

Oponentský posudek na diplomovou práci Jana Waltera: „Fylogeneze a biogeografie modrásků podtribu Everina“

Obhajovaná diplomová práce je fylogenetickou, biogeografickou a částečně také systematickou studií vybrané skupiny motýlů, konkrétně modrásků podtribu Everina (čeleď Lycaenidae). Těžištěm práce je fylogenetická analýza celého podtribu a na ní navazující fylogeografické analýzy pěti vybraných druhů nebo druhových komplexů. Práce je založena na analýze krátkého úseku mitochondriální DNA široce používaného jako determinační marker („molekulární barkód“) a podle vlastního vyjádření studenta má posloužit jako pilotní studie pro budoucí analýzy (str. 44).

Silnou stránkou práce je její zaměření na cílovou skupinu jako celek. Fylogenetická analýza celého podtribu poskytuje kontext pro analýzy jednotlivých taxonů a jejich srovnání umožňuje nahlédnout, nakolik jsou si jejich fylogeografické struktury podobné. Práce je také přehledně strukturovaná – zařazuje předmět studia do širšího kontextu, dopředu poukazuje na problematické body, se kterými se čtenář bude v textu setkávat, analytická část postupuje od celkové fylogeneze k dílčím fylogeografickým analýzám a taxonomické poznámky neodvádí pozornost od hlavního tématu. V neposlední řadě je třeba ocenit vědecký přínos práce, protože fylogeografických studií pokrývajících více příbuzných druhů na kontinentální škále není zas tak mnoho.

Slabou stránkou práce je nepoměr mezi složitostí analýz a jednoduchostí dat, na kterých jsou tyto analýzy založeny. Použití barkódového markeru umožnilo doplnit vlastní data podstatným počtem dříve publikovaných sekvencí. Na druhou stranu je potom potřeba počítat s omezeními, která s sebou přináší skutečnost, že barkódový marker je krátký a variabilní, takže může poskytovat jen málo informace pro vyřešení hlubokých divergencí. Tato skutečnost může být důvodem, proč v prezentovaných analýzách zůstala řada hlubších mezidruhových divergencí nevyřešených (Obr. 4 a 5) a proč je jejich datování zatíženo takovou nejistotou (Obr. 6).

Podobnou nejistotou jsou pravděpodobně zatíženy i rekonstrukce ancestrálních areálů, avšak tato analýza je velmi problematická i z jiných důvodů. Především pracuje s biogeografickými oblastmi, které byly vymezeny pro současné druhy a jejich areály, ale které v jiných dobách nemusely vůbec existovat. Je dobré si uvědomit, kolik věcí se odehrálo na této časové škále. Například vyschlo Středozemní moře a pro tu dobu pak zřejmě nedává smysl mluvit o Španělsku jako diskrétně vymezeném (bio)regionu. Naopak někdejší svébytné bioregiony v analýze chybí, pokud dnes neexistují – příkladem může být v práci několikrát zmíněná Beringie. I kdyby však biogeografická analýza byla omezena na kratší časové škály, bylo by potřeba vysvětlit, jakým způsobem byly vymezeny regiony v ní použité. V práci jsou také zmíněny modely porovnávané programem BioGeoBEARS, ale chybí jakýkoli popis těchto modelů. Čtenář se nikde nedozví, jaké procesy podchycují a na jakých předpokladech staví. Neví tím pádem ani co přesně plyne ze skutečnosti, že nějaký z těchto modelů byl vybrán jako nejlepší.

Posledně zmíněný problém se týká i shlukové analýzy v programu hierBAPS. Chybí jakýkoli popis modelu, na kterém je postaven jeho algoritmus ani není vysvětleno jak pracuje. Čtenář si musí jen domýšlet, co to znamená, že

v určitém druhu byl nalezen určitý počet shluků. Podobně bych uvítal i u haplotypových sítí krátké vysvětlení, co odhadují a jaký je tedy význam spojek mezi haplotypy a jak jsou tyto spojky anotovány.

Určité výhrady mám i k prezentaci výsledků. U fylogenetických analýz není v textu popsáno, které části stromů jsou vyřešené dobře a které ne. V textu jsou také komentovány pouze střední (průměrné či mediánové) doby divergencí, ale ne rozsah možných hodnot (tj. „highest posterior density“ interval). V tomto případě je to přitom zásadní informace, neboť u některých divergencí je rozsah hodnot tak široký, že o něj nelze opřít prakticky žádnou diskusi. Některé údaje jsou z obrázků špatně čitelné (např. přehled ancestrálních regionů), legendy obrázků jsou nedostatečné (např. význam barev v Obr. 4 a 5 je třeba hledat v hlavním textu). Fylogeografickým obrázkům (Obr. 8-12) by prospělo vynechat obarvení haplotypů podle místa původu. Takto se v nich překrývají dva barevné kódy, což je činí nepřehlednými. Hodnoty haplotypových a nukleotidových diverzit by bylo nejlépe shrnout do tabulky a na tu pak v textu odkazovat tam, kde je to účelné. Především však bylo potřeba vysvětlit, co tyto hodnoty říkají o daném druhu.

V diskusi mi chybí pokus o souhrnné srovnání vnitrodruhových struktur (jak moc se liší a jaké může být pozadí těchto rozdílů). Další připomínky jsou jen dílčího rázu. Je zde například diskutována fylogenetická pozice druhu *Elkalyce cogina*, která je v této práci nevyřešená a tak k ní prakticky nejde říct nic nového. Nesprávně je použit termín „genetický drift (genetic drift)“ namísto „genový tah (gene drive)“ (str. 39). V diskusi je zmíněna nejistota, kterou je zatížen odhad ancestrálních areálů. Nezmiňoval bych proto v závěrech práce odhad podle kterého byl předek potribu rozšířen na celé severní polokouli, neboť právě takový odhad tím pádem nejde brát úplně vážně.

Velkého vylepšení by bylo potřeba po jazykové stránce. Jednak obsahuje velmi mnoho malých chyb a nejasných nebo neobratných formulací (např. „(haplotypy) jsou majoritně tvořeny vzorky“, str. 31), takže působí dojmem jako kdyby autor odevzdal zhruba nahozený text, který si už neměl čas pročíst.

Na některých místech stylisticky vybočuje z žánru vědecké práce a připomíná poznámky z porady (např. popis cílů práce, str. 16) nebo vyprávění (např. „proto jsem vyzkoušel oba modely...“, str. 21). V některých případech text až příliš spoléhal na to, že čtenář „přece ví o co jde“ (např. popis nastavení analýzy v programu MrBayes, str. 21). Také některé zkratky (HPD, AICc) zůstaly nevysvětleny a některé informace byly komentovány dříve, než se s nimi čtenář mohl seznámit (např. skutečnost, že některé vzorky byly špatně určeny).

Uvádím zde v bodech několik podnětů, jak by bylo možno podobné analýzy provést jinak a v některých případech se domnívám, že i lépe.

- datování divergencí by se mělo opírat alespoň částečně o sekvence jaderných genů. Minimalistickou verzí takového datasetu by nemělo být těžké připravit, protože pro některé druhy jsou tyto sekvence dostupné a zároveň není nutné, aby byly přítomny u všech. Jednoznačně bych pak v takové analýze ponechal jen jednoho jedince na druh a použil birth-death prior.

- pro analýzu ancestrálních areálů bych zvážil použití postupů modelujících změnu areálu jako difúzi prostorovým grafem (Bouckaert et al. 2018, Nat. Ecol. & Evol. 2: 741-749) nebo takzvané biogeografické datování (Landis 2017, Syst. Biol. 66: 128-144)
- hierBAPS je podle mého názoru vhodný spíše pro multi-lokusová data, v případě jednolokusových bych se opřel o příslušný (zakořeněný) strom a použil ke shlukování metodu mPTP (Kapli et al. 2017, Bioinformatics 33: 1630-1638)
- alternativou vůči SPADS by mohla být metoda EEMS („estimated effective migration surfaces“, Petkova et al. 2016 Nat. Genet. 48: 94-100)

Nakonec mám dvě biogeografické otázky:

Jak dlouho a jak často existovala Beringie jakožto nezamrzlá souš a jak je vůbec možné, že nebyla pokryta ledovcem, když přitom leží v poměrně vysokých zeměpisných šířkách a ledovců bylo celkově dost na to, aby snížily hladinu moře?

Jak vypadala ona glaciální a interglaciální refugia, jaký typ krajiny a vegetace si máme představit?

Hodnocení:

Obhajovaná práce na mně působí rozporným dojmem. Na jednu stranu má bezesporu formát diplomové práce, přináší nová zjištění a student při jejím vypracování musel vstřebat mnoho různorodých informací a alespoň uživatelsky zvládnout řadu statistických analýz. Na druhou stranu působí značně nehotovým dojmem, ať už po formální nebo obsahové stránce, a v tomto směru by si zasloužila ještě jedno důkladné přepracování. Při zvážení obou těchto pohledů ji doporučuji k obhajobě s hodnocením 3 (dobře).

Ondřej Mikula

Studenec, 21. 5. 2021

