

Jihočeská univerzita v Českých Budějovicích
Přírodovědecká fakulta

Bakalářská práce

2019

Matěj Miláček

Jihočeská univerzita v Českých Budějovicích
Přírodovědecká fakulta

**Variabilita molekulární diversity a proteinové
struktury Toll-like receptorů v populacích
galapážských drozdců**

Bakalářská práce

Matěj Miláček

Školitel: Mgr. Jakub Vlček

České Budějovice 2019

Bakalářská práce

Miláček M. 2019. Variability of molecular diversity and protein structure of Toll-like receptors in Galapagos mockingbirds. (Variability of molecular diversity and protein structure of Toll-like receptors in Galapagos mockingbirds. Bc. Thesis, in Czech) 47p. Faculty of Science, The University of South Bohemia, České Budějovice, Czech Republic.

Anotace

Fragmentace a zmenšování populací ztrátou habitatů je typickým znakem probíhajícího 6. masové vymírání druhů, z hlediska ochránářské genetiky je nutné porozumět změnám, které s sebou přináší. Pro budoucí rozhodnutí v ochránářské genetice je podstatné porozumět vlivu velikosti populace na variabilitu imunitních genů. Abychom lépe porozuměli genetické variabilitě nestačí analyzovat pouze sekvence DNA, ale je potřeba zaměřit se na funkční proteinovou variabilitu. Tato práce zkoumá fyziologické vlastnosti a struktury jednotlivých haplotypů Toll-like receptorů v alopatrických populacích galapážských drozdců (*Mimus* sp.). Toll-like receptory hrají klíčovou roli v imunitním systému, kde tvoří můstek mezi vrozenou a adaptivní imunitní odpovědí, detekcí patogenů. Data v této práci odhalují variabilitu jejich diversity napříč malými galapážskými populacemi a velkou kontinentální populací drozdce mnohohlasého (*M. polyglottos*). Porovnání heterozygotnosti populací přineslo zjištění, že míra zastoupení heterozygotů je menší v galapážských populacích než v kontinentální populaci. To naznačuje větší vliv diversifikující selekce v kontinentální populaci, nebo redukci genetické variability v galapážských populacích vlivem efektu zakladatele a menší populační velikosti. Kromě popisu genetické diversity byly popsány fyziologické vlastnosti objevených haplotypů, vymodelovány a porovnány struktury všech haplotypů TLR1B, TLR4 a TLR15.

Annotation

Fragmentation of populations and habitat loss is a hallmark of ongoing 6th mass extinction, and in terms of conservation genetics, it is necessary to understand the changes that it entails. It is essential to understand the impact of population size on the variability of the immune genes for the future in conservation genetics. In order to better understand genetic variability, it is not enough to analyze DNA sequences only, but focus on functional protein variability. This paper examines the physiochemical properties and structures of the individual haplotypes of Toll-like receptors in allopatric populations of Galapagos mockingbirds (*Mimus* sp.). Toll-like receptors play a key role in the immune system, where they form the link between innate and adaptive immune responses, by the detection of pathogens. The data in this work reveal the variability of their diversity across the small Galapagos populations and the large continental populations of the northern mockingbird (*M. polyglottos*). Comparison of heterozygosity of populations revealed that the proportion of heterozygotes is smaller in galapagos populations than in the continental population. This suggests a greater impact of diversifying selection in the continental population, or a reduction in genetic variability in Galapagos populations due to founder effect and smaller population size. In addition to the description of genetic diversity, the physiochemical properties of the discovered haplotypes were described, the structures of all TLR1B, TLR4 and TLR15 haplotypes were modeled and compared.

Prohlašuji, že svoji bakalářskou práci jsem vypracoval samostatně pouze s použitím pramenů a literatury uvedených v seznamu citované literatury.

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

V Českých Budějovicích, 17. 4. 2019.

.....

Matěj Miláček

Poděkování

Tímto bych rád poděkoval svému školiteli Kubovi Vlčkovi za odborné vedení, přátelský přístup a cenné rady po celé tři roky. Chtěl bych také poděkovat Honzovi Štefkovi a Marušce Krausové za rady a pomoc v nouzi. Dále celé laboratoři molekulární ekologie a evoluce za dobrou náladu a příjemné pracovní prostředí. Také chci poděkovat rodině za podporu při studiu a v neposlední řadě Michalovi Vinklerovi za uhrazení nákladů na sekvenování vzorků.

Obsah

1	Úvod a cíle práce	1
1.1	Koevoluce mezi hostiteli a patogeny	1
1.2	Toll-like receptory	2
1.3	Galapágy a galapážští drozdci	4
1.4	Možnosti predikce proteinové struktury	6
1.5	Cíle práce	7
2	Metody	7
2.1	Sekvenační data	7
2.2	Analýza hlavních komponent (PCA)	8
2.3	Modelování proteinové struktury	8
2.4	Analýza povrchového náboje (PIPSA)	9
2.5	Strukturní alignment (super)	10
3	Výsledky	11
3.1	Sekvenační data	11
3.1.1	TLR1B	11
3.1.2	TLR4	11
3.1.3	TLR15	11
3.1.4	Porovnání galapážských populací s kontinentální populací	12
3.2	Analýza hlavních komponent (PCA)	12
3.3	Analýza povrchového náboje (PIPSA)	15
3.4	Strukturní alignment (super)	25
4	Diskuze	27
4.1	Nízká variabilita TLR v galapážských populacích	28
4.2	Odlišnosti TLR15	29
4.3	Mezidruhový polymorfismus (TSP) v galapážských populacích	29
4.4	Analýza struktury TLR	29
5	Závěr	31
6	Seznam použité literatury	31
7	Přílohy	39

7.1	Seznam tabulek	39
7.2	Seznam obrázků	40
7.3	Tabulky	41
7.4	Obrázky	46

1 Úvod a cíle práce

Jelikož v současné době čelíme trendu dalšího masového vymírání druhů (Barnosky et al. 2011; Pimm et al. 2014), fragmentaci populací vytvářením umělých bariér a zmenšování populací zabíráním přirozených habitatů, je důležité porozumět právě zákonitostem týkajících se malých populací. Takové populace se podobají izolovaným ostrovním populacím jako například populacím galapážských drozdců (*Mimus*. sp), kteří na Galapágách tvoří izolované populace. Analyzováním diversity imunitních genů jako jsou Toll-like receptory, či geny hlavního histokompatibilního komplexu poté můžeme zjistit, jaké mechanismy se podílejí na utváření variability v těchto malých populacích, která je zásadní pro jejich přežití (Reed & Frankham 2003). Analyzování fyzi chemických vlastností a struktury produktů zmíněných genů pomůže více porozumět zákonitostem souboje mezi patogeny a jejich hostiteli.

1.1 Koevoluce mezi hostiteli a patogeny

Jednou z hlavních evolučních sil v přirozených podmínkách je souboj mezi hostiteli a patogeny. Právě díky koevolučním procesům mezi hostiteli a jejich patogeny můžeme u vyšších organismů najít různorodé obranné mechanismy, které se tohoto boje účastní a zároveň jsou jeho produktem. Imunitní systém hostitelů se snaží patogeny rozpoznat a zničit. Při adaptaci na nového hostitele dochází v populaci patogenu k optimalizaci virulence směrem k maximalizaci fitness v závislosti na hustotě hostitelské populace a pravděpodobnosti přenosu (Ebert 2002). Molekulární změny u patogenů vyvolávají reciproční tlak na změnu hostitelského imunitního systému (Slev & Potts 2002). Tento efekt popisuje teorie červené královny. Ta se týká neustálé potřeby organismů se vyvíjet ne proto, aby získaly evoluční výhodu, ale proto, aby dokázaly přežít a odolat ostatním vyvíjejícím se organismům (Van Valen 1973). Van Valenova teorie uvádí, že z hlediska pravděpodobnosti vyhynutí populace záleží především na schopnosti daného organismu přizpůsobit se prostředí, ve kterém se nachází. To zahrnuje i patogeny, či predátory, kteří se vyvíjejí spolu s ním.

V současnosti se řada studií zabývá otázkou, zda v malých populacích převládá v evoluci imunitních genů patogeny řízená selekce, nebo je hlavní silou náhodný genetický drift. Se snižující se populační velikostí klesá efektivita selekce a roste vliv náhodného genetického driftu (Ohta 1973) a tudíž v malých populacích bude osud výhodných mutací spíše ovlivněn genetickým driftem. Zásadními faktory jsou síla selekce (selekční koeficient) a efektivní velikost populace. Osud nové mutace je dle „téměř neutrální teorie molekulární evoluce“ (Ohta 1992) určen právě součinem těchto dvou faktorů. Když se tedy objeví mutace

s velmi vysokým kladným selekčním koeficientem, tak s vysokou pravděpodobností dojde k její fixaci jak ve velké, tak malé populaci. Nicméně pokud se objeví mutace se slabším selekčním koeficientem, tak se ve velké populaci s jistou pravděpodobností udrží a případně zafixuje, kdežto v malé populaci nad selekci převáží náhoda a mutace se s větší pravděpodobností vytratí (Akashi et al. 2012).

Pro přežití populace je tedy důležitá velikost populace a s ní související genetická diversita (Reed & Frankham 2003). S nízkou genetickou diversitou populace nedokáže čelit novým patogenům např.: havajští šatovníci (Jarvi et al. 2004; Wikelski et al. 2004). Nicméně lze v ochranné genetice najít i případy, kdy se populace dokázaly zotavit z extrémně malých počtů a tím i extrémně nízké genetické diversity, např.: lejsčáci *Petroica traversi* (Ardern & Lambert 1997). Tato studie používala jako genetický marker mikrosatelity. To jsou nekódující oblasti DNA, které nejsou pod selekčním tlakem. A přestože představují mikrosatelity dobrý nástroj pro popis korelace fitness a heterozygotnosti populace (Forstmeier et al. 2012), nemusí věrně odrážet diversitu dalších oblastí genomu (Váli et al. 2008). Pro dosažení lepších výsledků je vhodné kombinovat selekčně neutrální marker s kódujícími oblastmi DNA, jako například imunitními geny (MHC, TLR), které podléhají selekci (Sommer 2005; Hartmann et al. 2014).

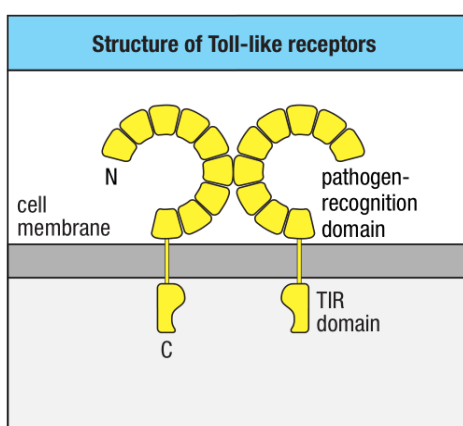
MHC – hlavní histokompatibilní komplex – představuje klíčovou součást adaptivního imunitního systému. Geny MHC jsou oblíbeným markerem populačních studií (Jarvi et al. 2004; Bollmer et al. 2010; Vlček et al. 2016), který přinesl mnoho důkazů o koevoluci mezi patogeny a hostiteli (Spurgin & Richardson 2010). Jejich nevýhodou je ale velký počet lokusů. Stejnými primery tedy lze amplifikovat obrovské množství alel, z nichž některé mohou být pseudogeny, které nemají žádný vliv na fitness populace a je potřeba tyto nefunkční alely vyfiltrovat dalšími metodami. Pro komplexnější pochopení složitého imunitního systému je důležité brát v potaz i ostatní imunitní geny, například geny kódující receptory NK buněk (natural killers), či geny chemokinových receptorů (CCR) (Acevedo-Whitehouse & Cunningham 2006). Mezi jednodušší, ale stále komplexní systémy imunitních genů se řadí Toll-like receptory.

1.2 Toll-like receptory

Toll-like receptory (TLR) patří mezi patern rozpoznávající receptory – PRRs (pattern-recognition receptors) – které rozeznávají konzervované struktury patogenů – PAMPs (pathogen-associated molecular patterns) – jako CpG motif DNA, lipopolysacharid nebo flagelin. Jsou odvozeny z Toll receptorů, které se při embryonálním vývoji podílí na

dorzoventrální orientaci embrya octomilky *Drosophila melanogaster* (Hashimoto et al. 1988; Hansson & Edfeldt 2005). Řadí se tedy do vrozené složky imunitního systému. U savců byly poprvé popsány v roce 1997 (Medzhitov et al. 1997). Předpokládá se, že k rozvoji TLR u obratlovců došlo před či během kambria (Roach et al. 2005).

Všechny TLR se skládají ze tří částí. Z ligand vázající extracelulární domény (ECD), jednoduché transmembránové šroubovice a intracelulární domény (TIR). To je nejkonzervovanější část proteinu, jejíž homology byly nalezeny i v rostlinách (Akira et al. 2001; O'Neill et al. 2003; Burch-Smith & Dinesh-Kumar 2007). ECD tvoří běžně 19-25 na leucin bohatých repetitivních – LRRs (leucine-rich repeats) – které udávají TLR jejich charakteristický tvar podkovy (Obr. 1). ECD je ohraničena dvěma subdoménami. N-konec



Obr. 1: Struktura Toll-like receptorů (převzato z Parham 2009).

proteinu tvoří LRR-NT motif složený z β -vláseček spojených disulfidickou vazbou a na C-konci ECD se nachází LRR-CT motif tvořený dvěma α -šroubovicemi (Botos et al. 2011). Oproti ostatním LRR proteinům, jako jsou variabilní lymfocytové receptory (VLR) či ribonukleázové inhibitory, je u TLR vazebné místo pro ligand na boku vzestupné části ECD (Bella et al. 2008; Liu et al. 2008; Kang et al. 2009; Park et al. 2009).

Při navázání ligandu TLR dimerizují a jejich signalizace spouští odpověď vrozené imunity (O'Neill & Bowie 2007; Jin & Lee 2008). TLR tvoří můstek mezi vrozenou a získanou složkou imunity. Spuštění odpovědi získané imunity je podmíněno stimulací antigen-prezentujících buněk skrze TLR (Akira 2009). Toll-like receptor vázající PAMP aktivuje dendritické buňky, které následně pomocí MHC receptorů prezentují antigen na svém povrchu a pomáhají tak k aktivaci a proliferaci T-lymfocytů, které jsou složkou získané imunity (Iwasaki & Medzhitov 2010).

U ptáků bylo objeveno 10 TLR z nichž 5 představuje přímé ortology savčích Toll-like receptorů. Jedná se o TLR2A, TLR2B, TLR4, TLR5 a TLR7 (Brownlie & Allan 2011). Mezi zbylých pět se řadí TLR1A a TLR1B, TLR3, TLR15 a TLR21. Každý z nich rozeznává jiné struktury patogenů. U ptáků došlo k duplikaci TLR1 na TLR1A a TLR1B, kdy TLR1B rozpoznává lipoproteiny (Kang et al. 2009; Werling et al. 2009; Alcaide & Edwards 2011; Velová et al. 2018). U savců dimerizuje TLR2 s TLR1 a dalšími kofaktory a váže triacylované mikrobiální lipoproteiny (Brownlie & Allan 2011). TLR4 váže pomocí proteinu MD-2 lipopolysacharid (LPS), to je hlavní součást vnější membrány gram-negativních bakterií

(Akira et al. 2001; O'Neill & Bowie 2007). TLR3 je exprimován na endosomálních membránách a detekuje dvouvláknovou RNA jenž je typická pro dsRNA viry. TLR5 váže flagelin, protein tvořící bičíky bakterií. Někteří ptáci, např.: *Chaetura pelagica*, *Nestor notabilis* mají pouze pseudogen TLR5, který je nefunkční. Stejně jako TLR3, TLR7 je exprimován endosomálně a váže RNA v její častější jednovláknové podobě. U některých ptáků jako například *Taeniopygia guttata*, *Cuculus canorus*, či zmíněných *Petroica traversi* došlo k duplikaci tohoto TLR (Akira 2009; Grueber et al. 2012; Keestra et al. 2013; Velová et al. 2018). Toll-like receptor 15 je unikátní pro ptáky a plazy a spouští signalizaci poté, co je štěpen mikrobiálními proteázami (de Zoete et al. 2011; Boyd et al. 2012). Endosomálně exprimovaný TLR21 váže CpG motif DNA (Brownlie & Allan 2011). Ukazuje se tedy, že ačkoliv každý TLR váže různé typy patogenních struktur, některé TLR se doplňují (TLR4, TLR15, TLR21) například v detekci bakteriálních struktur, kdy každý TLR váže jiný prvek typický pro bakterie.

Jelikož TLR5 se u některých druhů může vyskytovat pouze ve formě pseudogenu byl nevhodný pro tuto práci. Stejně tak TLR7, který je u některých druhů duplikovaný. Z důvodu funkčnosti primerů a předběžně analyzované diversity byly vybrány TLR1B, TLR4 a TLR15.

1.3 Galapágy a galapážští drozdci

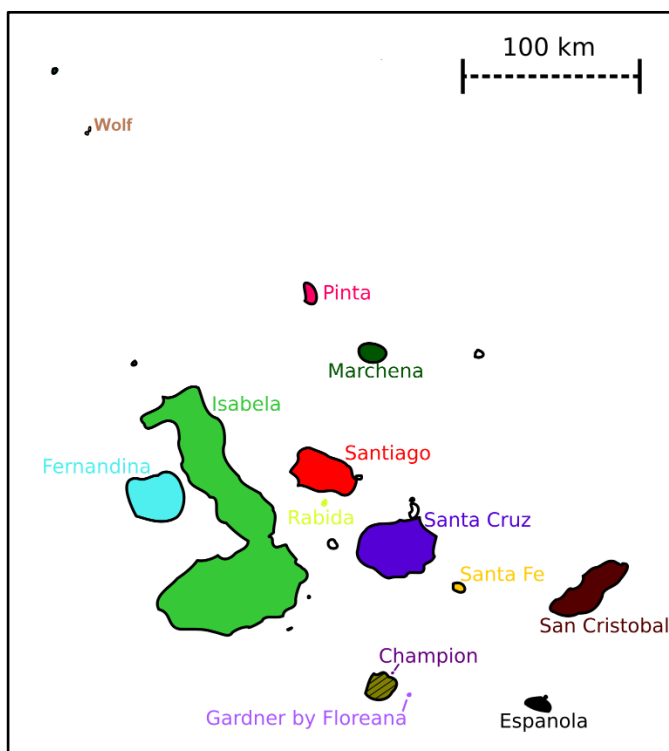
Galapágy jsou díky svým ostrovům a relativně velkým vzdálenostem mezi nimi vhodným systémem pro studium malých populací. Právě studium malých populací je důležité z hlediska ochránářské genetiky, poněvadž nyní čelíme trendu označovanému jako 6. masové vymírání, či vymírání v holocénu (Barnosky et al. 2011; Pimm et al. 2014). Malé populace jsou nejvíce ohrožené vyhynutím, neboť kvůli nízké genetické diversitě nedokáží čelit změnám jako jsou invazivní organismy, či působení člověka ve smyslu zavlečení cizích nemocí (Jarvi et al. 2004; Wikelski et al. 2004), zániku habitatů a dalších (Stork 2010).

Na Galapážském souostroví se vyskytují 4 druhy drozdců: *Mimus macdonaldi*, vyskytující se na ostrově Española na jihu souostroví; *M. melanotis*, endemit ostrova San Cristobal; *M. trifasciatus*, který vymřel na ostrově Floreana, se vyskytuje jen na ostrůvcích Champion a Gardner by Floreana; a *M. parvulus*, nejhojnější drozdec, vyskytující se na severozápadě

souostroví (Nietlisbach et al. 2013).

Drozdcí rodu *Mimus* na Galapágách tvoří alopatrické populace (Obr. 2). Migrace mezi ostrovy je velice malá, a proto jsou jednotlivé populace unikátními evolučními jednotkami. Výjimku tvoří sousední pobřeží ostrovů Fernandina a Isabela a ostrovy Santiago a Rabida (Hoeck et al. 2010; Štefka et al. 2011).

Dle mutační rychlosti kódujících oblastí mitochondriální DNA proběhla kolonizace souostroví drozdcí před 1,6-5,5 miliony lety (Arbogast et al. 2006). To znamená, že v té době došlo k bottlenecku, označovanému jako efekt zakladatele, kdy je genofond omezen na diversitu kolonizujících jedinců. Po kolonizaci došlo k alopatrické radiaci mezi jednotlivými částmi souostroví, přičemž se předpokládá, že pro



Obr. 2: Alopatrické populace galapážských drozdců.

Z ostrovů, u kterých je uvedeno jméno byly sebrány vzorky.

omnivorní drozdce prostředí Galapág nepředstavuje dostatečně rozrůzněné niky umožňující sympatrický výskyt více druhů (Arbogast et al. 2006). Analýza neutrálních markerů prokázala, že je genetická diversita posledních 100 let stabilní, to značí, že v populacích došlo k ustálení rovnováhy mutací a driftu (Hoeck et al. 2010). Hoeck a Keller (Hoeck & Keller 2012) se ve své práci věnovali vlivu inbreedingu – příbuzenského páření – na fitness galapážských drozdců. Fitness drozdců posuzovali z hlediska imunokompetence a neprokázali efekt inbreedingu ani na jednu ze čtyř zkoumaných imunokompetentních vlastností (poměr heterofilů a lymfocytů, lýzi a aglutinaci krve a počet ektoparazitů – vši). To naznačuje, že některé znaky fitness nemusí souviset s genetickou diversitou populace, ale s přítomností konkrétních alel (Spielman et al. 2004).

Studie kombinující mikrosatelitní sekvence a antigen-vázající doménu (β -doména) MHC II. třídy (Vlček et al. 2016) odhalila pozitivní korelaci velikosti ostrova a počtu alel (mikrosatelitních i MHC) na populaci. Protože počty mikrosatelitních alel byly proporcčně nižší než počty MHC alel, musí zde působit nějaký mechanismus potlačující redukční roli

driftu. Pomocí rozdílu nesynonymních a synonymních mutací (dN-dS) byla objevena balancující selekce operující na celém lokusu β -domény MHC II. třídy. Důležitost konkrétních MHC alel naznačuje fakt, že velká část alel byla přítomna ve více populacích, či druzích galapážských drozdců. Tento mezidruhový polymorfismus (TSP) není pouze fenomén MHC genů, ale týká se i dalších imunitních genů, kdy jedna výhodná alela může být sdílena vícero druhů (Těšický & Vinkler 2015).

Zatímco u MHC genů působí balancující selekce, u TLR byla objevena purifikující selekce, značící, že větší variabilita TLR negativně ovlivňuje fitness jedinců (Grueber et al. 2013; Hartmann et al. 2014). Při srovnání s neutrálním markerem tedy vykazují TLR menší diversitu. Hlavní silou při formování genetické diversity TLR v malých populacích je nicméně genetický drift (Grueber et al. 2013; Hartmann et al. 2014; Gonzalez-Quevedo et al. 2015; Gilroy et al. 2017; Nelson-Flower et al. 2018).

1.4 Možnosti predikce proteinové struktury

Mnohé studie týkající se ochránářské genetiky analyzují genetickou diversitu jen se znalostmi DNA sekvence. Takový přístup může ukázat více variability než skutečně existuje. Abychom věrně popsali funkční variabilitu genetické diversity, je potřeba přeložit DNA sekvence do proteinových, které eliminují synonymní změny v DNA sekvencích. Pro ještě optimálnější výsledky je analýza fyzichemických vlastností aminokyselin.

Pro přesné popsání sekundární a vyšší struktury proteinů je možno využít dvou experimentálních metod – rentgenové krystalografie a spektroskopie nukleární magnetické rezonance (NMR). Rentgenová krystalografie využívá ohyb paprsku rentgenového záření po průchodu krystalem. Takový ohyb – difrakce – je nenáhodný a díky tomu jsme schopni z úhlů difrakce a intenzity dopadu elektronů zjistit informace například o poloze atomů, či jejich vazbách (Svergun et al. 2001). Přestože má tato metoda vynikající výsledky, je experimentálně velmi náročná. NMR spočívá v interakci jader atomů s magnetickým polem, ve kterém se pohybují. Procházející jádro mění resonanci magnetického pole a ze změny signálu magnetického pole můžeme vyčíst informace o množství, složení, prostorové struktuře a dalších vlastnostech zkoumané látky (Wüthrich 2001).

V současné době převažuje při modelování struktury proteinů přístup *in silico*. Díky výkonným počítačům můžeme modelovat strukturu proteinů dvěma způsoby. První, stejně jako rentgenová krystalografie a NMR, modeluje strukturu *de novo*. Tato metoda předpokládá, že v přirozených podmínkách mají proteiny takovou strukturu, která disponuje minimální volnou energií. Vybírá z mnoha možných konformací krátkých úseků dané sekvence a vybírá

ty s nejnižší volnou energií (Baker & Sali 2001). Druhý přístup používá rozsáhlé databáze známých struktur a novou sekvenci buďto navléká (threading), či porovnává s templátem (Jilong et al. 2013; Yang et al. 2015). Na poli predikce proteinové struktury byl v posledních letech učiněn velký pokrok, a tak se tyto metody velmi zpřesnily, přesto však nedosahují přesnosti rentgenové krystalografie. *Z in silico* metod jsou přesnější predikce podle templátu (Moult et al. 2018)

1.5 Cíle práce

Tato práce si dává za cíl popsat funkční variabilitu Toll-like receptorů 1B, 4 a 15 v populacích galapážských drozdců (*Mimus* sp.) a v kontinentální populaci *M. polyglottos*. Jelikož variabilita TLR ovlivňuje schopnost vázat struktury patogenů a tím působit na fitness, tak takto získaná data mají potenciál přispět k objasnění mechanismů selekce v udržení životaschopnosti malých, či ohrožených populací obratlovců. Parciální cíle práce jsou:

1. Popsat diversitu sekvencí DNA.
2. Komplexně analyzovat fyziologické vlastnosti jednotlivých haplotypů TLR.
3. Sestavit 3D alignment proteinové struktury.
4. Porovnat objevené vlastnosti haplotypů mezi galapážskou a kontinentální populací.

2 Metody

2.1 Sekvenační data

Sekvence DNA jednotlivých TLR používaných v této práci byly získány amplikonovým sekvenováním na přístroji Illumina MiSeq. Amplifikované úseky byly vybrány tak, aby obsahovaly ligand vázající doménu (pokud byla známá), případně variabilní místa popsaná Velovou (Velová et al. 2018). V rámci cílových úseků délky 889 a 970 bp u TLR15 a TLR4 byly navrženy tři páry primerů (Tab. 2) tak, aby se výsledné produkty překrývaly minimálně 90 bázemi, přičemž výsledný produkt musel být kratší než 500 bp z důvodu limitace délky “pair-end” čtení na MiSeq. Pro TLR1B byl navržen jeden pár primerů (Tab. 2) amplifikující fragment délky 472 bp. Osekvenováno bylo přes 200 jedinců drozdců z 12 ostrovů pro každý receptor (Tab. 6). Sekvence byly následně očištěny od nekvalitně přečtených bází a adaptorů. Párové sekvence byly spojeny díky překryvu a v případě TLR4 a 15 byly haplotypy jedinců asemblovány ze třech překrývajících se fragmentů. Diploidní data byla následně fázována pomocí programu PHASE (Scheet & Stephens 2006). Výsledkem sekvenování byly tedy haplotypy extracelulárních domén tří TLR pro více než 200 jedinců

Galapážských drozců a kontinentální populace *M. polyglottos* z Kalifornie (Tab. 3-5). Genetická variabilita TLR v rámci jednotlivých populací (nukleotidová diverzita, počty haplotypů, heterozygotnost, Tajimovo D) byla analyzována pomocí vlastního skriptu a programu DNAsp v6 (Rozas et al. 2017). Poměr nesynonymních a synonymních mutací, tedy index, který zohledňuje jaký typ selekce působí na danou molekulu byl vypočten algoritmem SLACK (Kosakovsky Pond & Frost 2005) na serveru Datamonkey (Weaver et al. 2017). Výpočty heterozygotnosti proteinových haplotypů byly spočítány jako počet heterozygotů v populaci. Generování sekvenačních dat bylo provedeno mým školitelem, metodika mé vlastní práce je popsána v následujících odstavcích.

2.2 Analýza hlavních komponent (PCA)

Na sekundární struktuře se spolu s primární strukturou podílejí fyzi chemické vlastnosti aminokyselin. Avšak mnohé recentní práce týkající se ochránářské genetiky se zaměřují pouze na sekvence DNA. Přínos této práce je v analyzování fyzi chemických vlastností aminokyselin, které jsou důležitým faktorem při utváření vyšších stupňů struktury proteinů. Otázka byla, zda se haplotypy liší a zda existuje nějaký vzorec těchto vlastností mezi galapážskými a kontinentálními haplotypy.

Vlastnosti všech haplotypů byly shrnuty a vzájemně porovnány analýzou hlavních komponent – PCA (Principal Component Analysis). Tato analýza zjednodušuje n-rozměrný prostor n-proměnných do dvourozměrného, či třírozměrného prostoru s minimálním možným zkreslením a maximalizací variability na několika prvních osách (Lepš & Šmilauer 2016).

Haplotypy byly porovnány z hlediska náboje (formal charge), hydropatie, normalizovaného van der Waalsova objemu (Fauchere et al. 1988) a účinné energie dělení (Miyazawa & Jernigan 1985). Vlastnosti byly vybrány tak, aby byly nepodobné a co nejvíce charakterizovaly povahu aminokyselin. Porovnávání byly vlastnosti pouze segregujících míst v haplotypech. Soubor hodnot byl získán pomocí programu Quantiprot (Konopka et al. 2017), který používá databázi vlastností aminokyselin AAindex (Kawashima et al. 2008). Vlastnosti pak byly pomocí PCA v programu R (R Core Team 2016) převedeny do dvourozměrného grafu, který popisuje takto sumarizovanou fyzi chemickou vzdálenost různých haplotypů.

2.3 Modelování proteinové struktury

Pro porovnání rozdílů ve struktuře a dalších vlastnostech například povrchovém náboji či konzervovanosti aminokyselin bylo potřeba vytvořit proteinové 3D struktury TLR. K modelování sekvenovaných částí ECD doplněných o konce TLR lejska bělokrkého

(*Ficedula albicollis*) (Tab. 1) byl použit program I-TASSER (Zhang 2008; Roy et al. 2010; Yang et al. 2015). Doplnění zkoumané sekvence o lejsčí bylo podstatné z důvodu vyhledání odpovídajícího templátu a správného modelování sekundární struktury proteinu.

Tab. 1: Přístupové kódy databáze NCBI pro referenční Toll-like receptory lejska bělokrkého.

TLR	Přístupový kód NCBI
TLR1B	XP_005045420.1
TLR4	XM_005055449.1
TLR15	XM_005042682.1

Každý kód je jedinečný pro sekvenci lejsčího TLR. Pokud začíná na „XP“ jedná se o sekvenci aminokyselin, pokud začíná „XM“ představuje sekvenci mRNA. Sekvence mRNA byly přeloženy v programu Geneious (Geneious 10.2.2, <https://www.geneious.com>).

I-TASSER používá hierarchický přístup pro predikci proteinové struktury a funkce. Nejprve identifikuje templáty pomocí navlékání sekundární struktury zadané sekvence na potenciální templát z databáze proteinových struktur na webových stránkách I-TASSERu (<https://zhanglab.ccmb.med.umich.edu/library/>). Iteračně simuluje sestavení struktury. Poté zvolí vhodné modely podle volné energie a dalších faktorů, tyto modely jsou následně upraveny a anotovány na základě struktury (Yang et al. 2015).

Nejlepší strukturní model byl zvolen podle tzv. C-skóre, to je skóre spolehlivosti identifikovaných modelů, při počítání zohledňuje signifikanci alignmentů porovnávaných struktur a konvergenční parametry simulací sestavujících strukturu. Toto skóre se pohybuje v rozmezí (-5, 2), přičemž větší hodnota znamená větší spolehlivost (Yang et al. 2015). Všechny vybrané strukturní modely byly modelovány podle stejných templátů v rámci TLR.

Modely byly vizualizovány v programu PyMOL (The PyMOL Molecular Graphics System, Version 2.2.3 Schrödinger, LLC). Byly označeny aminokyseliny na N- a C-koncích sekvenované části a bylo vyznačeno ligand-vázající místo – TLR1B (Jin et al. 2007) a TLR4 (Velová et al. 2018) – či místo štěpení – TLR15 (de Zoete et al. 2011) – (Obr. 18).

2.4 Analýza povrchového náboje (PIPSA)

Jednou z cest, kterou lze porovnat podobnost různých haplotypů, je analýza povrchového náboje. Pro tento účel byl použit program „analýzy podobnosti vlastností proteinové interakce“ (PIPSA) (Wade et al. 2001; Gabdoulline et al. 2007; Richter et al. 2008). PIPSA porovnává elektrostatický potenciál struktur. Jako vstupní data byly použity proteinové 3D struktury vymodelované I-TASSERem (.PDB formát). 5 haplotypů pro TLR1B, 18 haplotypů pro TLR4 a 22 haplotypů pro TLR15. Byla vybrána možnost

superpozice „optimizesup2pdbs“. Elektrostatický potenciál struktur byl vypočítán „adaptivním Poisson-Boltzmannovým řešitelem“ (APBS). Parametry prostředí měly výchozí nastavení, teplota 300 K, iontová síla 50 mM. Porovnávány jsou elektrostatické potenciály povrchových obalů proteinů, kdy povrch je definován sondou o poloměru 2 Å a povrchový obal se rozpíná od této vzdálenosti další 3 Å. Software počítá Hodgkinovy a Carbovy indexy podobnosti (Gabdoulline et al. 2007). Indexy jsou v intervalu od <-1,+1>, kdy -1 značí anti-korelující, 0 nekorelující a +1 zcela-korelující potenciál. Tyto hodnoty jsou poté převedeny na elektrostatické vzdálenosti vztahem $\sqrt{2 - 2SI}$, kde SI je odpovídající index podobnosti. Elektrostatické vzdálenosti spadají do intervalu <0,2>, přičemž 0 znamená identičnost a 2 anti-korelaci. Klastrovací analýza a výpočet epogramu struktur je proveden v programu R (Ihaka & Gentleman 1996). Výsledky jsou následně vizualizovány jako epogram, tedy skupinový diagram, který seskupuje struktury s nejvíce podobnou elektrostatickou vzdáleností, a jako teplotní mapa, která znázorňuje podobnost struktur z hlediska elektrostatické distance v párovém porovnání (Richter et al. 2008).

Elektrostatický potenciál povrchu proteinů byl vizualizován v programu PyMOL (The PyMOL Molecular Graphics System, Version 2.2.3 Schrödinger, LLC).

2.5 Strukturální alignment (super)

Jelikož schopnost detekovat a vázat patogeny je přímo spřažena se strukturou bílkovin, byla porovnána sekundární struktura 3D modelů v programu PyMOL (The PyMOL Molecular Graphics System, Version 2.2.3 Schrödinger, LLC). Použitá funkce „super“ provede sekvenčně nezávislý strukturální alignment dvou struktur. Poté následuje několik cyklů doladění pro vylepšení zarovnání struktur. Byly zvoleny výchozí hodnoty pro doladění alignmentu (počet cyklů: 5, „cutoff“: 2.0). Pro každou dvojici struktur je vypočítáno RMSD, to je míra průměrné vzdálenosti mezi atomy struktur. RMSD identických struktur je 0 a s přibývajícím odlišností struktur RMSD roste. V rámci každého TLR byly proti sobě porovnány všechny haplotypy a jejich vzájemné RMSD bylo vizualizováno v tepelné mapě pomocí webu programu Heatmapper (<http://www2.heatmapper.ca/>) (Babicki et al. 2016). Každý řádek hodnot je považován za bod a vzdálenost mezi body byla spočítána podle Eukleidovské metriky, kde je vzdálenost vypočítána jako délka úsečky spojující dvě hodnoty. Alignment všech struktur v rámci každého TLR byl nakonec vizualizován v programu PyMOL.

3 Výsledky

3.1 Sekvenační data

Podářilo se osekvenovat část ECD tří TLR (Obr. 18) u více než 200 jedinců z 13 galapážských populací a jedné kontinentální populace příbuzného druhu (*M. polyglottos*). U TLR15 sekvenovaný úsek obsáhl štěpné místo a u TLR1B celé ligand vázající místo. U TLR4 přesahuje ligand vázající část receptoru sekvenovaný úsek, který končí GLN-438 (Obr. 18). Celkově bylo objeveno 39 segregujících aminokyselin, rozdělujících tři TLR na 45 proteinových haplotypů. Všechny haplotypy s výjimkou jednoho se nachází buďto v galapážských populacích, nebo v kontinentální populaci.

3.1.1 TLR1B

Pro TLR1B bylo nalezeno 5 proteinových haplotypů o délce 472 aminokyselin (Tab. 6) lišící se ve 4 pozicích. Ze všech haplotypů je zdaleka nejrozšířenější haplotyp 1, který se nachází jak na Galapágách, tak i na kontinentě. Ostatní haplotypy jsou specifické buďto pro jednu galapážskou populaci, či pro kontinentální populaci (Tab. 3). Nukleotidová diversita (π) haplotypů je 0,00096 a Tajimovo D se rovná -1,265 ($p > 0,05$), jelikož není průkazné, nedá se vyloučit rovnováha mutací a driftu.

3.1.2 TLR4

U TLR4 bylo nalezeno 18 proteinových haplotypů mezi 229 jedinci. Každý haplotyp je dlouhý 889 aminokyselin (Tab. 6) a haplotypy obsahují 19 segregujících míst. Nejvíce rozšířeným je haplotyp 10, který se nachází ve všech populacích s výjimkou kontinentu a populace Champion (Tab. 4). Nukleotidová diversita (π) je 0,00218 a hodnota Tajimova D se rovná -1,655 ($p > 0,05$), to je velmi blízko nuly a nelze vyloučit rovnováhu mutací a driftu.

3.1.3 TLR15

Sekvence TLR15 od 201 jedinců obsahuje 22 proteinových haplotypů dlouhých 970 aminokyselin (Tab. 6) s 16 segregujícími místy. Nejvíce rozšířeným haplotypem je č. 10, který se vyskytuje v 9 galapážských populacích (Tab. 5). Nukleotidová diversita (π) je 0,02742 a Tajimovo D se rovná -1,495 ($p > 0,05$), vzhledem k tomu, že není signifikantní, nedá se vyloučit rovnováha mutací a driftu.

3.1.4 Porovnání galapážských populací s kontinentální populací

U TLR1B byly objeveny 2 galapážské a 2 kontinentální proteinové haplotypy a jeden vyskytující se v obou populacích (Tab. 3). Přestože aminokyselinová sekvence tohoto haplotypu je stejná, DNA sekvence se liší. A ačkoliv počet proteinových haplotypů je stejný, nukleotidová diversita (π) je vyšší v kontinentální populaci (0,0008), než na Galapágách (0,0002). Poměr pn/ps nebylo možné spočítat pro nedostatečná data. Pozorovaná proteinová heterozygotnost zprůměrovaná pro všechny galapážské populace se výrazně lišila ($<0,01$) od té pozorované v kontinentální populaci (0,125) a byla nejnižší ze všech tří TLR (Tab. 7). Rozdíl pozorované DNA a proteinové heterozygotnosti není u galapážských populací velký ($>0,05$), nicméně u kontinentální populace se DNA a proteinová het. liší o 0,185.

Lokus TLR4 obsahuje 8 galapážských proteinových haplotypů a 10 kontinentálních (Tab. 3). Kontinentální populace disponuje vyšší nukleotidovou diversitou (0,0027) než galapážská (0,00079), poměr pn/ps (Tab. 7) není výrazně rozdílný, nicméně větší nadbytek synonymních mutací (větší možný vliv purifikující selekce) vykazují galapážské populace (0,0746). Velmi patrný rozdíl je v proteinové heterozygotnosti, kdy u kontinentální populace dosahuje vysoké hodnoty (0,588), zatímco galapážské méně než 0,01. Tomu může napomáhat také příbuzenské křížení v malých galapážských populacích. Rozdíl DNA a proteinové heterozygotnosti je markantní u obou populací. U galapážské dosahuje rozdíl $<0,39$, u kontinentální populace je to ještě více ($<0,41$).

Pro TLR15 bylo nalezeno 10 galapážských a 12 kontinentálních proteinových haplotypů (Tab. 7). π je v kontinentální populaci výrazně větší (0,0062) než v galapážských populacích (0,00073). Poměry pn/ps jsou velmi podobné (Tab. 7) a vykazují přebytek synonymních mutací. Proteinová heterozygotnost se liší velmi, kdy v kontinentální populaci se vyskytuje 90,9 % heterozygotů (het.=0,909), kdežto v galapážských populacích zastoupení heterozygotů dosahuje 19,9 % (het.=0,1988), to je očividně větší procento heterozygotů než u TLR1B a TLR4. Rozdíl DNA a proteinové heterozygotnosti je značný, kdy je jak u galapážské, tak u kontinentální $\sim 0,1$.

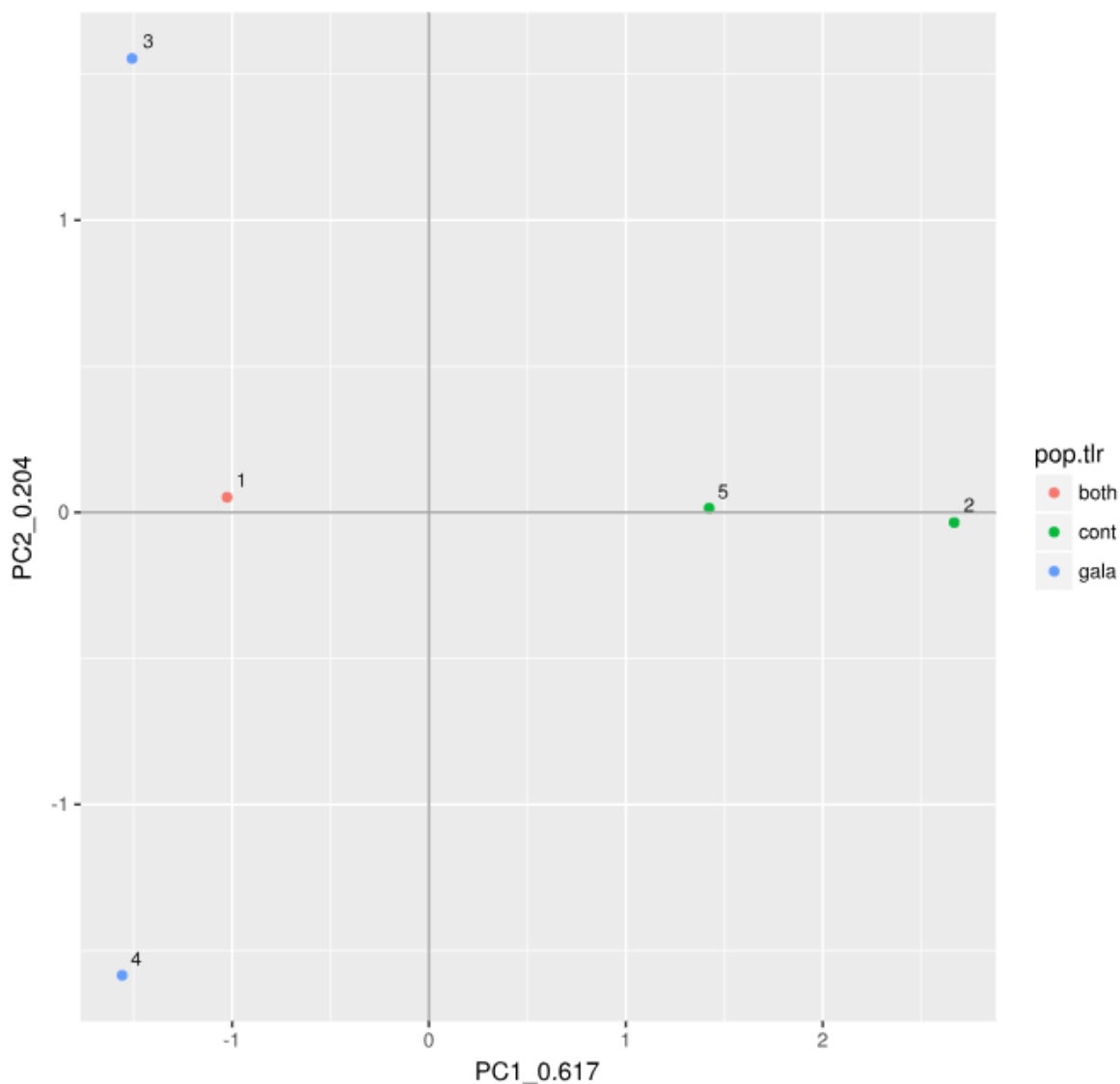
3.2 Analýza hlavních komponent (PCA)

Byly sestrojeny grafy obsahující údaje o náboji (formal charge), hydropatii, objemu a účinné energii dělení. Na základě těchto fyzichemických vlastností haplotypů TLR1B nelze pozorovat jasný vzorec podobnosti (Obr. 3). Graf PCA zde vysvětluje více než 82 % variability. Podle osy PC1 (61,7 % vysvětlené variability) jsou si nejpodobnější haplotypy 3 a 4. Z hlediska osy PC2 jsou si nejpodobnější haplotypy 1, 5 a 2. Největší fyzichemickou

vzdálenost najdeme mezi dvojicemi haplotypů 4, 2 a 2, 2. Nejmenší vzdálenost najdeme mezi haplotypy 5 a 2.

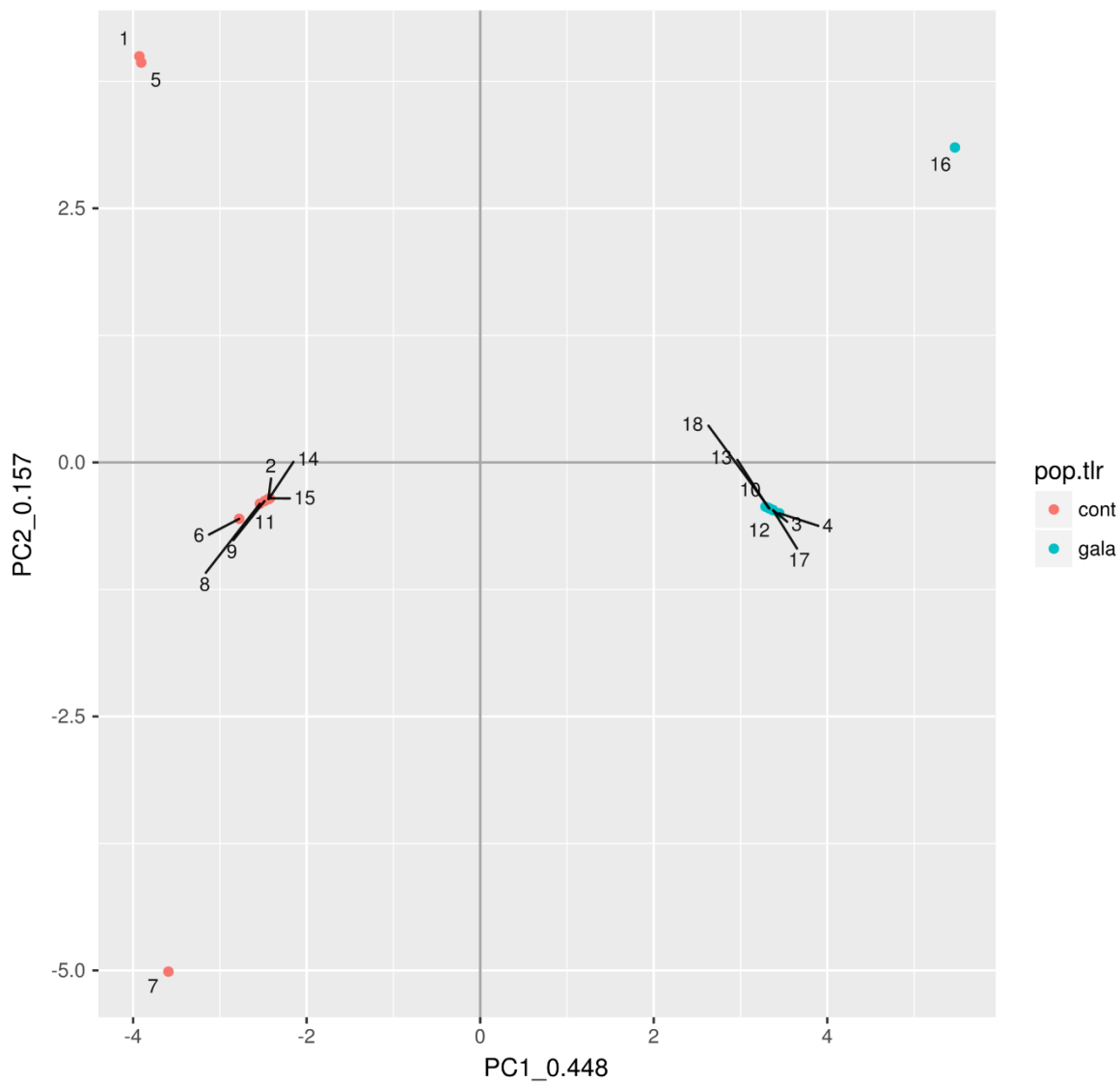
U TLR4 tvoří galapážské (3, 4, 10, 12, 13, 17, 18) a kontinentální (2, 6, 8, 9, 11, 14, 15) haplotypy jasné shluky (Obr. 4), které jsou od sebe odlišné. Z galapážského shluku se nejvíce liší haplotyp 6. Výjimkami jsou haplotypy 1, 5, 7 a 16. Haplotypy 1 a 5 jsou si však velmi podobné. Osy PC1 a PC2 dohromady vysvětlují více než 60 % variability. Hlavní rozdíl mezi shluky haplotypů je z hlediska osy PC1 (44,8 % vysvětlené variability).

Haplotypy TLR15 (Obr. 5) tvoří shluk všech galapážských (1, 2, 5, 6, 8, 10, 11, 13, 15, 18) a tří kontinentálních (4, 16 a 20). Podobné jsou si také dvě dvojice (7, 12; 9, 21) a jedna trojice (3, 17, 14) kontinentálních haplotypů. Haplotyp 19 je odlišný od všech ostatních, přičemž nejblíže má k haplotypu 17. Graf vysvětluje 54,8% variability a větší proměnlivost je z hlediska osy PC1 (34,9% vysv. variability). Galapážské haplotypy jsou si tedy více podobnější než jejich kontinentální protějšky.



Obr. 3: Všechny haplotypy TLR1B ve dvourozměrném fyziokemickém prostoru.

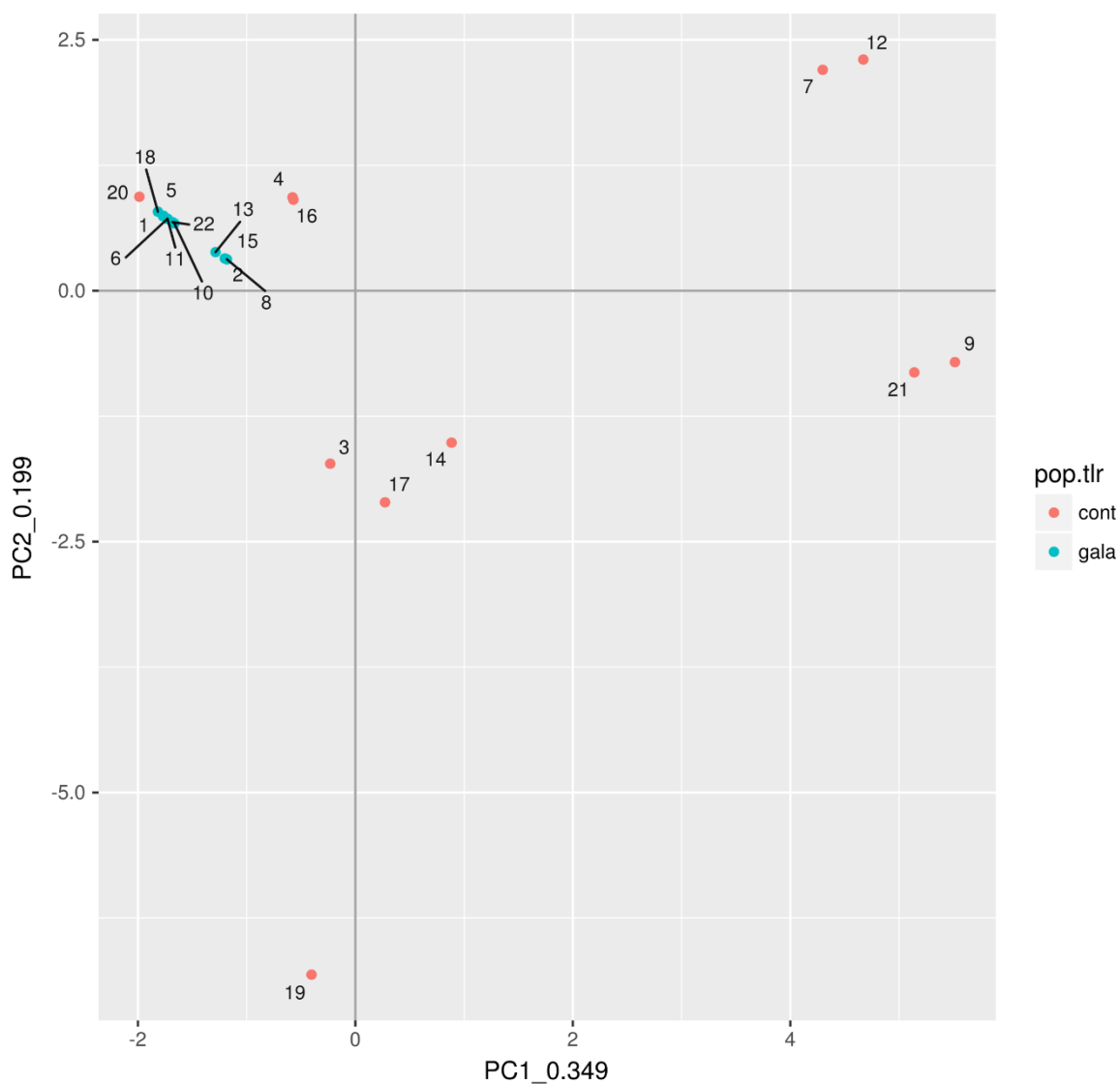
Haplotypy mají barvu podle toho, kde se nachází – galapágy (modrá), kontinent (zelená), obojí (oranžová). Číslo u teček vyjadřují příslušný haplotyp.



Obr. 4: Všechny haplotypy TLR4 ve dvourozměrném fyziologickém prostoru.

Haplotypy mají barvu podle toho, kde se nachází – galapágy (modrá), kontinent (oranžová).

Číslo u teček vyjadřují příslušný haplotyp.



Obr. 5: Všechny haplotypy TLR15 ve dvourozměrném fyzi chemickém prostoru. Haplotypy mají barvu podle toho, kde se nachází – galapágy (modrá), kontinent (oranžová). Čísla u teček vyjadřují příslušný haplotyp.

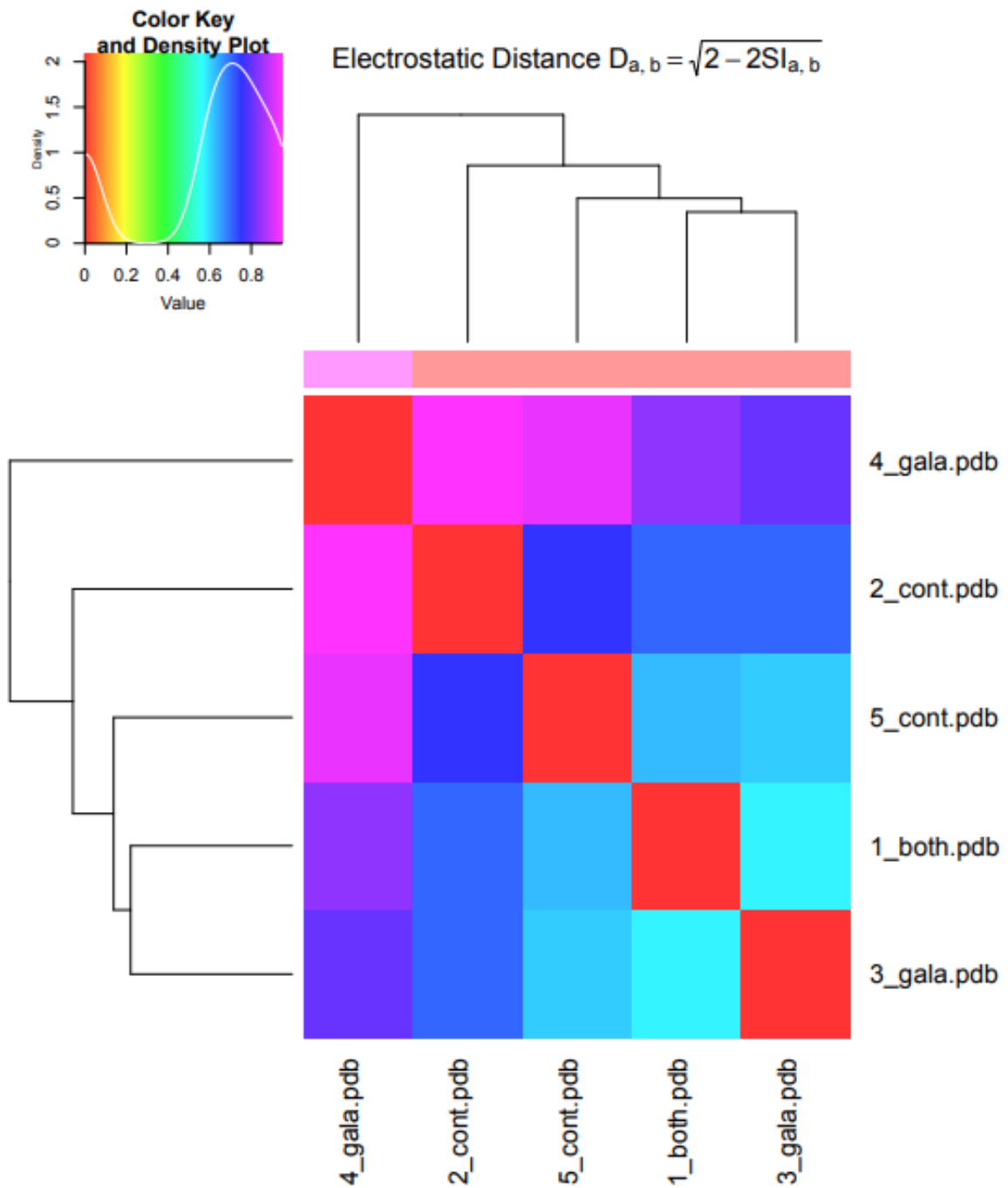
3.3 Analýza povrchového náboje (PIPSA)

U TLR1B dosahuje elektrostatická vzdálenost mezi dvěma haplotypy (4 a 2) hodnoty 0,94763 (Obr. 6), to znamená, že je lze označit za nepodobné. Odlišnost těchto dvou haplotypů koresponduje s grafem PCA, kde jsou si tyto haplotypy velmi vzdáleny (Obr. 3), tyto haplotypy jsou také velmi odlišné od všech ostatních (Obr. 7). Nejpodobnějšími jsou haplotypy 1 a 3 (el.vzdálenost 0,57793), jim nejpodobnějším haplotypem podle epogramu je haplotyp 5, v grafu PCA jsou tyto však daleko od sebe. Na Obr. 8 si lze všimnout velké variability jak struktury haplotypů tak jejich povrchového náboje. Nicméně všechny haplotypy

TLR1B mají na konvexní straně C-konce (blíže TIR) sekvenované oblasti záporně nabitou část.

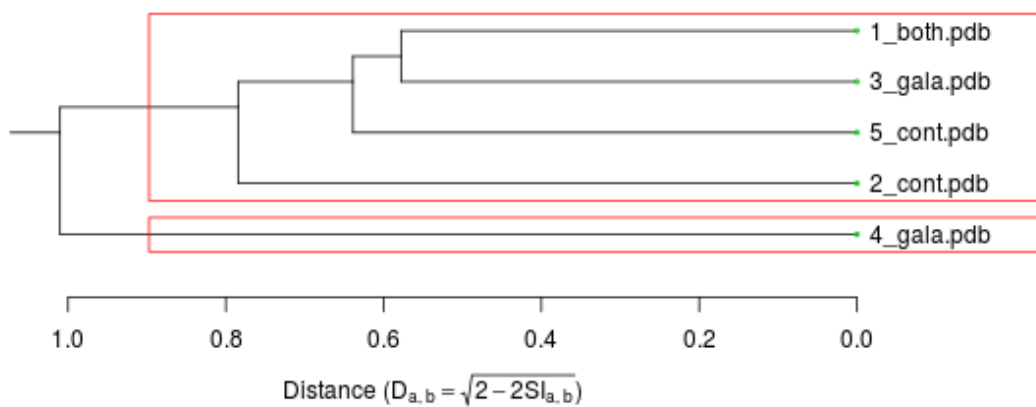
Nejmenší elektrostatická vzdálenost (0,50398) dvou haplotypů TLR4 je u 8 a 9 (Obr. 9). Podobné jsou si i z hlediska PCA, kde patří do stejného shluku (Obr. 4). Největší elektrostatickou vzdálenost (0,77201) mezi sebou mají haplotypy 1 a 15, které i podle PCA dělí velká elektrochemická vzdálenost. Epogram (Obr. 10) také ukazuje odlišnost haplotypů 10 a 13 od všech ostatních. Velmi podobné jsou si také haplotypy 2 a 6, jejichž podobnost také ukazuje PCA. Podle Obr. 11 má povrchový náboj gradient od středu sekvenované části (záporný) k oběma koncům (kladný). Střední část sekvenovaného segmentu je mezi haplotypy méně variabilní než konce.

V rámci TLR15 jsou si z hlediska povrchového náboje nejpodobnější haplotypy 5 a 11 (elektrostatická vzdálenost 0,45826) (Obr. 12). Jejich podobnost ukazuje i graf PCA, kde jsou v absolutní blízkosti (Obr. 5). Velmi podobné jsou si také haplotypy 1 a 12. Největší vzdálenost (0,73212) mezi sebou mají haplotypy 3 a 16. Podle PCA haplotypy nejsou moc vzdálené, ale přesto mají blíže k jiným haplotypům. Přestože epogram ukazuje dvě odlišné skupiny haplotypů, haplotypy 5 a 1, které jsou z jiných skupin, jsou si podobnější více než s jinými haplotypy uvnitř svých skupin (Obr. 13). Povrchový náboj sekvenovaných částí haplotypů TLR15 (Obr. 14) je z většiny záporný, avšak na konvexní straně C-konce má povrchový náboj u všech haplotypů hodnotu ≥ 0 . U téměř všech haplotypů je povrchový náboj N-konce sekvenované části také kladný, ovšem plocha je výrazně menší než u C-konce. Náboj na povrchu štěpného místa (štěpné místo - Obr. 18) je na většině haplotypů záporný, největší odchylku najdeme u haplotypů 18 a 21, které poblíž místa štěpení mají ostrůvky kladného náboje.

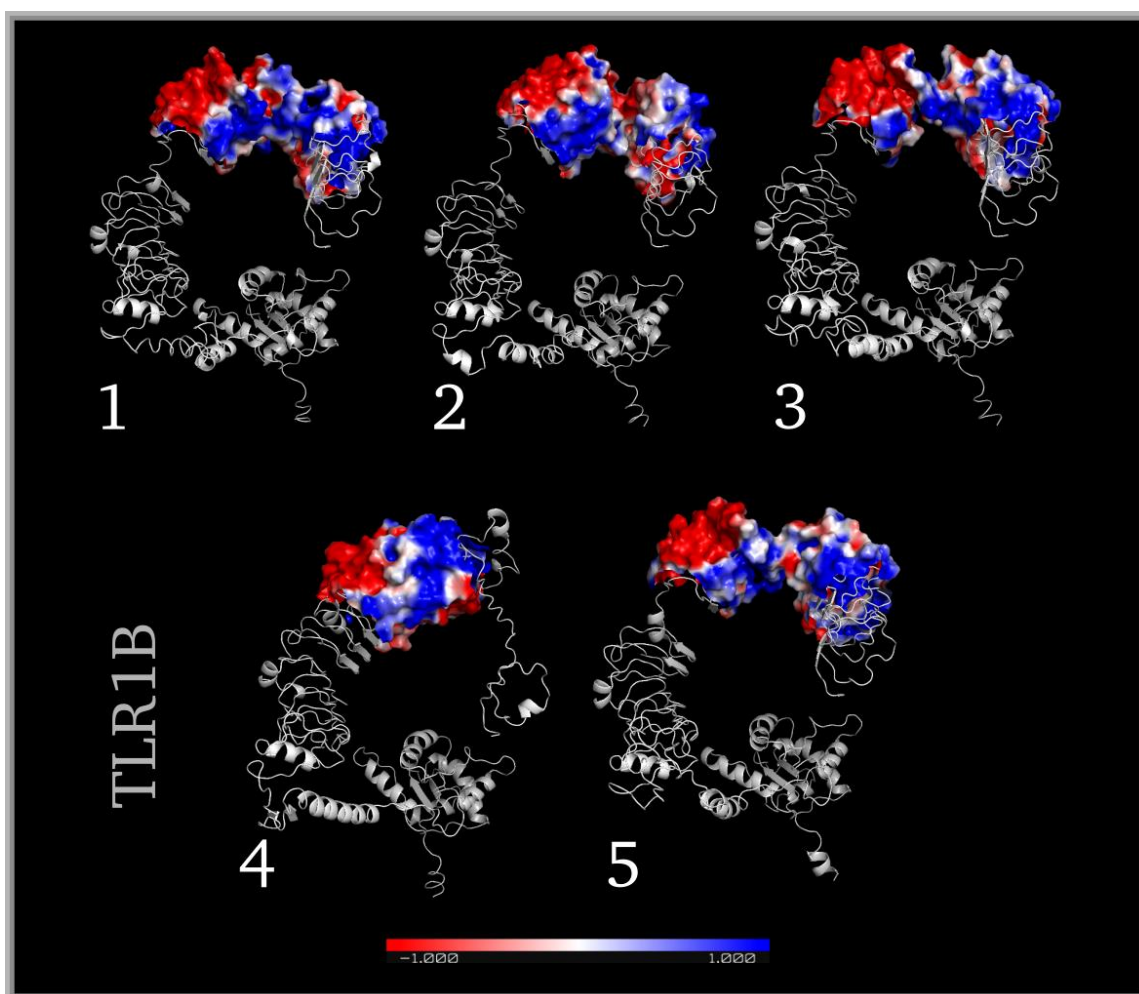


Obr. 6: Tepelná mapa podobnosti povrchového náboje haplotypů TLR1B podle programu PIPSA.

Míra podobnosti je vyjádřena barvou, kdy červená znamená identické haplotypy a směrem k růžové míra podobnosti klesá.

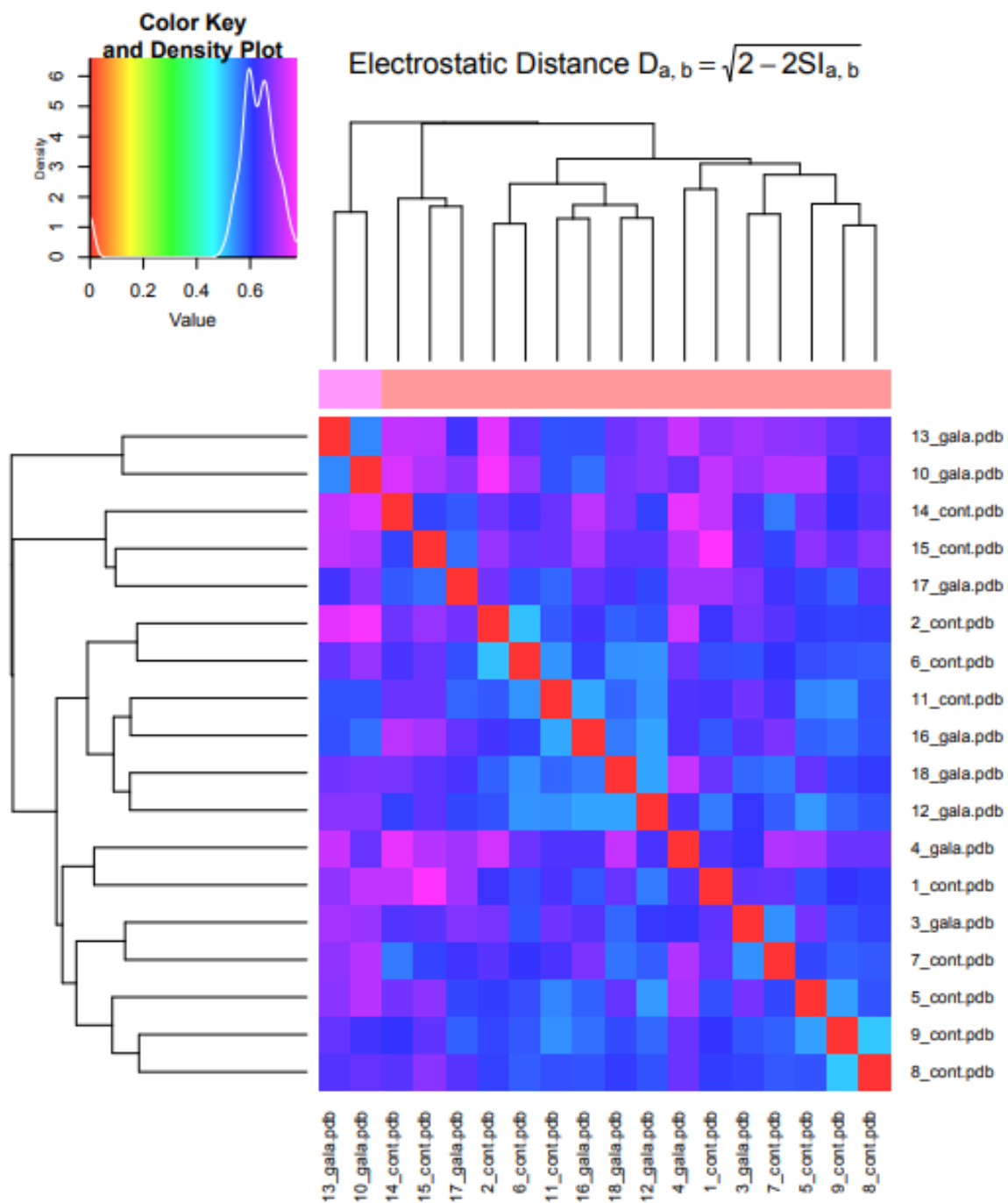


Obr. 7: Epogram elektrostatické vzdálenosti haplotypů TLR1B podle programu PIPSA.



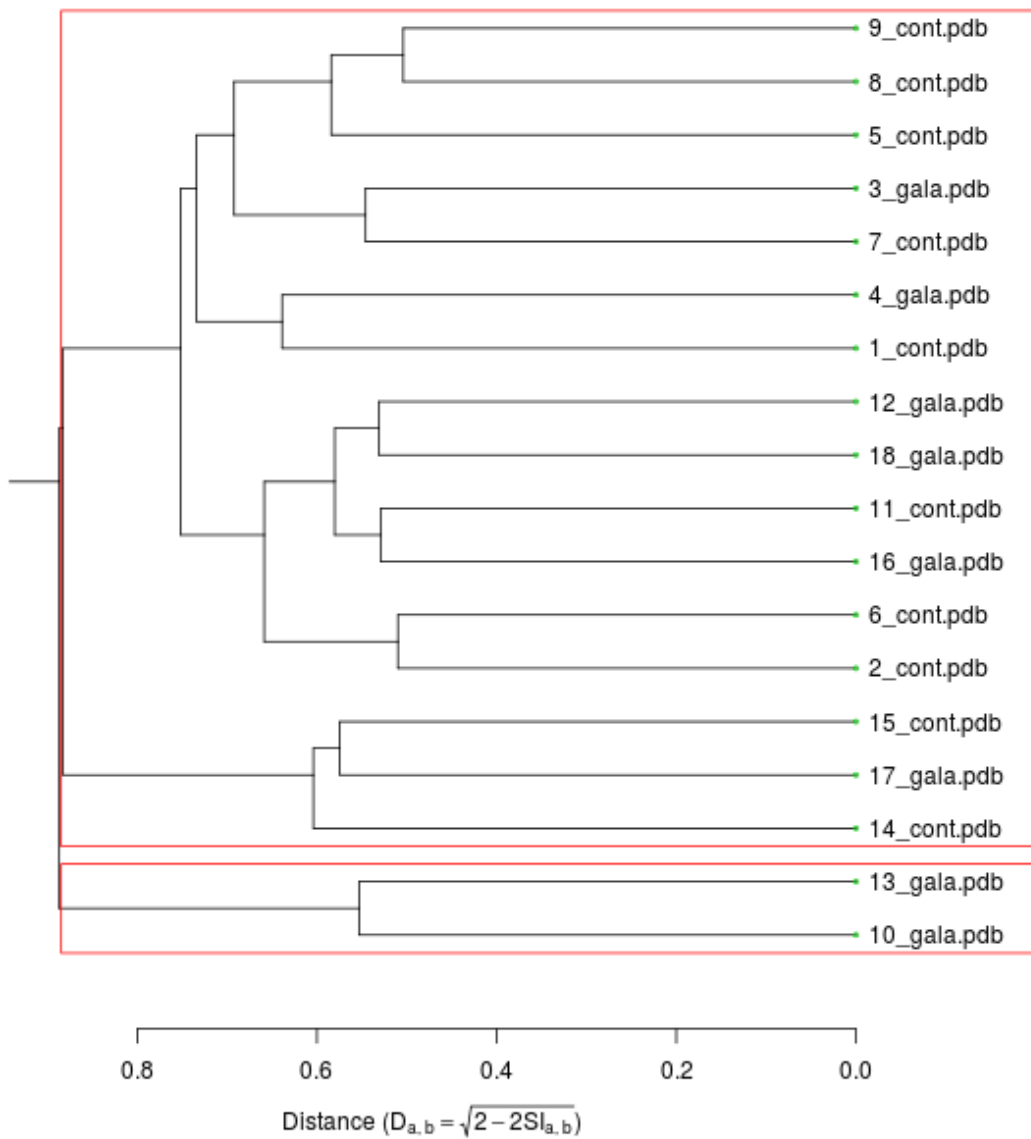
Obr. 8: Vizualizace povrchového náboje sekvenovaného segmentu TLR1B.

Jednotlivé haplotypy TLR1B a jejich povrchový náboj znázorněný barevným gradientem od negativního (červená) po pozitivní (modrá).

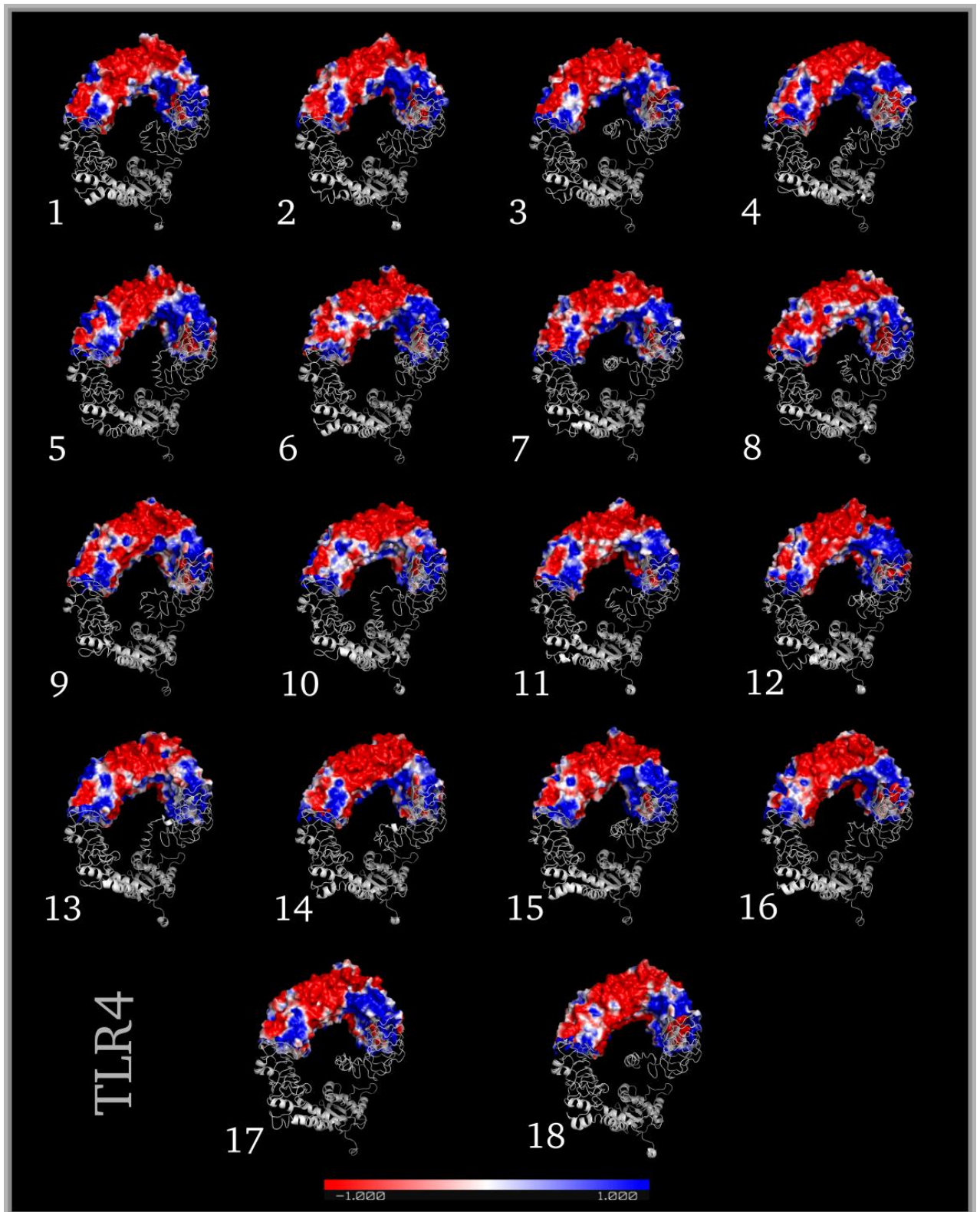


Obr. 9: Tepelná mapa podobnosti povrchového náboje haplotypů TLR4 podle programu PIPSA.

Míra podobnosti je vyjádřena barvou, kdy červená znamená identické haplotypy a směrem k růžové míra podobnosti klesá.

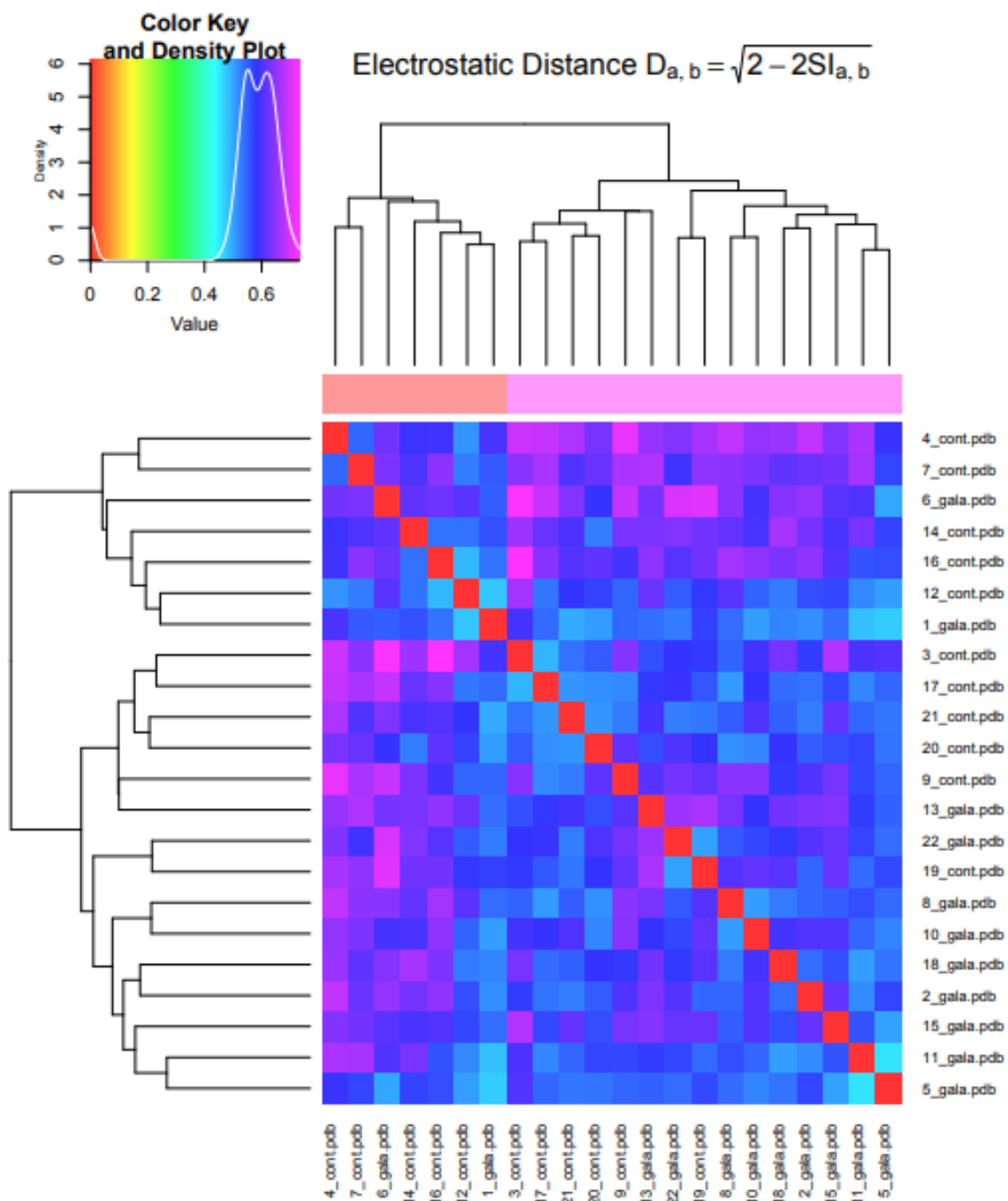


Obr. 10: Epogram elektrostatické vzdálenosti haplotypů TLR4 podle programu PIPSA.



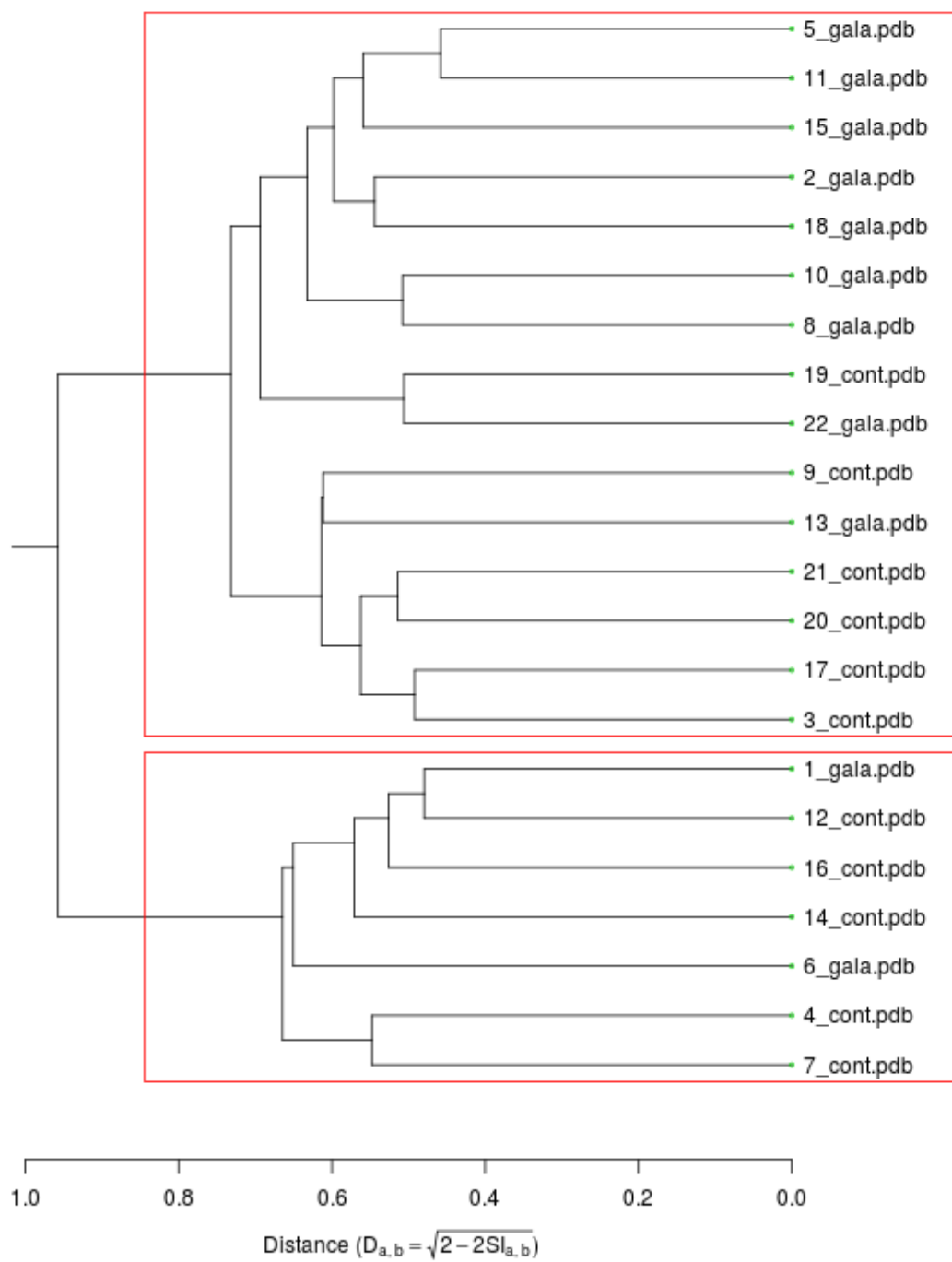
Obr. 11: Vizualizace povrchového náboje sekvenovaného segmentu TLR4.

Jednotlivé haplotypy TLR4 a jejich povrchový náboj znázorněný barevným gradientem od negativního (červená) po pozitivní (modrá).

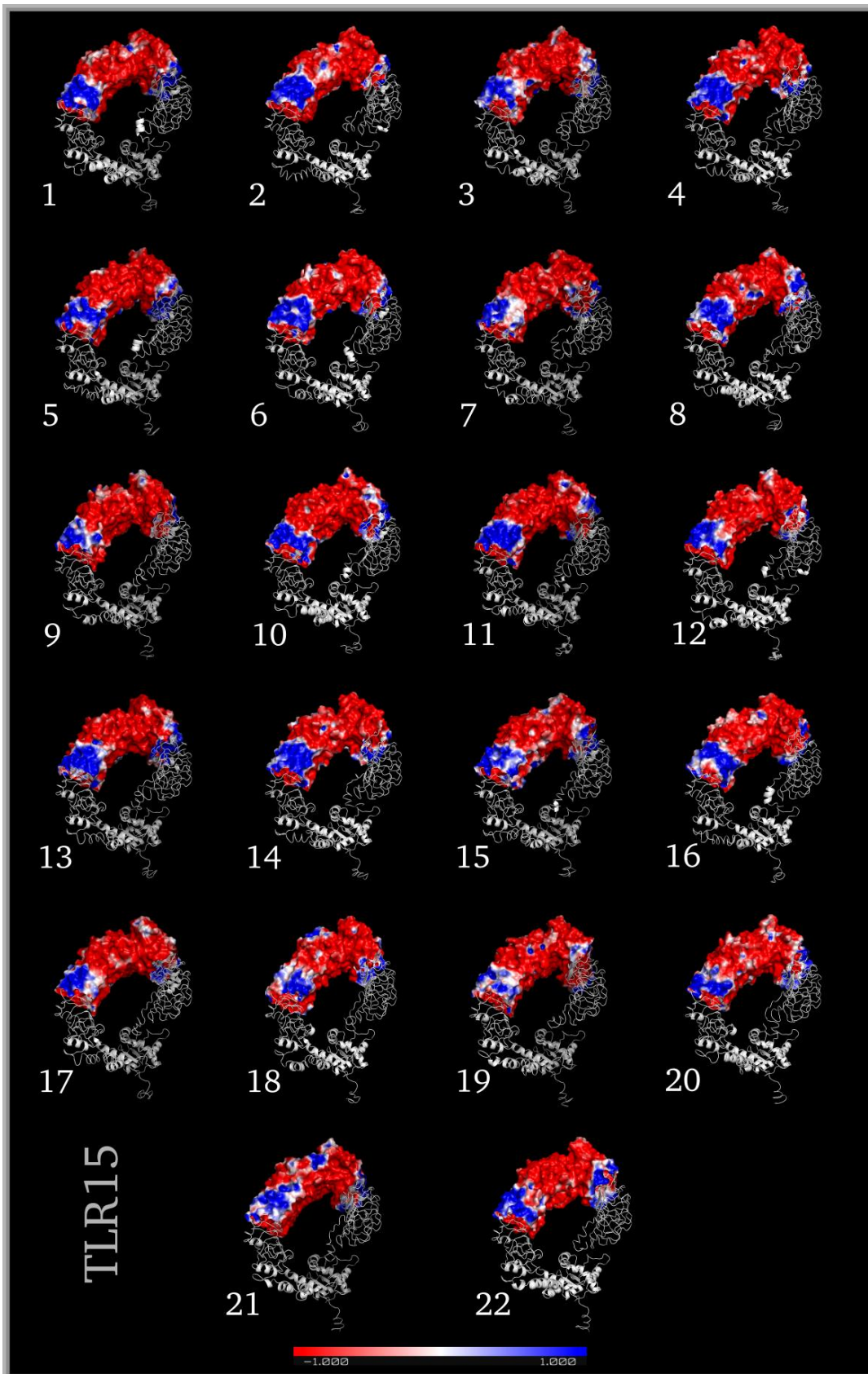


Obr. 12: Tepelná mapa podobnosti povrchového náboje haplotypů TLR15 podle programu PIPSA.

Míra podobnosti je vyjádřena barvou, kdy červená znamená identické haplotypy a směrem k růžové míra podobnosti klesá.



Obr. 13: Epogram elektrostatické vzdálenosti haplotypů TLR15 podle programu PIPSA.



Obr. 14: Vizualizace povrchového náboje sekvenovaného segmentu TLR15.

Jednotlivé haplotypy TLR15 a jejich povrchový náboj znázorněny barevným gradientem od negativního (červená) po pozitivní (modrá).

3.4 Strukturní alignment (super)

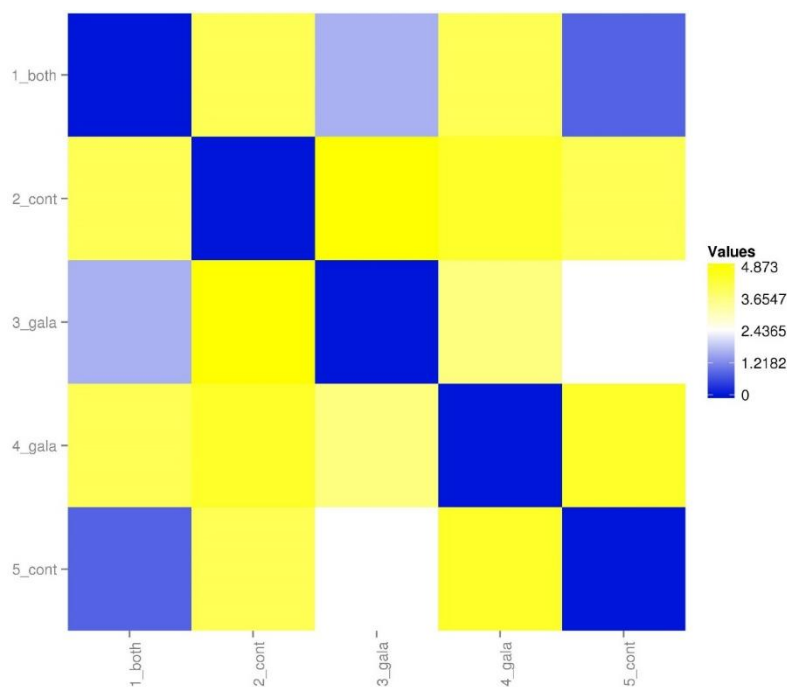
Tepelné mapy zobrazující podobnost haplotypů v rámci každého TLR ukazují spíše homogenní strukturu haplotypů, ve které lze najít několik výjimek.

V rámci TLR1B (Obr. 15) jsou si nejvíce podobné haplotypy 1 a 5 (vzdálenost 0,961), které sdílí podobnost v náboji (Obr. 6), ale podle PCA (Obr. 3) jsou poměrně vzdálené. Nejméně podobné jsou si haplotypy 2 a 3 (vzdálenost 4,873), to potvrzuje i PCA (Obr. 3). Podle mapy jsou od všech ostatních nejvíce odlišné haplotypy 4 a 2 stejně jako u porovnání povrchového náboje (Obr. 6).

Tepelná mapa TLR4 (Obr. 16) ukazuje největší podobnost haplotypů 1 a 16 (vzdálenost 0,319), to potvrzuje i podobný povrchový náboj (Obr. 9), nicméně v grafu PCA jsou tyto haplotypy velmi vzdálené. Nejmenší podobnost je u haplotypů 14 a 17 (vzdálenost 4,1185), přestože povrchovým nábojem jsou si spíše podobné. V grafu PCA jsou tyto haplotypy v jiných shlucích (Obr. 4). Zajímavá je podobnost haplotypů 9 a 14 (0,854), které jsou si s téměř všemi ostatními velmi nepodobné. Podle povrchového náboje jsou si spíše nepodobné (Obr. 9) a v grafu PCA patří do stejného shluku (Obr. 4).

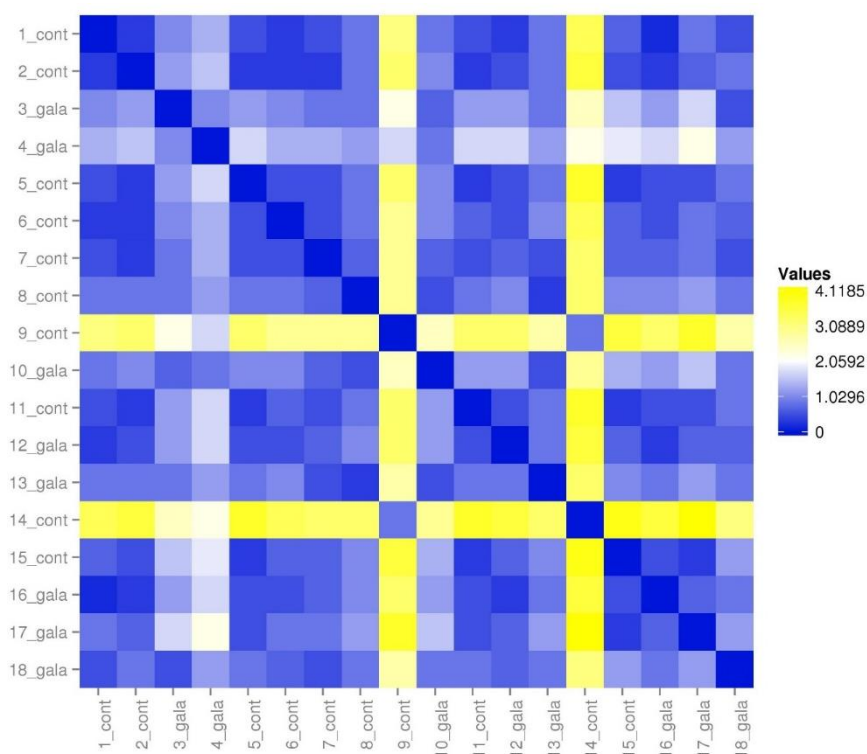
Haplotypy s nejmenší vzdáleností (0,391) u TLR15 jsou 4 a 14 (Obr. 17). Jejich povrchový náboj je spíše nepodobný (Obr. 12) a v grafu PCA mají blíže k jiným haplotypům (Obr. 5). Nejvíce odlišné (vzdálenost 4,5661) jsou haplotypy 15 a 18, které jsou v grafu PCA ve stejném shluku a jejichž náboj je také spíše odlišný. Haplotyp 18 je odlišný od všech ostatních s výjimkou haplotypu 21 (vzdálenost 1,078), který se také liší od téměř všech ostatních. Ačkoliv haplotyp 21 podle PCA dělí od ostatních velká vzdálenost (s výjimkou haplotypu 9), haplotyp 18 je ve shluku s velkým počtem haplotypů.

Vizualizace alignmentu (Obr. 19) přinesla zajímavý pohled na variabilní místa ve strukturách především u TLR15, kde štěpné místo (Obr.18) spadá do variabilního úseku uprostřed sekvenovaného segmentu. U TLR4 jsou variabilní místa 2, jedno uprostřed sekvenované části, kde začíná ligand vázající úsek proteinu, a druhé na jejím C-konci. Přestože prvně zmíněné místo ve středu sekvenované části vykazuje strukturní variabilitu, jeho povrchový náboj (Obr. 11) je u většiny haplotypů záporný. Druhé místo na C-konci sekvenované části je z hlediska náboje velmi variabilní. Haplotypy TLR1B vykazují velkou variabilitu sekundární struktury a např. haplotyp 4 (žlutá barva) má zcela odlišný tvar od ostatních.



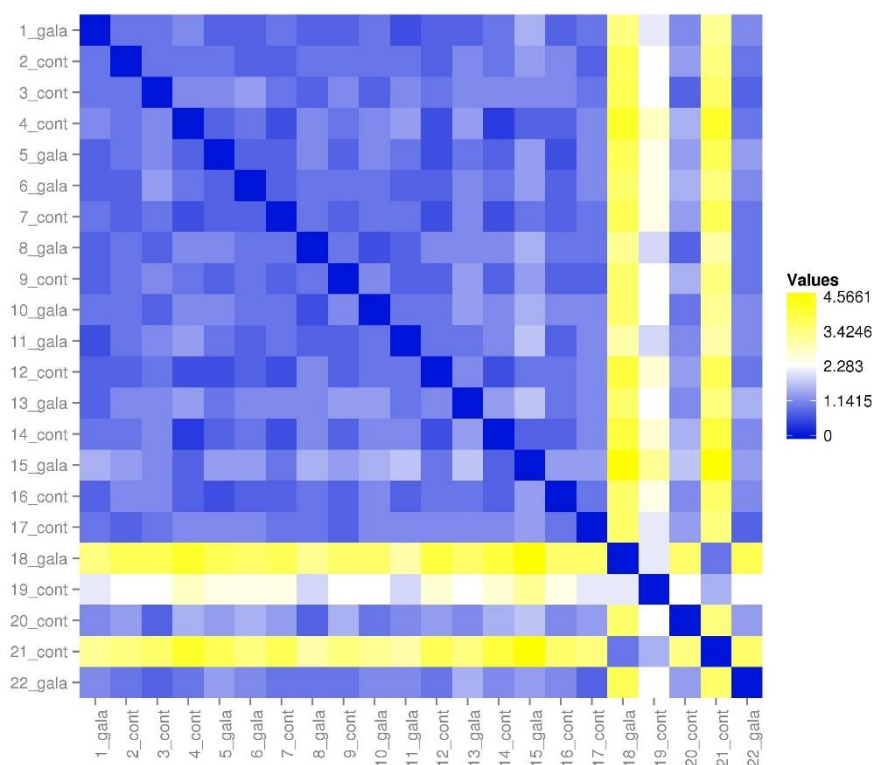
Obr. 15: Tepelná mapa podobnosti struktury haplotypů TLR1B.

Míru podobnosti (vzdálenost) značí příslušná barva, pro identické struktury – hodnota 0 (tmavě modrá), pro nejvíce odlišné struktury – hodnota 4,873 (sytě žlutá).



Obr. 16: Tepelná mapa podobnosti struktury haplotypů TLR4.

Míru podobnosti (vzdálenost) značí příslušná barva, pro identické struktury – hodnota 0 (tmavě modrá), pro nejvíce odlišné struktury – hodnota 4,1185 (sytě žlutá).



Obr. 17: Tepelná mapa podobnosti struktury haplotypů TLR15.

Míru podobnosti (vzdálenost) značí příslušná barva, pro identické struktury – hodnota 0 (tmavě modrá), pro nejvíce odlišné struktury – hodnota 4,5661 (sytě žlutá).

4 Diskuze

Porovnání sekvencí DNA a sekvencí aminokyselin přineslo zajímavé výsledky, kdy se počet haplotypů (Tab. 6) snížil téměř o polovinu. Velké rozdíly se také objevily při srovnání DNA a proteinové heterozygotnosti, kde největší rozdíly dosáhly více než 0,39 (Tab. 7). Pro objektivnější výsledky je tedy vhodné pro analýzu gen. informace přeložit sekvence DNA do proteinových. Další optimalizaci může nabídnout analyzování fyziokemických vlastností sekvence aminokyselin, v čemž spočívá přínos této práce. Nicméně i sekvence DNA jsou důležité pro rozlišení různých původů haplotypů, které proteinová sekvence uhlazuje, jako je tomu například u haplotypu 1 TLR1B.

Při porovnání galapážských populací (*Mimus* sp.) a kontinentální populace *M. polyglottos* byla objevena větší variabilita v rámci kontinentální populace (Tab. 7) z hlediska nukleotidové diversity, počtu haplotypů (kromě TLR1B, kde byl počet stejný). Největší rozdíl však byl objeven v heterozygotnosti (Tab. 7). Rozdíly mezi počtem heterozygotů v galapážské a kontinentální populaci byly různé podle lokusu – TLR1B ~12x, TLR4 ~58x a TLR15 ~5x větší počet heterozygotů na kontinentu. U populace *M. polyglottos* očekáváme větší efektivní velikost populace a širší spektrum patogenů (Wikelski et al. 2004; Hellgren et al. 2013) než

na Galapágách. Pro kontinentální populaci může být výhodnější heterozygotnost v daném lokusu (Charlesworth & Charlesworth 1990), jelikož i variabilita patogenů je větší, nicméně pro testování této hypotézy by bylo zapotřebí doplňujících údajů o fitness heterozygotních jedinců. Na Galapágách je však různorodost patogenů omezenější (Wikelski et al. 2004; Hellgren et al. 2013), klesá zde efektivita selekce a roste vliv driftu (Ohta 1992), vyskytuje se zde tedy více homozygotů. Podobný trend byl popsán u *A. pallidiceps*, kdy malá poslední populace tohoto druhu vykazovala výrazně menší genetickou variabilitu než porovnávané velké populace častějších druhů (Hartmann et al. 2014). Tento trend však může být podpořen i různou mírou inbreedingu na jednotlivých ostrovech (Spielman et al. 2004). Prokázání vlivu inbreedingu, potažmo homozygotity na imunokompetenci u galapážských drozdců, což je relevantní pro ochránářskou genetiku, se nepodařilo v práci Hoeck a Kellera (Hoeck & Keller 2012), kde bylo jako vysvětlení navržena teorie výhody specifické alely. Výhoda konkrétní alely byla popsána například u lejsčků (Grueber et al. 2013), kdy frekvence jedné alely TLR4 byla signifikantně vyšší než očekávaná frekvence podle Hardy-Weinbergovy rovnováhy. To podporuje i fakt, že u všech tří TLR existuje jeden hlavní galapážský haplotyp, který je dominantní a vyskytuje se na většině ostrovů (TLR1B hap. 1, Tab. 3; TLR4 hap. 10, Tab. 4; TLR15 hap. 10, Tab. 5). U TLR1B je tento haplotyp dokonce zastoupen i v kontinentální populaci, byť DNA sekvence jsou rozdílné. TSP u TLR1B naznačuje možné působení stabilizující selekce v udržování konkrétních haplotypů (alel) při boji s patogeny (Spielman et al. 2004; Sepil et al. 2013; Hartmann et al. 2014). Na druhou stranu, jelikož se haplotyp 1 TLR1B vyskytuje v kontinentální populaci pouze jednou – to může být i kvůli malému vzorku kontinentální populace – může se jednat o konvergenci (Jones 2010).

4.1 Nízká variabilita TLR v galapážských populacích

Ačkoliv u TLR1B není zřejmá podobnost galapážských haplotypů (pravděpodobně kvůli dominanci haplotypu 1), u TLR4 a TLR15 lze vyzorovat nízkou variabilitu v galapážských haplotypech. To je poněkud zvláštní, jelikož je většina populací izolovaných a migrace mezi nimi je nízká (Hoeck et al. 2010; Štefka et al. 2011). Vzhledem k vyšší variabilitě haplotypů v kontinentální populaci, ukazují takovéto výsledky na vliv velikosti populace při formování molekulární diversity TLR1B, TLR4 a TLR15 v galapážských populacích. Ačkoliv u TLR4 je variabilita obou populací srovnatelná, výrazná změna je v heterozygotnosti populací (Tab. 7). Nižší molekulární variabilita TLR v galapážských populacích může mít negativní vliv na detekci širšího spektra patogenů, podobné výsledky ukazují i studie na jiných malých ostrovních populacích (Grueber et al. 2013; Hartmann et al.

2014; Gonzalez-Quevedo et al. 2015; Grueber et al. 2015; Gilroy et al. 2017) nebo přímo na galapážských drozdcích, ale pomocí neutrálního markeru (Hoeck et al. 2010). Z toho tedy vyplývá, že galapážské populace můžou být díky své nízké variabilitě náchylnější na nové patogeny než diversifikovaná kontinentální populace (Wikelski et al. 2004).

4.2 Odlišnosti TLR15

Zatímco TLR1B i TLR4 mají jak v galapážské, tak v kontinentální populaci zastoupeny jeden haplotyp, který má extrémní frekvenci, TLR15 má v galapážské populaci jeden hlavní haplotyp (Tab. 5) s frekvencí 170 (hap. 10) a tři velmi početné (haplotypy 8, 11 a 22). Ze všech tří TLR má TLR15 nejvíce proteinových haplotypů. Také se liší svou heterozygotností, kdy na galapágách je 19,9 % heterozygotů a na kontinentě 90,1 % heterozygotů, to je velký rozdíl oproti TLR1B a TLR4. Tyto výsledky se shodují s trendem jak počtu haplotypů, tak s vysokou heterozygotností u TLR15 rákosníka seychelského *A. sechellensis* (Gilroy et al. 2017). Studie zabývající se lejsčiky (Grueber et al. 2013) naznačuje negativní korelaci (nesignifikantní) heterozygotnosti TLR1B a TLR4 s fitness, u TLR15 je tento trend pozitivní (nesignifikantní). Větší procento heterozygotnosti u TLR15 jak v galapážské, tak v kontinentální populaci by mohlo znamenat výhodu heterozygota (Charlesworth & Charlesworth 1990), nicméně pro ověření této teorie nejsou data o fitness heterozygotů pro TLR15.

4.3 Mezidruhový polymorfismus (TSP) v galapážských populacích

V předchozí studii věnující se galapážským drozdcům byl objeven TSP v MHC genech napříč galapážskými populacemi (Vlček et al. 2016). Podařilo se potvrdit, že TSP není záležitost výhradně MHC, ale obecnějším trendem (Klein et al. 2007; Těšický & Vinkler 2015), kdy byl TSP pozorován přes všechny zkoumané lokusy (Tab. 3-5) napříč většinou galapážských populací a všemi druhy drozdců. U TLR1B byl dokonce objeven shodný proteinový haplotyp i v populaci *M. polyglottos*, ačkoliv není vyloučená konvergence těchto haplotypů (Jones 2010), jelikož DNA sekvencemi se liší.

4.4 Analyzování struktury TLR

Analýza struktury byla založena na RMSD. To je mnohými studiemi považováno za neoptimální pro náchylnost k lokálním odlišnostem v alignmentu dvou proteinů (Zhang & Skolnick 2005; Kufareva & Abagyan 2012). Nicméně chyby jsou velké především při alignmentu různě velkých struktur (Carugo & Pongor 2001). Jednou z cest, která by vedla k optimálnějších výsledkům je TM-align, založený na TM-skóre, jež není tak náchylné

k lokálním chybám jako RMSD (Zhang & Skolnick 2005). Důvodem pro výběr funkce „super“ pro 3D alignment a výpočet RMSD byla možnost porovnat libovolné množství struktur k jedné. U TM-align lze porovnat pouze dvě struktury navzájem. Právě funkce „super“ v programu PyMOL (The PyMOL Molecular Graphics System, Version 2.2.3 Schrödinger, LLC) byla zvolena, jelikož je více robustní pro méně podobné struktury než jiné funkce, například „align“, „Cealign“ (<https://pymolwiki.org/index.php/Super>).

Strukturní 3D alignment přinesl zajímavé zjištění u TLR1B, jehož haplotyp 4 se vymyká svou strukturou (Obr. 19), ve které se opakuje pravidelná densní šroubovice odlišná od natažené struktury ostatních haplotypů TLR1B. Jediná pozorovaná změna byla substituce argininu 128 za histidin (Tab. 3). Taková substituce je v různých proteinech považována za konzervativní (Zhang 2000; Sanghavi et al. 2004; Jann et al. 2008). Nicméně v jedné studii byl prokázán zásadní efekt substituce těchto dvou aminokyselin v proteinu, který se podílí na vnitrobuněčné signalizaci u TLR 3, 7 a 9 (Brinkmann et al. 2007). Tento protein nebyl v případě dané substituce schopen podílet se na signalizaci. Avšak převaha studií ukazujících na konzervativitu této substituce, a to i u TLR1B, ale v jiném místě (Nelson-Flower et al. 2018), naznačuje, že příčina tak velké změny ve struktuře může být v odlišném postupu modelovacího programu (Yang et al. 2015), jelikož i na Obr. 8. vykazují identická místa – lejsší konce TLR – různou strukturu. Nicméně modelování probíhalo podle stejného templátu jako u ostatních haplotypů. Frekvence tohoto haplotypu je pouze 1 (Tab. 3), to může buďto znamenat, že je škodlivou alelou, která ještě nebyla odstraněna z populace, nebo že haplotyp nepředstavuje lepší řešení detekce patogenu než například haplotyp 1 s frekvencí 391. Haplotypu 1 je nejvíce podobný haplotyp 5, který je dominantní v kontinentální populaci (Tab. 3) a liší se pouze jednou aminokyselinou (T/A105, Tab. 4) v místě, kde se jejich struktury liší (Obr. 19 – hap. 1 zelená, hap. 5 hnědá). Rozdíl těchto aminokyselin je ve velikosti a polaritě (Bainová 2013). Argument pro vysvětlení dominance těchto dvou haplotypů by mohla být výhoda specifických alel (Spielman et al. 2004; Sepil et al. 2013; Hartmann et al. 2014).

U TLR4 se objevuje zajímavě strukturně variabilní místo blízko C-konci sekvenovaného segmentu (Obr. 19). Nicméně aminokyselinová variabilita v této oblasti pozorována nebyla a nejbližší segregující aminokyselina (R/H220, Tab. 4) je vzdálená několik otoček. Jelikož se strukturně variabilní místo nachází na konci sekvenovaného úseku, je možné, že se aminokyselinová variabilita nachází mimo sekvenovanou část.

Haplotypy TLR15 disponují ve své struktuře velkou variabilní částí přesně v místě štěpení (Obr. 19) (de Zoete et al. 2011). Tato část by mohla být spojena s vysokou variabilitou haplotypů napříč lokusem TLR15. V tomto místě leží segregující aminokyseliny (G/V164;

P/R172; E/K/D222; L/F225; P/S242; Q/K242, Tab 5), z nichž 3 byly popsány jako místa pod pozitivní selekcí (Wang et al. 2015). Zajímavá je také část oddělující velké variabilní místo a malé, které je směrem k C-konci. Tato část byla popsána jako velmi konzervovaná (Wang et al. 2015) a ve všech haplotypech je beze změny. Její konzervovanost a proximita ke štěpnému místu naznačuje význam této sekvence při štěpení mikrobiálními proteázami.

5 Závěr

Hlavním cílem práce bylo popsat funkční variabilitu Toll-like receptorů galapážských drozdců. Tento a další dílčí cíle byly splněny popisem jak diversity objevené v DNA, tak v sekvencích aminokyselin. Dále komplexním popisem fyzichemických vlastností objevených haplotypů populací galapážských drozdců (*Mimus* sp.) a populaci *M. polyglottos* v Kalifornii. Podařilo se sestavit a popsat alignment 3D struktury proteinových haplotypů. Nakonec byly porovnány vlastnosti haplotypů mezi galapážskou a kontinentální populací.

6 Seznam použité literatury

- Acevedo-Whitehouse K, Cunningham AA. 2006. Is MHC enough for understanding wildlife immunogenetics? *Trends Ecol Evol.* 21:433–438.
- Akashi H, Osada N, Ohta T. 2012. Weak selection and protein evolution. *Genetics.* 192:15–31.
- Akira S. 2009. Review Pathogen recognition by innate immunity and its signaling. *Discovery.* 85:143–156.
- Akira S, Takeda K, Kaisho T. 2001. Toll-like receptors: critical proteins linking innate and acquired immunity. *Nat Rev Immunol.* 2:675–680.
- Alcaide M, Edwards S V. 2011. Molecular evolution of the toll-like receptor multigene family in birds. *Mol Biol Evol.* 28:1703–1715.
- Arbogast BS, Drovetski S V., Curry RL, Boag PT, Seutin G, Grant PR, Grant BR, Anderson DJ. 2006. the Origin and Diversification of Galapagos Mockingbirds. *Evolution (N Y).* 60:370.
- Ardern SL, Lambert DM. 1997. Is the black robin in genetic peril? *Mol Ecol.* 6:21–28.
- Babicki S, Arndt D, Marcu A, Liang Y, Grant JR, Maciejewski A, Wishart DS. 2016. Heatmapper: web-enabled heat mapping for all. *Nucleic Acids Res.* 44:W147–W153.

- Baker D, Sali A. 2001. Protein Structure Prediction and Structural Genomics. *Science*. 294:93–96.
- Barnosky AD, Matzke N, Tomiya S, Wogan GOU, Swartz B, Quental TB, Marshall C, McGuire JL, Lindsey EL, Maguire KC, et al. 2011. Has the Earth's sixth mass extinction already arrived? *Nature*. 471:51–57.
- Bella J, Hindle KL, McEwan PA, Lovell SC. 2008. The leucine-rich repeat structure. *Cell Mol Life Sci*. 65:2307–2333.
- Bollmer JL, Dunn PO, Whittingham LA, Wimpee C. 2010. Extensive MHC class II B gene duplication in a passerine, the common yellowthroat (*Geothlypis trichas*). *J Hered*. 101:448–460.
- Botos I, David M. S, Davies DR. 2011. Structural biology of TLRs. *Structure*. 19:447–459.
- Boyd AC, Peroval MY, Hammond JA, Prickett MD, Young JR, Smith AL. 2012. TLR15 Is Unique to Avian and Reptilian Lineages and Recognizes a Yeast-Derived Agonist. *J Immunol*. 189:4930–4938.
- Brinkmann MM, Spooner E, Hoebe K, Beutler B, Ploegh HL, Kim YM. 2007. The interaction between the ER membrane protein UNC93B and TLR3, 7, and 9 is crucial for TLR signaling. *J Cell Biol*. 177:265–275.
- Brownlie R, Allan B. 2011. Avian toll-like receptors. *Cell Tissue Res*. 343:121–130.
- Burch-Smith TM, Dinesh-Kumar SP. 2007. The functions of plant TIR domains. *Sci STKE*. 2007.
- Carugo O, Pongor S. 2001. A normalized root-mean-square distance for comparing protein three-dimensional structures. *Protein Sci*. 10:1470–1473.
- Charlesworth D, Charlesworth B. 1990. Inbreeding depression with heterozygote advantage and its effect on selection for modifiers changing the outcrossing rate. *Evolution (N Y)*. 44:870–888.
- Ebert D. 2002. Experimental Evolution of Parasites. *Science*. 282:1432–1436.
- Fauchere J, Charton M, Kier L, Verloop A, Pliska V. 1988. Amino acid side chain parameters for correlation studies in biology and pharmacology. *Int J Pept Protein Res*. 32:269–278.
- Forstmeier W, Schielzeth H, Mueller JC, Ellegren H, Kempenaers B. 2012. Heterozygosity-fitness correlations in zebra finches: microsatellite markers can be better than their reputation. *Mol Ecol*. 21:3237–3249.
- Gabdoulline RR, Stein M, Wade RC. 2007. qPIPSA: Relating enzymatic kinetic parameters and interaction fields. *BMC Bioinformatics*. 8:373.
- Gilroy DL, van Oosterhout C, Komdeur J, Richardson DS. 2017. Toll-like receptor variation

in the bottlenecked population of the endangered Seychelles warbler. *Anim Conserv.* 20:235–250.

Gonzalez-Quevedo C, Spurgin LG, Illera JC, Richardson DS. 2015. Drift, not selection, shapes toll-like receptor variation among oceanic island populations. *Mol Ecol.* 24:5852–5863.

Grueber CE, Knafler GJ, King TM, Senior AM, Grosser S, Robertson B, Weston KA, Brekke P, Harris CLW, Jamieson IG. 2015. Toll-like receptor diversity in 10 threatened bird species: relationship with microsatellite heterozygosity. *Conserv Genet.* 16:595–611.

Grueber CE, Wallis GP, Jamieson IG. 2013. Genetic drift outweighs natural selection at toll-like receptor (TLR) immunity loci in a re-introduced population of a threatened species. *Mol Ecol.* 22:4470–4482.

Grueber CE, Wallis GP, King TM, Jamieson IG. 2012. Variation at Innate Immunity Toll-Like Receptor Genes in a Bottlenecked Population of a New Zealand Robin. *PLoS One.* 7.

Hansson GK, Edfeldt K. 2005. Toll To Be Paid at the Gateway to the Vessel Wall. *Arterioscler Thromb Vasc Biol.* 25:1085–1087.

Hartmann SA, Schaefer HM, Segelbacher G. 2014. Genetic depletion at adaptive but not neutral loci in an endangered bird species. *Mol Ecol.* 23:5712–5725.

Hashimoto C, Hudson KL, Anderson K V. 1988. The Toll gene of drosophila, required for dorsal-ventral embryonic polarity, appears to encode a transmembrane protein. *Cell.* 52:269–279.

Hellgren O, Križanauskienė A, Hasselquist D, Bensch S. 2013. Low Haemosporidian Diversity and One Key-Host Species in a Bird Malaria Community on a Mid-Atlantic Island (São Miguel, Azores). *J Wildl Dis.* 47:849–859.

Hoeck PEA, Bollmer JL, Parker PG, Keller LF. 2010. Differentiation with drift: a spatio-temporal genetic analysis of Galapagos mockingbird populations (*Mimus* spp.). *Philos Trans R Soc B Biol Sci.* 365:1127–1138.

Hoeck PEA, Keller LF. 2012. Inbreeding, immune defence and ectoparasite load in different mockingbird populations and species in the Galápagos Islands. *J Avian Biol.* 43:423–434.

Ihaka R, Gentleman R. 1996. R: A Language for Data Analysis and Graphics. *J Comput Graph Stat.* 5:299–314.

Iwasaki A, Medzhitov R. 2010. Regulation of adaptive immunity by the innate immune system. *Science.* 327:291–295.

Jann OC, Werling D, Chang JS, Haig D, Glass EJ. 2008. Molecular evolution of bovine Toll-like receptor 2 suggests substitutions of functional relevance. *BMC Evol Biol.* 8:288.

- Jarvi SI, Tarr CL, Mcintosh CE, Atkinson CT, Fleischer RC. 2004. Natural selection of the major histocompatibility complex (Mhc) in Hawaiian honeycreepers (Drepanidinae). *Mol Ecol.* 13:2157–2168.
- Jilong L, Bhattacharya D, Cao R, Adhikari B, Deng X, Eickholt J, Cheng J. 2013. The MULTICOM Protein Tertiary Structure Prediction System. *Methods Mol Biol.* 1137:29–41.
- Jin MS, Kim SE, Heo JY, Lee ME, Kim HM, Paik S-G, Lee H, Lee J-O. 2007. Crystal Structure of the TLR1-TLR2 Heterodimer Induced by Binding of a Tri-Acylated Lipopeptide. *Cell.* 130:1071–1082.
- Jin MS, Lee JO. 2008. Structures of the Toll-like Receptor Family and Its Ligand Complexes. *Immunity.* 29:182–191.
- Jones G. 2010. Molecular Evolution: Gene Convergence in Echolocating Mammals. *Curr Biol.* 20:R62–R64.
- Kang JY, Nan X, Jin MS, Ryu YH, Youn S-J, Han SH, Mah S, Paik S-G, Lee J-O, Lee H. 2009. Recognition of Lipopeptide Patterns by Toll-like Receptor 2-Toll-like Receptor 6 Heterodimer. *Immunity.* 31:873–884.
- Kawashima S, Pokarowski P, Pokarowska M, Kolinski A, Katayama T, Kanehisa M. 2008. AAindex: Amino acid index database, progress report 2008. *Nucleic Acids Res.* 36:202–205.
- Keestra AM, de Zoete MR, Bouwman LI, Vaezirad MM, Van Putten JPM. 2013. Unique features of chicken Toll-like receptors. *Dev Comp Immunol.* 41:316–323.
- Klein J, Sato A, Nikolaidis N. 2007. MHC, TSP, and the Origin of Species: From Immunogenetics to Evolutionary Genetics. *Annu Rev Genet.* 41:281–304.
- Konopka BM, Marciniak M, Dyrka W. 2017. Quantiprot - a Python package for quantitative analysis of protein sequences. *BMC Bioinformatics.* 18:1–6.
- Kosakovsky Pond SL, Frost SDW. 2005. Not so different after all: A comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol.* 22:1208–1222.
- Kufareva I, Abagyan R. 2012. Methods of protein structure comparison. *Methods Mol Biol.*:231–257.
- Liu L, Botos I, Wang Y, Leonard JN, Shiloach J, David M, Davies DR. 2008. Structural basis of toll-like receptor 3 signaling with double-stranded RNA. *Science.* 320:379–381.
- Medzhitov R, Preston-Hurlburt P, Janeway CJ. 1997. A human homologue of the *Drosophila* Toll protein signals activation of adaptive immunity. *Nature.* 388:394–397.
- Miyazawa S, Jernigan RL. 1985. Estimation of Effective Interresidue Contact Energies from Protein Crystal Structures: Quasi-Chemical Approximation. *Macromolecules.* 18:534–552.
- Moult J, Fidelis K, Kryshchak A, Schwede T, Tramontano A. 2018. Critical assessment

of methods of protein structure prediction (CASP)—Round XII. *Proteins Struct Funct Bioinforma.* 86:7–15.

Nelson-Flower MJ, Germain RR, MacDougall-Shackleton EA, Taylor SS, Arcese P. 2018. Purifying Selection in the Toll-Like Receptors of Song Sparrows *Melospiza melodia*. *J Hered.* 109:501–509.

Nietlisbach P, Wandeler P, Parker PG, Grant PR, Grant BR, Keller LF, Hoeck PEA. 2013. Hybrid ancestry of an island subspecies of Galápagos mockingbird explains discordant gene trees. *Mol Phylogenet Evol.* 69:581–592.

O’Neill LAJ, Bowie AG. 2007. The family of five: TIR-domain-containing adaptors in Toll-like receptor signalling. *Nat Rev Immunol.* 7:353–364.

O’Neill LAJ, Fitzgerald KA, Bowie AG. 2003. The Toll-IL-1 receptor adaptor family grows to five members. *Trends Immunol.* 24:286–289.

Ohta T. 1973. Slightly Deleterious Mutant Substitutions in Evolution. *Nature.* 246:96–97.

Ohta T. 1992. The Nearly Neutral Theory of Molecular Evolution. *Annu Rev Ecol Syst.* 23:263–286.

Park BS, Song DH, Kim HM, Choi B-S, Lee H, Lee J-O. 2009. The structural basis of lipopolysaccharide recognition by the TLR4–MD-2 complex. *Nature.* 458:1191–1195.

Pimm SL, Jenkins CN, Abell R, Brooks TM, Gittleman JL, Joppa LN, Raven PH, Roberts CM, Sexton JO. 2014. The biodiversity of species and their rates of extinction, distribution, and protection. *Science.* 344.

Reed DH, Frankham R. 2003. Correlation between Fitness and Genetic Diversity. *Conserv Biol.* 17:230–237.

Richter S, Wenzel A, Stein M, Gabdoulline RR, Wade RC. 2008. webPIPSA: a web server for the comparison of protein interaction properties. *Nucleic Acids Res.* 36:276–280.

Roach JC, Glusman G, Rowen L, Kaur A, Purcell MK, Smith KD, Hood LE, Aderem A. 2005. The evolution of vertebrate Toll-like receptors. - Big Phylogeny Paper Mainly Mammals. *Proc Natl Acad Sci U S A.* 102:9577–9582.

Roy A, Kucukural A, Zhang Y. 2010. I-TASSER: A unified platform for automated protein structure and function prediction. *Nat Protoc.* 5:725–738.

Rozas J, Ferrer-Mata A, Sanchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sanchez-Gracia A. 2017. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol Biol Evol.* 34:3299–3302.

Sanghavi SK, Shankarappa R, Reinhart TA. 2004. Genetic analysis of Toll/Interleukin-1 Receptor (TIR) domain sequences from rhesus macaque Toll-like receptors (TLRs) 1-10

reveals high homology to human TLR/TIR sequences. *Immunogenetics*. 56:667–674.

Scheet P, Stephens M. 2006. A Fast and Flexible Statistical Model for Large-Scale Population Genotype Data: Applications to Inferring Missing Genotypes and Haplotypic Phase. *Am J Hum Genet*. 78:629–644.

Sepil I, Lachish S, Sheldon BC. 2013. Mhc-linked survival and lifetime reproductive success in a wild population of great tits. *Mol Ecol*. 22:384–396.

Slev PR, Potts WK. 2002. Disease consequences of pathogen adaptation. *Curr Opin Immunol*. 14:609–614.

Sommer S. 2005. The importance of immune gene variability (MHC) in evolutionary ecology and conservation. *Front Zool*. 2:16.

Spielman D, Brook BW, Briscoe DA, Frankham R. 2004. Does inbreeding and loss of genetic diversity decrease disease resistance? *Conserv Genet*. 5:439–448.

Spurgin LG, Richardson DS. 2010. How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. *Proc Biol Sci*. 277:979–988.

Štefka J, Hoeck PEA, Keller LF, Smith VS. 2011. A hitchhikers guide to the Galápagos: co-phylogeography of Galápagos mockingbirds and their parasites. *BMC Evol Biol*. 11:284.

Stork NE. 2010. Re-assessing current extinction rates. *Biodivers Conserv*. 19:357–371.

Svergun DI, Petoukhov M V., Koch MHJ. 2001. Determination of domain structure of proteins from x-ray solution scattering. *Biophys J*. 80:2946–2953.

Těšický M, Vinkler M. 2015. Trans-Species Polymorphism in Immune Genes: General Pattern or MHC-Restricted Phenomenon? *J Immunol Res*. 2015:1–10.

Van Valen L. 1973. A New Evolutionary Law. *Evol Theory*.:1–30.

Väli Ü, Einarsson A, Waits L, Ellegren H. 2008. To what extent do microsatellite markers reflect genome-wide genetic diversity in natural populations? *Mol Ecol*. 17:3808–3817.

Velová H, Gutowska-Ding MW, Burt DW, Vinkler M, Yeager M. 2018. Toll-Like Receptor Evolution in Birds: Gene Duplication, Pseudogenization, and Diversifying Selection. *Mol Biol Evol*. 35:2170–2184.

Vlček J, Hoeck PEA, Keller LF, Wayhart JP, Dolinová I, Štefka J. 2016. Balancing selection and genetic drift create unusual patterns of MHCII β variation in Galápagos mockingbirds. *Mol Ecol*. 25:4757–4772.

Wade RC, Gabdoulline RR, De Rienzo F. 2001. Protein interaction property similarity analysis. *Int J Quantum Chem*. 83:122–127.

Wang J, Zhang Z, Chang F, Yin D. 2015. Bioinformatics analysis of the structural and evolutionary characteristics for toll-like receptor 15. *Nat Methods*. 4:e2079.

- Weaver S, Shank SD, Spielman SJ, Li M, Muse S V, Kosakovsky Pond SL. 2017. Datamonkey 2.0: A Modern Web Application for Characterizing Selective and Other Evolutionary Processes. *Mol Biol Evol.* 35:773–777.
- Werling D, Jann OC, Offord V, Glass EJ, Coffey TJ. 2009. Variation matters: TLR structure and species-specific pathogen recognition. *Trends Immunol.* 30:124–130.
- Wikelski M, Foufopoulos J, Vargas H, Snell H. 2004. Galápagos Birds and Diseases: Invasive Pathogens as Threats for Island Species. *Ecol Soc.* 9.
- Wüthrich K. 2001. The way to NMR structures of proteins. *Nat Struct Biol.* 8:923–925.
- Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. 2015. The I-TASSER Suite: protein structure and function prediction. *Nat Methods.* 12:7–8.
- Zhang J. 2000. Rates of conservative and radical nonsynonymous nucleotide substitutions in mammalian nuclear genes. *J Mol Evol.* 50:56–68.
- Zhang Y. 2008. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics.* 9:40.
- Zhang Y, Skolnick J. 2005. TM-align: A protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* 33:2302–2309.
- de Zoete MR, Bouwman LI, Kestra AM, van Putten JPM. 2011. Cleavage and activation of a Toll-like receptor by microbial proteases. *Proc Natl Acad Sci.* 108:4968–4973.

Knihy a diplomové práce

- Bainová Z. 2013. Polymorfismus heterodimerů TLR2/TLR1 a TLR2/TLR6 u inbredních linií myši domácí odvozených z přirozených populací. Magisterská práce. Univerzita Karlova, Praha.
- Lepš J, Šmilauer P. 2016. Biostatistika. Nakladatelství Jihočeské univerzity v Českých Budějovicích.
- Murphy K, Travers P, Walport M, & Janeway C. 2009. *Janeway's immunobiology* (8th ed.). New York: Garland Science.
- Parham Peter. 2009. *The immune system* (3rd ed.). New York: Garland Science.
- R Core Team. 2016. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.or>
- Těšický M. 2016. Trans-species polymorphism in selected innate immunity genes in tits (Paridea family). Magisterská práce. Univerzita Karlova, Praha.

Internetové zdroje

<https://www.geneious.com> [online]. 2005 [cit. 6. 3. 2019].

<https://zhanglab.ccmb.med.umich.edu/library> [online]. 2015 [cit. 6. 3. 2019].

<http://www2.heatmapper.ca> [online]. 2016 [cit. 31. 3. 2019].

<https://pymolwiki.org/index.php/Super> [online]. 2015 [cit. 31. 3. 2019].

7 Přílohy

7.1 Seznam tabulek

Tab. 1: Přístupové kódy databáze NCBI pro referenční Toll-like receptory lejska bělokrkého.....	9
Tab. 2: Sekvence navržených primerů pro amplikonové sekvenování TLR.	41
Tab. 3: Segregující pozice aminokyselin, frekvence haplotypů a výskyt proteinových haplotypů v populacích pro TLR1B.....	42
Tab. 4: Segregující pozice aminokyselin, frekvence haplotypů a výskyt proteinových haplotypů v populacích pro TLR4.	43
Tab. 5: Segregující pozice aminokyselin, frekvence haplotypů a výskyt proteinových haplotypů v populacích pro TLR15.	44
Tab. 6: Charakteristiky sekvenačních dat všech tří TLR.	44
Tab. 7: Porovnání populačně-genetických statistik mezi galapážskými populacemi a kontinentální populací.....	45

7.2 Seznam obrázků

Obr. 1: Struktura Toll-like receptorů (převzato z Parham 2009).....	3
Obr. 2: Alopatrické populace galapážských drozdců.....	5
Obr. 3: Všechny haplotypy TLR1B ve dvourozměrném fyzi chemickém prostoru.	13
Obr. 4: Všechny haplotypy TLR4 ve dvourozměrném fyzi chemickém prostoru.	14
Obr. 5: Všechny haplotypy TLR15 ve dvourozměrném fyzi chemickém prostoru.	15
Obr. 6: Tepelná mapa podobnosti povrchového náboje haplotypů TLR1B podle programu PIPSA.	17
Obr. 7: Epogram elektrostatické vzdálenosti haplotypů TLR1B podle programu PIPSA.....	18
Obr. 8: Vizualizace povrchového náboje sekvenovaného segmentu TLR1B.....	18
Obr. 9: Tepelná mapa podobnosti povrchového náboje haplotypů TLR4 podle programu PIPSA.....	19
Obr. 10: Epogram elektrostatické vzdálenosti haplotypů TLR4 podle programu PIPSA.	20
Obr. 11: Vizualizace povrchového náboje sekvenovaného segmentu TLR4.	21
Obr. 12: Tepelná mapa podobnosti povrchového náboje haplotypů TLR15 podle programu PIPSA.....	22
Obr. 13: Epogram elektrostatické vzdálenosti haplotypů TLR15 podle programu PIPSA.	23
Obr. 14: Vizualizace povrchového náboje sekvenovaného segmentu TLR15.	24
Obr. 15: Tepelná mapa podobnosti struktury haplotypů TLR1B.	26
Obr. 16: Tepelná mapa podobnosti struktury haplotypů TLR4.	26
Obr. 17: Tepelná mapa podobnosti struktury haplotypů TLR15.	27
Obr. 18: 3D modely Toll-like receptorů modelované serverem I-TASSER (vizualizace PyMOL).....	46
Obr. 19: Vizualizace strukturního alignmentu sekvenovaných segmentů TLR1B, TLR4 a TLR15 funkcí „super“ v programu PyMOL.....	47

7.3 Tabulky

Tab. 2: Sekvence navržených primerů pro amplikonové sekvenování TLR.

Cílové lokusy, jména primerů, jejich sekvence a směr (F – forward primer, R – reverzní primer).

Lokus	Jméno	Sekvence (5'-3')	F/R
TLR1B	TLR1B_1Ra	CTGGAGTTCAGACGTGTGCTCTTCCGATCTAATGTCTCCAGCTGAAGTAGGT	R
	TLR1B_1Fc	CTCTTTCCCTACACGACGCTCTTCCGATCTTATCTGATTTTCAGACAAGGAGC	F
TLR4	TLR4_1Fa	CTCTTTCCCTACACGACGCTCTTCCGATCTACCTCTCTGTCTGAGCTGC	F
	TLR4_1Rc	CTGGAGTTCAGACGTGTGCTCTTCCGATCTACTCCCCAGTATGAGCCT	R
	TLR4_2Fb	CTCTTTCCCTACACGACGCTCTTCCGATCTTGGCTCTCAGGTCTGCTT	F
	TLR4_2Ra	CTGGAGTTCAGACGTGTGCTCTTCCGATCTTGTCTGGGTAATACGAAGC	R
	TLR4_3Fa	CTCTTTCCCTACACGACGCTCTTCCGATCTTTGAGGAGATACCACAGGTT	F
	TLR4_3Rc	CTGGAGTTCAGACGTGTGCTCTTCCGATCTAGGGACAGAAAGACAGGGTA	R
TLR15	TLR15_1Fb	CTCTTTCCCTACACGACGCTCTTCCGATCTTACAGCCACGACCGAGCC	F
	TLR15_1Rb	CTGGAGTTCAGACGTGTGCTCTTCCGATCTGGTATTTGGCTGATGTTGGTGAAG	R
	TLR15_F2b	CTCTTTCCCTACACGACGCTCTTCCGATCTCCATTAACATCCTGGATGC	F
	TLR15_R2a	CTGGAGTTCAGACGTGTGCTCTTCCGATCTCAGTTCTTCCATTGTGTCAG	R
	TLR15_F3a	CTCTTTCCCTACACGACGCTCTTCCGATCTCCTGGAAGTGTACACATCT	F
	TLR15_R3b	CTGGAGTTCAGACGTGTGCTCTTCCGATCTTGTGATCAGGTTCCAGAA	R

Tab. 3: Segregující pozice aminokyselin, frekvence haplotypů a výskyt proteinových haplotypů v populacích pro TLR1B.

Hap	10	57	128	135	freq	g/c	pop
1	I	T	R	F	391	obojí	všechny
2	M	A	R	F	1	cont	Co
3	I	T	R	L	3	gala	St-Cru
4	I	T	H	F	1	gala	E
5	I	A	R	F	30	cont	Co

Hap – proteinový haplotyp, čísla 10, 57, 128... – aminokyselinovým kódem určené segregující pozice v sekvenované části TLR1B, freq – frekvence haplotypu, g/c – výskyt haplotypu, zda je galapážský či kontinentální, pop – výskyt v jednotlivých populacích. Zkratky pro populace: Santiago (Stg), Isabela (I), Champion (Ch), Santa Fe (St-Fe), Marchena (M), San-Cristobal (S-Cri), Santa Cruz (St-Cru), Wolf (W), Pinta (P), Rabida (R), Espanola (E), kontinent (Co), Gardner by Floreana (G), Fernandina (F).

Tab. 4: Segregující pozice aminokyselin, frekvence haplotypů a výskyt proteinových haplotypů v populacích pro TLR4.

Hap	33	56	65	68	69	80	86	99	104	105	113	130	133	174	180	189	191	213	220	freq	g/c	pop
1	T	L	L	S	L	E	G	E	H	T	L	Q	A	R	L	L	K	L	R	1	cont	Co
2	T	L	L	T	V	E	G	E	H	T	L	Q	E	R	L	L	K	L	R	1	cont	Co
3	T	L	L	T	L	E	G	E	R	T	L	Q	E	R	L	W	K	L	R	2	gala	St-Cru
4	S	L	V	T	L	E	G	E	H	T	L	Q	E	R	L	W	E	L	R	43	gala	Ch, G
5	T	L	L	T	L	E	G	E	H	T	L	Q	A	R	L	L	K	L	R	1	cont	Co
6	T	L	L	T	L	E	E	E	H	T	L	Q	E	R	L	L	K	L	R	1	cont	Co
7	T	L	L	T	L	E	G	V	H	T	L	Q	E	R	L	L	K	L	R	1	cont	Co
8	T	L	L	T	L	E	G	E	H	A	L	Q	E	R	L	L	K	L	R	1	cont	Co
9	T	F	L	T	L	E	G	E	H	T	L	Q	E	R	L	L	K	L	R	1	cont	Co
10	T	L	L	T	L	E	G	E	H	T	L	Q	E	R	L	W	K	L	R	360	gala	vše mimo Ch a Co
11	T	L	L	T	L	E	G	E	H	T	L	Q	E	R	I	L	K	V	R	2	cont	Co
12	T	L	L	T	L	E	G	E	H	T	F	Q	E	R	L	W	K	L	R	1	gala	F
13	T	L	L	T	L	E	G	E	H	T	L	K	E	R	L	W	K	L	R	9	gala	I, F
14	T	L	L	T	L	D	G	E	H	T	L	Q	E	R	L	L	K	L	R	1	cont	Co
15	T	L	L	T	L	E	G	E	H	T	L	Q	E	R	L	L	K	L	R	24	cont	Co
16	T	L	L	T	L	E	G	E	H	T	L	Q	E	L	L	W	K	L	R	1	gala	St-Cru
17	T	L	L	T	L	E	G	E	H	T	L	Q	E	R	L	W	K	L	H	2	gala	I, F
18	T	L	V	T	L	E	G	E	H	T	L	Q	E	R	L	W	K	L	R	6	gala	G

Hap – proteinový haplotyp, čísla 33, 56, 65... – aminokyselinovým kódem určené segregující pozice v sekvenované části TLR4, freq – frekvence haplotypu, g/c – výskyt haplotypu, zda je galapážský či kontinentální, pop – výskyt v jednotlivých populacích. Zkratky pro populace: Santiago (Stg), Isabela (I), Champion (Ch), Santa Fe (St-Fe), Marchena (M), San-Cristobal (S-Cri), Santa Cruz (St-Cru), Wolf (W), Pinta (P), Rabida (R), Espanola (E), kontinent (Co), Gardner by Floreana (G), Fernandina (F).

Tab. 5: Segregující pozice aminokyselin, frekvence haplotypů a výskyt proteinových haplotypů v populacích pro TLR15.

Hap	20	21	23	39	43	46	50	58	73	164	172	222	225	242	245	270	freq	g/c	pop
1	R	S	L	A	R	E	G	L	S	G	P	E	L	P	Q	F	2	gala	I
2	H	S	L	A	R	D	S	L	S	G	P	E	L	P	Q	F	2	cont	Co
3	H	S	L	V	R	E	S	L	S	G	P	E	L	P	Q	F	5	cont	Co
4	H	N	L	A	R	E	S	L	S	G	P	E	L	S	Q	F	2	cont	Co
5	R	S	L	A	R	E	S	L	S	G	P	D	L	P	Q	F	1	gala	St-Cru
6	R	S	L	A	R	E	S	L	S	G	P	E	L	P	Q	L	3	gala	I
7	H	N	L	A	R	E	S	L	S	V	P	E	L	P	Q	F	2	cont	Co
8	H	S	L	A	R	E	S	L	S	G	P	E	L	P	Q	F	59	gala	Stg, I, St-Fe, St-Cru, P, R, F
9	H	S	L	V	R	E	S	L	S	V	P	K	F	P	Q	F	1	cont	Co
10	R	S	L	A	R	E	S	L	S	G	P	E	L	P	Q	F	170	gala	Stg, I, St-Fe, S-Cri, St-Cru, W, R, F, M
11	R	S	L	A	R	E	S	F	S	G	P	E	L	P	Q	F	84	gala	I, Ch, G
12	H	N	L	A	R	E	S	L	S	V	P	E	F	P	Q	F	1	cont	Co
13	H	S	L	A	R	E	G	L	S	G	P	E	L	P	Q	F	6	gala	I, St-Cru, F
14	H	N	L	V	R	E	S	L	S	G	P	K	L	P	Q	F	1	cont	Co
15	H	S	L	A	R	E	S	L	S	G	P	E	L	P	K	F	1	gala	F
16	H	N	L	A	R	E	S	L	S	G	P	E	L	P	Q	F	4	cont	Co
17	H	S	L	V	R	E	S	L	S	G	P	K	L	P	Q	F	2	cont	Co
18	H	S	L	A	R	E	G	L	F	G	P	E	L	P	Q	F	2	gala	I
19	H	S	L	V	G	E	S	L	S	G	P	E	L	P	Q	F	1	cont	Co
20	R	S	L	A	R	E	S	L	S	G	R	E	L	P	Q	F	2	cont	Co
21	H	S	L	V	R	E	S	L	S	V	P	K	L	P	Q	F	1	cont	Co
22	R	S	V	A	R	E	S	L	S	G	P	E	L	P	Q	F	50	gala	E

Hap – proteinový haplotyp, čísla 20, 21, 23... – aminokyselinovým kódem určené segregující pozice v sekvenované části TLR15, freq – frekvence haplotypu, g/c – výskyt haplotypu, zda je galapážský či kontinentální, pop – výskyt v jednotlivých populacích. Zkratky pro populace: Santiago (Stg), Isabela (I), Champion (Ch), Santa Fe (St-Fe), Marchena (M), San-Cristobal (S-Cri), Santa Cruz (St-Cru), Wolf (W), Pinta (P), Rabida (R), Espanola (E), kontinent (Co), Gardner by Floreana (G), Fernandina (F).

Tab. 6: Charakteristiky sekvenačních dat všech tří TLR.

Lokus	Délka (aa)	Počet jedinců	Počet DNA haplotypů	Počet prot. haplotypů	π	TajD	dN-dS	dN	dS	ps	pn	pn/ps
TLR1B	472	213	8	5	0.00096	-1.265	-1.677	4	4	NA	NA	NA
TLR4	889	228	34	18	0.00218	-1.655	-0.749	21	18	0.00918	0.002534	0.27608
TLR15	970	214	39	22	0.02742	-1.495	-1.078	25	31	0.003535	0.007399	2.09314

Lokus, délka v aminokyselinách, počet sekvenovaných jedinců, počet DNA a proteinových haplotypů pro lokus, π – nukleotidová diversita, TajD – Tajimovo D, dN-dS – rozdíl nesynonymních a synonymních mutací počítaný programem SLACK, dN – počet nesynonymních mutací, dS – počet synonymních mutací, ps – poměr počtu synonymních mutací a počtu synonymních míst, pn – poměr počtu nesynonymních mutací a počtu nesynonymních míst, pn/ps – poměr pn a ps. Pro TLR1B nebyl dostatek dat pro spočítání ps, pn a pn/ps.

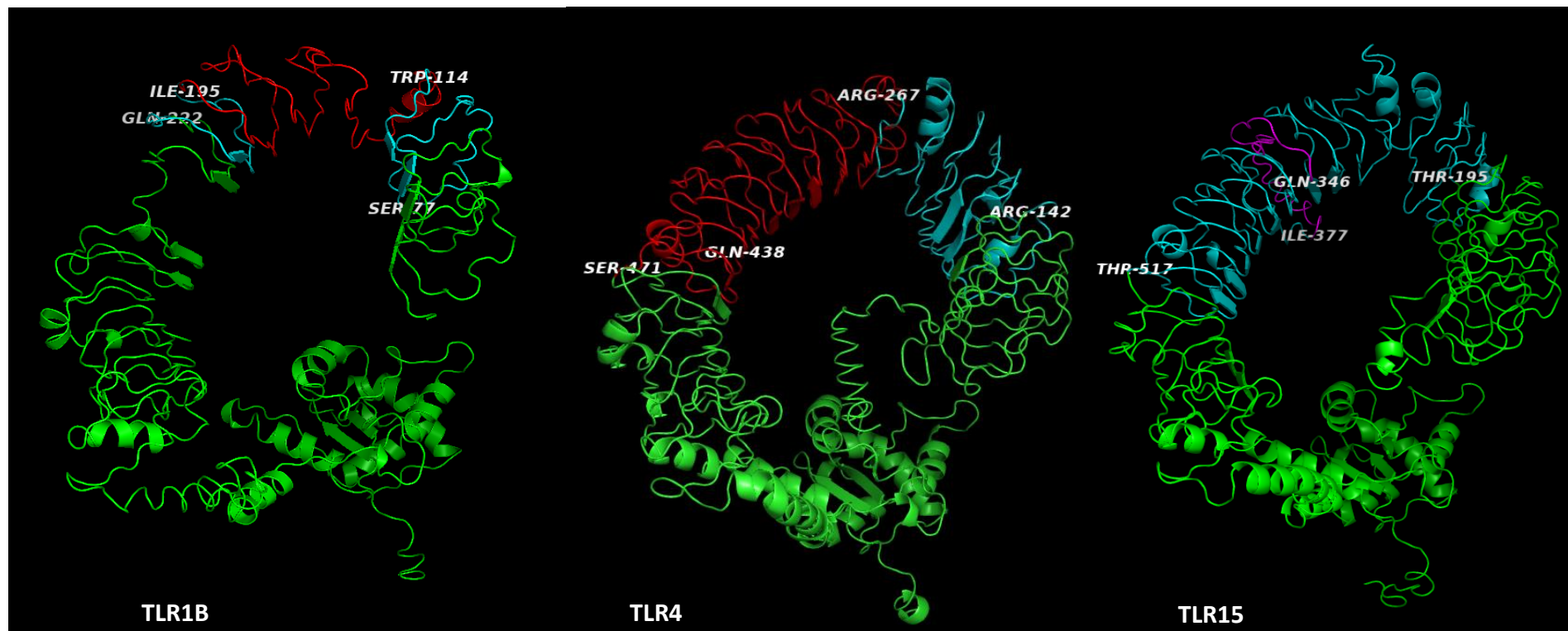
Tab. 7: Porovnání populačně-genetických statistik mezi galapážskými populacemi a kontinentální populací.

Lokus	Pop	π	Počet jedinců	Počet DNA hap	Počet prot. hap	pn/ps	Hs DNA	Hs prot
TLR1B	Gala	0.0002	197	4	3*	NA	0.0125	0.0087
	Cont	0.0008	16	4	3*	NA	0.31	0.125
TLR4	Gala	0.0008	195	16	8	0.0746	0.4058	0.0081
	Cont	0.0027	33	18	10	0.1306	1	0.5882
TLR15	Gala	0.0007	204	20	10	0.1679	0.33	0.1988
	Cont	0.0062	11	19	12	0.1851	1	0.9091

Lokus, Pop – galapážská či kontinentální populace, π – nukleotidová diversita, počty DNA a proteinových haplotypů, pn/ps – průměr pn/ps pro populace (ps – poměr počtu synonymních mutací a počtu synonymních míst, pn – poměr počtu nesynonymních mutací a počtu nesynonymních míst, pn/ps – poměr pn a ps), Hs DNA – průměr heterozygotnosti DNA haplotypů, Hs prot – průměr heterozygotnosti proteinových haplotypů.

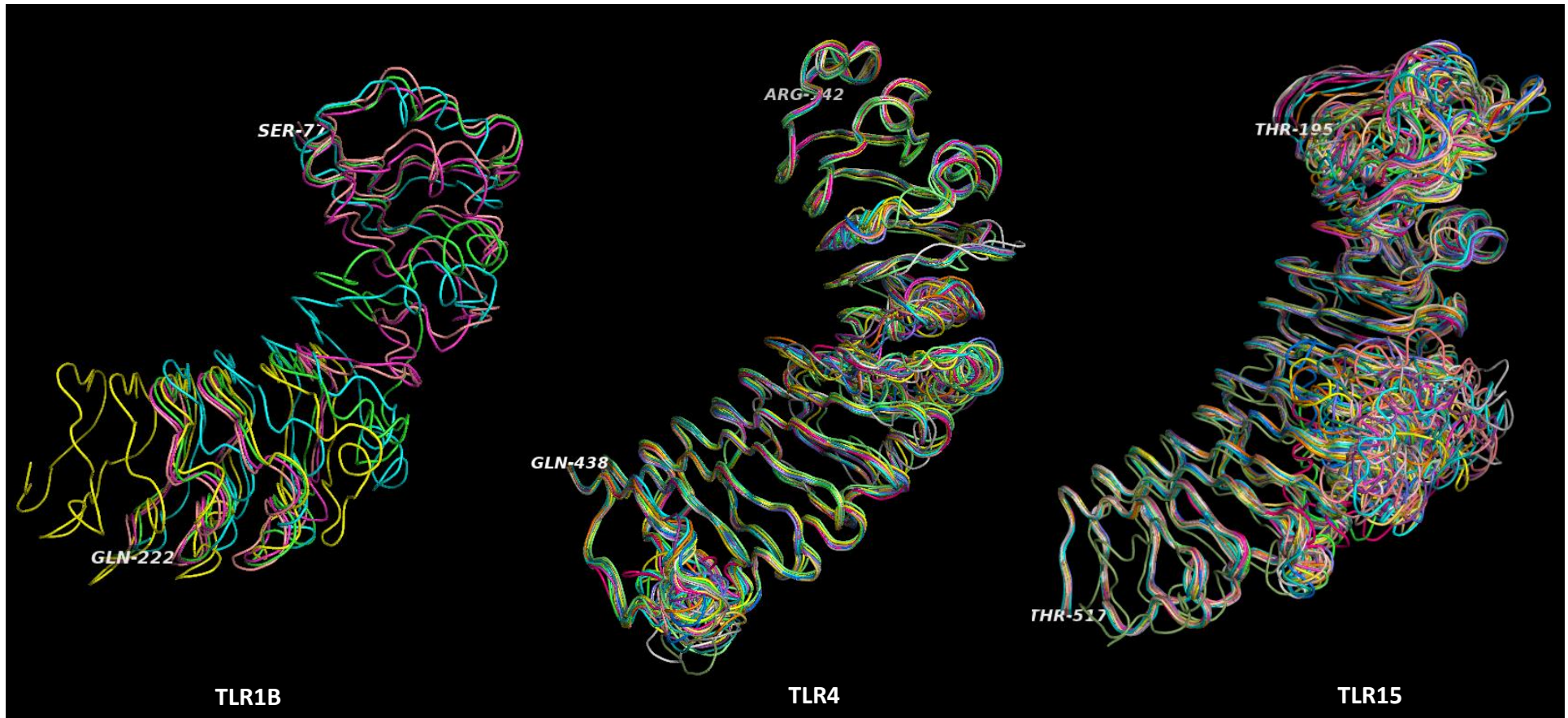
* - 2 galapážské a 2 kontinentální haplotypy a jeden společný pro obě populace, celkově tedy 3 v každé populaci

7.4 Obrázky



Obr. 18: 3D modely Toll-like receptorů modelované serverem I-TASSER (vizualizace PyMOL).

Sekvenované části TLR (tyrkysová) s ligand vázajícím místem (červená), či štěpným místem (fialová), doplněné o konce receptorů z *Ficedula albicollis* (zelená). Popsány jsou aminokyseliny ohraničující vyznačené části.



Obr. 19: Vizualizace strukturního alignmentu sekvenovaných segmentů TLR1B, TLR4 a TLR15 funkcí „super“ v programu PyMOL. První a poslední aminokyseliny jsou označeny pro orientaci struktury. Vyšší číslo je blíže k TIR a menší číslo je blíže k N-konci proteinu. Struktury haplotypů jsou barevně odlišeny pro znázornění strukturních rozdílů.