomu *Yponomeuta cagnagella* je 710 Mbp (Volenikova et al. 2022), což znamená, že genom byl pokryt (coverage) přibližně 17×. Po basecallingu byla kvalita sekvencí ověřena programem Nanoplot. Průměrná délka čtení byla 4381,9 bp a průměrné skóre kvality bylo 11,1.

První verze (v1) genomu byla sestavena bez filtrování krátkých sekvenačních čtení a s polishing procedurou pouze programem Flye, viz **kapitolu 3.2.3**. Tento genom byl dále používán pro k-merovou analýzu jako soubor F_ONT, viz **kapitolu 3.3.3**. Následně byl ze stejných ONT sekvenačních čtení sestaven genom znovu, viz skripty (**git/5-9**). Bylo využito filtrování krátkých sekvencí, a polishing programem Canu i Flye a vznikly tak dvě další verze genomu – jeden u něhož byla eliminována sekvenační čtení kratší než 5000 bp (v2) a u dalšího chyběla čtení kratší než 10000 (v3). Komplexita assembly byla vyhodnocena programem BUSCO. Použití BUSCO zároveň s modulem Augustus, což je genový prediktor patří k nepřesnějším programům svého druhu (Manni et al. 2021), nepřineslo komplexnější výsledky. Nevětší komplexity bylo dosaženo s referenčním datasetem endopterygota_odb10 (**git/04_BUSCO1**), a to pro první verzi genomu, viz **Tab. IV** a **V**. Programem BUSCO se stejným nastavením byl ohodnocen i genom M_ONT (Volenikova et al. 2022).

Tabulka IV: Výsledky z programu BUSCO. V tabulce jsou uvedeny počty ortologních genů BUSCO (endopterygota_odb10) bez použití modulu Augustus pro složený genom *Y. cagnagella* (v1, v2, v3 a M_ONT [Volenikova et al. 2022]).

Parametr	F_ONT (v1)	(v2)	(v3)	M_ONT
kompletních BUSCO genů (C)	2 057	2 004	1 984	2 028
kompletních BUSCO genů v jedné kopii (S)	1 569	1 673	1 565	1 826
kompletních duplikovaných BUSCO genů (D)	488	331	420	202
fragmentovaných BUSCO genů (F)	41	74	81	57
chybějících BUSCO genů (M)	26	46	59	39
prohledaných BUSCO skupin celkem	2 124	2 124	2 124	2 124

Srovnání výsledků všech BUSCO analýz ukazuje **Tabulka V.** Pro všechny verze assembly genomů bylo lepších výsledků dosaženo při použití referenčního datasetu endopterygota bez modulu Augustus.

Tabulka V: Srovnání výsledků z programu BUSCO v procentech.

Assembly	Referenční BUSCO dataset	Výsledek
F_ONT (v1)	endopterygota	C: 96,9 % [S: 73,9 %; D: 23,0 %]; F: 1,9 %; M: 1,2 %, n: 2 124
(v2)	endopterygota	C: 94,4 % [S: 78,8 %; D: 15,6 %]; F: 3,5 %; M: 2,1 %; n: 2 124
(v3)	endopterygota	C: 93,4 % [S: 73,6 %; D: 19,8 %]; F: 3,8 %; M: 2,8 %; n: 2 124
M_ONT	endopterygota	C: 95,5 % [S: 86,0 %; D: 9,5 %]; F: 2,7 %; M: 1,8 %; n: 2 124

Kvalita assembly byla ohodnocena také programem Quast, z jehož výsledků lze vidět, že první verze genomu byla sestavena do N = 22107 contigů a N50 dosahovala 124 kbp. Genom sestavený ze čtení delších než 5000 bp (v2) byl uspořádán do N = 15513 contigů (N50 = 165 kbp) a verze sestavená ze čtení delších než 10000 bp (v3) se skládala z N = 10701 contigů a hodnota N50 byla 178 kbp. Publikovaný genom M_ONT měl nejméně contigů (N = 11779) a nejnižší hodnotu N50, která dosahovala 96486 bp, viz **Tab. VI.**

Tabulka VI: Vyhodnocení vlastností poskládaného genomu *Y. cagnagella* programem Quast (v1, v2, v3 a M_ONT).

Parametr	F_ONT (v1)	(v2)	(v3)	M_ONT
contigů celkově	22 107	15 513	10 701	11 779
contigů \geq 50000	4 446	3 656	4 081	4 248
nejdelší contig (bp)	6 584 164	4 634 531	2 864 924	931 231
celková délka (bp)	799 750 997	735 953 823	788 679 736	626 271 432
N50 (bp)	124 538	165 257	178 975	96 486
neznámých, N (bp)	1 900	4 100	2 600	3 545
neznámých na 100 kbp	0,24	0,56	0,33	0,57

Ze srovnání uvedeného v **Tabulkách IV, V a VI** je patrné, že filtrování dat od krátkých sekvenačních čtení a následné polishing procedury (oprava čtení mezi sebou na základě podobnosti) programem Canu nepřinesly větší komplexitu sekvence genomu. Genomové verze 2 a 3 byly sestaveny do menšího množství contigů a dosáhly vyšší hodnoty N50 (viz **Tab. VI**), avšak genom bez úprav měl menší počet chybějících BUSCO genů (viz **Tab. IV**).

4.2. Výsledky YGS

K-merovou analýzou byl stanoven celkový počet 15 merů v genomu F_ONT, a to 202 087 576, z toho unikátních 90 859 741 (45,0 %) a repetitivních 111 227 835 (55,0 %)

Každé opakování k-merové analýzy postupem YGS (viz **kapitolu 3.3.3**) mělo za výsledek tabulku o jedenácti sloupcích pro každý fragment F_ONT genomu z celkových 22149, které byly detekovány, viz **Tab. VII**.

Tabulka VII: Ukázka části tabulky získané krokem final-run. Uvedený výběr reprezentuje výslednou tabulku pro jedince M3 (trace) a F1 (male_trace), dále jako M3F1. GI: název sekvence ve FASTA souboru; NUM: číselný index sekvence; MAX_K: maximální počet k-merů; K: počet k-merů nalezených v dané sekvenci; UK: počet k-merů bez shody se samčými čteními; SC_K: počet unikátních k-merů nalezených v sekvenci; SC_UK: počet unikátních k-merů bez shody se samčými čteními; P_SC_UK: procento unikátních k-merů bez shody se samčými čteními; VSC_K: počet validovaných unikátních k-merů; VSC_UK: počet validovaných unikátních k-merů bez shody se samčými čteními; P_VSC_UK: procento validovaných unikátních k-merů bez shody se samčými čteními.

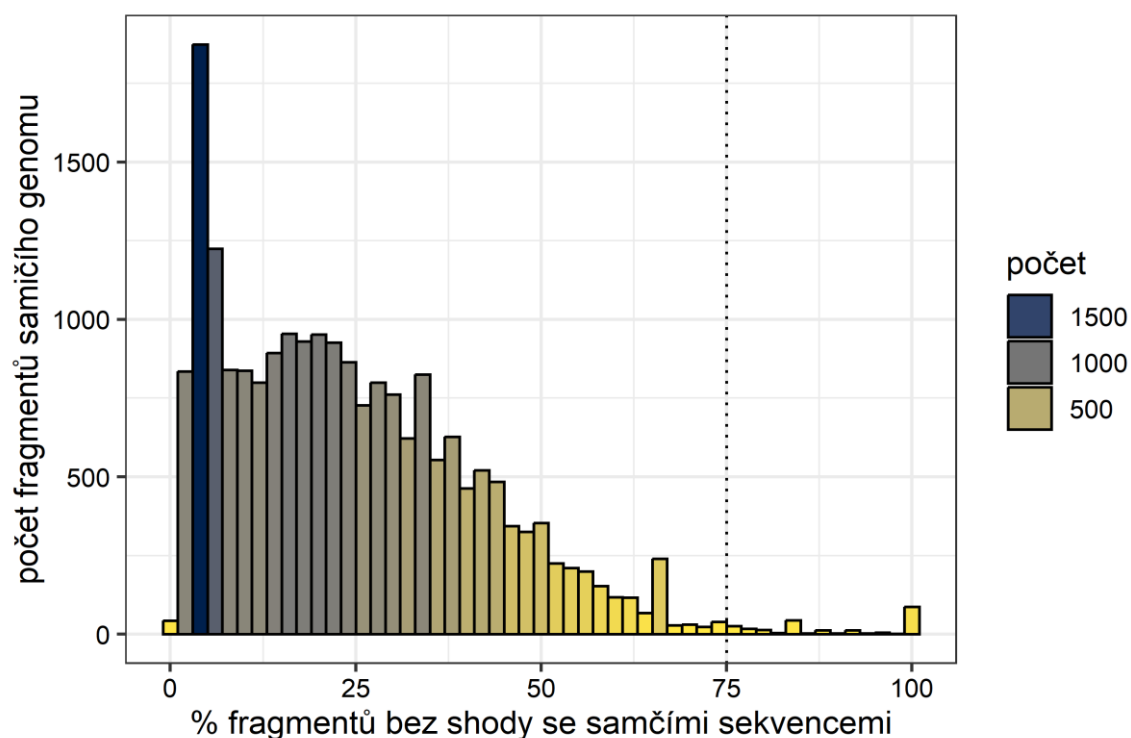
GI	NUM	MAX_K	K	UK	SC_K	SC_UK	P_SC_UK	VSC_K	VSC_UK	P_VSC_UK
contig_1	1	44 736	43 127	780	5 564	375	6,7	5 303	136	2,6
contig_10	2	17038	17021	185	111	87	78,4	20	4	20,0
contig_1000	3	320 793	299 377	8 606	60 627	6 808	11,2	58 175	5 008	8,6
contig_10000	4	20 688	20 479	1 002	992	681	68,6	362	111	30,7
contig_10001	5	40 287	39 587	1 082	1 322	720	54,5	635	164	25,8
contig_10002	6	8 255	8 213	310	302	210	69,5	123	50	40,7
contig_10003	7	11 648	11 610	117	1 825	106	5,8	1 572	17	1,1
contig_10004	8	37 510	36 971	662	4 923	518	10,5	4 402	130	3,0
contig_10006	9	38 334	37 117	876	5 526	708	12,8	4 994	336	6,7
contig_10007	10	35 733	34 987	907	1 497	648	43,3	857	192	22,4

Z těchto tabulek bylo získáno procento validovaných unikátních k-merů bez shody se samčými čteními (P_VSC_UK) pro každé opakování, viz **Tab. VIII**.

Tabulka VIII: Ukázka části finální tabulky získané analýzou YGS. Vynesené hodnoty odpovídají procentu unikátních samičích 15 merů, které byly přítomny na daném fragmentu assembly (P_VSC_UK). V popisku sloupce je zaznamenáno, z jaké analýzy výsledek vychází, např. M2F2 znamená, že jedinec M2 byl použit jako trace a jedinec F2 jako male_trace. V posledním sloupci je průměr hodnot P_VSC_UK z výsledků všech analýz.

fragment	M1F1	M1F2	M2F1	M2F2	M3F1	M3F2	průměr (P_VSC_UK)
contig_1	6,8	4,1	8,7	5,5	2,6	2,6	5,05
contig_10	20	32	25	28	20	24	24,83
contig_1000	7,2	4,2	9,5	6,1	8,6	4,9	6,75
contig_10000	8,3	21,1	16	17,1	30,7	32,1	20,88
contig_10001	25,2	43,4	29,1	28,5	25,8	30,1	30,35
contig_10002	6,5	14,8	37,4	24,6	40,7	23	24,50
contig_10003	7,1	14,3	4,5	7,8	1,1	1,4	6,03
contig_10004	2,4	4,2	7,6	7,8	3	3,3	4,72
contig_10006	5,6	3,5	4,5	4,4	6,7	4,2	4,82
contig_10007	61,7	61,8	12,8	15,9	22,4	12,8	31,23

Z hodnot ve sloupci průměr (P_VSC_UK) byl vytvořen histogram zobrazující distribuci fragmentů F_ONT genomu s ohledem na proporci namapovaných samičích sekvenačních čtení, viz **Obr. 5**.



Obrázek 5: Výsledek YGS: míra mapování samčích k-merů na samičí fragmenty. Genomové fragmenty v levé části grafu odpovídají autozomům a chromozomu Z, protože obsahují malé procento k-merů specifických pro samice, zatímco v pravé části grafu jsou sekvence s vysokým podílem k-merů specifických pro samice, tedy kandidátní W-specifické. Fragmenty dosahující hodnoty $\geq 75\%$ ve všech analýzách byly považovány za specifické pro samice.

4.3. Identifikace kandidátních samičích sekvencí

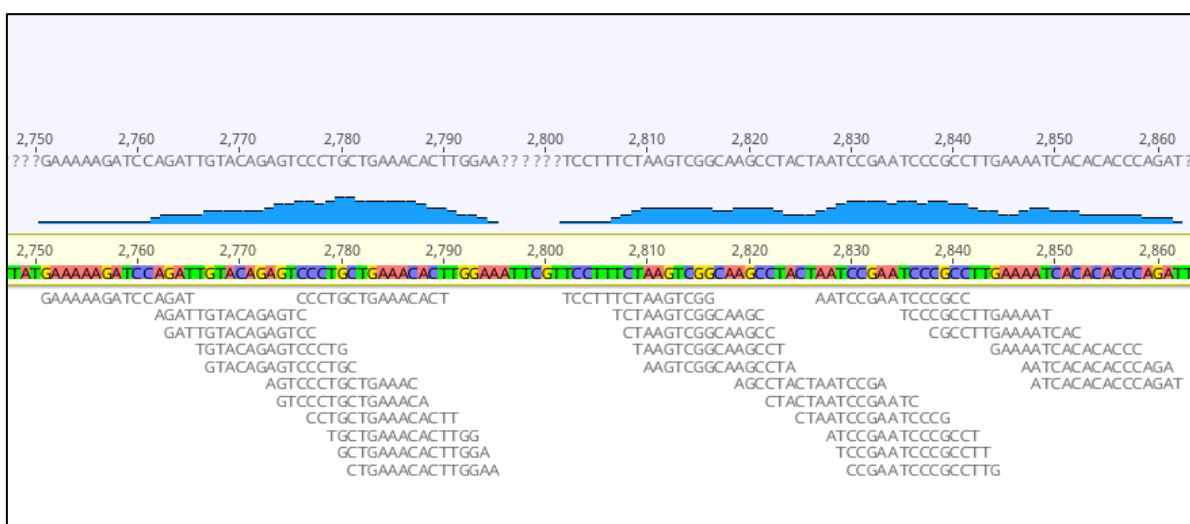
Výsledky YGS umožnily odlišit fragmenty pocházející z autozomů a chromozomu Z, které sdílí stejné k-mery u obou pohlaví, od fragmentů odpovídajících chromozomu W. Z finální tabulky byly získány fragmenty assembly F_ONT, které ve všech analýzách dosáhly hodnoty $P_VSC_UK \geq 75\%$. Postup YGS po kombinaci všech šesti analýz ukázal, že většina chromozomu W je repetitivní a pouhých **31 detekovaných contigů** o celkové velikosti 70569 bp je bohatých na unikátní samičí sekvence, viz **Tab. IX**. Sekvence těchto fragmentů je uvedena jako „YGS_RESULT_W-specific_fragments“ na github.com/aneta-pilikova/w-chromosome-markers.

Pro bližší určení oblastí, které by bylo možné použít jako pohlavně specifické markery, byla YGS pipeline rozvedena o další kroky, které v originálním postupu autoři neuvádějí. Zúžení výběru kandidátních sekvencí bylo dosaženo mapováním FASTA sekvence těchto fragmentů nástrojem BLAST na samčí genom M_ONT, který doposud nebyl v analýze

zahrnut. U 13 fragmentů byla nalezena shoda, a byly proto z analýzy vyřazeny (v **Tab. IX** sloupec „**BLAST M_ONT**“), **18 contigů** bylo považováno za fragmenty specifické pro samice.

Dalším přídatným krokem bylo ověření, zda kandidátní oblasti nejsou známé repetice, které unikly mechanismu eliminace repetice přímo v pipeline YGS. U zbývajících 18 fragmentů byly zamaskovány repetitivní sekvence programem RepeatMasker. Zamaskováno bylo celkem pouhých 3473 (7,04 %) bází z celkových 49329. Jedna z repetice byla identifikována jako retrotranspozon LTR, Ty1/copia (2257 bp), a dále šlo o jednoduché repetice (1128 bp) a oblasti s nízkou komplexitou (88 bp).

Následně byl porovnáním výstupů získaných programem Jellyfish získán soubor 759 421 k-merů specifických pro samice, které byly namapovány na 18 kandidátních fragmentů a mezi nimi bylo **7 oblastí obohacených o samičí k-mery**, tedy potenciálně specifických pro chromozom W (v **Tab. IX** sloupec „**Map F 15 mers**“). Jedna z oblastí obohacených o samičí k-mery je zobrazena na **Obrázku 6**.



Obrázek 6: Ukázka mapování samičích 15 merů na kandidátní oblast o délce zhruba 100 bp. Tato oblast byla považována za potenciálně W-specifickou.

V dalším kroku bylo těchto 7 kandidátních oblastí znovu porovnáno s genomem M_ONT (Voleniková et al. 2022) v programu Geneious nástrojem Map to reference, který využívá jiný algoritmus než BLAST. Byly vyřazeny další 3 oblasti, u kterých byla detekována shoda se sekvencemi v samčím genomu (**Tab. IX** sloupec „**Map M_ONT**“).

Po těchto in silico ověřovacích krocích byly **4 contigy** považovány za samičí markery, viz **Tab. IX**.

Tabulka IX: Proces identifikace pohlavně specifických oblastí pro samice *Y. cagnagella*. Z celkových 31 fragmentů detekovaných jako samičí sekvence v analýze YGS jich bylo 13 vyloučeno po porovnání s M_ONT nástrojem BLAST (označeno hnědou barvou) ve sloupci **BLAST M_ONT**. Dalších 11 fragmentů bylo vyřazeno, protože nebyly obohaceny o samičí k-mery, vyznačeno hnědě ve sloupci **Map F 15 mers**. Pro další 3 contigy byla nalezena shoda při mapování na M_ONT nástrojem Map to reference a jsou v posledním sloupci **Map M_ONT** vyznačeny hnědou barvou. V posledním sloupci Tabulky IX jsou zeleně vyznačeny 4 potenciální samičí markery, tedy fragmenty, o kterých lze předpokládat, že náleží chromozomu W.

#	název fragmentu	délka (bp)	Výsledek YGS – hodnoty P_VSC_UK							průměr	BLAST M_ONT	Map F 15 mers	Map M_ONT
			M1F 1	M1F 2	M2F 1	M2F 2	M3F 1	M3F 2					
1	contig_24514	5086	78,3	78,1	81,6	80,9	86,3	85,7	81,8				
2	contig_24139	1873	100	100	100	100	100	100	100				
3	contig_3883	3155	100	100	100	100	100	100	100				
4	contig_16238	4426	100	100	100	100	100	100	100				
5	contig_18336	950	85,7	81,8	100	90,9	100	90,9	91,5				
6	contig_19260	3493	100	100	100	100	100	100	100				
7	contig_29219	1763	77,8	83,3	88,9	83,3	88,9	100	87				
8	contig_22929	582	100	100	100	100	100	100	100				
9	contig_10146	14217	100	77,8	100	88,9	100	77,8	90,8				
10	contig_14116	1321	100	100	100	100	100	100	100				
11	contig_18699	2132	100	100	100	100	100	100	100				
12	contig_23962	2034	88,9	100	100	100	88,9	100	96,3				
13	contig_27735	1784	75	100	100	100	100	100	95,8				
14	contig_3914	523	100	100	90	87,5	96,7	100	95,7				
15	contig_5723	515	100	100	100	100	100	100	100				
16	contig_6352	1987	100	100	100	100	100	100	100				
17	contig_636	2398	100	100	100	100	100	100	100				
18	contig_22178	1090	100	100	100	80	100	80	93,3				

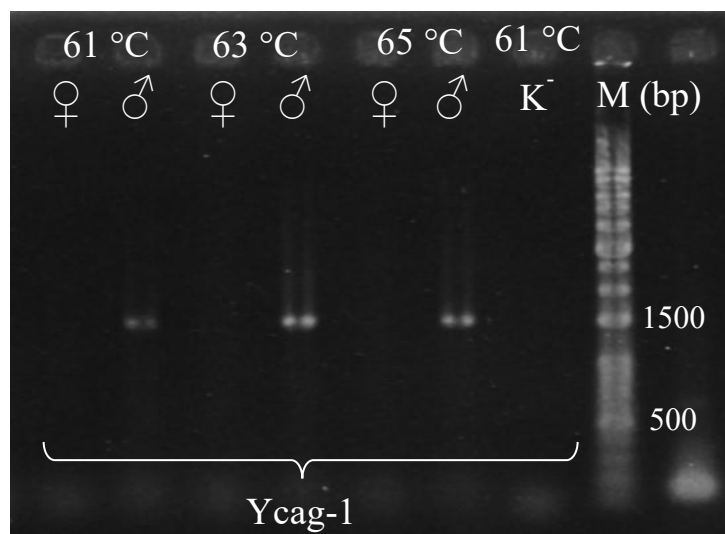
19	contig_14189	621	100	100	100	100	100	100	100			
20	contig_16895	2277	100	100	100	100	100	100	100			
21	contig_222	877	100	100	93,3	100	100	100	98,9			
22	contig_223	873	100	100	100	100	100	100	100			
23	contig_2262	786	100	100	100	100	100	100	100			
24	contig_25573	1301	100	100	100	75	100	75	91,7			
25	contig_27905	3402	100	100	100	100	100	100	100			
26	contig_28198	2181	100	100	100	100	100	100	100			
27	contig_28674	2087	100	100	100	100	100	100	100			
28	contig_3556	1766	100	100	100	100	100	100	100			
29	contig_401	824	100	100	100	100	100	100	100			
30	contig_5407	2004	85,7	100	100	100	100	83,3	94,8			
31	contig_8470	2241	83,3	81,8	75	81,8	75	81,8	79,8			

Kandidátní contig_24514 byl identifikován jako W-specifický s vysokou mírou pravděpodobnosti, protože jde o oblast, která se zdá obohacená u samic v analýze YGS, nevyskytuje se v dostupném samčím genomu a hojně na ni mapují samiči 15 mery. Na tento contig o délce 2542 bp bylo navrženo pět párů primerů (Ycag-1 až Ycag-5). Primery Ycag-6 byly navrženy na contig_16238, Ycag-7 na contig_24139 a Ycag-8 na contig_3883, viz **Tabulku II** v **kapitole 3.4**. Všechny primery byly nejprve otestovány in silico a následně byly ověřeny jako PCR markery pro pohlavně specifické sekvence.

4.4. Ověření PCR markerů

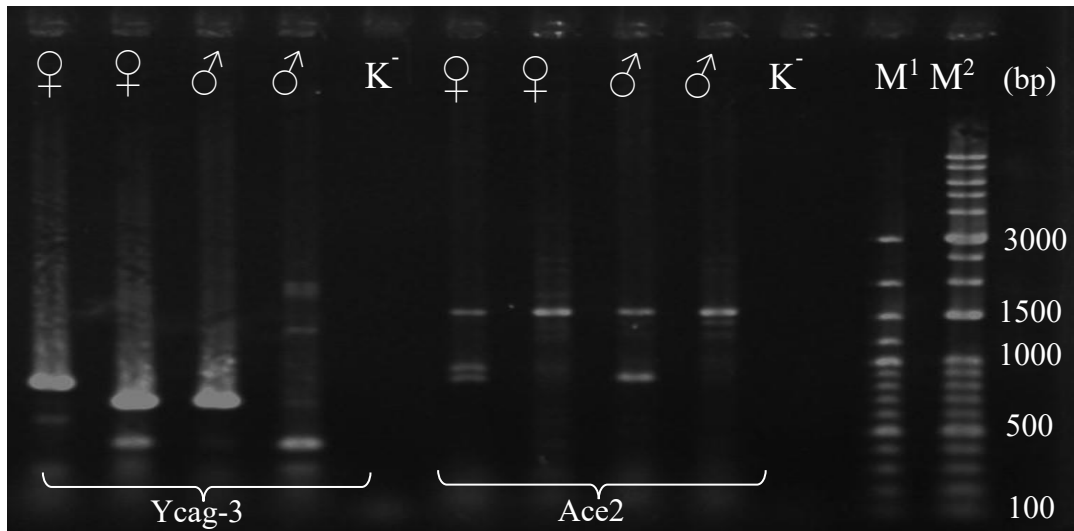
PCR a následná elektroforetická separace byla provedena podle postupu uvedeného v **kapitole 3.5**. Primery byly testovány na různých vzorcích DNA od každého pohlaví a při různých teplotách nasedání, zároveň byly ověřovány primery na části referenčních genů *Ace1*, *Ace2* a *EFl-a*. Přestože byly referenční primery navrhovány tak, aby mohly být společně s testovanými pohlavně specifickými primery použity jako vnitřní kontrola (duplex PCR), experimenty ukázaly, že většinou mají různou teplotu nasedání. Páry primerů byly vždy testovány na dvou různých jedincích od obou pohlaví a obou geografických populací.

Objednané primery se zdály při kontrole nesedacích teplot jako nespecifické, a tvořily produkty i u samčích vzorků, viz **Obr. 7**.



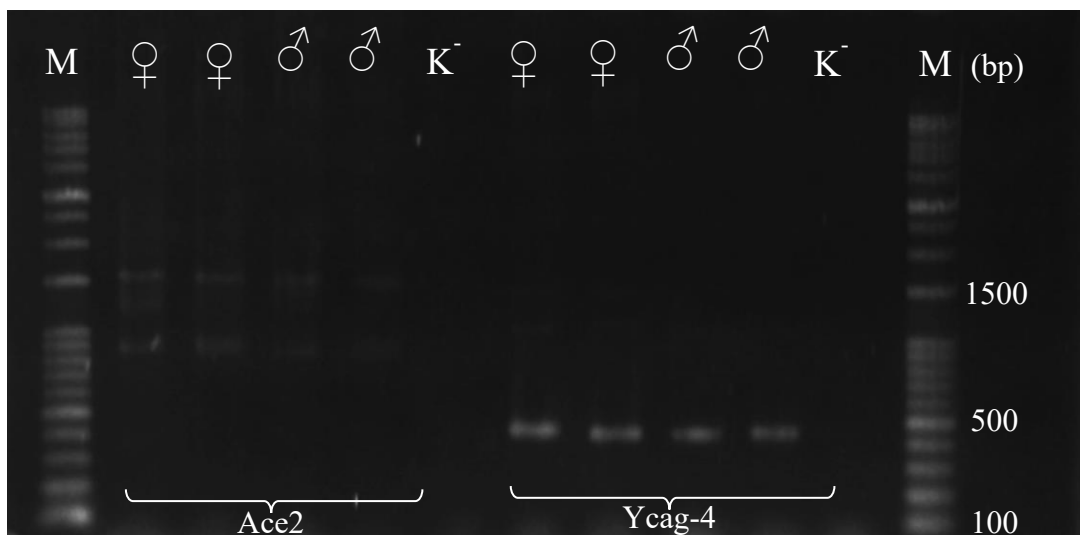
Obrázek 7: Výsledky PCR s primery Ycag-1. Tento pár primerů, navržen na délku produktu 643 bp tvoří nespecifický produkt zhruba o velikosti 1500 bp u samce. Písmenem M je označen velikostní standard PCR BIO Ladder I. K⁻ označuje negativní kontrolu, tedy PCR reakci bez DNA.

Také u páru primerů Ycag-3 vznikaly nespecifické produkty, jak je vidět na **Obrázku 8**. Zároveň se ukázalo, že primery navržené na část genu *Ace2* vytváří zřetelné produkty o velikosti zhruba 1500 bp a další kratší sekundární produkty, které však neodpovídají očekávané délce 542 bp, a ukázaly se tak jako nespecifické. Fragmenty o velikosti 1500 bp byly vždy dobře zřetelné, viz **Obr. 8, 9, 14**.



Obrázek 8: Výsledky PCR s primery Ycag-3 a Ace2. Pár primerů Ycag-3 hybridizuje při teplotě 63 °C a tvoří několik různě dlouhých produktů, ačkoli očekávaná velikost fragmentů byla 627 bp. Primery navržené na referenční gen *Ace2* hybridizují při teplotě 57 °C a u všech jedinců tvoří nespecifické produkty o velikosti okolo 1500 bp a několik dalších kratších. M¹ značí standard Gene Ruler 1 kb DNA Ladder a M² PCRBIO Ladder I.

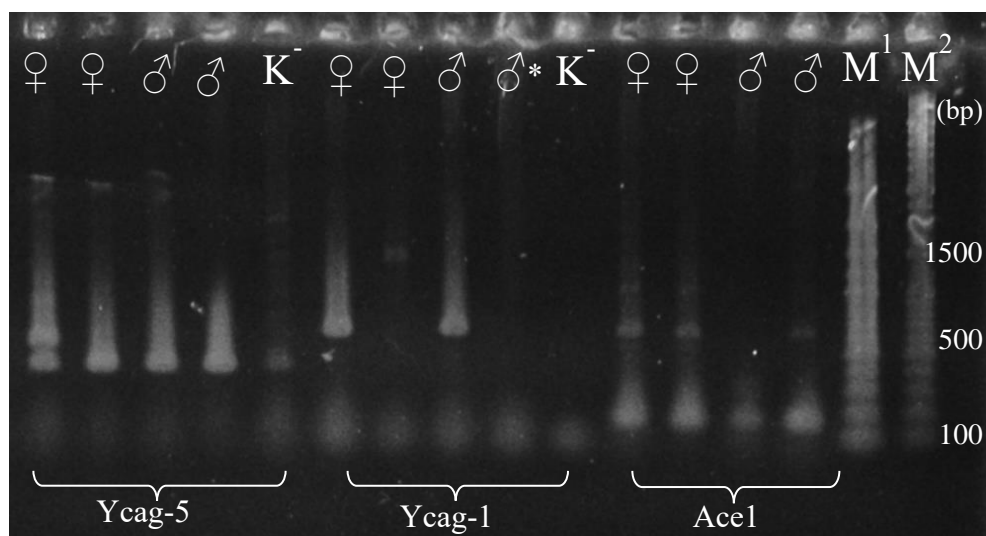
Pár primerů Ycag-4 byl navržen na produkt o velikosti 433 bp. Fragment této velikosti byl detekován jak u samic, tak u samců, viz **Obrázek 9**.



Obrázek 9: Výsledky PCR s primery Ace2 a Ycag-4. Primery pro část referenčního genu *Ace2* vytváří produkty o velikosti 1500 bp a řadu kratších fragmentů u obou pohlaví. Primery Ycag-4 vytváří produkt dlouhý 433 bp také u obou pohlaví.

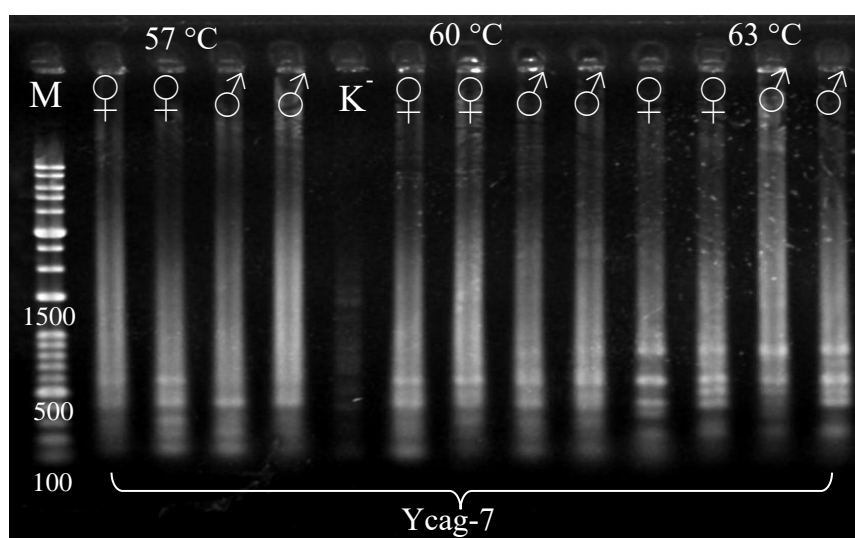
Primery Ycag-5 byly navrženy na sekvenci o délce 413 bází. DNA fragmenty byly detekovány u samců i samic, viz **Obr. 10**. Na témže obrázku lze pozorovat nespecifické

amplikony syntetizované primery Ycag-1, které u samce z nizozemské populace nevytváří žádný produkt. Pro srovnání byly použity primery Ace1, které měly tvořit krátký produkt o délce 121 bp, ale byly pozorovány i fragmenty DNA o jiné velikosti.



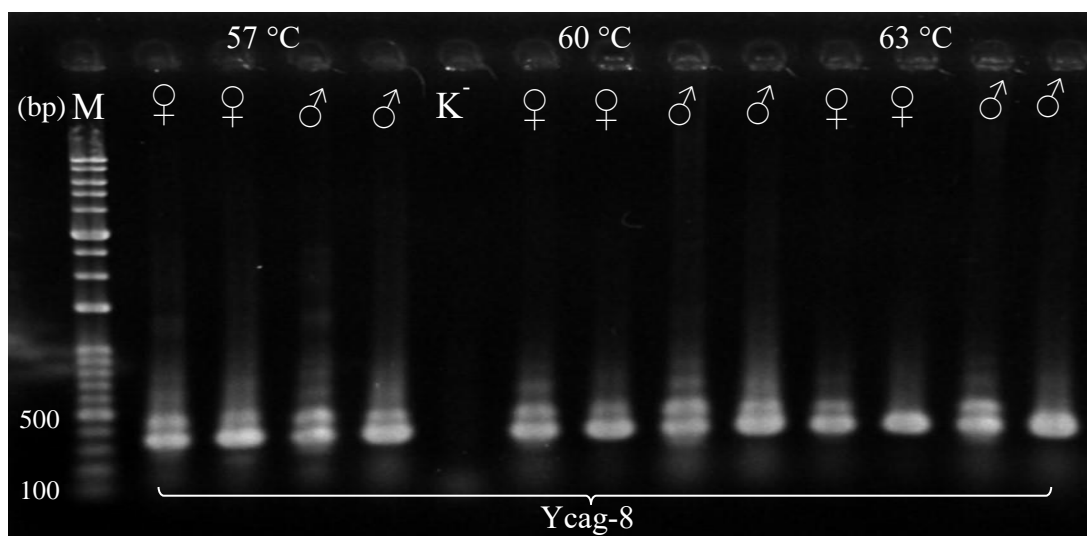
Obrázek 10: Výsledky PCR s primery Ycag-5, Ycag-1 a Ace1. Primery Ycag-5 vyvábí produkty o velikosti okolo 500 bp u všech testovaných jedinců bez rozdílu pohlaví. Primery Ycag-1 také nejsou specifické a vidíme DNA fragmenty u jedné samice a jednoho samce. U druhého samce, který byl z nizozemské populace (*) tento primer neamplifikoval žádnou DNA. Část genu *Ace1* o velikosti 121 bp byla detekována u všech testovaných vzorků, avšak vytváří i sekundární produkty o délce zhruba 700 bp.

Pár primerů Ycag-7 detekoval pravděpodobně úsek satelitní repeticce, a to u všech testovaných jedinců při různých teplotách, viz **Obr. 11**.



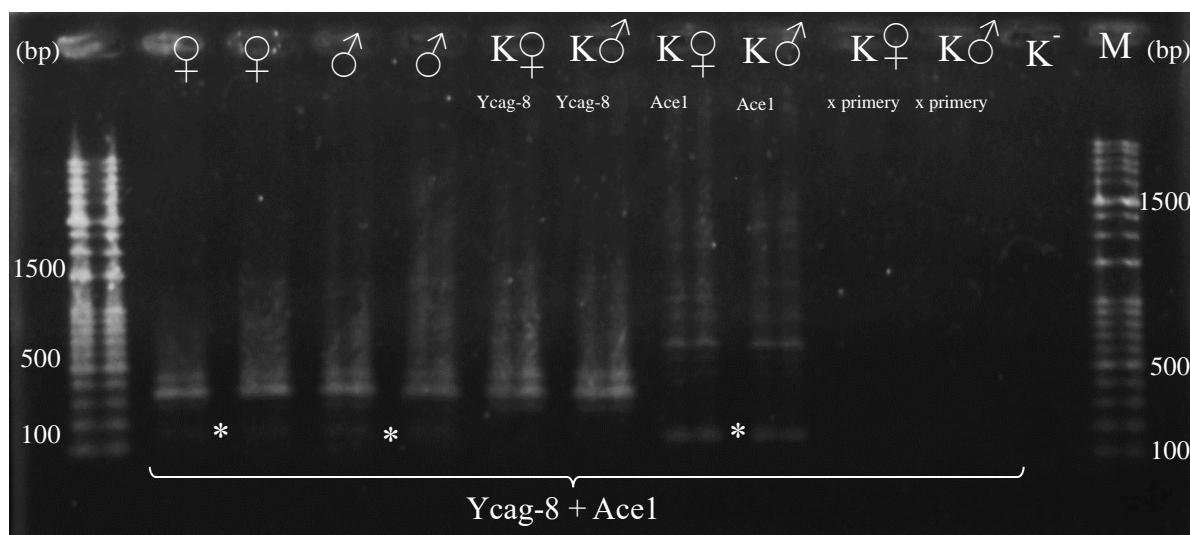
Obrázek 11: Výsledky PCR s primery Ycag-7. Produkty primerů Ycag-7 vytvářejí vzor typický pro repetitivní úseky DNA u obou pohlaví.

Primery Ycag-8 byly navrženy na oblast o velikosti 340 bp. Vedle toho ovšem vznikaly také větší amplikony, asi 450 bp, a to opět u vzorků obou pohlaví, viz **Obr. 12**.



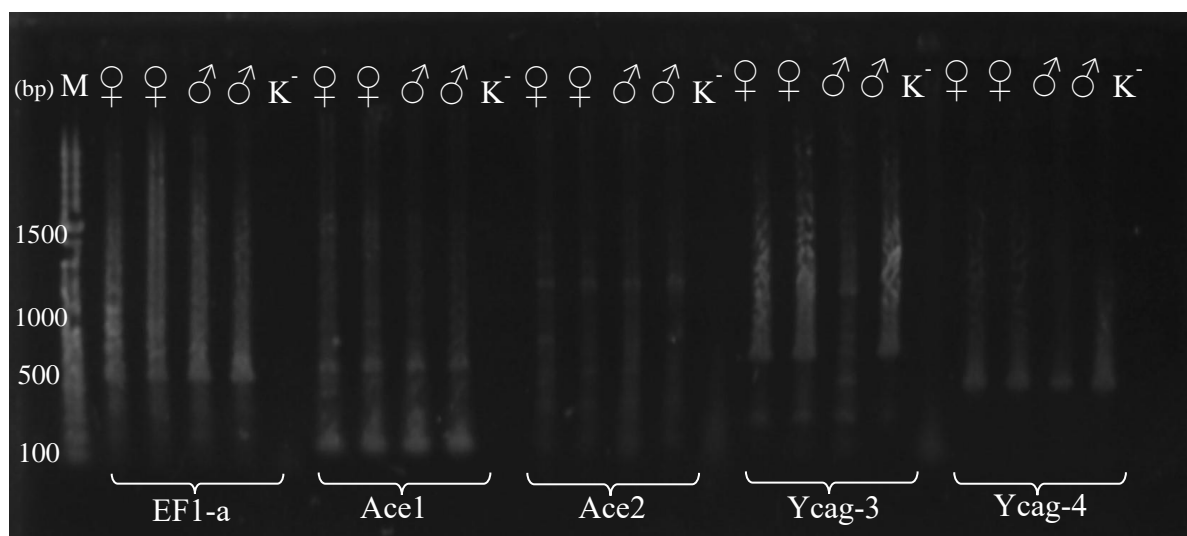
Obrázek 12: Výsledky PCR s primery Ycag-8. Při teplotách 57 °C a 60 °C vznikají dva zřetelné produkty (340 bp a 450 bp) u všech jedinců. Při teplotě 63 °C lze u jedné za samic a jednoho samce pozorovat pouze produkt očekávané velikosti (340 bp), a u zbylých dvou opět dva produkty. Tento primer není pohlavně specifický.

Byla provedena také duplex PCR reakce s primery Ycag-8 a Ace2. Oba páry primerů vytvářely více produktů u samců i samic, viz **Obr. 13**.



Obrázek 13: Výsledky duplex PCR s primery Ycag-8 společně s vnitřní kontrolou Ace1. Testované primery Ycag-8 amplifikují dva zřetelné produkty (340 bp a asi 450 bp) u všech vzorků bez rozdílu pohlaví. Primery Ace1 produkují krátký fragment DNA (121 bp; *), který je méně znatelný u všech jedinců a řadu dalších produktů s velikostí okolo 650 bp.

Srovnání výsledků PCR s referenčními geny a některými ověřovanými primery lze vidět na **Obrázku 14**. Referenční geny nevytváří zřetelné produkty, ale detekují sekvence u všech testovaných jedinců. Primery Ycag-3 a Ycag-4 neamplifikují pohlavně specifické sekvence.



Obrázek 14: Výsledky PCR s primery na detekci genů *EFl-a*, *Ace1* a *Ace2* společně s primery Ycag-3 a Ycag-4. Primery navržené na gen *EFl-a* vytváří produkt o velikosti 541 bp, gen *Ace1* má 121 bp dlouhý produkt a další nespecifický, *Ace2* syntetizuje úsek DNA o délce asi 1500 bp a také kratší fragmenty. U referenčních genů není rozdíl mezi jedinci různých pohlaví. Primery Ycag-3 vytváří několik produktů, které nejsou specifické. U jednoho ze samců vidíme jiný vzor než u ostatních, avšak ani tento primer neidentifikuje pouze oblast samičího genomu. Primery Ycag-4 poskytly stejný výsledek, jako na Obrázku 4, tedy produkty o délce 433 bp vznikají u samic i samců.

Nejsou ukázány všechny výsledky ověřování PCR markerů tedy primerů navržených na oblasti contig_24514, contig_16238, contig_24139 a contig_3883, ale celkový výsledek je prezentován formou **Tabulky X**. Žádný z navržených párů primerů nebyl pohlavně specifický a nelze jej použít pro molekulární určení pohlaví samic.

Tabulka X: Souhrn výsledků ověřování PCR markerů na jedincích obou pohlaví.

Název primeru	Fragment	Velikost produktu (bp)	Výsledek amplifikace
Ycag-1	contig_24514	643	nespecifické produkty o délkách 643 bp a 1500 bp u obou pohlaví
Ycag-2	contig_24514	600	nespecifické produkty o různých délkách u obou pohlaví
Ycag-3	contig_24514	672	nespecifické produkty o různých délkách u obou pohlaví
Ycag-4	contig_24514	433	produkt očekávané velikosti u obou pohlaví
Ycag-5	contig_24514	413	produkt očekávané velikosti u obou pohlaví
Ycag-6	contig_16238	311	nespecifické produkty o různých délkách u obou pohlaví
Ycag-7	contig_24139	444	nespecifické produkty repetitivní povahy o různých délkách u obou pohlaví
Ycag-8	contig_3883	340	dva produkty (340 bp a 450 bp) u obou pohlaví

5. Diskuze

5.1. Sestavení genomu

Modelovým druhem pro motýly je bourec morušový (*Bombyx mori*), domestikovaný druh motýla, který je význačným producentem hedvábí, a také detailně studovaným organismem z hlediska fyziologie, evoluce a genetiky. *B. mori* má 28 chromozomů v haploidní sadě a velikost genomu ~ 450 Mbp (Tong et al. 2022). V roce 2004 vznikla dvě nezávislá genomová assembly *B. mori*, a to pro kmen p50T s coverage 3× (Mita et al. 2004) a pro kmen Dazao s coverage 5,9× (N50 = 12,5 kbp) (Xia et al. 2004), čímž se tento druh stal vůbec prvním motýlem se známou sekvencí genomu. V obou studiích bylo použito whole-genome shotgun (WGS) Sangerovo sekvenování a bylo dosaženo celkového pokrytí genomu přes 90 %. Dalším milníkem bylo assembly *B. mori* sestavené v roce 2008 v rámci mezinárodního uskupení International Silkworm Genome Consortium, kdy autoři obohatili stávající assembly o nová sekvenační data, dosáhli coverage 8,5× a odhalili více než 3000 genů specifických pro bource (International Silkworm Genome Consortium 2008). V dnešní době je bourec morušový zdaleka nejprostudovanějším zástupcem řádu Lepidoptera, jeho genom byl sekvenován na úroveň chromozomů mnoha dostupnými technologiemi (Tang et al. 2021), je známo a anotováno přes 17 000 genů, avšak sekvence chromozomu W ani u tohoto druhu doposud sestavena nebyla (Kawamoto et al. 2022)

Genom druhu *Yponomeuta cagnagella* je tvořen 61 chromozomy u samic a 62 chromozomy u samců. První genom tohoto druhu byl publikován v práci Voleníkové et al. (2022). Jednalo se o assembly sestavené z dat ONT a Illumina stejného samce, který byl v této práci použit jako kontrola M_ONT, viz **kapitulu 4.3**. Autoři stanovili velikost genomu 709.54 ± 6.52 Mbp podle průtokové cytometrie, což lze považovat za přesnější údaj, než výpočet z k-merového spektra, které indikuje velikost genomu 508,47 Mbp. Při porovnání s genomy dalších motýlů je patrné, že genom předivky brslenové (*Y. cagnagella*) je jedním z největších motýlích genomů (Liu et al. 2020). Sekvenováním technologií ONT bylo získáno 26 Gbp sekvence (N50 = 19,1 kbp), což odpovídá hloubce čtení přibližně 37×. Po polishing procedurách dosáhl tento publikovaný sestavený genom N50 96,5 kbp, byl uspořádán do N = 11790 contigů o celkově velikosti 626,3 Mbp, viz také **Tab. VI**. Ohodnocení genomu pomocí nástroje BUSCO s využitím databáze endopterygota, které bylo pro srovnání provedeno se stejným nastavením jako v této práci, ukazuje na 95,5 % kompletních ortologů, 2,7 % fragmentovaných a 1,8 % chybějících BUSCO genů, viz **Tabulku V**.

Výsledné assembly F_ONT sestavené v této práci (viz **kapitolu 4.1**) mělo velmi uspokojivé parametry. Sekvenování bylo provedeno na již použité a promyté průtokové komůrce, která stále měla 1070 funkčních nanopórů a výtěžek činil 7,2 Gbp (N50 = 16,6 kbp). Nejdelší fragment měl délku přes 431 kbp. Každá pozice v genomu byla sekvenátorem pokryta zhruba 17×. Při přípravě sekvenační knihovny nebyly eliminovány krátké fragmenty. Pokud byla následně před samotným sestavením genomu odstraněna krátká čtení (v2, v3) nebyla komplexita výsledného assembly vyšší, viz **Tabulku IV a V**. Genom (v1) při srovnání s databází endopterygota dosáhl komplexity 96,9 %, fragmentovaných BUSCO ortologů bylo 1,9 % a chybějících bylo pouhých 1,2 %. Genom však dosahoval velkého počtu duplikovaných genů (23 %) a byl pokládán do poměrně velkého počtu contigů (N = 22107), což ovšem vzhledem k povaze YGS analýzy nebylo na závadu. Výsledné assembly genomu použité pro další analýzy (v1) mělo délku 800 Mpb a hodnota N50 byla 124,5 kbp (viz **Tab. VI**). Pokud uvažíme, že samice *Y. cagnagella* disponuje 61 chromozomy a velikost genomu je ± 710 Mbp, dostáváme se na hypotetickou délku jednoho chromozomu 11,6 Mbp. Nejdelší získaný scaffold dosahoval 6,5 Mbp, což převyšuje polovinu velikosti tohoto hypotetického chromozomu. Hodnota N50 genomové sekvence získané v této práci přesahuje tyto statistiky z genomů jiných nemodelových druhů motýlů, které byly postaveny na stejné sekvenační technologii ONT, přehled např. v publikacích Triant et al. (2018), Ellis et al. (2021) nebo Volenikova et al. (2022). Genom byl sekvenován v roce 2020 a vzhledem k tomu, že chybělo pouhých 1,2 % ortologů BUSCO (endopterygota), představovalo assembly tohoto samičího genomu *Y. cagnagella* cenný zdroj pro studium tohoto druhu.

Sekvenování před nástupem sekvenačních platforem třetí generace (viz **kapitolu 1.3**) bylo pomalé, velmi finančně náročné, a proto v podstatě omezené na druhy ekonomicky významné. V současnosti dochází ke snižování chybovosti a finanční a časové náročnosti masivně paralelního sekvenování, což společně se širokým spektrem bioinformatických nástrojů činí assembly genomů různých organismů v podstatě rutinní součástí výzkumu a genomická data neustále přibývají. V databázi National Center for Biotechnology Information (NCBI) je pro taxon Lepidoptera dostupných 1 839 genomů, z toho 376 na úrovni chromozomů (ncbi.nlm.nih.gov/data-hub/genome/?taxon=7088, ke dni 8. 4. 2023). Do konce roku 2020 bylo zveřejněných pouze 242 motýlích genomů, z toho 12 na úrovni chromozomů. V pozdější fázi této diplomové práce, v listopadu 2022, byly zveřejněny sekvence genomů *Y. cagnagella* a dalších čtyř příbuzných druhů získaných v rámci projektu **Darwin Tree of Life** (darwintreeoflife.org), který má za cíl sekvenovat kompletní genomy eukaryotních organismů ve Velké Británii a Irsku. V současnosti bylo v rámci tohoto projektu dostupných

701 genomů pro řád Lepidoptera (portal.darwintreeoflife.org/data, 8. 4. 2023). Genom *Y. cagnagella* byl získán ze samičího vzorku pocházejícího z území Velké Británie. Sekvence byla sestavena na základě dat z technologií PacBio a Arima2 Hi-C do 90 contigů, hloubka čtení byla 39× a velikost genomu 575,39 Mbp. Jeho sekvence byla stažena z databáze NCBI (ncbi.nlm.nih.gov/genome/121631?genome_assembly_id%20=%202017806) a dále je označována jako „F_DTOL“. Assembly bylo ohodnoceno stejnou metodou jako samičí genom F_ONT a samčí genom M_ONT, viz **kapitulu 3.2.3**. Assembly F_DTOL obsahovalo v porovnání s databází endopterygota 98,8 % komplexních BUSCO ortologů, fragmentovaných 0,7 % a chybějících 0,5 %. Dostupná je i část chromozomu W, ačkoli 39 sekvencí náležících chromozomu W o celkové délce 132 kbp se nepodařilo lokalizovat. F_DTOL assembly pro druh *Y. cagnagella* je tedy nejlepší dostupnou verzí genomu tohoto druhu.

Přestože lze již standardně dosáhnout assembly na úrovni jednotlivých chromozomů s nesrovnatelně menší mírou úsilí, než na počátku 21. století, sekvence pohlavních chromozomů W a Y jsou při skládání genomu stále problematické (Tomaszkiewicz et al. 2017). I v nejlepších genomových sekvencích motýlů představují sekvence pohlavních chromozomů patrně artefakty (Wan et al. 2019; Carey et al. 2022).

5.2. K-merová analýza YGS

Nástrojem, který byl zvolen pro identifikaci pohlavně vázaných sekvencí v této práci je „Y chromosome Genome Scan“ (YGS) (Carvalho a Clark 2013), viz **kapitulu 1.4**. Vstupními daty pro vývoj YGS byly celogenomové sekvence ze dvou jedinců *Drosophila viridis*. Po rozložení sekvencí na k-mery, jsou k-mery s příliš nízkou frekvencí výskytu odstraněny, neboť pravděpodobně představují sekvenační chyby. Autoři uvádí, že stejně důležitou součástí je také **identifikace všech typů repetitivních sekvencí**, protože ty jsou ze své podstaty sdílené mezi chromozomem Y (a W) a zbytkem genomu. Tento krok je přímo implementován v YGS. Přímým porovnáním k-merů referenčního genomu a sekvenačních čtení heterogametického pohlaví je dosaženo odstranění opakujících se sekvencí všech typů (např. transpozonů či segmentálních duplikací). Důležité je naopak zachování variant těchto repetitiv, tedy polymorfismů (nazývaných „sequence family variants“ nebo „singly unique nucleotide“). Mnoho unikátních oblastí chromozomu Y (a W) jsou ve skutečnosti vzácnými variantami některé z repetitiv, například poškozené mobilní elementy, které jsou pro tento chromozom unikátní (Traut et al. 2013) a mohou být využity pro identifikaci Y (nebo W) chromozomálních sekvencí pomocí k-merové analýzy (Carvalho

a Clark 2013; Li et al. 2018). Postup YGS byl úspěšně ověřen k detekci sekvencí specifických pro samice motýla *B. mori* (Li et al. 2018).

YGS stanovuje procento namapovaných k-merů unikátních pro každý fragment assembly heterogametického pohlaví, neidentifikuje však konkrétní oblasti DNA, které jsou na tyto unikátní k-mery bohaté. V předkládané práci bylo identifikace těchto sekvencí dosaženo dodatečným mapováním k-merů na kandidátní fragmenty. Na některých fragmentech, které byly YGS identifikovány jako potenciálně samičí, byly samičí k-mery distribuovány roztroušeně a nepřekrývaly se. Je pochopitelně pravděpodobné, že se u samic vyskytly i neobvyklé společné alely, které u samců chyběly vlivem náhody a nemusely být specifické pro chromozom W. Za potenciální pohlavně specifické markery byly považovány ty sekvence, na kterých byly pozorovány **píky namapovaných k-merů**, jak je vidět například na **Obrázku 6 (kapitola 4.3)**. Po ověření specificity kandidátních fragmentů na všech dostupných datech in silico byly domnělé samičí markery ověřeny PCR experimenty na DNA jedinců obou pohlaví a obou populací.

Výsledky ověření specificity PCR markerů (viz **kapitolu 4.4**) ale ukázaly, že kandidátní **oblasti nebyly pohlavně specifické**. Otestované primery amplifikovaly produkty u **jedinců obou pohlaví**, tudíž sekvence DNA, na které byly navrženy, nebylo možné použít jako markery pro molekulární sexing.

Po zveřejnění genomů několika zástupců rodu *Yponomeuta* získaných v projektu Darwin Tree of Life byly ověřované sekvence analyzovány nástrojem blastn (blast.ncbi.nlm.nih.gov) v databázi „nucleotide collection (nr/nt)“. Testované fragmenty (contig_24514 a contig_16238, viz **Tabulku IX**) byly identifikovány jako sekvence nacházející se jak na chromozomu W, tak na chromozomu Z a řadě autozomů, a to přímo u testovaného druhu nebo blízce příbuzných (*Y. rorelus*, *Y. padella*, *Y. sedellus*, *Y. evonylemela* a *Y. plumbellus*) s více než 90 % identity. Fragmenty contig_24139 a contig_3883 (viz **Tabulku IX**) byly namapovány na chromozom Z a autozomy těchto druhů.

Detekované sekvence jsou tedy pravděpodobně **části repetitivních sekvencí**, které se nachází v genomu několika druhů rodu *Yponomeuta*, včetně *Y. cagnagella*. Repetitivní povaze některých částí kandidátních fragmentů nasvědčují i výsledky PCR, viz **Obrázek 11 a Obrázek 12 (kapitola 4.4)**.

Přes všechny kroky provedené za účelem zpřesnění postupu YGS a snížení pravděpodobnosti falešně pozitivních výsledků se nepodařilo předejít detekci nespecifických repetitivních sekvencí, které **unikly detekci v pipeline YGS i maskování** pomocí RepeatMasker. Dále byla zkoumána povaha 18 fragmentů, které byly postupem YGS a následným porovnáním

s genomem M_ONT identifikovány jako kandidátní W-specifické oblasti (viz sloupec BLAST M_ONT v **Tabulce IX**). Nástrojem Minimap byla ověřena přítomnost 14 kandidátních sekvencí přímo na sekvenci chromozomu W získané z assembly F_DTOL. Namapovány byly všechny fragmenty kromě čtyř, včetně fragmentu contig_3883, který byl v této práci jedním z testovaných markerů. Totožnost všech 18 kandidátních sekvencí, které byly v této práci identifikovány jako potenciálně W-specifické, byla **ověřena také nástrojem DANTE** (Domain based ANnotation of Transposable Elements, Galaxy Version 1.1.0), což je nástroj pro anotaci mobilních elementů na základě podobnosti s databází konzervovaných proteinových domén, (**git/20_DANTE**). Celkově 4 repetice byly anotovány jako retrotranspozony (LINE či LTR), a jeden jako DNA transpozon (Wicker et al. 2007), viz **Tabulku XI**.

Tabulka XI: Výsledky ověření identity testovaných sekvencí nástrojem DANTE. Fragmenty contig_3883 a contig_16238 byly v této práci ověřovány jako kandidátní samičí markery.

Název fragmentu	Délka (bp)	Klasifikace repetice	Procento identity
contig_3883	3155	Třída_II, podtřída 1, TIR, <i>PIF-Harbinger</i>	100 %
contig_16238	4426	Třída_I, LINE	100 %
contig_10146	14217	Třída_I, LTR, Ty1/copia	5 shod, 86 - 100 %
contig_19260	3493	Třída_I, LINE	63 %
contig_636	2398	Třída_I, LINE	100 %

V této práci byly jako potenciální pohlavně specifické markery ověřovány 4 fragmenty samičího genomu (contig_24514, contig_16238, contig_24139 a contig_3883). Ani jeden z nich není pohlavně specifickou sekvencí, jak bylo ukázáno metodou PCR (viz **kapitolu 4.4**) a podle pozdějších ověření metodou blastn a nástrojem DANTE jde o části **mobilních elementů**. V případě contigu_16238 jde o retrotranspozon třídy I (LINE). Kandidátní fragment contig_3883 byl anotován jako DNA transpozon *PIF-Harbinger*. Fragmenty contig_24514 a contig_16238 jsou pravděpodobně součástí doposud neznámých repetit, které se nachází i na chromozomu W, ale nejsou pro něj specifické. Doposud není známa žádná repetitivní sekvence, která by byla pro chromozom W u *Y. cagnagella* specifická, což bylo studováno i v bakalářské práci autorky (Pilíková 2019).

Pravděpodobným vysvětlením těchto výsledků je **nedostatečná kontinuita** vstupního samičího assembly F_ONT. Pokud byly v této genomové sekvenci mezery místo repetitivních sekvencí, případně měly repetice artifiální zastoupení, mohlo to být zdrojem chyb v analýze YGS, která poskytla falešně pozitivní výsledky. Svou roli mohl případně hrát také **populačně specifický rozdíl v sekvenci nalezené repetice**. V provedených analýzách byli všichni zahrnutí jedinci samčího pohlaví z nizozemské populace, a to včetně M_ONT, který byl použit jako reference. Je tedy možné, že detekované fragmenty byly **varianty transpozonů**, které ve studovaném samčím vzorku chyběly. Tomu nasvědčuje výsledek z jednoho experimentu PCR. V **kapitole 4.4** lze na **Obrázku 10** vidět, že u jednoho z testovaných jedinců nevzniká žádný produkt, přičemž šlo o vzorek DNA samce z nizozemské populace.

Oblasti, které byly v této práci metodou YGS detekovány jako pohlavně specifické, mohly být také ve skutečnosti rozhraním dvou transpozonů. Pokud se mobilní elementy na chromozomu W vmezeří jeden do druhého, vzniká na jejich rozhraní unikátní sekvence (Abe et al. 2005; Traut et al. 2013), která se v analýze může zdát jako pohlavně specifická. Zároveň mobilní elementy, které jsou dostatečně divergované, obsahují značné množství jednonukleotidových polymorfismů (SNP) (Jin-Shan et al. 2005), které mohou být k-merovou analýzou také detekovány jako oblasti specifické pro samice. Chromozomy Y a W jsou převážně složeny z repetitivních sekvencí, které však až na výjimky nejsou výlučně vázané na tyto chromozomy a většina repetice má analogické sekvence na autozomech, či chromozomu X nebo Z (Schaack et al. 2010; Kim et al. 2012; Hall et al. 2013). I motýlí chromozomy W jsou zpravidla tvořeny právě mobilními elementy, a to zejména retrotranspozony, jak bylo dokázáno u *Bombyx mori* (Abe et al. 2005; 2010) a také u *Ephestia kuehniella* (Traut et al. 2013). U různorožce trnkového (*Peribatodes rhomboidaria*) bylo identifikováno deset repetice obohacených na chromozomu W, z nichž většina jsou retrotranspozony LTR nebo LINE (Hejníčková et al. 2023). Pohlavní chromozom W u *Y. cagnagella* je obohacen o DNA transpozon *Maverick* (Pilíková 2019). Především detekci repetitivních sekvencí při hledání pohlavně specifických markerů by bylo možné, pokud by bylo **maskování provedeno pomocí druhově specifické databáze** repetice pro *Y. cagnagella*, například s využitím nástroje RepeatExplorer (repeatexplorer.org, Novak et al. 2013).

Mobilní elementy mají hlavní roli ve vývoji a diferenciaci pohlavních chromozomů, což je v případě motýlů usnadněno absencí meiotické rekombinace u samic (Hejníčková et al. 2023). Přítomnost transpozonů v oblasti určující pohlaví může usnadnit translokaci pohlavní determinanty buďto prostřednictvím transpozice nebo ektopické rekombinace, což může

potenciálně vést ke vzniku nového páru pohlavních chromozomů (Tennesen et al. 2018; Kabir et al. 2022).

5.3. Další cesty k identifikaci pohlavně vázaných sekvencí

Doposud známé způsoby ověření pohlaví u motýlů pomocí PCR předpokládaly znalost konkrétních sekvencí vázaných na pohlavní chromozomy ať už konkrétních genů nebo satelitních repetitiv (Fuková et al. 2009; Aguirre et al. 2020; Deng et al. 2020) viz **kapitulu 1.2**. Kromě YGS, je známo několik dalších bioinformatických nástrojů, které umožňují odhalit pohlavně specifické sekvence bez předchozí znalosti specifických oblastí genomu (de novo), a které by se daly uvážit jako alternativní přístupy k vývoji molekulárních markerů pro určení pohlaví u motýlů. Na principu porovnání genomu heterogametického pohlaví s krátkými sekvenačními čteními genomu homogametického pohlaví prostřednictvím k-merů funguje také **nástroj DiscoverY** (Rangavittal et al. 2019), který byl navržen na detekci oblastí chromozomu Y u člověka a gorily. Autoři uvádí, že jde o efektivní metodu získání samčích fragmentů, ovšem oblasti, které jsou homologní k chromozomu X je stále obtížné detekovat. Kompletní sekvence lidského chromozomu Y byla získána až v roce 2022, čímž bylo dokončeno úplné osekvenování genomu člověka od telomery k telomeře (Nurk et al. 2022). DiscoverY dosahuje lepších výsledků než YGS, pokud má referenční assembly malou kontinuitu, ovšem kvalita sestaveného genomu je i v případě této metody hlavním limitem (Rangavittal et al. 2019).

K identifikaci pohlaví byla použita i metoda **K-mer Quotient (KQ)** založená na poměru frekvencí k-merů představená v práci Li et al. (2018), v níž byly identifikovány sekvence pohlavního chromozomu W u *Bombyx mori* z inbrední linie Dazao. Příslušnost patřičného genomického fragmentu je stanovena na základě frekvence výskytu k-merů, které byly získány ze sekvenačních dat. Pokud poměr frekvence k-meru u samců a samic odpovídá přibližně číslu 1, je daný k-mer součástí autozomů a poměr 2 je typický pro chromozom Z, který mají samci dvakrát. Chromozom W je přítomen pouze u samic, a proto je poměr frekvence k-merů, které jej reprezentují, 0 (Li et al. 2018). Tento postup byl ověřen i na *D. melanogaster*, kde byly úspěšně identifikovány oblasti chromozomu Y. K identifikaci pohlavně specifických markerů v předkládané práci nebyl nástroj KQ zvolen, protože vyžaduje využití genomických dat s vysokou hloubkou pokrytí, autoři doporučují coverage větší než 50×. Nevýhodou je také možnost zkreslení výsledků při použití více jedinců, kteří nebudou z inbrední linie. V takovém případě by byly frekvence k-merů ovlivněny polymorfismy, a proto je tento přístup u divokých populací nevhodný.

Jedním z používaných postupů pro zachycení skupinově-specifických sekvencí na základě specifických k-merů je **KmerGO** (Wang et al. 2020), který umožňuje rozdělení sekvencí na k-mery a určení frekvence jejich zastoupení, k čemuž je v případě YGS potřeba využít jiný nástroj (například Jellyfish). KmerGO detekuje skupinově-specifické k-mery, a to včetně informace o jejich příslušnosti ke genomické sekvenci. Oproti tomu u pipeline YGS jsou identifikovány fragmenty genomu, které jsou o unikátní k-mery obohaceny, ale pohlavně-specifické k-mery bylo na oblasti potřeba namapovat zpětně, viz kapitolu 3.3.4. Hrubá (2022) úspěšně použila KmerGO ve své diplomové práci pro identifikaci sekvencí specifických pro chromozom W u modráška jehlicového (*Polyommatus icarus*, Lycaenidae). V celkových 23 odhalených samičích genomových fragmentech se jí podařilo identifikovat jednu oblast o velikosti 263 bp, kterou použila jako samičí marker pro určení pohlaví u raných stádií modrášků (Hrubá 2022). Přes vynaložené úsilí se v rámci této předkládané práce nepodařilo nástroj KmerGO kvůli technickým překážkám použít a postup zopakovat.

Molekulární určení pohlaví je možné provést také na základě analýzy známých genů na chromozomu Z. U tří druhů motýlů různých nadčeledí (*Plutella xylostella*, Yponomeutoidea; *Galleria mellonella*, Pyraloidea a *Lymantria dispar*, Noctuoidea) byl metodou qPCR porovnáván počet kopií genu *kettin*, který leží na chromozomu Z, tudíž je u samčích vzorků amplifikováno dvojnásobné množství sekvencí odpovídacích části tohoto genu než u samic (Belousova et al. 2019). Přítomnost genu *kettin* na chromozomu Z byla u *Y. cagnagella* ověřena nástrojem blastn, a to jak u genomu sestaveného v této práci (F_ONT) tak u genomu publikovaného (F_DTOL). Na ověření pohlavně vázaných genů metodou qPCR je však potřeba značné množství DNA, neboť postup vyžaduje technické i biologické repliky, a proto jej prakticky nelze použít pro určení pohlaví raných vývojových stádií. Stanovení správných parametrů experimentů qPCR je časově náročné a přináší nejednoznačné výsledky, jak je vidět na příkladu testování známých samičích markerů u motýla *Tuta absoluta* v práci Carabajal Paladino et al. (2019).

Studium sekvenčního obsahu chromozomu W u motýlů je stále výzvou a široce uplatnitelný nástroj pro detekci pohlavně specifických sekvencí chybí. Kromě práce Hrubé (2022) a Li et al. (2018) nebyly, pokud víme, k-merovou analýzou detekovány fragmenty specifické pro chromozom W u motýlů. Přibývající množství komplexnějších sekvenačních dat nám však dává naději, že vlastnostem pohlavních chromozomů, jejich evoluci a pohlavnímu rozmnožování jako takovému brzy porozumíme lépe.

6. Závěr

Cíle diplomové práce byly částečně naplněny. Genom samice předivky *Yponomeuta cagnagella* byl sekvenován technologií Oxford Nanopore a podařilo se jej poskládat do poměrně kompletního assembly. Byly nalezeny k-mery specifické pro samice a postupem YGS byly identifikovány kandidátní sekvence odpovídající chromozomu W. Po in silico ověření specificity těchto sekvencí byly na čtyři predikované samičí markery navrženy primery. Při PCR byly amplifikovány produkty u jedinců obou pohlaví, tudíž sekvence DNA, na které byly primery navrženy, nebyly pohlavně specifické. Po zveřejnění komplexnějších genomů předivek byly domnělé samičí markery identifikovány buďto jako známé mobilní elementy, nebo doposud neanotované repetice.

7. Seznam použité literatury

- ABBOTT, J. K., A. K. NORDÉN a B. HANSSON, 2017. Sex chromosome evolution: Historical insights and future perspectives. *Proceedings of the Royal Society B: Biological Sciences*. 284(1854), 20162806. ISSN 14712954. Dostupné z: doi:10.1098/rspb.2016.2806
- ABE, H., T. FUJII, T. SHIMADA a K. MITA, 2010. Novel non-autonomous transposable elements on W chromosome of the silkworm, *Bombyx mori*. *Journal of Genetics*. 89(3), 375–387. ISSN 0973-7731. Dostupné z: doi:10.1007/s12041-010-0049-3
- ABE, H., T. FUJII, N. TANAKA, T. YOKOYAMA, H. KAKEHASHI, M. AJIMURA, K. MITA, Y. BANNO, Y. YASUKOCHI, T. OSHIKI et al., 2008. Identification of the female-determining region of the W chromosome in *Bombyx mori*. *Genetica*. 133(3), 269–282. ISSN 00166707. Dostupné z: doi:10.1007/s10709-007-9210-1
- ABE, H., K. MITA, Y. YASUKOCHI, T. OSHIKI a T. SHIMADA, 2005. Retrotransposable elements on the W chromosome of the silkworm, *Bombyx mori*. *Cytogenetic and Genome Research*. 110(1–4), 144–151. ISSN 14248581. Dostupné z: doi:10.1159/000084946
- AGUIRRE, C., N. OLIVARES a P. HINRICHSEN, 2020. An Efficient Duplex PCR Method for Sex Identification of the European Grapevine Moth *Lobesia botrana* (Lepidoptera: Tortricidae) at Any Developmental Stage. *Journal of Economic Entomology*. 113(5), 2505–2510. ISSN 0022-0493, 1938-291X. Dostupné z: doi:10.1093/jee/toaa155
- ALTSCHUL, S. F., W. GISH, W. MILLER, E. W. MYERS a D. J. LIPMAN, 1990. Basic local alignment search tool. *Journal of Molecular Biology*. 215(3), 403–410. ISSN 0022-2836. Dostupné z: doi:10.1016/S0022-2836(05)80360-2
- ANDREWS, S. (2010). FastQC: a Quality Control Tool for High Throughput Sequence Data. Dostupné z: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- BACHTROG, D., 2006. A dynamic view of sex chromosome evolution. *Current Opinion in Genetics and Development*. 16(6), 578–585. ISSN 0959437X. Dostupné z: doi:10.1016/j.gde.2006.10.007
- BELOUSOVA, I., N. ERSHOV, S. PAVLUSHIN, Y. ILINSKY a V. MARTEMYANOV, 2019. Molecular sexing of Lepidoptera. *Journal of Insect Physiology*. 114, 53–56. ISSN 1879-1611. Dostupné z: doi:10.1016/j.jinsphys.2019.02.005
- BERGERO, R. a D. CARLESWORTH, 2008. The evolution of restricted recombination in sex chromosomes. *Trends Ecol Evol*. 2009;24(2):94-102. Dostupné z: doi:10.1016/j.tree.2008.09.010
- BROWN, K. S., Jr, B. VON SCHOULTZ a E. SUOMALAINEN, 2004. Chromosome evolution in Neotropical Danainae and Ithomiinae (Lepidoptera). *Hereditas*. 141(3), 216–236. ISSN 1601-5223. Dostupné z: doi:10.1111/j.1601-5223.2004.01868.x
- BUREŠ, P. a F. ZEDEK, 2014. Holokinetic drive: centromere drive in chromosomes without centromeres. *Evolution; International Journal of Organic Evolution*. 68(8), 2412–2420. ISSN 1558-5646. Dostupné z: doi:10.1111/evo.12437
- CARABAJAL PALADINO, L. Z., I. PROVAZNÍKOVÁ, M. BERGER, C. BASS, N. S. ARATCHIGE, S. N. LÓPEZ, F. MAREC a P. NGUYEN, 2019. Sex Chromosome Turnover in Moths of the Diverse Superfamily Gelechioidea. *Genome Biology and Evolution*. 11(4), 1307–1319. ISSN 1759-6653. Dostupné z: doi:10.1093/gbe/evz075
- CAREY, S. B., J. T. LOVELL, J. JENKINS, J. LEEBENS-MACK, J. SCHMUTZ, M. A. WILSON a A. HARKESS, 2022. Representing sex chromosomes in genome assemblies. *Cell Genomics*. 2(5), 100132 ISSN 2666-979X. Dostupné z: doi:10.1016/j.xgen.2022.100132
- CARVALHO, A. B. a A. G. CLARK, 2013. Efficient identification of Y chromosome sequences in the human and *Drosophila* genomes. *Genome Research*. 23(11), 1894–1907. ISSN 1088-9051. Dostupné z: doi:10.1101/gr.156034.113
- CLARK, F. E. a T. D. KOCHER, 2019. Changing sex for selfish gain: B chromosomes of Lake Malawi cichlid fish. *Scientific Reports*. 9(1), 20213. ISSN 2045-2322. Dostupné z: doi:10.1038/s41598-019-55774-8
- COATES, B. S. a R. L. HELLMICH, 2003. Two sex-chromosome-linked microsatellite loci show geographic variance among North American *Ostrinia nubilalis*. *Journal of Insect Science*. 3, 29. ISSN 1536-2442. Dostupné z: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC524668/>
- DAI, W., J. E. MANK a L. BAN, 2022. Repeated origin of the W chromosome from the Z chromosome in *Lepidoptera*. 21. říjen 2022. B.m.: bioRxiv. Dostupné z: doi:10.1101/2022.10.21.512844
- DALÍKOVÁ, M., M. ZRZAVÁ, I. HLADOVÁ, P. NGUYEN, I. ŠONSKÝ, M. FLEGROVÁ, S. KUBÍČKOVÁ, A. VOLENÍKOVÁ, A. Y. KAWAHARA, R. S. PETERS a F. MAREC, 2017. New Insights into the Evolution of the W Chromosome in Lepidoptera. *Journal of Heredity*. 108(7), 709–719. ISSN 0022-1503. Dostupné z: doi:10.1093/jhered/esx063

- DE COSTER, W., S. D'HERT, D. T. SCHULTZ, M. CRUTS a C. VAN BROECKHOVEN, 2018. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics (Oxford, England)*. 34(15), 2666–2669. ISSN 1367-4811. Dostupné z: doi:10.1093/bioinformatics/bty149
- DENG, Z., Y. ZHANG, M. ZHANG, J. HUANG, C. LI, X. NI a X. LI, 2020. Characterization of the First W-Specific Protein-Coding Gene for Sex Identification in *Helicoverpa armigera*. *Frontiers in Genetics*. 11. ISSN 1664-8021. Dostupné z: doi:10.3389/fgene.2020.00649
- DOUGLAS, L. J., P. M. UNTALAN a D. S. HAYMER, 2004. Molecular sexing in the Mediterranean fruit fly, *Ceratitis capitata*. *Insect Biochemistry and Molecular Biology*. 34(2), 159–165. ISSN 0965-1748. Dostupné z: doi:10.1016/j.ibmb.2003.08.006
- ELLIS, E. A., C. G. STORER a A. Y. KAWAHARA, 2021. De novo genome assemblies of butterflies. *GigaScience*. 10(6), giab041. ISSN 2047-217X. Dostupné z: doi:10.1093/gigascience/giab041
- FRAÏSSE, C., M. A. L. PICARD a B. VICOSO, 2017. The deep conservation of the Lepidoptera Z chromosome suggests a non-canonical origin of the W. *Nature Communications*. 8(1), 1486. ISSN 2041-1723. Dostupné z: doi:10.1038/s41467-017-01663-5
- FUKOVÁ, I., L. G. NEVEN, N. M. BÁRCENAS, N. A. GUND, M. DALÍKOVÁ a F. MAREC, 2009. Rapid assessment of the sex of codling moth *Cydia pomonella* (Linnaeus) (Lepidoptera: Tortricidae) eggs and larvae. *Journal of Applied Entomology*. 133(4), 249–261. ISSN 09312048. Dostupné z: doi:10.1111/j.1439-0418.2008.01352.x
- FUKOVÁ, I., W. TRAUT, M. VÍTKOVÁ, P. NGUYEN, S. KUBÍČKOVÁ a F. MAREC, 2007. Probing the W chromosome of the codling moth, *Cydia pomonella*, with sequences from microdissected sex chromatin. *Chromosoma*. 116(2), 135–145. ISSN 00095915. Dostupné z: doi:10.1007/s00412-006-0086-0
- GOTTER, A. L., J. D. LEVINE a S. M. REPPERT, 1999. Sex-linked period genes in the silkworm, *Antheraea pernyi*: implications for circadian clock regulation and the evolution of sex chromosomes. *Neuron*. 24(4), 953–965. ISSN 0896-6273. Dostupné z: doi:10.1016/s0896-6273(00)81042-9
- GRIFFITHS, R., M. C. DOUBLE, K. ORR a R. J. DAWSON, 1998. A DNA test to sex most birds. *Molecular Ecology*. 7(8), 1071–1075. ISSN 0962-1083. Dostupné z: doi:10.1046/j.1365-294x.1998.00389.x
- GUREVICH, A., V. SAVELIEV, N. VYAHHI a G. TESLER, 2013. QUASt: quality assessment tool for genome assemblies. *Bioinformatics (Oxford, England)*. 29(8), 1072–1075. ISSN 1367-4811. Dostupné z: doi:10.1093/bioinformatics/btt086
- HALL, A. B., Y. QI, V. TIMOSHEVSKIY, M. V. SHARAKHOVA, I. V. SHARAKHOV a Z. TU, 2013. Six novel Y chromosome genes in *Anopheles* mosquitoes discovered by independently sequencing males and females. *BMC Genomics*. 14, 273. ISSN 1471-2164. Dostupné z: doi:10.1186/1471-2164-14-273
- HEJNÍČKOVÁ, M., M. DALÍKOVÁ, M. ZRZAVÁ, F. MAREC, P. LORITE a E. E. MONTIEL, 2023. Accumulation of retrotransposons contributes to W chromosome differentiation in the willow beauty *Peribatodes rhomboidaria* (Lepidoptera: Geometridae). *Scientific Reports*. 13(1), 534. ISSN 2045-2322. Dostupné z: doi:10.1038/s41598-023-27757-3
- HEJNÍČKOVÁ, M., P. KOUTECKÝ, P. POTOCKÝ, I. PROVAZNÍKOVÁ, A. VOLENÍKOVÁ, M. DALÍKOVÁ, S. VISSER, F. MAREC a M. ZRZAVÁ, 2019. Absence of W chromosome in psychidae moths and implications for the theory of sex chromosome evolution in lepidoptera. *Genes*. 10(12). ISSN 20734425. Dostupné z: doi:10.3390/genes10121016
- HILL, J., P. RASTAS, E. A. HORNETT, R. NEETHIRAJ, N. CLARK, N. MOREHOUSE, M. DE LA PAZ CELORIO-MANCERA, J. C. COLS, H. DIRCKSEN, C. MESLIN et al., 2019. Unprecedented reorganization of holocentric chromosomes provides insights into the enigma of lepidopteran chromosome evolution. *Science Advances*. 5(6), eaau3648. Dostupné z: doi:10.1126/sciadv.aau3648
- HÖÖK, L., K. NÄSVALL, R. VILA, C. WIKLUND a N. BACKSTRÖM, 2023. High-density linkage maps and chromosome level genome assemblies unveil direction and frequency of extensive structural rearrangements in wood white butterflies (Leptidea spp.). *Chromosome Research*. 31(1). ISSN 0967-3849. Dostupné z: doi:10.1007/s10577-023-09713-z
- HORA, K. H. F. MAREC, P. ROESSINGH a S. B. J. MENKEN, 2019. Limited intrinsic postzygotic reproductive isolation despite chromosomal rearrangements between closely related sympatric species of small ermine moths (Lepidoptera: Yponomeutidae). *Biological Journal of the Linnean Society*. 128(1), 44–58. ISSN 0024-4066. Dostupné z: doi:10.1093/biolinnean/blz090
- HRUBÁ, M., 2022. Vývoj molekulárních markerů pro chromozom W u modráška *Polyommatus icarus*. [Development of molecular markers for the W chromosome in the blue butterfly *Polyommatus icarus*. Mgr. Thesis, in Czech.], Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic. Dostupné z: invenio.nusl.cz/record/520562?
- CHARLESWORTH, D., B. CHARLESWORTH a G. MARAIS, 2005. Steps in the evolution of heteromorphic sex chromosomes. *Heredity*. 95(2), 118–128. ISSN 0018-067X. Dostupné z: doi:10.1038/sj.hdy.6800697
- CHO, S., A. MITCHELL, J. C. REGIER, C. MITTER, R. W. POOLE, T. P. FRIEDLANDER a S. ZHAO, 1995. A highly conserved nuclear gene for low-level phylogenetics: elongation factor-1 alpha recovers

- morphology-based tree for heliothine moths. *Molecular Biology and Evolution*. 12(4), 650–656. ISSN 0737-4038. Dostupné z: doi:10.1093/oxfordjournals.molbev.a040244
- INTERNATIONAL SILKWORM GENOME CONSORTIUM, 2008. The genome of a lepidopteran model insect, the silkworm *Bombyx mori*. *Insect Biochemistry and Molecular Biology*. 38(12), 1036–1045. ISSN 1879-0240. Dostupné z: doi:10.1016/j.ibmb.2008.11.004
- JIN-SHAN, X., X. QING-YOU, L. JUN, P. GUO-QING a Z. ZE-YANG, 2005. Survey of long terminal repeat retrotransposons of domesticated silkworm (*Bombyx mori*). *Insect Biochemistry and Molecular Biology*. 35(8), 921–929. ISSN 0965-1748. Dostupné z: doi:10.1016/j.ibmb.2005.03.014
- JUMAEV, R., A. GAZIBEKOV, O. SULAYMONOV a B. SOBIROV, 2021. Representatives of Lepidoptera groups occurred in forestry and agricultural crops and their effective entomophage types. *E3S Web of Conferences*. 244, 02020. ISSN 2267-1242. Dostupné z: doi:10.1051/e3sconf/202124402020
- KABIR, A., R. IEDA, S. HOSOYA, D. FUJIKAWA, K. ATSUMI, S. TAJIMA, A. NOZAWA, T. KOYAMA, et al., 2022. Repeated translocation of a supergene underlying rapid sex chromosome turnover in *Takifugu pufferfish*. *Proceedings of the National Academy of Sciences*. 119(23), e2121469119. Dostupné z: doi:10.1073/pnas.2121469119
- KAISER, V. B. a D. BACHTROG, 2010. Evolution of Sex Chromosomes in Insects. *Annual Review of Genetics*. 44(1), 91–112. ISSN 0066-4197. Dostupné z: doi:10.1146/annurev-genet-102209-163600
- KAWAMOTO, M., T. KIUCHI a S. KATSUMA, 2022. SilkBase: an integrated transcriptomic and genomic database for *Bombyx mori* and related species. *Database*. 2022, baac040. ISSN 1758-0463. Dostupné z: doi:10.1093/database/baac040
- KIM, Y., J. LEE a K. HAN, 2012. Transposable Elements: No More „Junk DNA“. *Genomics & Informatics*. 10(4), 226. ISSN 1598-866X. Dostupné z: doi:10.5808/gi.2012.10.4.226
- KOLMOGOROV, M., D. M. BICKHART, B. BEHSAZ, A. GUREVICH, M. RAYKO, S. B. SHIN, et al., 2020. metaFlye: scalable long-read metagenome assembly using repeat graphs. *Nature Methods*. 17(11), 1103–1110. ISSN 1548-7105. Dostupné z: doi:10.1038/s41592-020-00971-x
- KOREN, S., B. P. WALENZ, K. BERLIN, J. R. MILLER, N. H. BERGMAN a A. M. PHILLIPPY, 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Research*. 27(5), 722–736. ISSN 1088-9051, 1549-5469. Dostupné z: doi:10.1101/gr.215087.116
- KOUKIDOU, M. a L. ALPHEY, 2014. Practical Applications of Insects' Sexual Development for Pest Control. *Sexual Development*. 8(1–3), 127–136. ISSN 1661-5425, 1661-5433. Dostupné z: doi:10.1159/000357203
- KREKLOVÁ M., 2020. Analýza pohlavních chromozomů modrásků (*Lycaenidae*) [Analysis of sex chromosomes of blue butterflies (*Lycaenidae*). Mgr. Thesis, in Czech], Faculty of Science, University of South Bohemia, Česká Budějovice, Czech Republic. Dostupné z: theses.cz/id/mw8kjjz/Hruba-Bakalarska_prace.pdf
- LANGMEAD, B. a S. L. SALZBERG, 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods*. 9(4), 357–359. ISSN 1548-7105. Dostupné z: doi:10.1038/nmeth.1923
- LEGGETT, R. M. a D. MACLEAN, 2014. Reference-free SNP detection: dealing with the data deluge. *BMC Genomics*. 15(4), S10. ISSN 1471-2164. Dostupné z: doi:10.1186/1471-2164-15-S4-S10
- LEWIS, J. J., F. CICONARDI, S. H. MARTIN, R. D. REED, C. G. DANKO a S. H. MONTGOMERY, 2021. The *Dryas iulia* Genome Supports Multiple Gains of a W Chromosome from a B Chromosome in Butterflies. *Genome Biology and Evolution*. 13(7), evab128. ISSN 1759-6653. Dostupné z: doi:10.1093/gbe/evab128
- LI, H., B. HANDSAKER, A. WYSOKER, T. FENNEL, J. RUAN, N. HOMER, G. MARTH, G. ABECASIS, R. DURBIN a 1000 GENOME PROJECT DATA PROCESSING SUBGROUP, 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*. 25(16), 2078–2079. ISSN 1367-4811. Dostupné z: doi:10.1093/bioinformatics/btp352
- LI, S., M. AJIMURA, Z. CHEN, J. LIU, E. CHEN, H. GUO, V. TADAPATRI, C. G. REDDY, J. ZHAN, H. KISHINO et al., 2018. A new approach for comprehensively describing heterogametic sex chromosomes. *DNA Research*. 25(4), 375–382. ISSN 17561663. Dostupné z: doi:10.1093/dnares/dsy010
- LIN, B., J. HUI a H. MAO, 2021. Nanopore Technology and Its Applications in Gene Sequencing. *Biosensors*. 11(7), 214. ISSN 2079-6374. Dostupné z: doi:10.3390/bios11070214
- LIU, G., Z. CHANG, L. CHEN, J. HE, Z. DONG, J. YANG, S. LU, R. ZHAO, W. WAN, G. MA et al., 2020. Genome size variation in butterflies (Insecta, Lepidoptera, Papilionoidea): a thorough phylogenetic comparison. *Systematic Entomology*. 45(3), 571–582. ISSN 1365-3113. Dostupné z: doi:10.1111/syen.12417
- LIU, Q., Y. GUO, Y. ZHANG, W. HU, Y. LI, D. ZHU, Z. ZHOU, J. WU, N. CHEN a X. ZHOU, 2019. A chromosomal-level genome assembly for the insect vector for Chagas disease, *Triatoma rubrofasciata*. *GigaScience*. 8. Dostupné z: doi:10.1093/gigascience/giz089
- LUKHTANOV, V. A., 2000. Sex chromatin and sex chromosome systems in nonditrysian Lepidoptera (Insecta). *Journal of Zoological Systematics and Evolutionary Research*. 38(2), 73–79. ISSN 1439-0469. Dostupné z: doi:10.1046/j.1439-0469.2000.382130.x

- LUKHTANOV, V. A., 2015. The blue butterfly *Polyommatus (Plebicula) atlanticus* (Lepidoptera, Lycaenidae) holds the record of the highest number of chromosomes in the non-polyploid eukaryotic organisms. *Comparative Cytogenetics*. 9(4), 683–690. ISSN 1993-0771. Dostupné z: doi:10.3897/compcytogen.v9i4.5760
- MANEKAR, S. C. a S. R. SATHE, 2018. A benchmark study of k-mer counting methods for high-throughput sequencing. *GigaScience*. 7(12), giy125. ISSN 2047-217X. Dostupné z: doi:10.1093/gigascience/giy125
- MANK, J. E. a H. ELLEGREN, 2009. Sex-linkage of sexually antagonistic genes is predicted by female, but not male, effects in birds. *Evolution*. 63(6), 1464–1472. ISSN 00143820. Dostupné z: doi:10.1111/j.1558-5646.2009.00618.x
- MANNI, M., M. R. BERKELEY, M. SEPPEY a E. M. ZDOBNOV, 2021. BUSCO: Assessing Genomic Data Quality and Beyond. *Current Protocols*. 1(12), e323. ISSN 2691-1299. Dostupné z: doi:10.1002/cpz1.323
- MARÇAIS, G. a C. KINGSFORD, 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics (Oxford, England)*. 27(6), 764–770. ISSN 1367-4811. Dostupné z: doi:10.1093/bioinformatics/btr011
- MAREC, F. a M. J. B. VREYSEN, 2019. Advances and Challenges of Using the Sterile Insect Technique for the Management of Pest Lepidoptera. *Insects*. 10(11), 371. ISSN 2075-4450. Dostupné z: doi:10.3390/insects10110371
- MCCOMBIE, W. R., J. D. MCPHERSON a E. R. MARDIS, 2019. Next-Generation Sequencing Technologies. *Cold Spring Harbor Perspectives in Medicine*. 9(11), a036798. ISSN 2157-1422. Dostupné z: doi:10.1101/cshperspect.a036798
- MENKEN, S. B. J., W. M. HERREBOUT a J. T. WIEBES, 1992. Small Ermine Moths (Yponomeuta): Their Host Relations and Evolution. *Annual Review of Entomology*. 37(1), 41–66. Dostupné z: doi:10.1146/annurev.en.37.010192.000353
- MITA, K., M. KASAHARA, S. SASAKI, Y. NAGAYASU, T. YAMADA, H. KANAMORI, N. NAMIKI, M. KITAGAWA, H. YAMASHITA, Y. YASUKOCHI et al., 2004. The genome sequence of silkworm, *Bombyx mori*. *DNA research: an international journal for rapid publication of reports on genes and genomes*. 11(1), 27–35. ISSN 1340-2838. Dostupné z: doi:10.1093/dnares/11.1.27
- MIYAO, A., J. KIYOMIYA, K. IIDA, K. DOI a H. YASUE, 2019. Polymorphic edge detection (PED): two efficient methods of polymorphism detection from next-generation sequencing data. *BMC Bioinformatics*. 20. Dostupné z: doi:10.1186/s12859-019-2955-6
- MONGUE, A. J., P. NGUYEN, A. VOLENÍKOVÁ a J. R. WALTERS, 2017. Neo-sex Chromosomes in the Monarch Butterfly, *Danaus plexippus*. *G3 (Bethesda, Md.)*. 7(10), 3281–3294. ISSN 2160-1836. Dostupné z: doi:10.1534/g3.117.300187
- NGUYEN, P., M. SÝKOROVÁ, J. ŠÍCHOVÁ, V. KŮTA, M. DALÍKOVÁ, R. ČAPKOVÁ FRYDRYCHOVÁ, L. G. NEVEN, K. SAHARA a F. MAREC, 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. ISSN 00278424. Dostupné z: doi:10.1073/pnas.1220372110
- NILSSON, N., C. LÖFSTEDT a L. DÄVRING, 1988. Unusual sex chromosome inheritance in six species of small ermine moths (Yponomeuta, Yponomeutidae, Lepidoptera). *Hereditas*. 108(2), 259–265. ISSN 00180661. Dostupné z: doi:10.1111/j.1601-5223.1988.tb00311.x
- NOVAK, P., P. NEUMANN, J. PECH, J. STEINHAIŠL a J. MACAS, 2013. RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. *Bioinformatics*. 29(6), 792–793. ISSN 1367-4803. Dostupné z: doi:10.1093/bioinformatics/btt054
- NURK, S., S. KOREN, A. RHIE, M. RAUTIAINEN, A. V. BZIKADZE, A. MIKHEENKO, M. R. VOLLGER, N. ALTEMOSE, L. URALSKY, A. GERSHMAN et al., 2022. The complete sequence of a human genome. *Science*. 376(6588), 44–53. Dostupné z: doi:10.1126/science.abj6987
- PENNELL, M. W., M. KIRKPATRICK, S. P. OTTO, J. C. VAMOSI, C. L. PEICHEL, N. VALENZUELA aj. KITANO, 2015. Y Fuse? Sex Chromosome Fusions in Fishes and Reptiles. *PLoS Genetics*. 11(5), e1005237. ISSN 1553-7390. Dostupné z: doi:10.1371/journal.pgen.1005237
- PILÍKOVÁ A., 2019. Analýza vybraných W obohacených repetitivních sekvencí u předivky brslenové *Yponomeuta cagnagella* (Lepidoptera). [Analysis of selected W-enriched repetitive sequences in the spindle ermine moth *Yponomeuta cagnagella* (Lepidoptera). Bc. Thesis, in Czech.], Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic. Dostupné z: theses.cz/id/5jgta/
- PROVAZNIKOVÁ I., 2022. Drivers of karyotype evolution in Lepidoptera. Ph.D. Thesis. University of South Bohemia, Faculty of Science, School of Doctoral Studies in Biological Sciences, České Budějovice, Czech Republic. Dostupné z: invenio.nucl.cz/record/521047/?
- RANGAVITTAL, S., N. STOPA, M. TOMASZKIEWICZ, K. SAHLIN, K. D. MAKOVA a P. MEDVEDEV, 2019. DiscoverY: A classifier for identifying y chromosome sequences in male assemblies. *BMC Genomics*. 20(1), 1–11. ISSN 14712164. Dostupné z: doi:10.1186/s12864-019-5996-3

- RISHI, S., G. SAHNI a K. RISHI, 1999. Inheritance of unusual sex chromosome evidenced by AAWZ sex trivalent in *Trabala vishnu* (Lasiocampidae, Lepidoptera). *Cytobios.* 100, 85–94. Dostupné z: <http://ci.nii.ac.jp/naid/10009419503/en/>
- SAHARA, K., A. YOSHIDO, N. KAWAMURA, A. OHNUMA, H. ABE, K. MITA, T. OSHIKI, T. SHIMADA, S. I. ASANO, H. BANDO a Y. YASUKOCHI, 2003. W-derived BAC probes as a new tool for identification of the W chromosome and its aberrations in *Bombyx mori*. *Chromosoma.* 112(1), 48–55. ISSN 00095915. Dostupné z: doi:10.1007/s00412-003-0245-5
- SAHARA, K., A. YOSHIDO a W. TRAUT, 2012. Sex chromosome evolution in moths and butterflies. *Chromosome Research.* 20(1), 83–94. ISSN 0967-3849. Dostupné z: doi:10.1007/s10577-011-9262-z
- SANGER, F., S. NICKLEN a A. R. COULSON, 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America.* 74(12), 5463–5467. ISSN 0027-8424. Dostupné z: doi:10.1073/pnas.74.12.5463
- SENARATNE, A. P., H. MULLER, K. A. FRYER, M. KAWAMOTO, S. KATSUMA a I. A. DRINNENBERG, 2021. Formation of the CenH3-Deficient Holocentromere in Lepidoptera Avoids Active Chromatin. *Current Biology.* 31(1), 173–181.e7. ISSN 0960-9822. Dostupné z: doi:10.1016/j.cub.2020.09.078
- SCHAACK, S., C. GILBERT a C. FESCHOTTE, 2010. Promiscuous DNA: Horizontal transfer of transposable elements and why it matters for eukaryotic evolution. *Trends in Ecology and Evolution.* 25(9), 537–546. ISSN 01695347. Dostupné z: doi:10.1016/j.tree.2010.06.001
- SCHATZ, M., 2017. Nanopore sequencing meets epigenetics. *Nat Methods* 14, 347–348. Dostupné z: <https://doi.org/10.1038/nmeth.4240>
- SIMÃO, F. A., R. M. WATERHOUSE, P. IOANNIDIS, E. V. KRIVENTSEVA a E. M. ZDOBNOV, 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics (Oxford, England).* 31(19), 3210–3212. ISSN 1367-4811. Dostupné z: doi:10.1093/bioinformatics/btv351
- SMITH, STANLEY G., 1945. HETEROPYCNOSIS AS A MEANS OF DIAGNOSING SEX*. *Journal of Heredity.* 36(7), 195–196. ISSN 0022-1503. Dostupné z: doi:10.1093/oxfordjournals.jhered.a105498
- SUETSUGU, Y., R. FUTAHASHI, H. KANAMORI, K. KADONO-OKUDA, S. SASANUMA, J. NARUKAWA, M. AJIMURA, A. JOURAKU, N. NAMIKI, M. SHIMOMURA et al., 2013. Large Scale Full-Length cDNA Sequencing Reveals a Unique Genomic Landscape in a Lepidopteran Model Insect, *Bombyx mori*. *G3: Genes/Genomes/Genetics.* 3(9), 1481–1492. ISSN 2160-1836. Dostupné z: doi:10.1534/g3.113.006239
- SUOMALAINEN, E., 1969. On the sex chromosome trivalent in some Lepidoptera females. *Chromosoma.* 28(3), 298–308. ISSN 1432-0886. Dostupné z: doi:10.1007/BF00284928
- ŠÍCHOVÁ, J., A. VOLENÍKOVÁ, V. DINCA, P. NGUYEN, R. VILA, K. SAHARA a F. MAREC, 2015. Dynamic karyotype evolution and unique sex determination systems in Leptidea wood white butterflies Speciation and evolutionary genetics. *BMC Evolutionary Biology.* 15(1), 89. ISSN 14712148. Dostupné z: doi:10.1186/s12862-015-0375-4
- TANG, M., S. HE, X. GONG, P. LÜ, R. H. TAHA a K. CHEN, 2021. High-Quality de novo Chromosome-Level Genome Assembly of a Single *Bombyx mori* With BmNPV Resistance by a Combination of PacBio Long-Read Sequencing, Illumina Short-Read Sequencing, and Hi-C Sequencing. *Frontiers in Genetics.* 12, 718266. ISSN 1664-8021. Dostupné z: doi:10.3389/fgene.2021.718266
- TENNESSEN, J. A., N. WEI, S. C. K. STRAUB, R. GOVINDARAJULU, A. LISTON a T. ASHMAN, 2018. Repeated translocation of a gene cassette drives sex-chromosome turnover in strawberries. *PLOS Biology.* 16(8), e2006062. ISSN 1545-7885. Dostupné z: doi:10.1371/journal.pbio.2006062
- TOMASZKIEWICZ, M., P. MEDVEDEV a K. D. MAKOVA, 2017. Y and W Chromosome Assemblies: Approaches and Discoveries. *Trends in Genetics.* 33(4), 266–282. ISSN 0168-9525. Dostupné z: doi:10.1016/j.tig.2017.01.008
- TONG, X., M. HAN, K. LU, S. TAI, S. LIANG, Y. LIU, H. HU, J. SHEN, A. LONG, C. ZHAN et al., 2022. High-resolution silkworm pan-genome provides genetic insights into artificial selection and ecological adaptation. *Nature Communications.* 13(1), 5619. ISSN 2041-1723. Dostupné z: doi:10.1038/s41467-022-33366-x
- TRAUT, W., K. SAHARA a F. MAREC, 2007. Sex Chromosomes and Sex Determination in Lepidoptera. *Sexual Development.* 1(6), 332–346. ISSN 1661-5425. Dostupné z: doi:10.1159/000111765
- TRAUT, W., U. EICKHOFF a J. SCHORCH, 2001. Identification and analysis of sex chromosomes by comparative genomic hybridization (CGH). In: Arun Kumar SHARMA a Archana SHARMA, ed. *Chromosome Painting: Principles, Strategies and Scope.* Dordrecht: Springer Netherlands, s. 155–161 ISBN 978-94-010-0330-8. Dostupné z: doi:10.1007/978-94-010-0330-8_16
- TRAUT, W. a F. MAREC, 1996. Sex Chromatin in Lepidoptera. *The Quarterly Review of Biology.* 71(2), 239–256. ISSN 0033-5770. Dostupné z: doi:10.1086/419371

- TRAUT, W., H. VOGEL, G. GLÖCKNER, E. HARTMANN a D. G. HECKEL, 2013. High-throughput sequencing of a single chromosome: a moth W chromosome. *Chromosome Research*. 21(5), 491–505. ISSN 0967-3849. Dostupné z: doi:10.1007/s10577-013-9376-6
- TRIAANT, D. A., S. D. CINEL a A. Y. KAWAHARA, 2018. Lepidoptera genomes: current knowledge, gaps and future directions. *Current Opinion in Insect Science*. 25, Insect genomics * Development and regulation, 99–105. ISSN 2214-5745. Dostupné z: doi:10.1016/j.cois.2017.12.004
- TURNER, H., N. LIESHOUT, W. E. VAN GINKEL a S. B. J. MENKEN, 2010. Molecular phylogeny of the small ermine moth genus *Yponomeuta* (Lepidoptera, Yponomeutidae) in the Palearctic. *PLoS ONE*. 5(3). ISSN 19326203. Dostupné z: doi:10.1371/journal.pone.0009933
- UDAONDO, Z., K. SITTIKANKAEW, T. UENGWETWANIT, T. WONGSURAWAT, C. SONTHIROD, P. JENJAROENPUN, W. POOTAKHAM, N. KAROONUTHAISIRI a I. NOOKAEW, 2021. Comparative Analysis of PacBio and Oxford Nanopore Sequencing Technologies for Transcriptomic Landscape Identification of *Panaeus monodon*. *Life*. 11(8), 862. ISSN 2075-1729. Dostupné z: doi:10.3390/life11080862
- ULENBERG, S. A., 2009. Phylogeny of the *Yponomeuta* species (Lepidoptera, Yponomeutidae) and the history of their host plant associations. *Tijdschrift voor Entomologie*. 152(2), 187–207. ISSN 0040-7496. Dostupné z: doi:10.1163/22119434-900000275
- VAN'T HOF, A. E., P. NGUYEN, M. DALÍKOVÁ, N. EDMONDS, F. MAREC a I. J. SACCHERI, 2013. Linkage map of the peppered moth, *Biston betularia* (Lepidoptera, Geometridae): A model of industrial melanism. *Heredity*. 110(3), 283–295. ISSN 0018067X. Dostupné z: doi:10.1038/hdy.2012.84
- VESELÁ J., 2012. Molekulární diferenciace pohlavních chromosomů předivek rodu *Yponomeuta* [Molecular differentiation of sex chromosomes in ermine moths of genus *Yponomeuta*. Mgr. Thesis, in Czech.], Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic. Dostupné z: dspace.jcu.cz/handle/123456789/24048
- VLAŠÁNEK, P., A. SUCHACKOVA BARTONOVA, F. MAREC a M. KONVICKA, 2017. Elusive *Parnassius mnemosyne* (Linnaeus, 1758) larvae: habitat selection, sex determination and sex ratio (Lepidoptera: Papilionidae). *SHILAP*. *SHILAP Rev. Lepidopt.* 45:561–569. ISSN: 0300-5267.
- VOLENÍKOVÁ A., 2015. Analýza pohlavních chromosomů vybraných druhů bazálních motýlů z čeledi hrotnokřídlecovití (Lepidoptera: Hepialidae). [Karyotype and sex chromosomes analysis of two species from basal lepidopteran family Hepialidae (Lepidoptera: Hepialidae), Mgr. Thesis, in Czech.], Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic. Dostupné z: theses.cz/id/6p6ue4
- VOLENIKOVA, A., P. NGUYEN, P. DAVEY, H. SEHADOVA, B. KLUDKIEWICZ, P. KOUTECKY, J. R. WALTERS, P. ROESSINGH, I. PROVAZNIKOVA, M. SERY et al., 2022. Genome sequence and silkomics of the spindle ermine moth, *Yponomeuta cagnagella*, representing the early diverging lineage of the ditrysian Lepidoptera. *Communications biology*. 5(1), 1281. ISSN 2399-3642. Dostupné z: doi:10.1038/s42003-022-04240-9
- VUCICEVIC, M., M. STEVANOV-PAVLOVIC, J. STEVANOVIC, J. BOSNJAK, B. GAJIC, N. ALEKSIC a Z. STANIMIROVIC, 2013. Sex determination in 58 bird species and evaluation of CHD gene as a universal molecular marker in bird sexing. *Zoo Biology*. 32(3), 269–276. ISSN 1098-2361. Dostupné z: doi:10.1002/zoo.21010
- WAN, F., C. YIN, R. TANG, M. CHEN, Q. WU, C. HUANG, W. QIAN, O. ROTA-STABELLI, N. YANG, S. WANG et al., 2019. A chromosome-level genome assembly of *Cydia pomonella* provides insights into chemical ecology and insecticide resistance. *Nature Communications*. 10(1), 4237. ISSN 2041-1723. Dostupné z: doi:10.1038/s41467-019-12175-9
- WANG, Y., Q. CHEN, C. DENG, Y. ZHENG a F. SUN, 2020. KmerGO: A Tool to Identify Group-Specific Sequences With k-mers. *Frontiers in Microbiology*. 11, 2067. ISSN 1664-302X. Dostupné z: doi:10.3389/fmicb.2020.02067
- WANG, Y., Y. ZHAO, A. BOLLAS, Y. WANG a K. F. AU, 2021. Nanopore sequencing technology, bioinformatics and applications. *Nature Biotechnology*. 39(11), 1348–1365. ISSN 1087-0156, 1546-1696. Dostupné z: doi:10.1038/s41587-021-01108-x
- WICKER, T., F. SABOT, A. HUA-VAN, J. L. BENNETZEN, P. CAPY, B. CHALHOUB, A. FLAVELL, P. LEROY, M. MORGANTE, O. PANAUD et al., 2007. A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics*. 8(12), 973–982. ISSN 1471-0056. Dostupné z: doi:10.1038/nrg2165
- WICKHAM, H., 2016. *ggplot2: Elegant Graphics for Data Analysis*. B.m.: Springer-Verlag New York. ISBN 978-3-319-24277-4. Dostupné z: <https://ggplot2.tidyverse.org>
- XIA, Q., Z. ZHOU, C. LU, D. CHENG, F. DAI, B. LI, P. ZHAO, X. ZHA, T. CHENG, C. CHAI et al., 2004. A draft sequence for the genome of the domesticated silkworm (*Bombyx mori*). *Science (New York, N.Y.)*. 306(5703), 1937–1940. ISSN 1095-9203. Dostupné z: doi:10.1126/science.1102210

- YOSHIDO, A., F. MAREC a K. SAHARA, 2005. Resolution of sex chromosome constitution by genomic in situ hybridization and fluorescence in situ hybridization with (TTAGG) n telomeric probe in some species of Lepidoptera. *Chromosoma*. 114(3), 193–202. ISSN 00095915. Dostupné z: doi:10.1007/s00412-005-0013-9
- YOSHIDO, A., Y. YAMADA a K. SAHARA, 2006. The W Chromosome Detection in Several Lepidopteran Species by Genomic *in situ* Hybridization (GISH). *Journal of Insect Biotechnology and Sericology*. 75(3), 147–151. Dostupné z: doi:10.11416/jibs.75.147
- ZALUCKI, M. P., A. SHABBIR, R. SILVA, D. ADAMSON, L. SHU-SHENG a M. J. FURLONG, 2012. Estimating the Economic Cost of One of the World's Major Insect Pests, *Plutella xylostella* (Lepidoptera: Plutellidae): Just How Long Is a Piece of String? *Journal of Economic Entomology*. 105(4), 1115–1129. ISSN 00220493, 00220493. Dostupné z: doi:10.1603/EC12107
- ZEDEK, F. a P. BUREŠ, 2018. Holocentric chromosomes: from tolerance to fragmentation to colonization of the land. *Annals of Botany*. 121(1), 9–16. ISSN 0305-7364. Dostupné z: doi:10.1093/aob/mcx118
- ZHANG, J., J. LI a Q. ZHOU, 2017. Genomic and Transcriptomic Analyses of Avian Sex Chromosomes and Sex-Linked Genes. In: Guojun SHENG, ed. *Avian and Reptilian Developmental Biology*. New York, NY: Springer New York, Methods in Molecular Biology, s. 69–85. ISBN 978-1-4939-7215-9. Dostupné z: doi:10.1007/978-1-4939-7216-6_4
- ZHANG, L., R. A. STEWARD, C. W. WHEAT a R. D. REED, 2021. High-Quality Genome Assembly and Comprehensive Transcriptome of the Painted Lady Butterfly *Vanessa cardui*. *Genome Biology and Evolution*. 13(7), evab145. ISSN 1759-6653. Dostupné z: doi:10.1093/gbe/evab145
- ZIMMER, C., 2009. On the Origin of Sexual Reproduction. *Science*. 324(5932), 1254–1256. Dostupné z: doi:10.1126/science.324_1254