

Jihočeská univerzita v Českých Budějovicích

Přírodovědecká fakulta

Vývoj molekulárních markerů pro chromozom W u modráška
Polyommatus icarus

Diplomová práce

Bc. Monika Hrubá

Školitel: RNDr. Petr Nguyen, Ph.D.

České Budějovice 2022

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.] – 50 p., Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic.

Annotation

This thesis focuses on development of molecular markers for the W chromosome in the blue butterfly *Polyommatus icarus*. In this blue butterfly, karyotype races with different constitution of the neo-sex chromosomes were reported. The markers were used for PCR sexing of the early developmental stages to obtain sex-specific total RNA samples for testing a role of sexually antagonistic selection in fixation of neo-sex chromosomes.

Finanční podpora

Tento projekt byl financován grantem Grantové agentury České republiky reg. č. 20-20650Y.

Prohlášení

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

V Českých Budějovicích 8. 12. 2022

.....

Bc. Monika Hrubá

Poděkování

Především děkuji mému školiteli, který mi zadal velmi zajímavé téma a vždy si našel čas na konzultaci dalšího postupu nebo výsledků. Dále děkuji také Martině Dalíkové, která mi velmi pomohla s bioinformatickou částí práce, a i přes velkou vzdálenost a časový posun mi vše ochotně vysvětlila. Martina mi také ukázala, jak se sekvenuje technologií Oxford Nanopore, což pro mě bylo opravdu velice zajímavé a jsem ráda, že jsem měla možnost si to vyzkoušet. Velký dík patří také Monče Hospodářské, z jejichž výsledků moje diplomová práce vychází, a která mi také ochotně poskytla vzorky modrásků. Také děkuji všem ostatním kolegům z laboratoře, v jejichž společnosti je radost pracovat. A v neposlední řadě děkuji celé své rodině, mému příteli a jeho rodině. Všichni mi jsou velkou oporou.

Obsah

1.	Úvod.....	1
1.1.	Chromozomy motýlů	1
1.2.	Genomová nestabilita modrásků (Lycaenidae).....	1
1.3.	Pohlavní chromozomy motýlů.....	2
1.4.	Neo-pohlavní chromozomy	3
1.5.	Určení pohlaví u raných vývojových stádií hmyzu	6
2.	Cíle práce	9
3.	Materiál a metody	10
3.1.	Izolace vysokomolekulární genomové DNA a výběr dlouhých fragmentů.....	10
3.2.	Sekvenace DNA technologií Oxford Nanopore	10
3.3.	Bioinformatická analýza	11
3.3.1.	Bioinformatická analýza dat získaných sekvenací na platformě Oxford Nanopore	11
3.3.1.1.	Basecalling	11
3.3.1.2.	Kontrola kvality a příprava dat.....	11
3.3.1.3.	Sestavení genomu	12
3.3.1.4.	Odstranění duplikací.....	12
3.3.2.	Zpracování sekvencí získaných technologií Illumina	12
3.3.3.	K-merová analýza.....	13
3.4.	Navržení primerů	16
3.5.	Izolace DNA z modrásků druhu <i>P. icarus</i>	18
3.6.	Otestování navržených primerů pomocí PCR	18
3.7.	Použití nástroje Minimap2 pro zjištění oblasti nasedání primerů.....	21
3.8.	Použití primerů pro určení pohlaví čerstvě vylíhlých housenek.....	21
4.	Výsledky	23
4.1.	Sekvenace genomu samice modráska <i>P. icarus</i> technologií Oxford Nanopore	23

4.2.	K-merová analýza	24
4.2.1.	Výběr pohlavně specifických k-merů.....	24
4.3.	Navržení a otestování primerů	26
4.4.	Použití programu Minimap2 pro zjištění oblasti nasedání primerů.....	31
4.5.	Určení pohlaví u čerstvě vylíhlých housenek	32
5.	Diskuze	33
5.1.	Bioinformatická analýza	33
5.2.	Navržení a otestování primerů	37
5.3.	Nalezení sekvencí z chromozomu neo-W.....	38
6.	Závěr	40
7.	Literatura.....	41

1. Úvod

1.1. Chromozomy motýlů

Motýli (řád Lepidoptera) mají holocentrické chromozomy s delokalizovanou centromerickou aktivitou. Takové chromozomy na rozdíl od monocentrických chromozomů mohou snáze podléhat přestavbám, aniž by došlo k problémům s rovnoměrnou segregací chromozomů během buněčného dělení. Při fúzi holocentrických chromozomů totiž nevznikají dicentrické chromozomy, a naopak při fragmentaci se netvoří acentrické chromozomální fragmenty, které by byly následně ztraceny (Murakami a Imai, 1974; Mola a Papeschi, 2006; Bureš a Zedek, 2014). I přesto však u většiny motýlů nalézáme konzervovaný počet chromozomů blízký ancestrálnímu počtu $n = 31$ (Ahola a kol., 2014). Zároveň ale můžeme u motýlů nalézt největší variabilitu v chromozomálních počtech v celé živočišné říši. V haploidním stavu se u motýlů vyskytuje od $n = 7$ do $n = 226$ chromozomů (White, 1973; Lukhtanov, 2015).

U motýlů je obtížné identifikovat jednotlivé chromozomy kvůli jejich malé velikosti, velkému počtu a značné uniformitě dané absencí centromery a jiných morfologických struktur. U motýlích chromozomů nefungují ani klasické proužkovací techniky, takže v karyotypu motýlů nelze identifikovat jednotlivé páry chromozomů klasickými cytogenetickými metodami (Traut a kol., 2007). Dodnes byla karyotypována jen malá část z dosud popsaných druhů motýlů a jednotlivé chromozomy byly identifikovány jen u bource morušového (*Bombyx mori*; Traut, 1976). Dnes však lze jednotlivé chromozomy motýlů identifikovat díky rozvoji molekulárně-cytogenetických metod pomocí fluorescenční *in situ* hybridizace a různých variant této metody (Goldsmith a Marec, 2010).

1.2. Genomová nestabilita modrásků (Lycaenidae)

Řád Lepidoptera zahrnuje asi 13 % popsaných druhů organismů (Mora a kol., 2011; Stork, 2018). Přestože se jedná o rozmanitou skupinu, zachovala si konzervovanou genomovou strukturu i mezi vzdálenými taxony, přičemž většina zástupců tohoto řádu má v haploidním stavu $n = 31$ chromozomů (Robinson, 1971), což je, jak už bylo uvedeno výše, považováno za ancestrální stav. Mezi zástupci řádu Lepidoptera není konzervovaný pouze počet chromozomů, ale také syntenie genů na jednotlivých chromozomech a jejich pořadí (Van't Hof a kol., 2013; Ahola a kol., 2014).

U čeledi modráskovití (Lycaenidae) můžeme však pozorovat hned několik odlišností. Významným rozdílem je, že u modrásků je nejčastější počet chromozomů $n = 23-24$ (Robinson, 1971). Modrásci jsou zároveň skupinou, ve které nalezneme největší rozdíly v chromozomálních počtech, které se pohybují od $n = 10$ po $n = 226$. I blízkce příbuzné druhy se tedy výrazně liší počtem chromozomů, což je způsobeno různými chromozomálními přestavbami. Většinu modrásků s vysokými počty chromozomů řadíme do tří skupin tribu Polyommagini, které dříve představovaly tři samostatné rody *Agrodiaetus* ($n = 10-125$), *Lysandra* ($n = 24-92$) a *Plebicula* ($n = 134-226$; Kandul a kol., 2004; Lukhtanov, 2015). Dle nejnovější taxonomické revize jsou rody *Agrodiaetus* a *Plebicula* řazeny do rodu *Polyommatus* (Talavera a kol., 2013). Vysoký počet chromozomů nalezený u některých druhů modrásků pravděpodobně není způsoben polyploidizací, ale vzniká díky opakovaným fragmentacím chromozomů, protože chromozomy jsou tím menší, čím více jich daný druh má (Lorković, 1990 citovaný v Kandul a kol., 2007) To bylo potvrzeno také měřením velikosti genomu průtokovou cytometrií (Kreklová, 2020).

Fragmentaci však zřejmě nepodléhají všechny chromozomy stejně, jelikož Lukhtanov a kol. (1997) našli u modráška *Polyommatus (Agrodiaetus) damone damone*, který má vysoký počet chromozomů, dva chromozomy, které byly výrazně větší než ostatní. Tyto velké chromozomy mohly být vytvořeny fúzí autozomů (Ennis, 1976). Podle alternativních hypotéz však velké chromozomy nalezené u modrásků s fragmentovaným karyotypem představují pohlavní chromozomy (Carabajal Paladino a kol., 2019).

U modráška *Lysandra coridon* byla dokonce nalezena vnitrodruhová variabilita v počtu chromozomů, přičemž počty chromozomů se lišily mezi jednotlivými evropskými populacemi. U jedinců ze Španělska a Francie bylo v haploidním stavu nalezeno 87-88 chromozomů, kdežto u balkánské populace bylo chromozomů více, $n = 90-92$ (Coutsis a kol., 2001).

1.3. Pohlavní chromozomy motýlů

Motýli (řád Lepidoptera) a také sesterská skupina chrostíků (Trichoptera) mají chromozomální určení pohlaví typu *Abraxas* s heterogametickými samicemi, které mají pohlavní chromozomy Z a W, nebo jen jeden chromozom Z. Naproti tomu samci mají dva pohlavní chromozomy Z a jedná se tedy o homogametické pohlaví. Motýli a chrostíci představují jediné dva řády hmyzu se striktně heterogametickými samicemi (Traut a kol., 2007). Tento systém determinace pohlaví se pravděpodobně vyvinul u společného předka

těchto dvou skupin před více než 300 miliony let (Kawahara a kol., 2019). Za ancestrální stav u motýlů a chrostíků je považována konstituce pohlavních chromozomů ZZ/ZO, kdežto systém ZZ/ZW je odvozený a byl nalezen u většiny zástupců skupiny Ditrysia, do kterého řadíme 99 % všech druhů řádu Lepidoptera (Traut a Marec, 1996; Kristensen a Skalski, 1999). Kromě těchto základních konstitucí, jsou u motýlů také časté různé odvozené varianty, kdy jsou pohlavní chromozomy zmnoženy, například ZZ/ZW₁W₂ a Z₁Z₁Z₂Z₂/Z₁Z₂W, které jsou výsledkem chromozomálních přestaveb. Může dojít také ke ztrátě pohlavního chromozomu W, čímž může druhotně vzniknout systém ZZ/ZO (Traut a Marec, 1996). Pokud je chromozom W přítomen, pak je u většiny druhů tvořen heterochromatinem. Dosud na něm bylo nalezeno jen málo genů a z velké části je tvořen repeticemi. Naopak pohlavní chromozom Z je u většiny druhů tvořen euchromatinem a nachází se na něm mnoho protein kódujících genů. Sekvence pohlavních chromozomů W a Z se od sebe značně liší a chromozomy se mohou lišit i velikostí (Traut a kol., 2007).

1.4. Neo-pohlavní chromozomy

Během evoluce organismů dochází k chromozomálním přestavbám, kterým podléhají jak autozomy, tak pohlavní chromozomy. Fúzí pohlavních chromozomů s autozomy vznikají tzv. neo-pohlavní chromozomy. U obratlovců bylo zjištěno, že k fúzím mezi pohlavními chromozomy a autozomy dochází obecně častěji u druhů s heterogametickými samci (XX/XY), než u druhů s heterogametickými samicemi (ZZ/ZW; Pokorná a kol., 2014; Pennell a kol., 2015). Nejpočetnější skupinu obratlovců s heterogametickými samicemi představují ptáci, avšak neo-pohlavní chromozomy se u nich nacházejí sporadicky (Pennell a kol., 2015). Zatím byly objeveny například v nadčeledi Sylvioidea (Pala a kol., 2012; Dierickx a kol., 2020) a komparativní analýzou čeledí Alaudidae a Panuridae byl nalezen největší pohlavní chromozom u ptáků, který je tvořen ze čtyř chromozomů kuřete (jedná se o chromozomy Z, 3, 4A a 5; Sigeman a kol., 2019). Další skupinou ptáků s neo-pohlavními chromozomy jsou papoušci, kteří mají chromozomy neo-Z a neo-W, které vznikly fúzí pohlavních chromozomů s autozomem 11 už u společného předka papoušků. U papouška mnišího pak byla nalezena fúze těchto neo-pohlavních chromozomů s autozomem 25 (Huang a kol., 2022).

Neo-pohlavní chromozomy byly intenzivně studovány například u octomilek *Drosophila pseudoobscura*, *Drosophila miranda* a *Drosophila albomicans*, u kterých byly popsány neo-pohlavní chromozomy různého stáří. Autozomy, které zfúzovaly s pohlavními chromozomy, přestaly díky chiasmatické meióze samců rekombinovat a představují tak různá

stádia molekulární degradace pohlavních chromozomů (Bachtrog, 2013). Neo-pohlavní chromozomy jsou tedy vhodné ke studiu diferenciaci pohlavních chromozomů u taxonů, které mají poměrně staré chromozomální systémy určení pohlaví a jejich pohlavní chromozomy jsou tedy plně diferencované (Pennell a kol., 2015).

U motýlů je však situace jiná. U všech zatím testovaných druhů čeledi Tortricidae byly pozorovány dva velké elementy, které odpovídají páru pohlavních chromozomů (Šíchová a kol., 2013). U obaleče jablečného *Cydia pomonella* odhalilo komparativní fyzické mapování, že se jedná o neo-pohlavní chromozomy, které byly vytvořeny fúzí pohlavního chromozomu Z s autozomem odpovídajícím chromozomu 15 bource morušového, *B. mori*. Neo-pohlavní chromozomy pravděpodobně vznikly už u společného předka podčeledi Tortricinae a Olethreutinae, jelikož byla tato fúze detekována také u dalších zástupců čeledi Tortricidae (Nguyen a kol., 2013).

Také u čeledí Pyralidae, Oecophoridae a Gelechiidae byl nalezen velký chromozomální pár (Ennis, 1976), který se později podařilo identifikovat jako pár pohlavních chromozomů u makadlovky *Tuta absoluta*, což je zástupce čeledi Gelechiidae. U zástupců nadčeledi Gelechioidea, do které spadají zmíněné čeledi Oecophoridae a Gelechiidae, se podařilo potvrdit fúzi pohlavních chromozomů a autozomů. Na neo-pohlavních chromozomech *T. absoluta* byly identifikovány dva původně autozomální syntenní bloky homeologní k chromozomům 7 a 27 bource morušového. Při analýze zástupců zkoumaných skupin se zjistilo, že už u společného předka nadčeledi Gelechioidea došlo k fúzi autozomu 7 s pohlavním chromozomem Z. Touto fúzí tedy vznikl neo-Z pohlavní chromozom, jež dále u zástupců tribu Gnoreschemini (podčeleď Gelechiinae) prodělal fúzi s autozomem 27 (Carabajal Paladino a kol., 2019).

Dalším příkladem mohou být neo-pohlavní chromozomy neo-Z a neo-W objevené u monarchy *Danaus plexippus* a také u dalších druhů rodu *Danaus* pomocí komparativní genomové a cytogenetické analýzy. Jelikož jsou neo-pohlavní chromozomy společné celému rodu *Danaus*, můžeme odhadovat, že vznikly před více než 5 miliony let. Neo-Z chromozom vznikl fúzí ancestrálního pohlavního chromozomu Z s autozomem, jež je homeologní s chromozomem 16 *B. mori*. U neo-W chromozomu bylo pomocí cytogenetických metod zjištěno, že dvě třetiny tohoto chromozomu jsou silně heterochromatizované, kdežto zbylá třetina je tvořena euchromatinovou oblastí. Toto zjištění podporuje teorii o vzniku neo-pohlavního chromozomu fúzí ancestrálního pohlavního chromozomu W s autozomem (Mongue a kol., 2017).

Můžeme tedy vidět, že k fúzím pohlavních chromozomů s autozomy dochází nejen u obratlovců, ale i u bezobratlých. Konkrétně u motýlů jsou takové fúze časté a může k nim docházet i opakovaně. Často dochází k fúzi s autozomem jak u chromozomu W, tak u chromozomu Z, a tudíž pak vzniká velký pár neo-pohlavních chromozomů (Carabajal Paladino a kol., 2019). Příkladem motýla s neo-pohlavními chromozomy je i modrásek *Polyommatus icarus*. Mezi samicemi tohoto druhu byly nalezeny rozdíly ve stavbě pohlavního chromozomu W. Při studiu genomu tohoto modráška pomocí genomové *in situ* hybridizace na samičích pachytenních bivalentech u většiny samic samičí celogenomová sonda hybridizovala jen na intersticiální část chromozomu W. U jedné samice však použitá sonda hybridizovala na celou polovinu chromozomu W. Pomocí metody array CGH byly porovnávány genomové sekvence modráška *P. icarus* a hnědáška kostkovaného (*Melitaea cinxia*), který byl zvolen jako zástupce s ancestrálním genomem, jelikož má v haploidním stavu 31 chromozomů (Aholá a kol., 2014). Tato metoda však neidentifikovala žádný syntenní blok, jež by byl autozomální v ancestrálním karyotypu a Z-vázaný u *P. icarus*, což naznačuje, že pokud jsou u *P. icarus* skutečně přítomny neo-pohlavní chromozomy, tak jsou velmi mladé a dosud nediferencované. U *P. icarus* byla také provedena fluorescenční *in situ* hybridizace s telomerickou sondou, která hybridizovala na konce všech chromozomů a také do jedné oblasti na rozhraní heterochromatinové a euchromatinové části u bivalentu tvořeného chromozomem Z a neo-W, který obsahoval dvě euchromatinové oblasti. Je tedy pravděpodobné, že u tohoto modráška nejprve došlo k fúzi pohlavních chromozomů Z a W s autozomem za vzniku neo-Z a neo-W a následně neo-W chromozom zfúzoval s dalším autozomem a jeho homolog se stal Z₂ chromozomem (Kreklová, 2020).

Fixace neo-pohlavních chromozomů u motýlů může být usnadněna pohlavně antagonistickou selekcí, tj. selekcí alel prospěšných pouze pro jedno pohlaví a potenciálně škodlivých pro druhé, pro které je vazba na pohlavní chromozom prospěšná (Charlesworth a Charlesworth, 1980). U *B. mori* byly nalezeny klastry genů na autozomech 2, 10, 15 a 16, které obsahují 75 % genů s expresí specifickou pro vaječníky (Suetsugu a kol., 2013). Fúze chromozomu Z s autozomem odpovídajícím chromozomu 15 u *B. mori* byla nalezena u obaleče *Cydia pomonella* (Nguyen a kol., 2013). U běláška *Pieris napi* byla nalezena fúze chromozomu Z s autozomem odpovídajícím chromozomu 2 u *B. mori* (Hill a kol., 2019) a stejná fúze byla nalezena také u předitky *Yponomeuta evonymella* (Provazníková a Nguyen, nepublikované výsledky). U monarchy *Danaus plexippus* byla nalezena fúze chromozomu Z s autozomem odpovídajícím chromozomu 16 u *B. mori* (Mongue a kol., 2017). Z těchto výsledků lze usuzovat, že pohlavně antagonistická selekce, která zvýhodňuje vazbu genů

důležitých pro reprodukční úspěšnost samic na pohlavní chromozomy, může být mechanismem podporujícím fixaci fúzí autozomů s pohlavními chromozomy.

Modrásek *P. icarus* by mohl být dobrým druhem pro testování hypotézy o roli pohlavně antagonistické selekce ve fixaci neo-pohlavních chromozomů. Jelikož u tohoto druhu byla nalezena vnitrodruhová variabilita v konstituci pohlavních chromozomů (Kreklová, 2020), bylo by zajímavé zjistit, zda autozomy, které fúzovaly s pohlavními chromozomy, obsahují geny výhodné pro jedno pohlaví, jako jsou například geny pro fertilitu, které se exprimují v gonádách. Pro testování těchto hypotéz by bylo vhodné mít u *P. icarus* dostupné markery pro jednoduché určení pohlaví a ideálně také markery pro rychlé odlišení karyotypových ras.

1.5. Určení pohlaví u raných vývojových stádií hmyzu

Jak bylo zmíněno výše, pro testování pohlavně antagonistické selekce u *P. icarus* by bylo vhodné mít k dispozici markery pro jednoduché určení pohlaví. I v jiných případech je pro práci s hmyzem často nutné znát pohlaví daného jedince. Možnost určit pohlaví jedinců by bylo možné využít také při ochraně proti škůdcům. Kromě jiných metod ochrany proti škůdcům se stále častěji používá technika sterilního hmyzu pro potlačení škůdců z řádu Lepidoptera (Bloem a Bloem, 2000). Možnost využít markery pro identifikaci pohlaví by jednoznačně usnadnila přípravu sterilních samců využívaných při technice sterilního hmyzu (Marec a kol., 2005; 2007). U některých organismů, se vyskytuje heterochromatinové tělísko, které lze pozorovat v interfázních buňkách jednoho pohlaví. Takovým příkladem může být Baarovo tělísko u žen, tvořené inaktivovaným pohlavním chromozomem X, které lze pozorovat například v buňkách odebraných bukálním stěrem (Miller, 2006). Také u řádu Lepidoptera lze u většiny druhů nalézt v interfázních jádrech samic heterochromatinové tělísko, které je tvořeno pohlavním chromozomem W. Toto tělísko je tedy dobrým znakem pro určení pohlaví. U druhů se systémem pohlavních chromozomů ZZ/Z0 však toto tělísko chybí (Traut a Marec, 1996). A také u druhů, kde se vytvořil neo-W chromozom může heterochromatinové tělísko chybět (Vlašánek a kol., 2017). Kromě výše zmíněných komplikací je navíc tato metoda velmi pracná a časově náročná.

Další možností, jak určit pohlaví, je průtoková cytometrie, založená na odlišném obsahu DNA v buňkách různých pohlaví, tedy na odlišné velikosti pohlavních chromozomů Z a W (nebo X a Y), případně na absenci chromozomu W (Nakamura a kol., 1990; Doležel a Göhde, 1995). Tato metoda byla úspěšně použita u třech zástupců čeledi vakonošovití, u kterých chybí

pohlavní chromozom W. U všech testovaných druhů bylo průtokovou cytometrií zjištěno, že genom samic je průkazně menší než genom samců (Hejníčková a kol., 2019).

U zástupců řádu Lepidoptera byla navržena také molekulární metoda pro identifikaci pohlaví pomocí kvantitativní PCR, při které se amplifikuje část genu *kettin* ležícího na chromozomu Z a sledují se rozdíly v počtu kopií mezi jedinci. Metoda je tedy založena na faktu, že u samců jsou přítomné dva chromozomy Z, zatímco u samic se nachází pouze jeden chromozom Z. Tato metoda umožnila správně určit pohlaví u zástupců řádu Lepidoptera pocházejících z různých skupin, a to v jakémkoli vývojovém stádiu (Belousova a kol., 2019). Většina jiných testů pro určení pohlaví využívá amplifikaci pohlavně vázaných úseků DNA pomocí PCR, jelikož tato metoda je rychlá a jednoduchá.

U všech ptáků s výjimkou běžců je používán univerzální test identifikace pohlaví, který využívá amplifikaci dvou různě velkých konzervovaných alel genu CHD. Jedna z nich se nachází na pohlavním chromozomu W a je tedy specifická pro samice, zatímco druhou, ležící na chromozomu Z, je možné detekovat u obou pohlaví (Griffiths a kol., 1998). Na stejném principu je založena také metoda umožňující určení pohlaví u vrtule velkohlavé (*Ceratitis capitata*), kde se využívá amplifikace sekvencí ležících na pohlavních chromozomech X a Y (Douglas a kol., 2004).

Pouze u několika druhů řádu Lepidoptera existují molekulární markery specifické pro jednotlivá pohlaví. Mezi takové druhy patří například bourec morušový (*B. mori*), u kterého jsou k dispozici markery specifické pouze pro samice (Abe a kol., 1998). Nicméně, tyto markery se nepoužívají, protože u bource je možné pohlaví odlišit na základě fenotypu daného jedince, a to během embryonálního i larválního stádia a také ve stádiu kukly (Nagaraju, 1996). Dalším druhem, u kterého byly nalezeny molekulární markery pro určení pohlaví, je zavíječ kukuřičný (*Ostrinia nubilalis*), u kterého byly nalezeny mikrosatelity vázané na chromozomy W a Z, které byly použity pro populační genetickou analýzu (Coates a Hellmich, 2003).

Podle dosud získaných dat je pohlavní chromozom W u zástupců řádu Lepidoptera tvořen převážně repetitivními sekvencemi a transpozony (Sahara a kol., 2003; Abe a kol., 2005; Vítková a kol., 2007; Yoshido a kol., 2007). Z tohoto důvodu nelze použít standartní molekulární analýzu pro jeho identifikaci. U obaleče jablečného (*C. pomonella*) se podařilo osekvenovat fragmenty chromozomu W získané laserovou mikrodisekcí tělísek sex-chromatinu z polyploidních buněk samic. S použitím získaných sekvencí byly nalezeny dvě sekvence, CpW2 a CpW5, vázané na chromozom W. Nicméně samy o sobě nebyly dostatečně dlouhé, aby je bylo možné využít pro spolehlivou metodu určení pohlaví, která by byla dostatečně robustní a reprodukovatelná (Fuková a kol., 2007). Avšak sousední oblast jedné

z nich byla využita jako molekulární marker pro určení samičího pohlaví. Tento marker (CpW2-*EcoRI*) byl následně použit v multiplex PCR v kombinaci s primery pro gen *Cpper* (gen orthologní ke genu *period*), který sloužil jako pozitivní kontrola v PCR. Takto bylo možno určit pohlaví i u vajíček a čerstvě vylíhlých housenek obaleče jablečného z různých zemí Evropy a Severní Ameriky (Fuková a kol., 2009). Laserová mikrodisekce byla použita také při sekvenaci chromozomu W u *Ephesia kuehniella*, která objevila řadu rodin mobilních elementů, mikrosatelity a místa nedávné inserce mitochondriální DNA. I přesto, že nebyly nalezeny žádné protein-kódující geny, mohly by být získané sekvence použity pro návržení PCR markerů pro určení pohlaví u tohoto druhu (Traut a kol., 2013).

Specifické primery pro určení pohlaví v jakémkoli životním stádiu byly získány také u obaleče mramorovaného (*Lobesia botrana*; Aguirre a kol., 2020). Pro jejich návržení byly použity sekvence homologní k sekvencím CpW2 a CpW5, které byly popsány u obaleče jablečného (Fuková a kol., 2009). Pro spolehlivé určení pohlaví u obaleče mramorovaného a vyloučení falešně negativních výsledků byla navržena metoda duplexní PCR, ve které dochází k amplifikaci sekvence dlouhé 453 párů bazí (bp) specifické pro chromozom W a zároveň k amplifikaci 151 bp dlouhého úseku genu pro elongační faktor 1 alfa (*EF-1 α*), který funguje jako pozitivní kontrola (Aguirre a kol., 2020).

Dalším zástupcem řádu Lepidoptera, u kterého byl nalezen marker specifický pro chromozom W umožňující určení pohlaví u ranných vývojových stádií, je černopáska bavlníková (*Helicoverpa armigera*). U tohoto významného škůdce nelze rozpoznat pohlaví jedince ve stádiu vajíčka ani larvy, a proto bylo potřeba vytvořit molekulární marker k určení pohlaví u těchto ranných stádií. K tomu byl použit PCR screening 17 transkriptů neobsahujících transpozony. Tyto transkripty byly vybrány z celkem 4855 sekvenačních čtení vázaných na chromozom W, jež byly identifikovány mapováním veřejně dostupného transkriptomu z vajíček obou pohlaví na samčí genom. Tímto postupem byla nalezena sekvence, která se u samce vůbec nenacházela a je tedy specifická pro samice. V této sekvenci byl odhalen marker GUW1, který umožňuje spolehlivě detekovat jedince samičího pohlaví (Deng a kol., 2020).

2. Cíle práce

U modráška *Polyommatus icarus* bylo dosud možné zjistit pohlaví jedince pouze u dospělců, díky pohlavnímu dimorfismu. Další možností určení pohlaví je pitva larev. Tato metoda je však zdlouhavá a náročná a lze ji použít pouze pro větší larvy 3.-5. instaru. Proto by práci s modráskem *P. icarus* výrazně zjednodušila dostupnost PCR markerů, které by umožnily rychlou a jednoduchou identifikaci pohlaví v jakémkoli vývojovém stádiu jedince, a to za použití malého množství tkáně. Dostupnost takového markeru by umožnila například studovat roli pohlavně antagonistické selekce ve fixaci neo-pohlavních chromozomů.

Cílem této diplomové práce je u modráška *P. icarus* identifikovat sekvence pocházející z pohlavního chromozomu W pomocí srovnávací analýzy resekvenančních genomových dat samců a samic tohoto druhu. Dalším cílem je navrhnout a otestovat markery pro určení pohlaví tohoto modráška a funkční markery použít pro genotypování čerstvě vylíhlých housenek.

3. Materiál a metody

3.1. Izolace vysokomolekulární genomové DNA a výběr dlouhých fragmentů

Genomová DNA byla izolována ze samice housenky modráška *Polyommatus icarus* sebrané ve Františkově v září roku 2020, která byla následně uchovávána v -80 °C. Pro izolaci byl použit kit Nanobind Tissue Big DNA Kit (Pacific Biosciences, USA) s výrobcem dodanými pufrů pro extrakci z hmyzích tkání. Při izolaci bylo postupováno podle protokolu High Molecular Weight Insect DNA Extraction Protocol s těmito odchylkami: ve 2. kroku bylo použito 400 µl pufru CT, v 8. kroku bylo použito 300 µl pufru PL1, v 16. kroku bylo přidáno 60 µl pufru BL3 a v následujícím kroku bylo použito 500 µl isoporpanolu, v kroku číslo 29 bylo přidáno 100 µl pufru EB a v kroku 33 byl vzorek ponechán přes noc v lednici. Koncentrace DNA byla změřena pomocí přístroje Qubit 4 Fluorometer (ThermoFisher Scientific, USA) pomocí kitu Qubit dsDNA BR Assay Kit (ThermoFisher Scientific, USA). Kvalita získané DNA byla ověřena také přístrojem NanoDrop One (ThermoFisher Scientific, USA).

Ze získané DNA byly vybrány pouze dlouhé fragmenty pomocí kitu Short Read Eliminator Kit XS (Pacific Biosciences, USA) podle protokolu výrobce. Získaná DNA očištěná od krátkých fragmentů byla inkubována 20 minut při pokojové teplotě, ponechána přes noc v lednici a po rozpuštění byla změřena její koncentrace pomocí kitu Qubit 1x dsDNA HS Assay Kit na zařízení Qubit 4 Fluorometer (ThermoFisher Scientific, USA).

3.2. Sekvence DNA technologií Oxford Nanopore

Před samotnou sekvenací byla genomová DNA opravena pomocí kitu NEBNext Companion Module for Oxford Nanopore Technologies Ligation Sequencing (New England BioLabs, USA), na konce molekul byly navázány adaptory a směs byla přečištěna. Vše bylo provedeno podle protokolu Genomic DNA by Ligation (SQK-LSK110) za použití chemikálií uvedených v tomto protokolu (New England BioLabs, USA; Oxford Nanopore Technologies, Velká Británie). Podle stejného protokolu byla také provedena samotná sekvence DNA na FlowCell R10 přístrojem MinION Mk1B (Oxford Nanopore Technologies, Velká Británie). Sekvence probíhala při -180 mV po dobu 72 hodin bez automatického basecallingu a přístroj byl ovládán pomocí aplikace MinKNOW (Ip a kol., 2015). Pro získání většího množství dat byl zbytek vyizolované DNA, který nebyl dosud použit, připraven stejným způsobem, jak bylo zmíněno výše, a nová knihovna byla sekvenována na stejné flow cell po jejím promytí pomocí Flow Cell Wash Kit (EXP-WSH004; Oxford Nanopore Technologies, Velká Británie).

Tentokrát bylo nastaveno -240 mV a po necelých 24 hodinách byla sekvenační reakce zastavena.

3.3. Bioinformatická analýza

Sekvenační data byla zpracována v operačním systému Linux v prostředí České Národní Gridové Infrastruktury Metacentrum (<https://www.metacentrum.cz/cs/>), které umožňuje pracovat i s daty o velkém objemu. Pro poslední část bioinformatické analýzy bylo použito prostředí RStudio (Boston, USA). Použité skripty jsou dostupné na: <https://github.com/hrubam/w-markers-polyommatus-icarus>.

3.3.1. Bioinformatická analýza dat získaných sekvenací na platformě Oxford Nanopore

Dva datasety získané během dvou sekvenací byly nejprve sloučeny a následně zpracovávány dohromady.

3.3.1.1. Basecalling

Pro basecalling byl použit program Guppy verze 4.4.1, který převádí formát FAST5 získaný sekvenací na formát FASTQ. Formát FAST5 obsahuje informace o změnách elektrického proudu při průchodu jednotlivých molekul skrz proteinový pór. Během procesu basecallingu se tato data převádí na sekvenci jednotlivých bazí a každé bazi je přiřazeno skóre kvality. Výstupem je pak formát FASTQ. Během basecallingu bylo také provedeno filtrování na kvalitu a sekvence se skórem kvality menším než 7, tedy sekvence, u nichž byla pravděpodobnost správného přečtení menší než přibližně 80 %, byly odstraněny.

3.3.1.2. Kontrola kvality a příprava dat

U získaných dat ve formátu FASTQ byly vyhodnoceny základní statistiky pomocí softwarového nástroje NanoPlot, který umožňuje získat například hodnotu N50, průměrnou délku a kvalitu sekvenačních čtení, počet získaných sekvenačních čtení a celkový počet osekvenovaných bazí (De Coster a kol., 2018).

Na základě výsledků získaných nástrojem NanoPlot byly získané sekvence filtrovány na délku pomocí nástroje NanoFilt (De Coster a kol., 2018). Byla použita dvě různá nastavení: pro odstranění sekvenačních čtení kratších než 5000 bp a kratších než 10000 bp. Byly tedy

získány dva datasets, pro které byly opět vyhodnoceny základní statistiky pomocí nástroje NanoPlot.

Dále byl pro zlepšení kvality sekvencí použit program Canu verze 2.1.1 (Koren a kol., 2017), který umožňuje opravit sekvence na základně překryvů mezi nimi.

3.3.1.3. Sestavení genomu

Pro sestavení genomu byl použit nástroj Flye verze 2.8 (Kolmogorov a kol., 2020), který umožňuje *de novo* složení genomu z dlouhých sekvenačních čtení s vysokou chybovostí. Jelikož byly použity dva různé vstupní datasets (sekvenační čtení s minimální délkou 5000 bp a 10000 bp), byly získány dva složené genomy o různé kvalitě. Nástroj Flye zároveň provedl 1 kolo tzv. „polishingu“, tedy opravy sekvencí sami sebou.

Pro vyhodnocení kvality získaných genomů byly použity nástroje BUSCO (Seppey a kol., 2019; Simão a kol., 2015) a Quast (Gurevich a kol., 2013).

3.3.1.4. Odstranění duplikací

Pomocí programu BUSCO bylo zjištěno, že oba složené genomy obsahují velké množství duplikovaných genů. Proto byl použit nástroj PurgeDups verze 1.0.1 (Guan a kol., 2020) pro odstranění duplikací a vytvoření nového složeného genomu. Kvalita získaných genomů byla opět ověřena pomocí programů BUSCO a Quast. Jako nejlepší složený genom byl vybrán ten, který byl vytvořen ze sekvencí o minimální délce 5000 bp a byl upraven pomocí programu PurgeDups. Tento složený genom byl použit pro další kroky této diplomové práce.

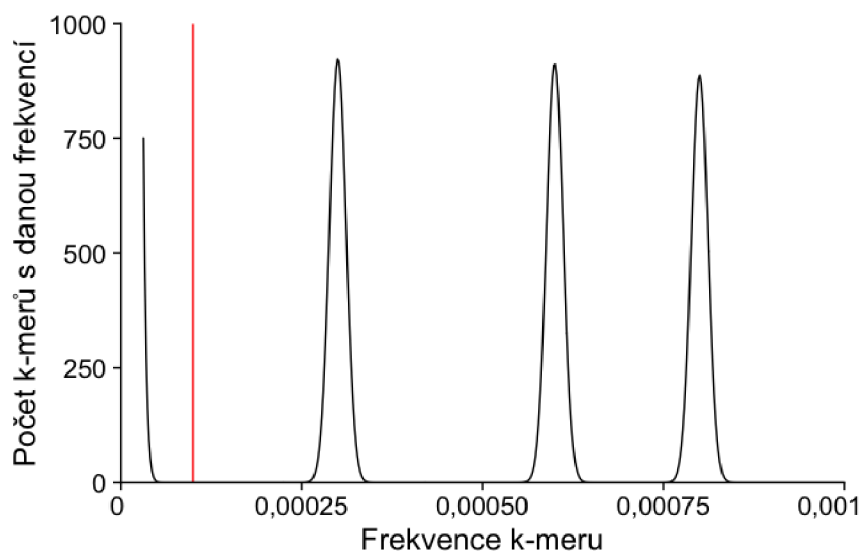
3.3.2. Zpracování sekvencí získaných technologií Illumina

V této práci byly použity také sekvence získané technologií Illumina, které mi poskytl můj školitel Petr Nguyen. Jednalo se o genomy šesti samců a čtyř samic modráška *P. icarus*, které byly sekvenovány technologií Illumina při hloubce čtení 20x (Illumina, USA). Byla použita metoda párových čtení („paired end“) a délka jednotlivých sekvenačních čtení byla 150 nukleotidů. U těchto sekvencí byla nejprve ověřena jejich kvalita pomocí nástroje FastQC verze 0.11.5 (Babraham Bioinformatics, Babraham Institute, Velká Británie, dostupný z: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Na základě získaných výsledků bylo odstraněno 5 bazí na začátku a na konci každé sekvence pomocí nástroje

Trimmomatic verze 0.36 (Bolger a kol., 2014), který byl také použit k odstranění adaptorů, tedy technických sekvencí. Zároveň byla pomocí tohoto programu provedena filtrace na minimální délku sekvencí 50 bp. Dále byly odstraněny poly-G kontaminace pomocí nástroje Trim Galore verze 0.6.2 (Babraham Bioinformatics, Babraham Institute, Velká Británie, dostupný z: https://www.bioinformatics.babra-ham.ac.uk/projects/trim_galore/). Po těchto úpravách byly všechny soubory náležící k jednomu jedinci konkatenovány do jednoho pomocí Linuxového příkazu cat.

3.3.3. K-merová analýza

Takto připravené sekvence byly následně rozděleny na k-mery pomocí nástroje KmerGO (Wang a kol., 2020). Tento program byl použit se dvěma různými nastaveními – pro vytvoření k-merů dlouhých 21 a 55 bp. Oba výstupy se skládaly z několika souborů a dále bylo pracováno s oběma datasey samostatně. Z každé z výsledných tabulek se získanými k-mery a jejich frekvencemi u jednotlivých jedinců byl proveden náhodný výběr 20000 řádků, které byly následně sloučeny do jednoho souboru. Tento soubor byl následně použit pro vytvoření k-merového spektra v prostředí RStudio (Boston, USA). K-merové spektrum je graf znázorňující počet různých k-merů o konkrétní frekvenci ve vzorku. Z tohoto grafu je možné odhalit k-mery obsahující sekvenační chyby, jelikož takové k-mery vznikají náhodně, sekvenčně se liší a mají nízkou frekvenci. Proto v grafu tvoří jeden vrchol oddělený od ostatních k-merů, u kterých jejich frekvence ve vzorku odpovídá počtu pročtení sekvenovaného genomu. Na základě k-merových spekter vytvořených pro jednotlivé jedince byla stanovena hranice oddělující chybové k-mery (viz Obrázek 1). Jako hraniční byla stanovena frekvence 0,0001 a všechny k-mery s nižší frekvencí byly z další analýzy odstraněny. Tato hranice byla stejná pro spektra vytvořená z 21-merů i 55-merů.



Obrázek 1: Stanovení hranice pro odstranění chybových k-merů. Na obrázku je ukázáno k-merové spektrum vytvořené z 55-merů u jednoho samce. Červená svislá linka označuje hranici pro odstranění chybových k-merů, která byla stanovena na frekvenci 0,0001 a byla zvolena tak, aby byla stejná u všech jedinců.

Dalším krokem byl výběr pohlavně specifických k-merů. Z tabulek vytvořených programem KmerGO, které obsahují pro každý k-mer frekvenci u jednotlivých jedinců, byly vybrány k-mery specifické pro samice. Tedy ty k-mery, které měly u všech samic frekvenci větší než 0,0001 (hranice oddělující k-mery obsahující sekvenční chyby a informativní k-mery) a zároveň byla jejich frekvence u všech samců rovna nule. Stejným způsobem byly vybrány k-mery specifické pro samce, tedy k-mery s frekvencí u všech samců větší než 0,0001 a zároveň u všech samic rovnou nule. Takto vybrané k-mery specifické pro jednotlivá pohlaví byly mapovány na referenční genom, tedy na složený genom vytvořený ze sekvencí získaných technologií Oxford Nanopore (viz kapitoly 3.2. a 3.3.1.). K tomuto mapování byl použit nástroj Bowtie2 verze 2.3.5.1 (Langmead a Salzberg, 2012) s nastavením, že pro namapování k-meru na genom je nutná 100% shoda sekvencí (--score-min 'L,0,0'). Dále byl použit parametr „-a“ umožňující namapovat jeden k-mer na více míst v referenčním genomu. Výsledkem byly čtyři soubory ve formátu SAM: kontigy s namapovanými 21-mery specifickými pro samice, kontigy s namapovanými 21-mery specifickými pro samce a podobně dva soubory pro 55-mery. Tyto soubory byly za použití softwaru Samtools verze 1.11-intel-19.0.4-wzth4e4 zkomprimovány do formátu BAM (Li a kol., 2009). Dále byl využit software Bedtools verze 2.26.0 (Quinlan a Hall, 2010), který umožňuje projít jednotlivé kontigy po částech a určit, kolik k-merů mapuje do jednotlivých oblastí. Nejprve bylo nutné

vytvořit soubor, kde byl složený genom rozdělen do “oken” o velikosti 2000 bp. K tomu byl využit nástroj Bedtools makewindows a první dva sloupečky indexovaného souboru vytvořeného nástrojem Samtools faidx, které obsahují informace o jménech a velikostech jednotlivých souborů. Další použitý nástroj z balíčku Bedtools byl Bedtools coverage, který využil výstup z předchozího kroku a postupně na něj namapoval každý ze 4 BAM souborů vytvořených dříve. Výsledkem tedy byly 4 tabulky ve formátu BED: 21-mery specifické pro samice namapované na kontigy rozdělené do úseků dlouhých 2000 bp, podobně namapované 21-mery specifické pro samce a odpovídající dva soubory pro 55-mery.

Tabulky v BED formátu pro obě pohlaví a stejnou délkou k-merů byly sloučeny do jedné tabulky ve formátu TSV, tak aby v jednom řádku byl název kontigu, počátek a konec daného „okna“ a počet k-merů specifických pro jednotlivá pohlaví, který se na daný úsek kontigu podařilo namapovat.

Takto vytvořené soubory byly dále zpracovány v prostředí RStudio (Boston, USA), kde byly nejprve sloučeny tabulky pro 21-mery a 55-mery. Poté byl z počtu namapovaných k-merů v jednotlivých „oknech“ vypočítán průměr pro daný kontig. K tomuto souboru byly následně přidány dva sloupce, jeden pro 21-mery a druhý pro 55-mery, s hodnotami logaritmu o základu 2 z poměru mezi počtem namapovaných k-merů specifických pro samice a počtem namapovaných k-merů specifických pro samce v daném kontigu. Pomocí příkazu „pivot_longer“ byla tabulka převedena na dlouhý formát. Následně byly vybrány kontigy, kde byl průměrný počet namapovaných 21-merů i 55-merů specifických pro samice větší než 0 a zároveň počet namapovaných 21-merů i 55-merů specifických pro samce byl roven 0.

Pro výběr kontigů specifických pro samice byl použit ještě jeden postup. Při něm nebyl počítán průměrný počet k-merů specifických pro jednotlivá pohlaví namapovaných na daný kontig, ale počty k-merů v jednotlivých oknech daného kontigu byly sečteny. Další postup byl stejný jako při předchozí analýze. Výhoda metody využívající součet spočívá v tom, že odhalí i kontigy, kde jsou namapované k-mery specifické pro samice pouze v určité malé oblasti. Při použití metody využívající průměrné počty k-merů by se takový kontig jevil jako nevhodný pro další použití.

Z vybraných kontigů specifických pro samice (viz Tabulku V) byly vybrány ty, které obsahovaly nejvíce namapovaných k-merů. Tyto zvolené kontigy byly zkontrolovány pomocí programu Geneious Prime 2022.0.2 (Biomatters, Nový Zéland).

3.4. Navržení primerů

Pro navržení primerů byly vybrány pouze kontigy, na které mapovaly 21-mery i 55-mery do podobných oblastí a kde bylo takových oblastí nejvíce. Po zobrazení k-merů specifických pro samice namapovaných na daný kontig pomocí programu Geneious Prime 2022.0.2 (Biomatters, Nový Zéland) bylo vizuálně patrné, že všechny k-mery v dané oblasti genomu mají společnou pouze malou oblast několika málo bází, případně jen jednu bazi, ve které se daná sekvence liší od sekvencí vyskytujících se u obou pohlaví. Při zobrazení pokrytí kontigu k-mery vznikly ostré píky. Do oblastí těchto píků byly navrhovány primery tak, aby se 3' konce primerů, pokud možno, překrývaly s píky, tj. místy specifickými pro samice.

Navrhované primery byly určeny pro použití v metodě duplexní PCR, kdy každá reakční směs obsahuje zároveň primery pro amplifikaci konzervované sekvence přítomné u obou pohlaví a tento produkt bude sloužit jako pozitivní kontrola dokládající správný průběh reakce. Pro amplifikaci jako pozitivní kontrola byla zvolena část genu pro 18S rRNA a použity byly primery, které již byly u druhu *P. icarus* otestovány (Hrubá, 2020). S ohledem na tyto primery, u nichž byla optimální teplota nasedání na cílovou sekvenci 62 °C, byly i nové primery navrhované na podobnou teplotu. Všechny navržené páry primerů byly také otestovány v kombinaci s primery pro pozitivní kontrolu, zda společně nevytvářejí sekundární struktury, pomocí nástroje Multiple Primer Analyzer (ThermoFisher Scientific, USA, dostupné z: <https://www.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.html>). Všechny navržené primery byly také zpětně namapovány na složené genomové sekvence samice. Primery, které mapovaly na více než jedno místo byly z dalšího postupu vyřazeny. Z navržených primerů (viz Tabulku I) bylo vybráno 10 párů (viz Tabulku VI), které byly pořízeny od firmy Generi Biotech s.r.o. (Hradec Králové, Česká republika).

Tabulka I: Navržené primery pro amplifikaci části chromozomu W.

Název primeru	Sekvence primeru	Pozice počátku primeru na kontigu
8238_F1	GGGACAACAATAAGGAACGTAGC	1096
8238_F2	ACGTGAGTCGGATTTTAGGAGTC	1127
8238_F3	CAATCAGAGTATGGTCGTGATTC	1168
8238_R1,2,3	AGGGACCTGAGAAGATACGG	1294
8238_F4,5,6	ATATTGCGTGTCCAGTTTCTCTG	2892
8238_R4	GTTTCGATTACTATGCGGTATACTC	3239
8238_R5	AAAACCTGGTCGTCCTATTTGTTATAAATTG	3207
8238_R6	AGTGAAAATCTCAAGGAAAATAACACATTC	3728
8238_F7	CCTTGAGATTTTCACTTAAATAACTTGTTTC	3713
8238_R7	GGCTTTGCGTTTTGAATAAGTCC	3917
8238_F8	CACGCGAAATTGAATGTGTTATTTTCC	3688
8238_R8	ACATCAATTATTACAAGGGGCCAG	3890
632_F1	GCGAAACGAAAAGAGCCTGTTC	4446
632_R1	ATTACAGCAAACGAAAGCTCAGG	5208
632_F2	AGGAAGTTGTACCAGTCGCG	4429
632_R2	GCAAACGAAAGCTCAGGAGAATA	5202
<i>7321_F1</i>	<i>CAAACAAAGCTGTACTATCAAACG</i>	<i>32566</i>
<i>7321_R1</i>	<i>CGTATTTAACAACCAAAAAGAAAAGGAAG</i>	<i>33407</i>
11217_F1,2	GTCTGTCTGCCTTTTTGTGCTG	2596
11217_R1	GTCAGTAATGTCAGTATATTCTATATCTATGC	3994
11217_R2,3	CTGTCAAAGTTGCATATTATTTGTTTCTTAG	4170
11217_F3	GCCTGACGGTACAGCATTAAG	3943
<i>11217_F4</i>	<i>TTTTATTCTTAAAACACTATTAAAACCTATCGC</i>	<i>4445</i>
<i>11217_R4</i>	<i>AGGTCGTGAACAAGTCGTCAC</i>	<i>5488</i>

První číslo v názvu primeru označuje kontig, na který byl primer navrhován. Primery podbarvené šedě byly vyřazeny, jelikož mapovaly i na jiné oblasti samičího genomu. Primery zapsané kurzivou nebyly testovány.

3.5. Izolace DNA z modrásků druhu *P. icarus*

DNA izolovaná pro ověření navržených primerů byla získána pomocí kitu NucleoSpin DNA Insect (Macherey-Nagel, Německo). Pro izolaci byly použity vzorky dospělců i housenek, které byly skladovány v -80 °C na sucho nebo v etanolu. Tkáň z motýlů byla vložena do zkumavky, ve které bylo připraveno 100 µl pufru „Elution Buffer“ a 40 µl pufru „Buffer MG“. U vzorků uchovaných v etanolu, byla tkáň nejprve vysušena. Tkáň ve zkumavkách byla homogenizována pomocí sterilních tlouček. Do každé zkumavky bylo následně přidáno 10 µl dodané proteinázy K, vzorky byly promíchány a ponechány 20 minut inkubovat při pokojové teplotě. Dále bylo postupováno podle přiloženého protokolu od výrobce. Finální eluce byla provedla do 50 µl pufru „Elution Buffer“. Další vzorky DNA byly získány od Moniky Hospodářské (viz Tabulku II).

3.6. Otestování navržených primerů pomocí PCR

Navržené primery byly testovány pomocí PCR při různých teplotách nasedání (62, 64, 66 a 68 °C). Jako optimální byla vybrána teplota 62 °C. Při současném použití primerů pro část chromozomu W i pro 18S rRNA byly vyzkoušeny různé kombinace koncentrací primerů. Toto testování bylo provedeno pouze s páry primerů číslo 3, 4 a 6. Jako optimální byla vybrána reakční směs o objemu 20 µl s tímto složením: 1–15 ng DNA; 1x OneTaq Quick-Load Reaction Buffer a 1 U OneTaq Quick-Load DNA Polymerase (New England BioLabs, USA); 1 µM každý primer pro část chromozomu W; 0,1 µM každý primer pro část genu pro 18S rRNA; 0,2 mM každý dNTP (TaKaRa, Japonsko).

Teplotní profil PCR reakce sestával z počáteční denaturace při 94 °C po dobu 3 minut a dále 25 cyklů, které zahrnovaly denaturaci trvající 30 sekund při 94 °C, nasedání primerů trvající 30 sekund při teplotě 62 °C a elongaci při teplotě 68 °C. Doba elongace závisela na délce amplifikovaného produktu PCR (1 minuta na 1000 bazí). Posledním krokem byla finální elongace, trvající 3 minuty při teplotě 68 °C. Po provedení všech těchto kroků byly vzorky zchlazeny na 4 °C. Pro provedení těchto kroků byl použit termocycler Biometra TAdvanced (Analytik Jena, Německo).

Páry navržených primerů byly testovány na DNA ze samce a samice modráška *P. icarus*. Výsledky reakcí byly analyzovány pomocí gelové elektroforézy v 1% agarózovém gelu v 1x TAE pufru (Tris-Acetate-EDTA: 40 mM Tris, 20 mM kyseliny octové, 1 mM EDTA; pH 8). Po separaci fragmentů DNA ve stejnosměrném elektrickém poli byl gel barven v 0,05% vodném roztoku ethidium bromidu po dobu 15 minut za stálého míchání a následně

vyfotografován v UV světle. Jako velikostní marker byl použit PCRBIO Ladder I (PCR Biosystems, Velká Británie).

Fungující primery byly dále otestovány na větším množství jedinců (5 samců a 4 samice) a pár číslo 3 pak na DNA z řady dalších jedinců obou pohlaví z pěti zemí Evropy (viz Tabulku II).

Tabulka II: Použití modrásci druhu *P. icarus*.

Označení jedince	Pohlaví	Místo sběru	Datum sběru	Izolace DNA
B	♀	ČR, fakultní zahrada	14.6.2021	Monika Hospodářská
F1B2	♂	odchov	odchov	Tato práce
F1B6	♀	odchov	odchov	Tato práce
C	♀	ČR, fakultní zahrada	14.6.2021	Monika Hospodářská
C	♂	ČR, fakultní zahrada	14.6.2021	Tato práce
D	♀	ČR, Františkov	25.7.2021	Monika Hospodářská
E	♀	ČR, Františkov	25.7.2021	Tato práce
E	♂	ČR, Františkov	25.7.2021	Tato práce
F	♀	ČR, Mašovice	26.7.2021	Tato práce
G	♀	ČR, Vyšné	29.7.2021	Tato práce
H	♀	ČR, Vyšné	29.7.2021	Monika Hospodářská
F1H6	♀	odchov	odchov	Tato práce
F1H8	♂	odchov	odchov	Tato práce
CH	♀	ČR, Vyšné	29.7.2021	Monika Hospodářská
I	♀	ČR, Třeboň	10.8.2021	Tato práce
J	♀	ČR, Třeboň	10.8.2021	Monika Hospodářská
K	♀	ČR, Třeboň	10.8.2021	Monika Hospodářská
F1K1	♂	odchov	odchov	Tato práce
F1K4	♀	odchov	odchov	Tato práce
L	♀	ČR, Lipnice	11.8.2021	Monika Hospodářská

F1L1	♂	odchov	odchov	Tato práce
F1L5	♀	odchov	odchov	Tato práce
M	♀	ČR, Lipnice	11.8.2021	Monika Hospodářská
N	♀	ČR, Horní Radouň	13.8.2021	Tato práce
12A	♀	ČR, Levín	8.8.2019	Monika Hospodářská
13B	♀	ČR, Tušť	2.8.2019	Monika Hospodářská
A	♀	Polsko, Mścichy	4.6.2022	Monika Hospodářská
A	♂	Polsko, Mścichy	4.6.2022	Monika Hospodářská
A	♀	Bulharsko, Madžarovo	6.5.2022	Monika Hospodářská
A	♂	Bulharsko, Poda	5.5.2022	Tato práce
B	♀	Bulharsko, Poda	4.5.2022	Monika Hospodářská
F1B2	♂	odchov	odchov	Monika Hospodářská
F1B6	♀	odchov	odchov	Monika Hospodářská
C	♀	Bulharsko, Kožuch	8.5.2022	Monika Hospodářská
F1C2	♂	odchov	odchov	Tato práce
D	♀	Bulharsko, Madžarovo	6.5.2022	Monika Hospodářská
A	♀	Španělsko, Olot	24.7.2022	Tato práce
B	♀	Španělsko, Olot	24.7.2022	Tato práce
C	♀	Španělsko, El Brull	24.7.2022	Tato práce
A	♀	Estonsko, Laulasmaa	7.6.2022	Monika Hospodářská
C	♀	Estonsko, Paldiski	8.6.2022	Monika Hospodářská
F1C1	♂	odchov	odchov	Tato práce
D	♀	Estonsko, Paldiski	8.6.2022	Monika Hospodářská
F1D5	♂	odchov	odchov	Tato práce

Jedinci jejichž označení začíná „F1“ se vylíhli v laboratoři a jsou potomky matky, která je v tabulce uvedena v řádce nad nimi. DNA z jedinců podbarvených šedě byla použita k testování primerů, které provedla Monika Hospodářská.

3.7. Použití nástroje Minimap2 pro zjištění oblasti nasedání primerů

Cílem této práce bylo nalézt sekvence pocházející z chromozomů W a neo-W. Pro zjištění, z jaké oblasti pocházejí kontigy, které byly identifikovány jako specifické pro samice, byl použit program Minimap2 verze 2.24 (Li, 2018) s parametrem (-x asm10). Pomocí tohoto programu byly kontigy specifické pro samice, na které byly navrženy primery, mapovány na složený genom samce modráška *P. icarus* (Monika Hospodářská, nepublikováno).

3.8. Použití primerů pro určení pohlaví čerstvě vylíhlých housenek

Řádně otestovaný pár primerů číslo 3 byl použit pro určení pohlaví u čerstvě vylíhlých housenek, u kterých není možné určit pohlaví na základě morfologie. Housenky byly uchovány homogenizované v 500 µl TRI reagentu (Sigma-Aldrich, USA) zmražené na -80 °C Monikou Hospodářskou. Z takto připravených jedinců byla izolována současně DNA i RNA. Homogenizované housenky v TRI reagentu byly rozmrazeny při pokojové teplotě, poté bylo ke každému vzorku přidáno 100 µl chloroformu a vzorky byly vortexovány po dobu 15 s. Po inkubaci 5 minut při pokojové teplotě byly zkumavky centrifugovány na 12000 g při 4 °C po dobu 15 minut. Následně bylo přeneseno 250 µl vodné fáze do nové zkumavky pro izolaci RNA. Organická fáze byla ponechána pro izolaci DNA.

Izolace DNA pokračovala z organické fáze, která zbyla v původních zkumavkách, odkud byly nejprve odstraněny poslední zbytky vodné fáze, a následně bylo přidáno 300 µl Back Extraction pufru (4M guanidium thiokyanát, 50mM citrát sodný a 1M Tris base v MilliQ H₂O; pH 8,5 – 9). Vzorky byly promíchány vortexováním 15 s, inkubovány 10-15 minut při pokojové teplotě a následně centrifugovány na 12000 g 10 minut. Poté bylo 300 µl vodné fáze obsahující DNA přesunuto do nové zkumavky, byl přidán 1 µl glykogenu (ThermoFisher Scientific, USA) a obsah zkumavek byl vortexován při nízké rychlosti. Pak bylo přidáno 300 µl isopropanolu a směs byla promíchána převrácením zkumavky 10-15x. Po inkubaci 5 min při pokojové teplotě byly vzorky centrifugovány na 12000 g 10 minut a supernatant byl odstraněn. Pelet byl 2x promyt 500 µl 70% etanolu vychlazeného na -20 °C. Po odstranění druhé dávky byl vzorek krátce stočen, aby mohly být odstraněny poslední zbytky etanolu. Pelet byl krátce vysušen při pokojové teplotě v otevřené zkumavce a následně byl rozpuštěn ve 20 µl PCR H₂O. Takto připravené vzorky byly přes noc ponechány v lednici a následující den řádně vortexovány. Pro změření koncentrace DNA byl použit kit Qubit 1X dsDNA HS

Assay Kit a přístroj Qubit 4 Fluorometer (ThermoFisher Scientific, USA). Koncentrace DNA byla v řádu desetin až setin $\text{ng}/\mu\text{l}$. Pro určení pohlaví jedince byl použit $1 \mu\text{l}$ této DNA a metoda PCR byla provedena stejným způsobem, jak je popsáno v kapitole 3.6.

4. Výsledky

4.1. Sekvenace genomu samice modráška *P. icarus* technologií Oxford Nanopore

Izolace genomové DNA z housenky byla prováděna z 41 mg tkáně a bylo získáno 4,15 µg. Výtěžek DNA poté, co byla zbavena krátkých fragmentů byl 2,17 µg. Koncentrace připravených knihoven byla 24,6 ng/µl a 56,6 ng/µl.

Sekvenací bylo získáno celkem 9,64 Gbp v 586460 sekvenačních čteních, jejichž hodnota N50 byla 32470 bp. Po filtraci na minimální délku sekvenačních čtení 5000 bp zbylo 393001 sekvenačních čtení a hodnota N50 byla 34373 bp. V případě filtrace na minimální délku sekvenačních čtení 10000 bp zbylo 274738 sekvenačních čtení a hodnota N50 byla 37626 bp. Z dat po filtraci byly vytvořeny složené genomy, jejichž kvalita byla zkontrolována pomocí programů BUSCO a Quast. Pro další použití byl vybrán složený genom vytvořený ze sekvencí o minimální délce 5000 bp a upravený pomocí programu PurgeDups. I přesto, že program BUSCO v tomto složeném genomu našel pouze 62,2 % kompletních BUSCO genů přítomných v jedné kopii, byl tento složený genom nejvíce kompletní ze všech, které se podařilo vytvořit. Počet duplikovaných BUSCO genů se po použití nástroje PurgeDups snížil z 927 na 83 (viz Tabulku III). Z výsledků programu Quast můžeme vidět, že sekvence se podařilo poskládat do 3538 kontigů a hodnota N50 složeného genomu činila 385084 bp (viz Tabulku IV).

Tabulka III: Výsledek z programu BUSCO pro složený genom *P. icarus* použitý pro k-merovou analýzu.

Počet kompletních BUSCO genů (C)	3373
Počet kompletních BUSCO genů v jedné kopii (S)	3290
Počet kompletních duplikovaných BUSCO genů (D)	83
Počet fragmentovaných BUSCO genů (F)	485
Počet chybějících BUSCO genů (M)	1428
Celkový počet prohledaných BUSCO skupin	5286
C: 63,8 % [S: 62,2 %; D: 1,6 %]; F: 9,2 %; M: 27,0 %	

Tabulka IV: Výsledek z programu Quast pro složený genom *P. icarus* použitý pro k-merovou analýzu.

Počet kontigů	3538
Počet kontigů \geq 50000 bp	2114
Největší kontig [bp]	5621148
Celkový počet bází	538017602
N50 [bp]	385084
Počet neznámých bází	1914
Počet neznámých bází na 100 kbp	0,36

4.2. K-merová analýza

4.2.1. Výběr pohlavně specifických k-merů

Po stanovení hranice mezi chybovými a informativními k-mery byly vybrány k-mery specifické pro samce a pro samice. Bylo nalezeno 147457 21-merů a 344432 55-merů specifických pro samce. Pro samice bylo specifických 158673 21-merů a 222427 55-merů.

Pomocí obou metod popsaných v kapitole 3.3.3. bylo nalezeno 23 kontigů specifických pro samice a v našem případě byly výsledky získané pomocí obou metod srovnatelné, takže v žádném kontigu pravděpodobně nebyly výrazné rozdíly v počtu namapovaných k-merů specifických pro samice mezi jednotlivými „okny“. O těchto kontigech lze předpokládat, že pocházejí z chromozomu W (viz Tabulku V) a z těchto kontigů byly později vybrány ty, na které byly navrhovány primery (viz kapitolu 4.3.).

Tabulka V: Nalezené kontigy z chromozomu W.

Název kontigu	Délka kontigu [bp]	Průměr		Součet	
		21-mery	55-mery	21-mery	55-mery
kontig_11217	9887	13,4	33,2	67	166
kontig_11999	49054	0,32	1,84	8	46
kontig_12001	35224	0,222	3,333	4	60
kontig_1215	74591	0,368	0,789	14	30
kontig_1282	22120	1,75	1,833	21	22
kontig_13746	55161	0,036	0,107	1	3
kontig_2964	57854	0,103	0,310	3	9
kontig_3060	21857	0,818	0,909	9	10
kontig_3184	112569	0,316	1,403	18	80
kontig_3497	31821	0,438	0,063	7	1
kontig_3941	141521	0,014	0,113	1	8
kontig_3967	46532	2,042	7,667	49	184
kontig_5231	84017	0,186	0,279	8	12
kontig_5292	82617	0,024	0,095	1	4
kontig_5432	155347	0,013	0,179	1	14
kontig_625	154635	0,577	1	45	78
kontig_632	87374	2,409	3,909	106	172
kontig_7321	64907	1,636	3,242	54	107
kontig_7326	33460	0,176	1,059	3	18
kontig_7756	67000	0,265	0,029	9	1
kontig_8181	8581	1,2	1,4	6	7
kontig_8238	12983	26,571	37,286	186	261
kontig_8381	15404	1,125	1,125	9	9

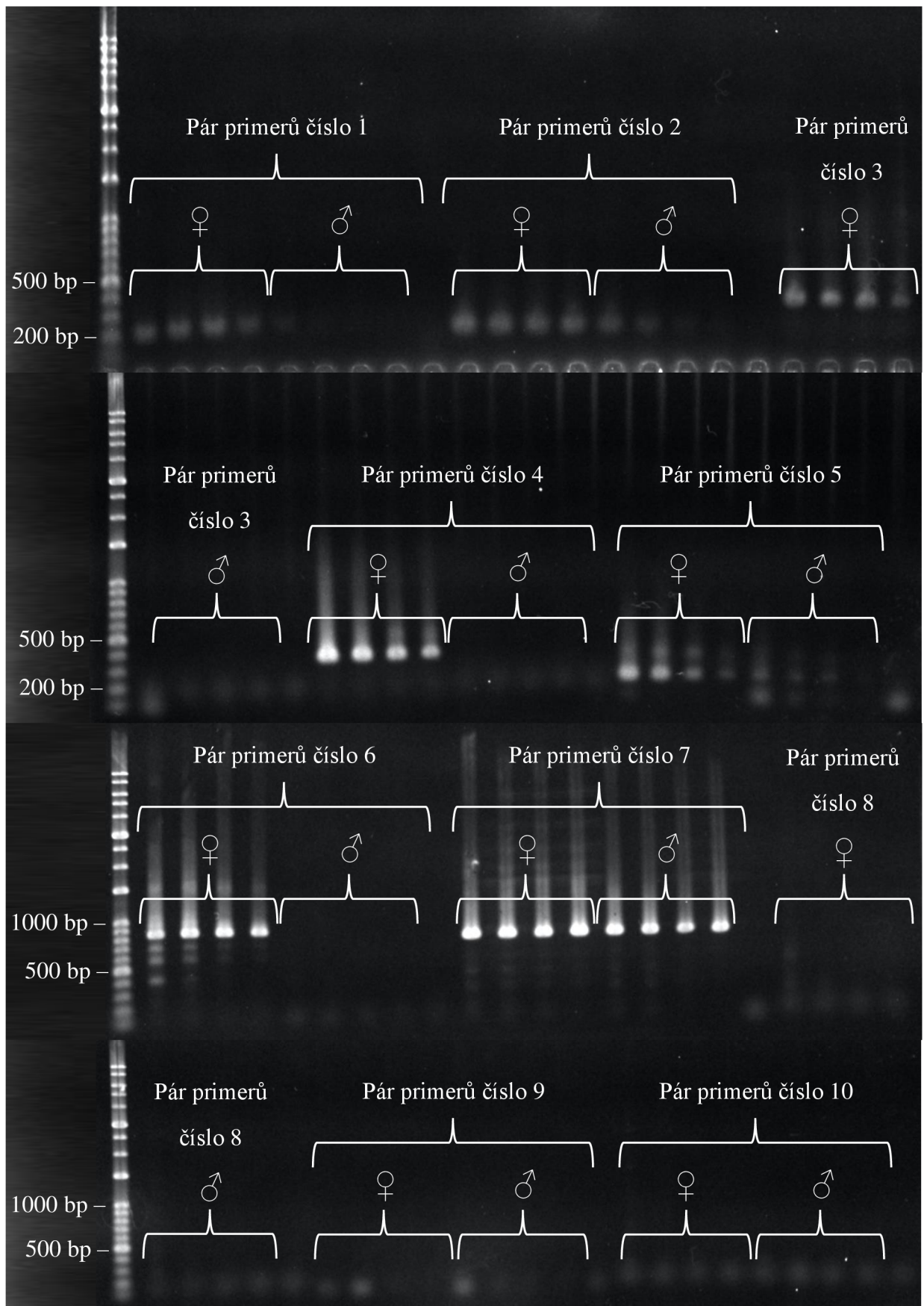
Sloupce označené „Průměr“ obsahují průměrný počet namapovaných 21-merů a 55-merů specifických pro samice vypočtený z jednotlivých „oken“ daného kontigu. Sloupce označené „Součet“ obsahují celkový počet k-merů specifických pro samice namapovaných na daný kontig.

4.3. Navržení a otestování primerů

Pro navržení primerů byly vybrány celkem 4 kontigy (označené čísla: 632, 7321, 8238, 11217). Pomocí programu Geneious Prime 2022.0.2 (Biomatters, Nový Zéland) se podařilo navrhnout celkem 24 primerů (15 párů), přičemž některé primery byly použity v různých párech. Po zpětném namapování na složené genomové sekvence samice bylo pět primerů vyloučeno, neboť tyto primery mapovaly na více oblastí. Ze zbývajících 12 párů bylo 10 otestováno (viz Tabulku I a VI).

Tabulka VI: Použité kombinace primerů a výsledek testování na DNA ze samce a samice.

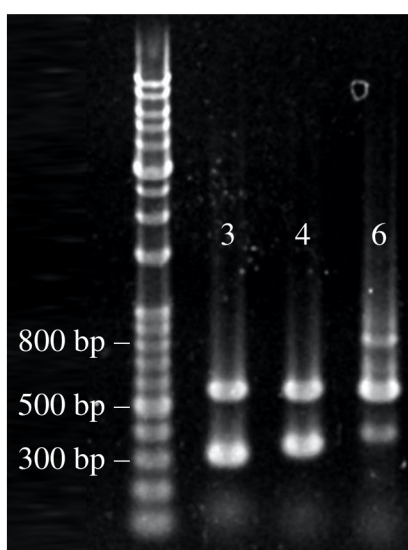
Číslo páru primerů	Použité primery	Délka produktu [bp]	Výsledek amplifikace
1	8238_F1 8238_R1,2,3	155	Při nižších teplotách produkt i u samce
2	8238_F2 8238_R1,2,3	125	Při nižších teplotách produkt i u samce
3	8238_F4,5,6 8238_R5	263	Produkt očekávané velikosti u samice
4	8238_F4,5,6 8238_R4	300	Produkt očekávané velikosti u samice
5	8238_F7 8238_R7	152	Při nižších teplotách produkt i u samce, nespecifické produkty
6	8238_F4,5,6 8238_R6	784	Produkt očekávané velikosti a nespecifické produkty u samice
7	632_F1 632_R1	717	Produkt očekávané velikosti u obou pohlaví
8	11217_F1,2 11217_R2,3	1522	Žádný produkt u obou pohlaví
9	11217_F3 11217_R2,3	176	Žádný produkt u obou pohlaví
10	11217_F1,2 11217_R1	1345	Žádný produkt u obou pohlaví



Obrázek 2: Výsledky testování objednaných primerů na DNA ze samice a samce při teplotách nasedání 62, 64, 66 a 68 °C.

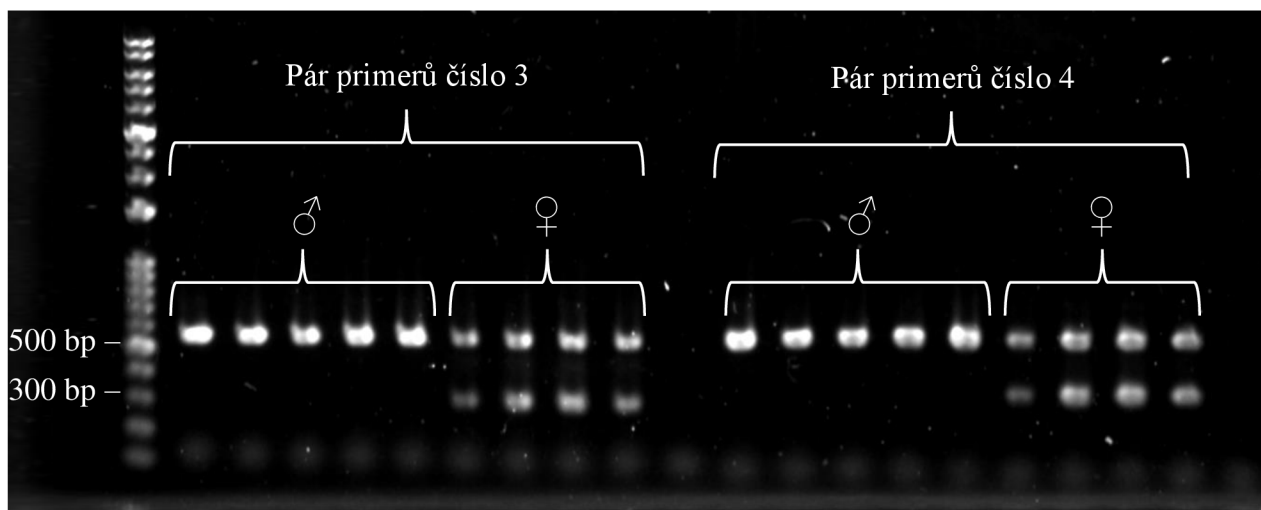
Objednané páry primerů byly nejprve testovány na DNA z jedné samice a jednoho samce při různých teplotách nasedání (62, 64, 66, 68 °C; viz Obrázek 2).

V dalším kroku byla metoda PCR optimalizována pro současné použití primerů pro část chromozomu W a pro část genu pro 18S rRNA. Tato optimalizace byla provedena s páry primerů číslo 3, 4 a 6. U primerů 3 a 4 byl optimální průběh reakce popsán v kapitole 3.6. U páru 6 však při tomto postupu nebyl produkt pocházející z chromozomu W vůbec patrný. Při použití stejného postupu, avšak s 30 cykly, produkt patrný byl, ale oproti referenčnímu produktu byl velmi slabý a špatně rozeznatelný, a navíc v reakci vznikl i nespecifický produkt. Proto byl pár primerů číslo 6 z dalšího testování vyřazen (viz Obrázek 3).



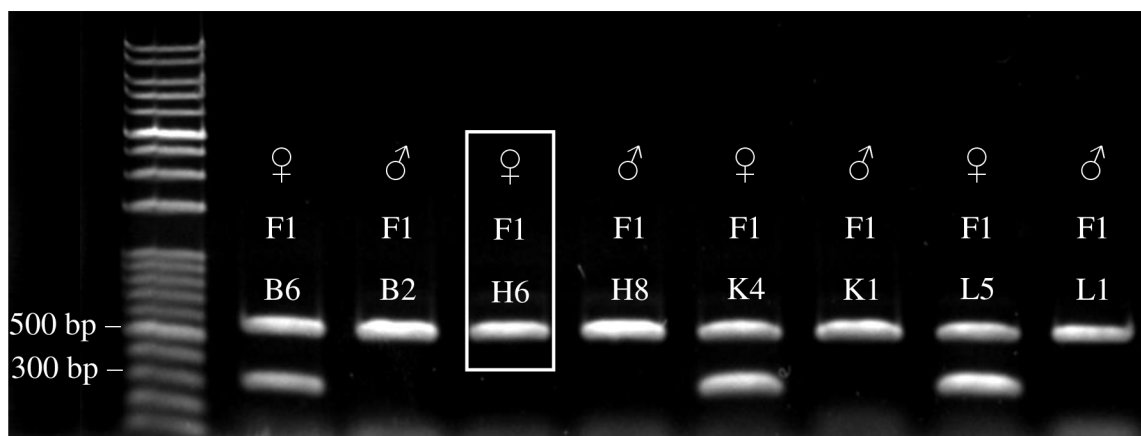
Obrázek 3: Výsledky optimalizace PCR pro současné použití primerů pro amplifikaci části chromozomu W a části genu pro 18S rRNA jako vnitřní kontroly. Čísla 3, 4 a 6 označují použitý pár primerů pro amplifikaci části chromozomu W. Očekávaná velikost vnitřní kontroly: 571 bp.

Páry primerů označené čísly 3 a 4 byly dále otestovány na DNA z dalších jedinců (viz Obrázek 4).

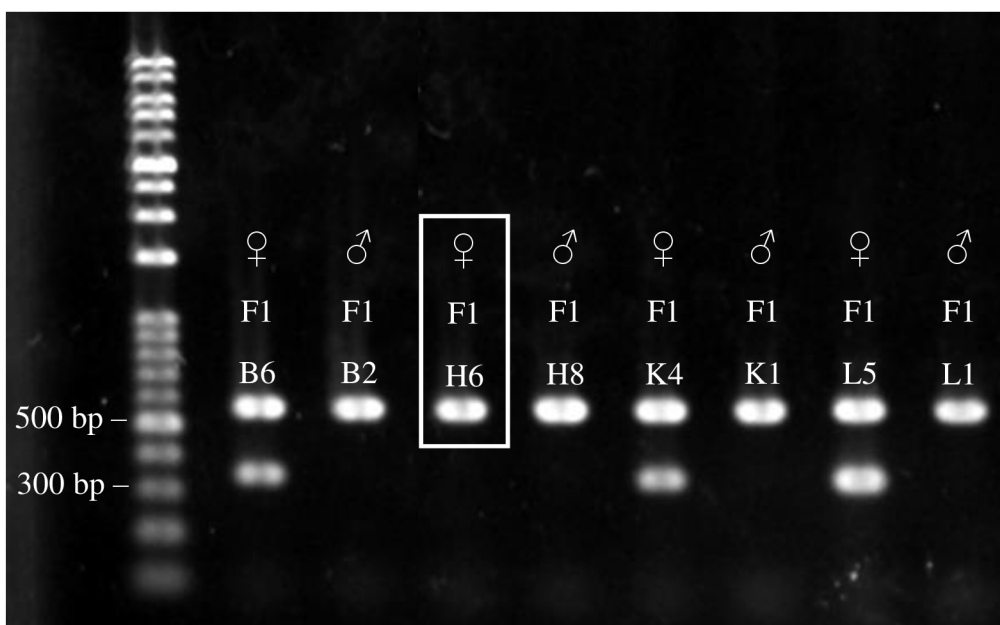


Obrázek 4: Výsledky testování párů primerů číslo 3 a 4 na DNA z 5 samců a 4 samic. Očekávaná velikost amplifikovaného elementu párem primerů 3: 265 bp, párem primerů 4: 300 bp.

Při dalším testování byla nalezena jedna samice (označená F1H6), u které primery na část chromozomu W neamplifikovaly žádný produkt. Takový výsledek byl získán při použití primerů číslo 3 i 4 (viz Obrázky 5 a 6). Stejným postupem otestovala primery Monika Hospodářská také na matce této samice, u které byl získán totožný výsledek.



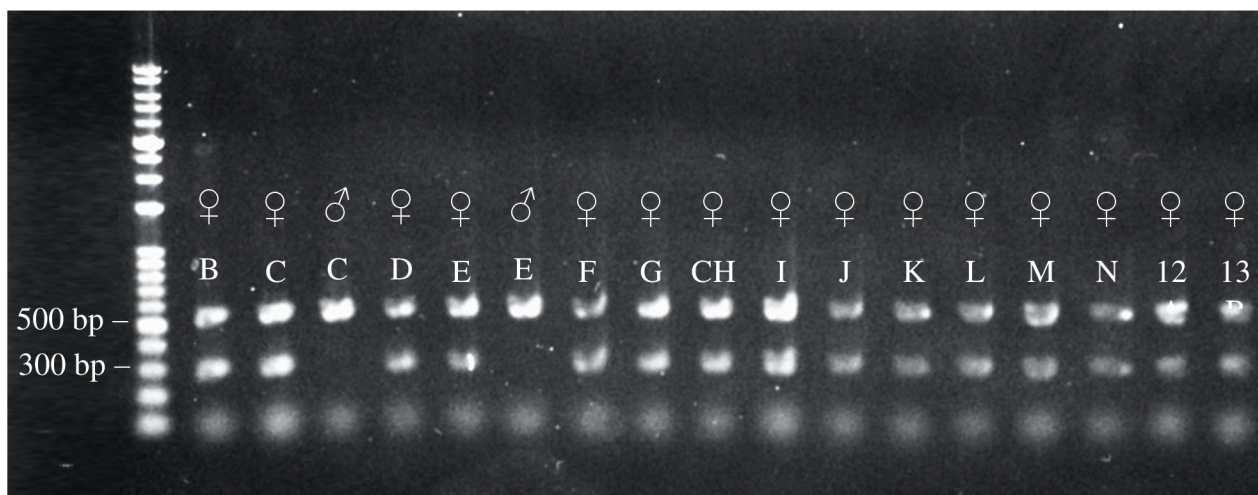
Obrázek 5: Výsledky testování páru primerů číslo 3 ukazující samici F1H6, u které tyto primery nevytváří žádný produkt. Vzorek od této samice je označen rámečkem.



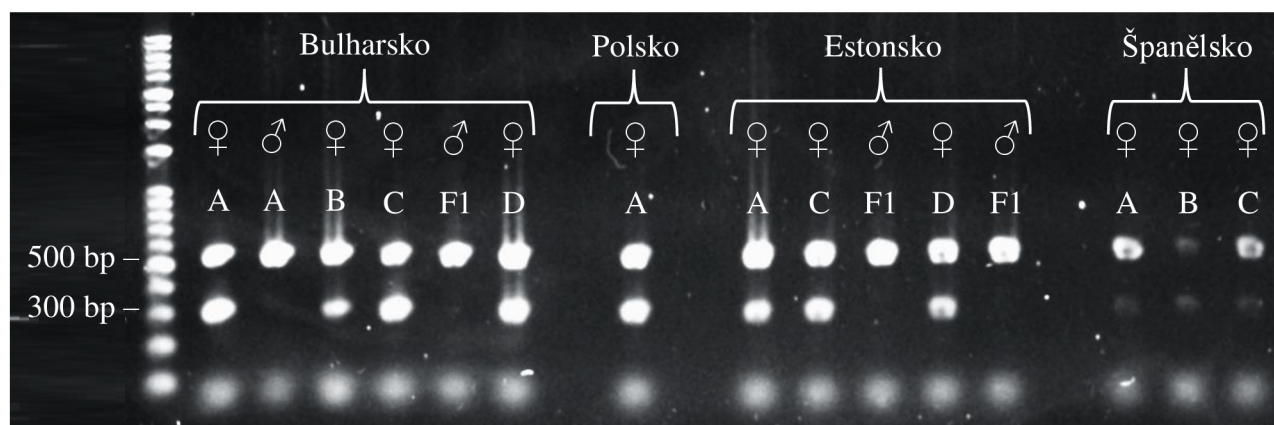
Obrázek 6: Výsledky testování páru primerů číslo 4 ukazující samici F1H6, u které tyto primery nevytváří žádný produkt. Vzorek od této samice je označen rámečkem.

Po získání tohoto výsledku bylo provedeno testování nižších teplot nasedání primerů (50–62 °C). Nicméně ani po této úpravě metody nebyl u samice F1H6 amplifikován žádný produkt pomocí primerů číslo 3.

Při testování páru primerů číslo 3 na dalších jedincích už žádný podobný výsledek získán nebyl. U všech testovaných samic primery amplifikovaly produkt o očekávané délce a u žádného z testovaných samců tento produkt vytvořen nebyl. Výsledky testování primerů na jedincích z České republiky ukazuje Obrázek 7 a výsledky testování na jedincích z ostatních populací Obrázek 8.



Obrázek 7: Výsledky testování páru primerů číslo 3 na jedincích z České republiky.
Pro informace o jedincích viz Tabulku II.



Obrázek 8: Výsledky testování páru primerů číslo 3 na jedincích z ostatních zemí.
Pro informace o jedincích viz Tabulku II.

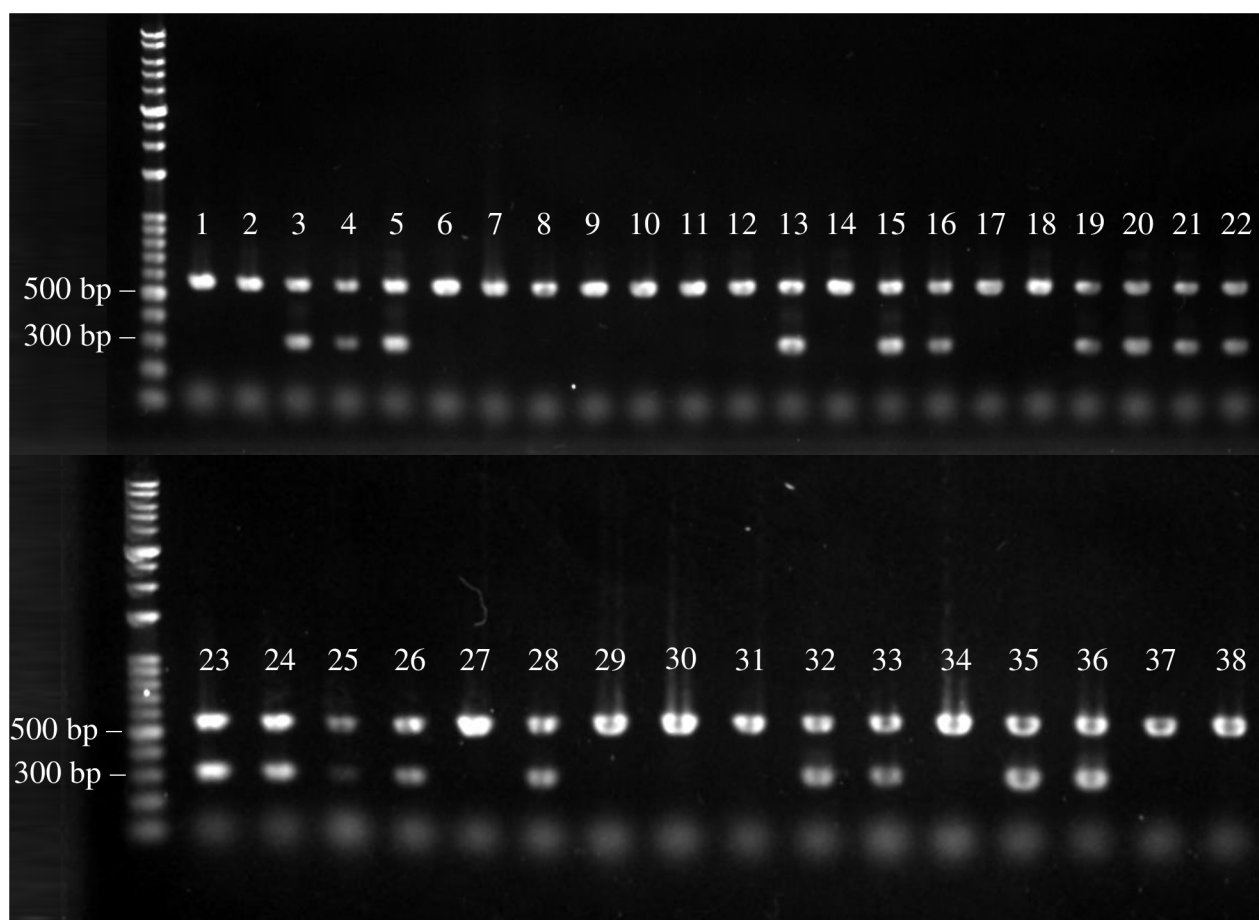
4.4. Použití programu Minimap2 pro zjištění oblasti nasedání primerů

Nástroj Minimap2 byl použit pro zjištění, zda navržené primery nasedají na oblast původního chromozomu W, nebo do oblasti neo-W, která má stále homologii k neo-Z chromozomu. Pokud by kontig použitý pro navržení primerů pocházel z původního chromozomu W, při mapování na samčí genom by se ho nepodařilo nikam namapovat. Pokud by se jednalo o kontig pocházející z neo-W, jeho část by v samčím genomu mapovala k homolognímu úseku chromozomu neo-Z.

U kontigů použitých pro navržení primerů se podařilo na samčí genom namapovat vždy jen krátké úseky, které mapovaly na různé autozomy, nebo i na pohlavní chromozom Z. Komplementarita mezi takto malými úseky je pravděpodobně způsobena výskytem repetice. Lze tedy předpokládat, že všechny navržené primery nasedají na původní chromozom W.

4.5. Určení pohlaví u čerstvě vylíhlých housenek

Otestovaný pár primerů číslo 3 byl použit pro určení pohlaví u čerstvě vylíhlých housenek pomocí PCR. Pro tuto metodu byla použita DNA izolována z housenek pomocí trizolu za současné izolace RNA (viz kapitolu 3.8.). Touto metodou bylo určeno pohlaví u 38 housenek, 19 z nich bylo určeno jako samice a 19 jako samci (viz Obrázek 9).



Obrázek 9: Výsledky určení pohlaví u čerstvě vylíhlých housenek modráška *P. icarus*. Jednotliví jedinci jsou označeni čísly 1-38, u samic jsou na snímku z gelové elektroforézy patrné dva proužky, u samců jen jeden.

5. Diskuze

Cílem této diplomové práce bylo identifikovat sekvence pocházející z pohlavního chromozomu W u modráška *Polyommatus icarus*, navrhnout a otestovat markery pro určení pohlaví a použít tyto markery ke genotypování čerstvě vylíhlých housenek. K tomuto účelu byla použita srovnávací analýza resekvenačních genomových dat 6 samců a 4 samic, při které byly získané sekvence rozděleny na k-mery mapované k referenčnímu samičímu genomu. V identifikovaných sekvencích specifických pro samice pak byly navrženy primery. Pomocí použité metodiky se podařilo nalézt funkční marker pro určení pohlaví pomocí PCR u modráška *P. icarus*.

5.1. Bioinformatická analýza

Pro vytvoření k-merů byl použit nástroj KmerGO, který nevyžaduje instalaci ani žádné prerekvizity pro spuštění (Wang a kol., 2020), což je výhoda oproti jiným nástrojům jako jsou MetaGO (Wang a kol., 2018), HAWK (Rahman a kol., 2018), Kover (Drouin a kol., 2016) nebo Kevlar (Standage a kol., 2019). Tento nástroj je možné využít i pro výběr k-merů specifických pro jakékoli skupiny, tedy ne pouze pro pohlaví (Wang a kol., 2020). Nicméně v této práci tato funkce použita nebyla, protože pro výběr pohlavně specifických k-merů bylo nutné znát hranici pro odstranění k-merů obsahujících sekvenační chyby a tuto funkci nástroj KmerGO nenabízí (Wang a kol., 2020). Bylo nalezeno výrazně více pohlavně specifických 55-merů než 21-merů (viz kapitolu 4.2.1.). To je dáno tím, že při zvolení delších k-merů bude více k-merů zasahovat do určité pohlavně specifické sekvence. Jelikož 55-mery mají přibližně dvojnásobnou délku než 21-mery, bylo nalezeno přibližně dvojnásobné množství pohlavně specifických 55-merů. V případě 21-merů jich bylo více specifických pro samice, zatímco u 55-merů bylo výrazně více 55-merů specifických pro samce. To může být způsobeno tím, že samci mají 2 pohlavní chromozomy Z, takže tento chromozom má u samců větší efektivní velikost populace než u samic. Za pohlavně specifické byly vybrány pouze ty k-mery, které zcela chyběly u všech jedinců opačného pohlaví. Jelikož pro tuto práci byly použity pouze 4 samice, velmi pravděpodobně u nich řada alel z chromozomu Z chyběla vlivem náhody. Tyto alely, které mohly být přítomné u samců, se pak jevily jako specifické pro samce, ačkoliv se v populaci mohou běžně vyskytovat i u samic. Jelikož, jak již bylo zmíněno výše, 55-merů je více než 21-merů, bylo toto zkreslení výsledků vlivem použití malého vzorku jedinců výraznější v případě 55-merů.

Analýza k-merů byla již použita k identifikaci pohlavně specifických sekvencí u drozofily a člověka, tedy u druhů s heterogametickým samčím pohlavím. K-merová analýza byla tedy použita k identifikaci sekvencí pocházejících z pohlavního chromozomu Y (Carvalho a Clark, 2013), který je stejně jako chromozom W u samic motýlů tvořen převážně repeticemi a je obtížné jeho sekvence skládat do delších úseků (Traut a kol., 2007). Carvalho a Clark (2013) použili pro identifikaci nových sekvencí z chromozomu Y u octomilky *Drosophila virilis* k-merovou analýzu, při které srovnávali 15-mery vytvořené z krátkých sekvenačních čtení samičího genomu a dostupných složených sekvencí genomu. Pro tuto analýzu byly vybrány jen unikátní k-mery, které na složené genomové sekvence mapovaly pouze jednou, což umožnilo odstranit k-mery pocházející z repetice a odstraněny byly také chybové k-mery. Výsledky této analýzy umožnily jasně odlišit „scaffoldy“ pocházející z autozomů a chromozomu X od „scaffoldů“ pocházejících z chromozomu Y. U „scaffoldů“ přítomných u jedinců obou pohlaví bylo nenamapovaných k-merů 0-5 %, zatímco u „scaffoldů“ z chromozomu Y to bylo 95-100 % (Carvalho a Clark, 2013). Tato metoda byla také použita k identifikaci sekvencí z chromozomu W u emu hnědého (*Dromaius novaehollandiae*; Zhang a kol., 2017). Na rozdíl od postupu použitého v této práci byly tedy pro vytvoření k-merů použity pouze sekvence z genomu jedné samice. Nalezené sekvence pocházející z pohlavního chromozomu Y by ale nebylo možné použít pro navržení markerů pro určení pohlaví pomocí PCR, protože tato metoda nijak neidentifikuje konkrétní úseky sekvencí, které jsou specifické pouze pro samce. Pro vývoj PCR markerů by tedy byla nutná další analýza sekvencí z chromozomu Y, která by identifikovala vhodné oblasti pro navržení primerů. Proto byl v této diplomové práci použit jiný postup, který zároveň identifikoval sekvence z chromozomu W i konkrétní úseky specifické pouze pro samice.

Dalším rozdílem oproti této diplomové práci bylo, že jediná samice byla sekvenována na výrazně vyšší hloubku pročtení (přibližně 85x po filtraci dat; Carvalho a Clark, 2013) oproti jedincům v této diplomové práci, jejichž genomy byly sekvenovány při hloubce pročtení přibližně 20x. Výsledky prezentované v této diplomové práci tedy ukazují, že nalezení sekvencí z chromozomu W je možné i při výrazně nižším množství dat od jednoho jedince, což může výrazně snížit náklady na sekvenování genomu jednoho jedince, a proto je možné do analýzy zahrnout více jedinců obou pohlaví. Navíc použití více jedinců od obou pohlaví umožňuje vyhnout se zkreslení výsledků vlivem alel polymorfních genů (viz níže).

Carvalho a Clark (2013) použili stejný postup také u dalších druhů octomilek a u lidského genomu. U většiny lidských „scaffoldů“ se podařilo jasně určit, zda pochází z chromozomu Y či nikoli, avšak několik „scaffoldů“ nebylo možné jasně zařadit. Jak bylo

později zjištěno, všechny tyto „scaffoldy“ pocházely z chromozomu Y a jejich problematická identifikace byla ve většině případů způsobena chybami v poskládání sekvencí (Carvalho a Clark, 2013). Skládání genomových sekvencí tedy může být zdrojem chyb, které zkomplikují následnou analýzu. Postup použitý v této práci skládání genomu také zahrnoval, a to mohlo způsobit, že některé sekvence z chromozomu W nebyly nalezeny, protože byly během skládání genomu špatně poskládány. Nicméně i přesto se podařilo najít funkční markery pro určení pohlaví pomocí PCR. Pokud by tomu tak nebylo, nebo by byla potřeba najít více různých markerů, bylo by možno optimalizovat postup a jako referenci použít neposkládané sekvence. Jelikož v této diplomové práci byla pro získání referenčního genomu použita platforma Oxford Nanopore, která vytváří dlouhá sekvenační čtení, byl by tento postup možný. Příkladem nástroje, který umožňuje identifikaci pohlavně specifických sekvencí ze sekvenačních dat bez skládání genomových sekvencí je Redkmer, který byl vyvinut pro identifikaci specifických sekvencí z chromozomu X z minimálně zpracovaných dlouhých sekvenačních čtení (Papathanos a Windbichler, 2018).

Další použití k-merové analýzy pro identifikaci pohlavně specifických sekvencí pochází z roku 2018, kdy byla tato metoda použita pro identifikaci sekvencí pocházejících z chromozomu W u bource morušového (*B. mori*). Pro analýzu byla použita krátká sekvenační čtení získaná sekvenací genomů samce a samice na platformě Illumina při hloubce pročtení genomu přibližně 73x. Jedinci pocházeli z inbrední linie. Ze sekvencí byly vytvořeny 15-mery a pro každý k-mer byla spočítána frekvence v samčím a samičím genomu. Pro každý k-mer byl vypočten poměr mezi frekvencí u samce a u samice. Pro k-mery pocházející z autozomů byl tento poměr roven přibližně 1, pro k-mery pocházející z pohlavního chromozomu Z byl tento poměr přibližně 2, jelikož samci mají dva chromozomy Z, zatímco samice jen jeden. A pro k-mery z chromozomu W byl poměr frekvencí přibližně 0, jelikož tento chromozom je přítomen pouze u samic v jedné kopii (Li a kol., 2018).

Autoři předpokládali, že takto identifikované k-mery pocházející z pohlavního chromozomu W (případně Y) mohou jednoznačně označit kontigy pocházející z těchto chromozomů a usnadnit tak skládání genomu. Tuto hypotézu ověřili na *D. melanogaster* a výsledky porovnali s výsledky dvou jiných metod pro určení „scaffoldů“ pocházejících z chromozomu Y (jedna metoda identifikuje „scaffoldy“ z chromozomu Y na základě počtu alignmentů vytvořených ze samičích a samčích sekvencí a druhá metoda využívá porovnání složeného genomu s krátkými sekvenačními čteními samičího genomu). Zjistili, že metoda založená na poměru frekvencí k-merů mezi pohlavími byla schopná identifikovat nejvíce „scaffoldů“ z chromozomu Y. Přičemž tato metoda byla výrazně úspěšnější při identifikaci

„scaffoldů“ kratších než 5 kbp, což je výhodné, protože sekvence z pohlavního chromozomu Y (případně W) bývají obvykle velmi fragmentované (Li a kol., 2018). V této diplomové práci byla pro výběr pohlavně specifických k-merů také použita metoda založená na rozdílné frekvenci k-merů mezi pohlavími, tedy na tom, že k-mery pocházející z chromozomu W se nacházejí pouze u samic. Pomocí mapování takto nalezených pohlavně specifických k-merů byly nalezeny pohlavně specifické kontigy z nichž nejkratší měl délku 8581 bp (viz Tabulku V).

Li a kol. (2018) tedy použili sekvence z genomu samce i samice, avšak na rozdíl od této diplomové práce byl sekvenován jeden jedinec od každého pohlaví, byť při výrazně vyšší hloubce čtení stejně jako v práci Carvalho a Clark (2013). I v práci Li a kol. (2018) tak mohlo dojít ke zkreslení výsledků vlivem polymorfismů specifických pro sekvenované jedince, což bylo v této diplomové práci eliminováno použitím více jedinců obou pohlaví. I při použití 4 samic a 6 samců k jistému zkreslení vlivem polymorfismů dochází, jak naznačují počty nalezených pohlavně specifických k-merů (viz kapitolu 4.2.1.).

Je tedy patrné, že k-merovou analýzu lze použít mnoha různými způsoby pro nalezení sekvencí pocházejících z pohlavního chromozomu W (nebo Y). V této diplomové práci bylo dokázáno, že pro nalezení takových pohlavně specifických sekvencí je dostačující použití složeného genomu vytvořeného z dlouhých sekvenačních čtení s poměrně vysokou chybovostí při pokrytí genomu menším než 1. I když je pravděpodobné, že nebyly nalezeny všechny sekvence z chromozomu W. Přesto se však podařilo vyvinout funkční PCR markery pro určení pohlaví.

Pro zlepšení výsledků by bylo možné vynechat tvorbu složeného genomu (viz výše). Zlepšení výsledků by také mohlo přinést získání dlouhých sekvencí s nižší chybovostí. Metodou, která produkuje dlouhá sekvenační čtení stejně jako použitá technologie Oxford Nanopore avšak s nižší chybovostí, je Pacbio HiFi (Wenger a kol., 2019; De Coster a kol., 2021). Jinou možností, jak zvýšit kvalitu referenčních sekvencí, je použít tzv. „polishing“ pomocí krátkých čtení s nízkou chybovostí. Pro tuto metodu by bylo potřeba získat od jedné samice jak dlouhá a chybová sekvenační čtení Oxford Nanopore, tak krátké a přesné Illumina sekvence, což by však výrazně zvýšilo náklady (De Coster a kol., 2021).

Metody používané pro molekulární určení pohlaví u řádu Lepidoptera využívají například známé geny na chromozomu Z, které se liší počtem kopií mezi pohlavími, což může být identifikováno pomocí kvantitativní PCR (Belousova a kol., 2019), nebo rozdílné množství DNA v buňkách různého pohlaví, detekované průtokovou cytometrií (Nakamura a kol., 1990; Doležel a Göhde, 1995; Hejníčková a kol., 2019). Tyto metody jsou však

technicky náročné, jelikož vyžadují speciální přípravu vzorů a také speciální měřicí přístroje. Metodika použitá v této práci umožňuje nalézt markery pro jednoduché určení pohlaví z DNA jedinců pomocí PCR. Jiné techniky určení pohlaví, také založené na metodě PCR, však vyžadují znalost sekvencí pocházejících z pohlavních chromozomů. Například mikrosatelitů (Coates a Hellmich, 2003) nebo konkrétních genů (Fuková a kol., 2009). Postup použitý v této diplomové práci naproti tomu umožňuje identifikovat pohlavně specifické sekvence *de novo*. To je velkou výhodou, protože u zástupců řádu Lepidoptera je pohlavní chromozom W tvořen převážně repetitivními sekvencemi a transpozony (Traut a kol., 2007), a proto bývá obtížné data získaná jeho sekvenací zpracovat a poskládat do delších úseků, případně složit sekvenci celého chromozomu. To je důvodem, proč jsou sekvence z tohoto chromozomu často nekompletní a špatně zpracovatelné (Tomaszkiewicz a kol., 2017).

5.2. Navržení a otestování primerů

V programu Geneious Prime 2022.0.2 bylo použito grafické zobrazení k-merů specifických pro samice namapovaných na referenční samičí genom. Při použití tohoto zobrazení bylo patrné, že oblast specifická pro samice je tvořena několika málo bazemi a v některých případech jen jednou bazí. Mohou to být varianty repetice, které se na chromozomu W mohou snadno hromadit, protože tento chromozom na rozdíl od zbytku genomu nerekombinuje, díky samičí achiasmatické meióze (Maeda, 1939; Traut a kol., 2007; Fuková a kol., 2007). Primery byly navrhovány tak, aby jejich 3' konec ležel přesně na jedné nebo více bazích specifických pro samice. Stejný postup navrhování primerů použili také Li a kol. (2018), kteří na některé identifikované W-vázané kontigy navrhli primery a pomocí PCR s DNA ze samce a samice otestovali, že se příslušná sekvence nachází pouze u samice.

Nicméně ne ve všech případech bylo v této diplomové práci možné navrhnout primer tímto způsobem. To bylo způsobeno buďto nevyhovující teplotou nasedání takového primeru, nebo takto navržený primer tvořil sekundární struktury. Podobně jako v práci Li a kol. (2018) byla při PCR použita relativně vysoká teplota, aby se zabránilo nasedání primerů na nespecifické oblasti. Jak je vidět z Obrázku 1, při použití páru primerů číslo 6 se samičí DNA, teplota nasedání primerů výrazně ovlivňuje, zda se budou tvořit nespecifické produkty.

Otestované páry primerů, které nebylo možné použít pro určení pohlaví buďto amplifikovaly příslušný produkt u jedinců obou pohlaví, takže sekvence, na kterou byly navrženy nebyla dostatečně pohlavně specifická. Nebo neamplifikovaly žádný produkt u obou pohlaví, což mohlo být způsobeno například inverzí nebo mutací v cílové sekvenci u samice.

Pár primerů číslo 3 byl vybrán jako nejlépe fungující pro určení pohlaví pomocí PCR, jelikož amplifikoval produkt o očekávané velikosti u samic a netvořily žádný produkt u samců (viz obrázky 4, 7 a 8). Jedinou výjimku tvořila samice H pocházející z České republiky a její potomci. U samic z této rodiny primery žádný produkt nevytvořily (viz Obrázek 5). Tento neočekávaný výsledek může být způsoben mutací, inverzí nebo delecí cílové sekvence na chromozomu W, která se přenáší i na potomky samičího pohlaví.

5.3. Nalezení sekvencí z chromozomu neo-W

U modráška *P. icarus* pravděpodobně existují neo-pohlavní chromozomy a karyotypové rasy s odlišnou konstitucí pohlavních chromozomů (Kreklová, 2020). U kontigů specifických pro samice nalezených v této práci bylo tedy zjišťováno, zda pocházejí z oblasti původního W chromozomu, nebo z autozomu, který fúzí s chromozomem W vytvořil neo-W. Nalezené kontigy specifické pro samice byly mapovány na složený samčí genom (viz kapitoly 3.7. a 4.4.), přičemž na samčí genom mapovaly vždy jen malé úseky, což bylo pravděpodobně způsobeno repeticemi, které se na chromozomu W hojně vyskytují (Traut a kol., 2007). V této práci se tedy nepodařilo najít sekvence pocházející z autozomu, který fúzoval s chromozomem W. Jelikož výsledky práce Kreklová (2020) naznačují, že neo-W chromozom u *P. icarus* je velmi mladý, a tudíž dosud nediferencovaný, kontig, který by pocházel z pohlavního chromozomu neo-W, by svojí velkou částí mapoval na autozom v samčím genomu, což ukázali např. Gan a kol. (2019). Ti použili podobný postup pro nalezení sekvencí z neo-W chromozomu u lejsčika žlutého (*Eopsaltria australis*), kdy byly sekvence, o kterých autoři předpokládali, že pocházejí z chromozomu W (a Z), mapovány na referenční genom zebřičky pestré (*Taeniopygia guttata castanotis*). Pro tuto analýzu použili nástroj RaGoo v1.0 (Alonge a kol., 2019) a zjistili, že velká část sekvencí z pohlavních chromozomů mapuje nejen na pohlavní chromozomy referenčního druhu, ale také na autozom 1A. To tedy naznačuje přítomnost neo-pohlavních chromozomů u lejsčika žlutého vzniklých fúzí s autozomem (Gan a kol., 2019).

Jinou možností pro nalezení kontigu z neo-W by bylo identifikovat autozom, který se podílel na tvorbě neo-W u *P. icarus* a vybrat kontigy, které na tento chromozom mapují. Poté u těchto kontigů zjistit, zda na některou jejich část nemapuje větší množství k-merů specifických pro samice, což by mohlo odhalit diferencovanou oblast. Další možností by bylo identifikovat kontigy pokrývající místo fúze mezi chromozomem W a autozomem. Pro vytvoření PCR markerů by pak bylo možné navrhnout primery kolem místa fúze.

Při následném provedení PCR by u samic pravděpodobně vznikl produkt o odpovídající velikosti a u samců, nebo samic bez neo-W chromozomu, by se produkt netvořil.

6. Závěr

V této diplomové práci se podařilo pomocí srovnávací analýzy resekvenčních genomových dat samců a samic modráška *Polyommatus icarus* nalézt sekvence pocházející z pohlavního chromozomu W a navrhnout na tyto sekvence primery, které by bylo možné použít jako markery pro určení pohlaví pomocí PCR. Po otestování navržených primerů byl nalezen funkční pár, který umožňuje spolehlivě odlišit samice od samců jednoduše pomocí PCR.

Tato metoda byla použita k určení pohlaví u čerstvě vylíhlých housenek *P. icarus*. Z těchto housenek byla izolována současně také RNA, kterou bude možno dále využít pro sekvenování a studium role pohlavně antagonistické selekce ve fixaci neo-pohlavních chromozomů u modráška *P. icarus*. Nalezení PCR markerů pro určení pohlaví tedy výrazně usnadní další studium tohoto modráška.

7. Literatura

- Abe H, Mita K, Yasukochi Y, Oshiki T, Shimada T (2005) Retrotransposable elements on the W chromosome of the silkworm, *Bombyx mori*. *Cytogenetic and Genome Research* 110: 144–151.
- Abe H, Ohbayashi F, Shimada T, Sugasaki T, Kawai S, Oshiki T (1998) A complete full-length non-LTR retrotransposon, BMC1, on the W chromosome of the silkworm, *Bombyx mori*. *Genes & Genetic Systems* 73: 353–358.
- Aguirre C, Olivares N, Hinrichsen P (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.
- Ahola V, Lehtonen R, Somervuo P, Salmela L, Koskinen P, Rastas P, Välimäki N, Paulin L, Kvist J, Wahlberg N, Tanskanen J, Hornett EA, Ferguson LC, Luo S, Cao Z, de Jong MA, Duploux A, Smolander OP, Vogel H, McCoy RC, Qian K, Chong WS, Zhang Q, Ahmad F, Haukka JK, Joshi A, Salojärvi J, Wheat CW, Grosse-Wilde E, Hughes D, Katainen R, Pitkänen E, Ylinen J, Waterhouse RM, Turunen M, Vähärautio A, Ojanen SP, Schulman AH, Taipale M, Lawson D, Ukkonen E, Mäkinen V, Goldsmith MR, Holm L, Auvinen P, Frilander MJ, Hanski I (2014) The *Glanville fritillary* genome retains an ancient karyotype and reveals selective chromosomal fusions in Lepidoptera. *Nature Communications* 5: 4737.
- Alonge M, Soyk S, Ramakrishnan S, Wang X, Goodwin S, Sedlazeck FJ, Lippman ZB, Schatz MC (2019) RaGOO: fast and accurate reference-guided scaffolding of draft genomes. *Genome Biology* 20(1): 224.
- Bachtrog D (2013) Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nature Reviews Genetics* 14: 113-124.
- Belousova I, Ershov N, Pavlushin S, Ilinsky Y, Martemyanov V (2019) Molecular sexing of Lepidoptera. *Journal of Insect Physiology* 114: 53–56.

- Bloem KA, Bloem S (2000) SIT for codling moth eradication in British Columbia, Canada. In: Area-wide control of fruit flies and other insect pests. *Penerbit Universiti Sains Malaysia*, Malaysia.
- Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics* 30: 2114-2120.
- Bureš P, Zedek F (2014) Holokinetic drive: centromere drive in chromosomes without centromeres. *Evolution* 68(8): 2412–2420.
- Carabajal Paladino LZ, Provazníková I, Berger M, Bass C, Aratchige NS, López SN, Marec F, Nguyen P (2019) Sex chromosome turnover in moths of the diverse superfamily Gelechioidea. *Genome Biology and Evolution* 11(4): 1307-1319.
- Carvalho AB, Clark AG (2013) Efficient identification of Y chromosome sequences in the human and *Drosophila* genomes. *Genome Research* 23(11): 1894-1907.
- Coates BS, Hellmich RL (2003) Two sex-chromosome-linked microsatellite loci show geographic variance among North American *Ostrinia nubilalis*. *Journal of Insect Science* 3: 29.
- Coutsis J, De Prins J, De Prins W (2001) The chromosome number and karyotype of the two morphs of *Polyommatus (Lysandra) coridon* from Greece (Lepidoptera: Lycaenidae). *Phegea* 29: 63-71.
- De Coster W, D’Hert W, Schultz DT, Cruts M, Van Broeckhoven C (2018) NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* 34(15): 2666–2669.
- De Coster W, Weissensteiner MH, Sedlazeck FJ (2021) Towards population-scale long-read sequencing. *Nature Reviews Genetics* 22: 572–587.
- Deng Z, Zhang Y, Zhang M, Huang J, Li C, Ni X, Li X (2020) Characterization of the First W-Specific Protein-Coding Gene for Sex Identification in *Helicoverpa armigera*. *Frontiers in Genetics* 11: 649.

- Dierickx EG, Sin SYW, van Veelen HPJ, Brooke ML, Liu Y, Edwards SV, Martin SH (2020) Genetic diversity, demographic history and neo-sex chromosomes in the Critically Endangered Rasolark. *Proceeding of the Royal Society B* 287(1922): 20192613.
- Doležel J, Göhde W (1995) Sex determination in dioecious plants *Melandrium album* and *M. rubrum* using high-resolution flow cytometry. *Cytometry* 19: 103–106.
- Douglas LJ, Untalan PM, Haymer DS (2004) Molecular sexing in the Mediterranean fruit fly, *Ceratitis capitata*. *Insect Biochemistry and Molecular Biology* 34(2): 159–165.
- Drouin A, Giguère S, Déraspe M, Marchand M, Tyers M, Loo VG, Bourgault AM, Laviolette F, Corbeil J (2016) Predictive computational phenotyping and biomarker discovery using reference-free genome comparisons. *BMC Genomics* 17(1): 754.
- Ennis TJ (1976) Sex chromatin and chromosome numbers in Lepidoptera. *Canadian Journal of Genetics and Cytology* 18(1): 119-130.
- Fuková I, Neven LG, Bárcenas NM, Gund NA, Dalíková M, Marec F (2009) Rapid assessment of the sex of codling moth *Cydia pomonella* (Linnaeus) (Lepidoptera: Tortricidae) eggs to larvae. *Journal of Applied Entomology* 133(4): 249-261.
- Fuková I, Traut W, Vítková M, Nguyen P, Kubičková S, Marec F (2007) Probing the W chromosome of the codling moth, *Cydia pomonella*, with sequences from microdissected sex chromatin. *Chromosoma* 116: 135-145.
- Gan HM, Falk S, Morales HE, Austin CM, Sunnucks P, Pavlova A (2019) Genomic evidence of neo-sex chromosomes in the eastern yellow robin. *Gigascience* 8(9): 111.
- Goldsmith MR, Marec F (2010) *Molecular Biology and Genetics of the Lepidoptera*. CRC Press, Boca Raton, USA, 368 pp.
- Griffiths R, Double MC, Orr K, Dawson RJ (1998) A DNA test to sex most birds. *Molecular Ecology* 7(8): 1071–1075.
- Guan D, McCarthy SA, Wood J, Howe K, Wang Y, Durbin R (2020) Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics* 36(9): 2896–2898.

- Gurevich A, Saveliev V, Vyahhi N, Tesler G (2013) QUASt: quality assessment tool for genome assemblies. *Bioinformatics* 29(8): 1072-1075.
- Hejníčková M, Koutecký P, Potocký P, Pro vazníková I, Voleníková A, Dalíková M, Visser S, Marec F, Zrzavá M (2019) Absence of W Chromosome in Psychidae Moths and Implications for the Theory of Sex Chromosome Evolution in Lepidoptera. *Genes* 10(12): 1016.
- Hill J, Rastas P, Hornett EA, Neethiraj R, Clark N, Morehouse N, de la Paz Celorio-Mancera M, Cols JC, Dirksen H, Meslin C, Keehnen N, Pruisscher P, Sikkink K, Vives M, Vogel H, Wiklund C, Woronik A, Boggs CL, Nylin S, Wheat CW (2019) Unprecedented reorganization of holocentric chromosomes provides insights into the enigma of lepidopteran chromosome evolution. *Science Advances* 5(6): e3648.
- Hrubá M (2020) Analýza repetitív v genomech vybraných druhů modrásků rodů *Polyommatus* a *Lysandra*. Bakalářská práce. Jihočeská univerzita, Přírodovědecká fakulta, České Budějovice.
- Huang Z, De O, Furo I, Liu J, Peona V, Gomes AJB, Cen W, Huang H, Zhang Y, Chen D, Xue T, Zhang Q, Yue Z, Wang Q, Yu L, Chen Y, Suh A, de Oliveira EHC, Xu L (2022) Recurrent chromosome reshuffling and the evolution of neo-sex chromosomes in parrots. *Nature Communications* 13: 944.
- Charlesworth D, Charlesworth B (1980) Sex differences in fitness and selection for centric fusions between sex-chromosomes and autosomes. *Genetical Research* 35(02): 205.
- Ip CLC, Loose M, Tyson JR, de Cesare M, Brown BL, Jain M, Leggett RM, Eccles DA, Zalunin V, Urban JM, Piazza P, Bowden RJ, Paten B, Mwaigwisya S, Batty EM, Simpson JT, Snutch TP, Birney E, Buck D, Goodwin S, Jansen HJ, O'Grady J, Olsen HE (2015) MinION Analysis and Reference Consortium: Phase 1 data release and analysis. *F1000Research* 4: 1075.
- Kandul NP, Lukhtanov VA, Dantchenko AV, Coleman JWS, Sekercioglu CH, Haig D, Pierce NE (2004) Phylogeny of *Agrodiaetus* Hübner 1822 (Lepidoptera: Lycaenidae) inferred from mtDNA sequences of *COI* and *COII*, and nuclear sequences of *EFL-a*: karyotype diversification and species radiation. *Systematic Biology* 53: 278-298.

- Kandul NP, Lukhtanov VA, Pierce NE (2007) Karyotypic Diversity and Speciation in *Agrodiaetus* Butterflies. *Evolution* 61(3): 546-559.
- Kawahara AY, Plotkin D, Espeland M, Meusemann K, Toussaint EFA, Donath A, Gimnich F, Frandsen PB, Zwick A, dos Reis M, Barber JR, Peters RS, Liu S, Zhou X, Mayer C, Podsiadlowski L, Storer C, Yack JE, Misof B, Breinholt JW (2019) Phylogenomics reveals the evolutionary timing and pattern of butterflies and moths. *Proceedings of the National Academy of Sciences* 116(45): 22657-22663.
- Kolmogorov M, Bickhart DM, Behsaz B, Gurevich A, Rayko M, Shin SB, Kuhn K, Yuan J, Polevikov E, Smith TPL, Pevzner PA (2020) metaFlye: scalable long-read metagenome assembly using repeat graphs. *Nature Methods* 17: 1103–1110.
- Koren S, Walenz BP, Berlin K, Miller JR, Phillippy AM (2017) Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Research* 27(5): 722–736.
- Kreklová M (2020) Analýza pohlavních chromozomů modrásků (Lycaenidae) Magisterská práce. Jihočeská univerzita, Přírodovědecká fakulta, České Budějovice.
- Kristensen N, Skalski A (1999) Phylogeny and palaeontology. In Handbook of Zoology Vol. IV, Arthropoda: Insecta. *Walter de Gruyter*, Berlin 1: 7–25.
- Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9: 357-359.
- Li H (2018) Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34(18): 3094–3100.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16): 2078-2079.
- Li S, Ajimura M, Chen Z, Liu J, Chen E, Guo H, Tadapatri V, Reddy CG, Zhang J, Kishino H, Abe H, Xia Q, Arunkumar KP, Mita K (2018) A new approach for comprehensively describing heterogametic sex chromosomes. *DNA Research* 25(4): 375-382.

- Lorković Z (1990) The butterfly chromosomes and their application in systematics and phylogeny. *Butterflies of Europe: Introduction to Lepidopterozoology* 2: 332-396.
- Lukhtanov (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.
- Lukhtanov V A, Dantchenko A V, Kandul NR (1997) Die Karyotypen von *Polyommatus (Agrodiaetus) damone damone* und *P. (A.) damocles rossicus* nebst einigen Problemen bei *Agrodiaetus* (Lepidoptera: Lycaenidae). — *Nachrichten des Entomologischen Vereins Apollo* 16: 43-48.
- Maeda T (1939) Chiasma studies in the silkworm *Bombyx mori*. *The Japanese Journal of Genetics* 15: 118-127.
- Marec F, Neven LG, Fukova I (2007) Developing transgenic sexing strains for the release of non-transgenic sterile male codling moths *Cydia pomonella*. *Area-Wide Control of Insect Pests* 103–111.
- Marec F, Neven LG, Robinson AS, Vreysen M, Goldsmith MR, Nagaraju J, Franz G (2005) Development of genetic sexing strains in Lepidoptera: from traditional to transgenic approaches. *Journal of Economic Entomology* 98: 248–259.
- Miller FA (2006) “Your true and proper gender”: the Barr body as a good enough science of sex. *Studies in History and Philosophy of Biological and Biomedical Sciences* 37: 459–483.
- Mola LM, Papeschi AG (2006) Holokinetic chromosomes at a glance. *Journal of Basic and Applied Genetics* 17(1): 17–33.
- Mongue AJ, Nguyen P, Voleníková A, Walters JR (2017) Neo-sex chromosomes in the Monarch butterfly, *Danaus plexippus*. *G3* 710: 3281-3294.
- Mora C, Tittensor DP, Adl S, Simpson AGB, Worm B (2011) How many species are there on earth and in the ocean? *PLOS Biology* 9(8): e1001127.

- Murakami A, Imai HT (1974) Cytological evidence for holocentric chromosomes of the silkworms, *Bombyx mori* and *B. mandarina* (Bombycidae, Lepidoptera). *Chromosoma* 47(2): 167-178.
- Nagaraju J (1996) Sex determination and sex-limited traits in the silkworm, *Bombyx mori*: their application in sericulture. *Indian Journal of Sericulture* 35: 83–89.
- Nakamura D, Tiersch TR, Douglass M, Chandler RW (1990) Rapid identification of sex in birds by flow cytometry. *Cytogenetics and Cell Genetics* 53: 201–205.
- Nguyen P, Sýkorová M, Šichová J, Kůta V, Dalíková M, Čapková Frydrychová R, Neven LG, Sahara K, Marec F (2013) Neo-sex chromosomes and adaptive potential in tortricid pests. *Proceedings of the National Academy of Sciences of the United States of America* 110: 6931-6936.
- Pala I, Naurin S, Stervander M, Hasselquist D, Bensch S, Hansson B (2012) Evidence of a neo-sex chromosome in birds. *Heredity* 108: 264-272.
- Papathanos PA, Windbichler N (2018) Redkmer: An Assembly-Free Pipeline for the Identification of Abundant and Specific X-Chromosome Target Sequences for X-Shredding by CRISPR Endonucleases. *The CRISPR Journal* 1(1): 88-98.
- Pennell MW, Kirkpatrick M, Otto SP, Vamosi JC, Peichel CL, Valenzuela N, Kitano J (2015) Y fuse? Sex chromosome fusions in fishes and reptiles. *PLoS Genetics* 11(5): e1005237.
- Pokorná M, Altmanová M, Kratochvíl L (2014) Multiple sex chromosomes in the light of female meiotic drive in amniote vertebrates. *Chromosome Research* 22(1): 35-44.
- Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26: 841-842.
- Rahman A, Hallgrímsdóttir I, Eisen M, Pachter L (2018) Association mapping from sequencing reads using k-mers. *Elife* 7: e32920.
- Robinson R (1971) *Lepidoptera Genetics*. Pergamon Press, Oxford, 698 pp.

- Sahara K, Marec F, Eikhoff U, Traut W (2003) Moth sex chromatin probed by comparative genomic hybridization (CGH). *Genome* 46: 339–342.
- Seppy M, Manni M, Zdobnov EM (2019) BUSCO: Assessing Genome Assembly and Annotation Completeness. *Methods in Molecular Biology* 1962: 227-245.
- Sigeman H, Ponnikas S, Chauhan P, Dierickx E, Brooke ML, Hansson B (2019) Repeated sex chromosome evolution invertebrates supported by expanded avian sex chromosomes. *Proceedings of the Royal Society B* 286: 20192051.
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31(19): 3210-3212.
- Standage DS, Brown CT, Hormozdiari F (2019) Kevlar: a mapping-free framework for accurate discovery of de novo variants. *iScience* 18: 28–36.
- Stork NE (2018) How many species of insects and other terrestrial arthropods are there on Earth? *Annual Review of Entomology* 63(1): 31–45.
- Suetsugu Y, Futahashi R, Kanamori H, Kadono-Okuda K, Sasanuma S, Narukawa J, Ajimura M, Jouraku A, Namiki N, Shimomura M, Sezutsu H, Osanai-Futahashi M, Suzuki MG, Daimon T, Shinoda T, Taniai K, Asaoka K, Niwa R, Kawaoka S, Katsuma S, Tamura T, Noda H, Kasahara M, Sugano S, Suzuki Y, Fujiwara H, Kataoka H, Arunkumar KP, Tomar A, Nagaraju J, Goldsmith MR, Feng Q, Xia Q, Yamamoto K, Shimada T, Mita K (2013) Large scale full-length cDNA sequencing reveals a unique genomic landscape in a lepidopteran model insect, *Bombyx mori*. *G3 (Bethesda)* 3(9): 1481-1492.
- Šichová J, Nguyen P, Dalíková M, Marec F (2013) Chromosomal evolution in tortricid moths: conserved karyotypes with diverged features. *PLoS One* 8: e64520.
- Talavera G, Lukhtanov VA, Pierce NE, Vila R (2013) Establishing criteria for higher-level classification using molecular data: the systematics of *Polyommatus* blue butterflies (Lepidoptera, Lycaenidae). *Cladistics* 29(2): 166-192.
- Tomaszkiewicz M, Medvedev P, Makova KD (2017) Y and W Chromosome Assemblies: Approaches and Discoveries. *Trends in Genetics* 33(4): 266-282.

- Traut W (1976) Pachytene mapping in the female silkworm *Bombyx mori* L. (Lepidoptera). *Chromosoma* 58: 275-284.
- Traut W, Marec F (1996) Sex chromatin in Lepidoptera. *The Quarterly Review of Biology* 71: 239-256.
- Traut W, Sahara K, Marec F (2007) Sex chromosomes and sex determination in Lepidoptera. *Sexual Development* 1(6): 332-346.
- Traut W, Vogel H, Glöckner G, Hartmann E, Heckel DG (2013) High-throughput sequencing of a single chromosome: a moth W chromosome. *Chromosome Research* 21(5): 491-505.
- Van't Hof AE, Nguyen P, Dalíková M, Edmonds N, Marec F, Saccheri IJ (2013) Linkage map of the peppered moth, *Biston betularia* (Lepidoptera, Geometridae): a model of industrial melanism. *Heredity* 110(3): 283-295.
- Vítková M, Fuková I, Kubíčková S, Marec F (2007) Molecular divergence of the W chromosomes in pyralid moths (Lepidoptera). *Chromosome Research* 15: 917–930.
- Vlašánek P, Bartonová A, Marec F, Konvigka M (2017) Elusive *Parnassius mnemosyne* (Linnaeus, 1758) larvae: habitat selection, sex determination and sex ratio (Lepidoptera: Papilionidae). *SHILAP Revista de lepidopterología* 45(180): 561-569.
- Wang Y, Fu L, Ren J, Yu Z, Chen T, Sun F (2018) Identifying group-specific sequences for microbial communities using long k-mer sequence signatures. *Frontiers in Microbiology* 9: 872.
- Wang Y, Chen Q, Deng C, Zheng Y and Sun F (2020) KmerGO: A Tool to Identify Group-Specific Sequences With k-mers. *Frontiers in Microbiology* 11: 2067.
- Wenger AM, Peluso P, Rowell WJ, Chang PC, Hall RJ, Concepcion GT, Ebler J, Fungtammasan A, Kolesnikov A, Olson ND, Töpfer A, Alonge M, Mahmoud M, Qian Y, Chin CS, Phillippy AM, Schatz MC, Myers G, DePristo MA, Ruan J, Marschall T, Sedlazeck FJ, Zook JM, Li H, Koren S, Carroll A, Rank DR, Hunkapiller MW (2019) Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nature biotechnology* 37(10): 1155-1162.

White MJD (1973) *Animal cytology and evolution*. Cambridge University Press, Cambridge.

Yoshido A, Yasukochi Y, Marec F, Abe H, Sahara K (2007) FISH analysis of the W chromosome in *Bombyx mandarina* and several other species of Lepidoptera by means of *B. mori* W-BAC probes. *Journal of Insect Biotechnology and Sericology*. 76: 1–7.

Zhang J, Li J, Zhou Q (2017) Genomic and Transcriptomic Analyses of Avian Sex Chromosomes and Sex-Linked Genes. *Methods in Molecular Biology* 1650: 69-85.