



BIOLOGY CENTRE CAS

address: Branišovská 1160/31, 370 05 České Budějovice, Czech Republic

IBAN – CZ24 0800 0000 0000 0606 3942 | SWIFT CODE – GIBACZPX | VAT No.: CZ60077344

phone: +420 387 771 111 (telephone exchange) | www.bc.cas.cz | e-mail: bc@bc.cas.cz

Posudek magisterské práce Martina Těšického „**Trans-species polymorphism in selected innate immunity genes in tits (Paridae family)**“.

Obhajovaná práce se zabývá velmi aktuálním tématem evoluce imunitních genů v populacích volně žijících zvířat. Úvod práce je velmi dobře a podrobně proveden, autor přibližuje téma od komplexních ekologicko-evolučních hypotéz, přes molekulární úroveň fungování imunitní odpovědi v buňce, po popis modelového organismu - sýkorek čeledi Paridae. V úvodu se zjevně autor mohl opřít již o vlastní publikaci (Těšický a Vinkler, 2015), což není u studentů tohoto stupně určitě běžné a ukazuje to na velmi dobré autorské schopnosti. Text anglicky psané práce je téměř bez chyb. I když občasné překlepy, chyby v citacích způsobené špatným formátováním, nebo vzácně i formulační nejasnosti by se našly, žádné z nich nejsou na překážku porozumění textu.

Komplexitu zvoleného tématu práce demonstruje fakt, že téměř každá hypotéza o formování či udržování diverzity imunitních genů si s postupem času nalézá empirickou oporu. Posléze není jednoduché nalézt vhodný model, který umožní od sebe odlišit relativní vliv jednotlivých faktorů. Autor si k prozkoumání tématu mezidruhového sdílení (trans-species polymorfismu, TSP) TLR imunitních genů vybral modelovou skupinu s poměrně komplikovanou ale na druhou stranu dobře prozkoumanou evoluční historií. Autor aplikoval velmi pokročilé metody sekvenování na Illumina platformě pomocí multiplexů genů a jedinců, i pokročilé metody analýzy dat. Byť, jak je uvedeno v poděkování, některé postupy byly provedeny spolupracovníky, rozsah prací provedených autorem a rozpětí osvojených metod je spíše na úrovni doktorandských diplomantů.

Vedle analýzy divergence TLR a neutrálních lokusů a jejich sdílení mezi různými druhy, se autor pustil i do hledání aminokyselin podléhajících selekci a do odhadu míry konzervativnosti či proměnlivosti jednotlivých pozic v evoluční historii.

Analýzy historické migrace a změn velikostí populací založené na neutrálních lokusech pak měly poskytnout informace o míře introgrese jako jednoho z možných zdrojů TSP. Tyto analýzy jsou velmi náročné na dostatek informativních lokusů. S nárůstem porovnávaných párů druhů nebo populací strmě roste i počet odhadovaných parametrů a tak nepřekvapilo, že zejména u modelu porovnávacího čtyři druhy nastal problém s ESS statistikou a s velkým rozpětím v odhadech některých parametrů (např. stáří druhové divergence a míry genového toku). Přesto analýza přinesla několik důležitých údajů pro porovnání s rozsahem mezidruhového polymorfismu TLR genů.

Autor použil také pokročilé metody rekonstrukce 3D modelů TLR proteinů a jejich povrchového napětí, které jsou mimo moje vlastní zkušenosti, ale svým provedením působí velmi dobře zvládnuty.

Práce patří svým celkovým rozsahem spíše k delším, proto mě trochu překvapila poměrně úsporně napsaná diskuze. Scházela mi například diskuze výsledků IMA2 analýzy ve vztahu k ekologii zvolených druhů (např. velikost areálu vs. efektivní velikost populací a převládající směr introgrese, posléze vliv těchto faktorů na TSP). Tyto ekologické předpoklady mohly být rovněž formulovány již v cílech práce jako předpoklady hypotéz k testování. Diskuze výsledků IMA2 se omezovala spíše na to, zda byla introgrese detekována a u kterých druhů (příbuzné vs. vzdálené taxony), což je důležité, ale scházelo mi ekologické pozadí diskutovaných druhů.

Vzhledem k poměrně inovativnímu sekvenačnímu přístupu mi v diskuzi také scházela aspoň krátká zmínka o kvalitě sekvenačních dat ve vztahu k rekonstrukci alel a odhadu jejich počtů u jedinců apod. Na druhou stranu chápu, že u TLR je toto téma méně ožehavé než u multiplikovaných lokusů jako jsou třeba MHC geny.

Z charakteru i kvality diplomové práce je zjevný záměr o následné převedení do formy manuskriptu (možná i dvou) určeného k publikování. Je zjevné, že po korekcích některých drobných chyb a po zkrácení některých částí bude práce jistě přijata v kvalitním časopise. Gratuluji diplomantovi k dobře odvedenému dílu, práci doporučuji k obhajobě a hodnotím stupněm jedna.

K práci mám ale také několik dotazů:

- 1) Multiplexové sekvenování na platformě s velkou hloubkou čtení představuje vyšší riziko kontaminace (znám z vlastní zkušenosti :). Metodika diplomové práce je většinou velmi podrobná a srozumitelná, ale chybí info, zda byly učiněny nějaké preventivní/kontrolní kroky? Bioinformatická část ohledně filtrování pravých a falešných alel poskytuje některá vodítka, ale není v tomto až tak podrobná.
- 2) Na str. 42 autor popisuje workflow ohledně assembly kontigů a fázování TLR alel. V jakém procentu výsledných sekvencí byla délka a variabilita dostatečná k manuální assembly alel a kdy bylo nutno použít PHASE algoritmy?
- 3) Mutační rychlosti pro program IMA2 byly odhadovány pomocí Neiiovských distancí (D_{xy}). Nebylo by správnější pokusit se o odhad pomocí sofistikovanějších metod (např. Bayesiánské algoritmy v BEAST), které umožňují jednak pracovat s nerovnoměrností v rychlostech v odlišných větvích a hlavně poskytnou lepší obraz o nejistotě v odhadovaných parametrech (posteriorní probabilita, konfidenční interval, atd.)?
- 4) Zdá se, že jeden z multiplexů (C ve Fig. 8, což zřejmě odpovídá multiplexu 3 v Tab. 7) fungoval špatně, s nízkou coverage. Čím si to autor vysvětluje a mohlo to ovlivnit výsledky?
- 5) Jakou roli mohl mít při detekci introgrese sampling u druhů s velkým areálem – např. pokud existuje populační struktura uvnitř některých druhů, může zakrýt introgresi, v případě že data nebyla nasbírána přímo z parapatrických populací hybridizujících druhů.

Jan Štefka.

V Č. Budějovicích 07.09. 2016

Drobné připomínky (určeno ke korekcím při přípravě publikace; náhodný, nikoli kompletní výčet):

Str. 47 uprostřed, duplikována věta o rozdílu mezi IM a IMA2 modely.

Str. 50 výsledky, v textu zřejmě záměna 190 specimens vs. 190 species, místy nejasný význam number of sequences vs. specimens.

Popisek Fig. 26, mělo by být model 2 ne model 3.