

## Oponentský posudek – Vlček: Variability of MHC class II $\beta$ gene in Galápagos mockingbirds

Diplomová práce Bc. Jakuba Vlčka se zabývá studiem variability MHCII  $\beta$  genů galapážských drozdců rodu *Mimus*. Práce má rozsah 61 stran včetně příloh a je psána formou manuskriptu v anglickém jazyce. Mimo vytvoření souhrnu současných znalostí o struktuře, funkci a evoluci MHC komplexu si práce klade za cíl charakterizovat variabilitu uvedených genů u 203 jedinců pocházejících z 15 populací 5 druhů drozdců (4 ostrovní a jeden pevninský) a její rozložení s použitím sekvenačních technologií druhé generace, a dále pak analyzovat vliv evolučně-genetických a imunologických /ekologických faktorů na uvedený komplex.

Z formálního hlediska je práce psána srozumitelně, má jasnou a logickou strukturu a obsahuje jen malé množství překlepů. V práci jsou chybné některé odkazy na obrázky (9, 10 místo 8, 9 – str. 31), odkazy na obrázky a přílohy nejsou referovány ve vzestupném pořadí a tab. 4 chybí. Jednotlivé druhy jsou nazývány v obrázcích a tabulkách spíše latinsky a v textu většinou anglicky, což čtenáře poněkud zdržuje. U tab. 6 a tabulky v příloze 1 bych uvítal seřazení populací jinak než podle názvu ostrova – buď standardně dle taxonomie s outgroupem v prvním nebo posledním řádku nebo podle některého z parametrů korelujících s alelickou variabilitou (velikost ostrova, velikost populace, celkový počet alel v populaci), čtenáři by to usnadnilo porovnání vlivu ekologických a imunologických faktorů na AR. Podobně u tab. 3 by asi bylo přehlednější uvádět primery a jejich fúzní deriváty s odpovídajícími si bázemi pod sebou.

Úvodní souhrn o MHC a představení modelového systému je velmi zdařilý; přehledný a čtivý byť se jedná o problematiku obsáhlou a složitou. V metodické části postrádám aspoň jednu větu s charakteristikou sledovaných parazitů z hlediska potenciálu vytvářet selekční tlak, to je ale z velké části napraveno v diskusi. Chybí i představení imunitních faktorů.

Dále bych měl k práci několik metodických výhrad které mají často společného jmenovatele, a tím je použití dle mého názoru nejméně vhodné sekvenační platformy, a to jak z kvalitativního tak kvantitativního hlediska. Ion Torrent byl pravděpodobně jedinou možností z důvodu omezených financí, takže se zřejmě nejednalo o svobodnou volbu autora. Výsledná hloubka čtení ale je, jak autor na několika místech přiznává, evidentně nedostatečná a navíc nerovnoměrně rozložená mezi vzorky. Osobně si sice myslím, že saturační graf v příloze 3 může být z velké části arteficiální a nemusí nutně znamenat neodhalení velkého množství alel, několikanásobně větší sekvenační kapacita je ale bezpodmínečně nutná minimálně pro zkvalitnění celého filtrovacího procesu. Reproducibilita vzorků mezi dvěma sekvenačními běhy je kritická, byť mi není z popisu jasné, zda se jedná o výpočet

provedený jen ze všech finálně klasifikovaných alel anebo některého z meziproductů. V prvním případě by to pravděpodobně znamenalo velké nadhodnocení výsledné variability a tím i možné ovlivnění vyslovených závěrů.

Souhlasím s tvrzením, že šance že by genotypovací chyby ovlivnily výslednou silnou korelaci mezi velikostí ostrova, SSR AR a celkovým počtem MHC alel v populaci je malá (str. 35; dle autora nejvýznamnější výsledek práce), neboť dle mého názoru i významné pod- nebo nadhodnocení variability MHC by tento vztah zachovalo. Autor ale z tohoto faktu v následující větě vyvozuje, že tento výsledek potvrzuje spolehlivost genotypovacího procesu, s čímž si dovoluji velmi nesouhlasit.

Při tvorbě amplikonů není zjevné, zda byla před použitím fúzních primerů nejprve provedena amplifikace některých vzorků s běžnými primery 325F/325R a výsledky porovnány (přechod na fúzní primery vyžaduje optimalizaci a může vést k ochuzení o některé PCR produkty). HPLC purifikace primerů mi přijde jako zbytečně velká investice navíc; při použití 10-místného barkódu pro takto malé množství vzorků je tento odolný proti velkému počtu chyb. V bioinformatickém procesu mi chybí zmínka o kroku pro odhalení potenciálních kontaminantů (porovnání sekvencí s referenční databází), ač je tento krok nedílnou součástí postupu v autorem použitým programem Sesame. Naopak, není vyřazení sekvencí s chybou v oblasti primeru v dané fázi zbytečným ochuzením o téměř 30 % dat? Uvítal bych jednotný postup při zpracování 454 a Ion Torrent dat, porovnání variability by pak možná dávalo více informací o možném zdroji chyb; k tomu by rovněž přispělo zařazení sekvenačních (a nejen amplifikačních) replikátů vzorků do procesu. Obecně mi celá bioinformatická roura přijde příliš závislá na explicitně definovaných parametrech s malou mírou autoadaptability na vlastnosti konkrétního datasetu; jedná se ale o problematiku mnohem složitější než se na první pohled zdá a v současné době zjevně není mnoho funkčních alternativ.

Diskuse je velmi zdařilá a vypovídá o hlubokých znalostech problematiky a schopnosti autora interpretovat získané výsledky; bohužel právě výsledky jsou z již uvedených důvodů nejnapadnutelnější částí práce. Oceňuji, že autor si je většiny problematických míst ve své práci vědom a a aktivně na ně sám reaguje.

Autor při vytváření své diplomové práce jednoznačně prokázal že se orientuje v dané problematice, zvládá učit se a aplikovat nejnovější postupy (jak laboratorní tak bioinformatické) a je schopen samostatné vědecké práce. Místo obvyklého odeslání vzorků k sekvenaci komerčním firmám si navíc dokázal osvojit základy hned dvou NGS platforem a aktivně se podílel na zpracování DNA knihoven, provedení emPCR a nasazení sekvenačních běhů. Jeho práce bezesporu splňuje veškeré nároky

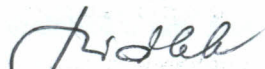
kladené na diplomové práce na PŘF JU a já ji plně doporučuji k obhajobě a budu hodnotit stupněm výborně.

Na autora mám následující dotazy:

1) Pokud by dnes autor byl na začátku své experimentální práce a tentokrát nebyl příliš limitován finančními prostředky, jaký postup by zvolil: a) zůstal by věrný sekvenování ampliconů (s jakými modifikacemi) nebo zvolil cílené sekvenování MHC komplexu (např. NimbleGen) či jiný postup; b) jakou platformu a sekvenační systém by zvolil (případně zaujal vyčkávací taktiku); c) jak by zhruba vypadal celý bioinformatický postup selekce klasifikovaných alel?

2) Za velmi zajímavý výsledek považuji úroveň alelických divergencí u jednotlivých populací ve shodě s hypotézou Spurgin & Richardson (2010). Zároveň zcela souhlasím s autorem, že tento parametr bude více robustní vůči genotypovacím chybám než vlastní AR. Proč tedy není s tímto výsledkem dále pracováno? Má autor představu, jak by asi vycházely korelace divergencí s ekologickými parametry nebo jejich saturační grafy (příloha 3)?

V Českých Budějovicích dne 11.1. 2015

  
Lubomír Piálek



Oponentský posudek na magisterskou diplomovou práci

### Variability of MHC class II $\beta$ gene in Galápagos mockingbirds

Autor: Bc. Jakub Vlček

Oponent: Doc. Mgr. et Mgr. Josef Bryja, PhD.

Magisterská práce Jakuba Vlčka analyzuje variabilitu MHC Class II genů u galapážských drozdců. Studovaný systém je velmi vhodný pro hledání odpovědí na otázky spojené s evolucí adaptivní variability ostrovních populací, ale získané výsledky mohou být velice užitečné i v aplikované ochranářské praxi. Pěvci jsou obecně známí svou vysokou komplexitou MHC, a proto není možno pro genotypizaci použít tradiční postupy využívané třeba u savců. V diplomové práci proto byla MHC data získána metodami tzv. sekvenování druhé generace. Přestože se tyto metody již několik let vcelku rutinně používají, tak získané výsledky ukazují, že jejich interpretace je občas komplikovaná a přeměna rozsáhlých sekvenčních dat do smysluplných genotypů není rozhodně samozřejmá. Jakub Vlček si však s metodickými problémy poradil velmi dobře. Jednoznačně si uvědomuje možné zdroje chyb v získaných datech a k tomu přizpůsobuje své interpretace. Zjištěná variabilita MHC pak byla následně analyzována v souvislosti s velikostí populace (tj. intenzitou driftu), imunologickými parametry a intenzitou napadení ektoparazity. Za nejzajímavější zjištění považuji (1) významný vliv náhodného genetického driftu na populační variabilitu MHC, srovnatelný s neutrální variabilitou, a (2) výrazný rozdíl ve vnitropopulační divergenci MHC molekul mezi kontinentální a ostrovními populacemi.

Celá práce je psána srozumitelnou angličtinou a věřím, že její převedení do kvalitní publikace bude velmi přímočaré (i když finální jazyková korekce rodilým mluvčím bude určitě prospěšná). Při čtení mne však napadaly různé otázky, které specifikuji níže a které možná mohou pomoci vylepšit některé analýzy či interpretace. Formát práce je velmi dobrý, pouze Literární rešerše mohla být rozdělena více podnadpisy a citace obrázků nejdou vždy po sobě (na str. 31 jsou navíc špatné odkazy). Velmi oceňuji kapitolu Diskuze, která bývá většinou nejslabší částí diplomových prací. Řada otázek, které jsem si poznačil při čtení výsledků a na které jsem se chtěl zeptat při obhajobě, byla velmi uspokojivě vysvětlena právě v Diskuzi. Níže uvádím některé koncepční připomínky, které jsou zároveň námětem pro diskuzi u obhajoby:

(1) Celá práce je založena na DNA sekvencích MHC ("response variable = total number of MHC alleles"), nicméně funkci MHC určuje výsledný protein. Bylo by určitě žádoucí uvést, zda se každá alela překládá do odlišné aminokyselinové sekvence a pak provést analýzy znovu s variabilitou proteinových sekvencí. Rovněž je známo, že chemicky velice podobné proteinové sekvence mohou mít podobnou funkci a určitě by stálo za to seskupit jednotlivé proteinové varianty to tzv. „supertypes“.

(2) Je pěkné, že autor se snažil asociovat vztah individuální MHC variability s vybranými imunologickými parametry a s výskytem ektoparazitů, ale v práci chybí vysvětlení možných proximálních mechanismů. Skutečně může diverzita MHC ovlivnit třeba intenzitu napadení ektoparazity či aglutinaci?

(3) Výrazně nižší vnitropopulační divergence MHC na ostrovech může být způsobena nižším parazitárním tlakem (jak je zmíněno v diskuzi). Ví se něco o parazitech, kteří skutečně mohou přímo selekčně působit na MHC? Na str. 36 je třeba zmíněno, že na ostrovech je "impoverished insular parasite and pathogen fauna", ale bylo by dobré to rozvinout do větších detailů.

(4) Pozitivní a negativní selekce na úrovni sekvencí byly mimo jiné analyzovány metodou SLAC. Pro získání robustních výsledků bych doporučoval i použití dalších "codon-based" metod, např. REL či FEL. Rovněž relativně málo informativní je srovnání s ABS definovanými na základě lidského strukturního

modelu. Zajímavější by bylo srovnat, zda se místa pod pozitivní selekcí shodují s obdobně detekovanými místy u jiných pěvců (viz např. naše práce Promerová et al. 2009 v Immunogenetics) – pokud by tam byla shoda (a já věřím, že ano), tak by to znamenalo, že interakce s antigenem se u pěvců účastní jiné aminokyseliny.

(5) Poněkud mně překvapila opravdu nízká opakovatelnost jak v rámci opakovaných genotypizací stejných jedinců (jen 38,2 % identických genotypů – pokud jsem dobře pochopil, jak se k tomuto číslu došlo), tak i mezi jednotlivými technologiemi (průměrně jen 52,3 % alel bylo shodných mezi 454 a IonTorrent). Autor v diskuzi uvádí možné příčiny, nicméně mně to pořád vrtá hlavou. Můžeme tedy těmto datům skutečně věřit a případně co udělat, abychom získali opravdu důvěryhodné genotypy? Nevybavuji si, že by v předchozích studiích byla zaznamenána tak nízká opakovatelnost (viz např. opět naše práce Promerová et al. 2012, Mol Ecol Res). Má třeba technologie od Illuminy lepší vlastnosti?

Do textu práce jsem přímo označil některé překlepy či nejasné formulace, některé drobnější připomínky uvádím níže (při obhajobě se třeba můžeme bavit o těch, které **jsou uvedeny tučně**):

Anotace – "specific form of natural selection" je poněkud vágní

str. 1 i jinde – často je používán termín "immunocompetence" – vždy je potřeba vysvětlit, co přesně se tímto termínem myslí, neboť imunitní odpověď může být skutečně rozmanitá. Například určité není pravda, že "MHC variation is a measure of immunocompetence" (str. 40).

str. 15 – fylogenetický strom na základě mtDNA není příliš informativní (je spíše matoucí vzhledem k hybridizacím atd.) a pro popis genetické struktury galapážských drozdců a evolučních vztahů mezi jednotlivými populacemi by bylo lepší asi použít nějaké jaderné znaky?

str. 8, 36 i jinde – funkce MHC asi nemůže být úplně nahrazena jinou částí imunity, je to nepřesné  
str. 23, konec 1. odstavce – "all reads shorter than 164 bp" je trochu zavádějící a několik stran mně trvalo, než jsem pochopil, že to znamená "all 163 bp reads".

str. 24, konec 2. odstavce – chiméry byly definovány na základě sekvencí nalezených u daného jedince. Jak pak může existovat možnost "(3) no obvious chimeric origin was found"?

**str. 28 a jinde – extrémní počty MHC variant na jedince byly 2 a 28. Byli tito jedinci genotypováni opakovaně? Protože to jsou "outliers", tak by to bylo žádoucí.**

str. 34, Fig. 10 – na základě jakého klíče byly vybírány sekvence ostatních druhů pěvců do této analýzy?

**str. 42 a jinde – pseudogeny tvořily 1/10 coverage, což je dosti podstatná část (nefunkční) MHC variability. Jak rozšířená je taková situace u jiných pěvců? Toto zjištění by si určitě zasloužilo detailnější diskuzi.**

str. 43 – jeden ze závěrů je, že evoluce MHC u galapážských drozdců není ovlivněna selekcí způsobenou patogeny. Při neznalosti informací o patogenech je to myslím dosti předčasné tvrzení.

#### Závěr:

Jakub Vlček v rámci své diplomové práce nasbíral unikátní data týkající se MHC variability u ohrožené skupiny ostrovních pěvců. Prokázal, že je schopen využívat moderní genetické technologie, správně vyhodnotit získaná data a interpretovat je v širším kontextu. Práci jednoznačně **doporučuji k obhajobě** na PŘF JU v Českých Budějovicích a navrhuji nejlepší možné hodnocení.

Ve Studenci 16.1.2015

Josef Bryja