MUNI | RECETOX
SCI

Posudek oponenta bakalářské/diplomové práce

Název studijního programu: Matematická biologie a biomedicína

Jméno autora/ky: Bc. Michaela Havlíčková

UČO: 500150

Jméno vedoucí/ho práce: Mgr. Markéta Harazim, Ph.D.

Jméno oponenta/tky: RNDr. Martin Těšický, Ph.D.

Název práce: Detekce selekce v genetických markerech používaných pro fylogenetickou rekonstrukci u obratlovců

	Kritérium hodnocení	max. bodů	Body
1	Úvodní přehledová kapitola práce (úplnost relevantních zdrojů a zpracování publikovaných poznatků, vyvození závěrů pro navazující práci)	10	7
2	Tvůrčí část práce (správnost použitých postupů – z hlediska cílů práce i z hlediska principů a podmínek jejich použití, srozumitelnost jejich popisu)	10	9
3	Výsledky práce (prezentace experimentálních výsledků z hlediska jejich výpovědní hodnoty, úroveň diskuze nad získanými výsledky, úroveň a správnost závěrů vzhledem k zadání)	10	6
4	Jazyková a stylistická úroveň, gramatická správnost	5	4
5	Grafická úprava práce, obrázky, tabulky a grafy (legendy a popis, srozumitelnost bez zřetele k ostatnímu textu, vysvětlení značek, odkazy v textu)	5	4
	Celkem bodů		30

Slovní hodnocení, dotazy a připomínky (možno připojit i samostatný list):
Autorka si pro zpracování své bakalářské práce vybrala téma "Detekce selekce v genetických markerech používaných pro fylogenetickou rekonstrukci u obratlovců", přičemž se snažila zjistit, zdali se v nejčastěji používaných "neutrálních" markerech vyskytuje signál pozitivní či negativní selekce, který by mohl zkreslit výsledky fylogenetických analýz.

Toto je zajímavé a často přehlížena téma. Autorka si zvládla osvojit velké množství "state-of-the-art" bioinformatických metod (např. PAML), připravila robustní automatizovanou pipelinu a analyzovala velké množství dat. Výsledky teto práce přináší mnoha doporučeni pro budoucí fylogenetické studie a jak už to bývá, přináší vice otázek do budoucna než odpovědi.

Celkově se jedná o kvalitní práci s dobrou grafickou a formální úpravou a velmi detailním popisem metodiky, který umožňuje práci snadno reprodukovat.

Bohužel se však velké úsilí vložené do samotných bioinformatických analýz nepodařilo zcela přenést do textu samotné práce. Určitě bych uvítal širší úvod a následnou diskuzi, jak selekce může ovlivnit fylogenetické analýzy (v úvodu jen 1,5 strany), a dále pak, jaké z toho plynou závěry pro další výzkum.

Na můj vkus jsou v práci často popisovány běžné termíny a koncepty (např. FASTA formát, degenerovanost kodu apod.), což někdy odvádí pozornost od hlavních cílů práce. Text tak místy místo přehledného kritického review působí jako učebnicový text, což však může být dáno také zvyklostmi daného oboru.

Některé kapitoly by bylo vhodnější lépe strukturovat – začít popisem obecných, shrnujících sdělení, a poté přistoupit k detailům, pro lepší plynutí „příběhu“. Občas se také důležité body ztrácejí v textu. Napr. v kapitole 3.6 autorka začíná popisem, v jakých tabulkách čtenář najde výsledky, ale vhodnější by bylo začít popisem výsledků a přirozeně odkazovat na tabulky – to se děje až od 4. odstavce. Některé kapitoly z výsledků jsou také spíše metodického charakteru (např. výběr markerů, taxonů, stahování z NCBI) a hlavní výsledky jsou shrnuty jen v relativně útlé kapitole 3.6 „Detekce fylogenetického signálu“. V samotném úvodu diskuze (kapitola 4.1) – často nejdůležitější části DP, kde lze vyzdvihnout hlavní výsledky a zasadit je do širšího literárního kontextu – autorka začíná popisem konstrukce fylogenetického stromu pro selekční analýzy, ačkoliv sama později připouští (také na základě jiných studií), že jeho topologie nemá až tak výrazný vliv na detekci selekce. Co mi však citelně chybí, je přehledné, zjednodušené shrnutí výsledků na základě konsenzu metod a jaké z nich plyne doporučení pro budoucí studie.

Další drobné připomínky (není úplný výčet):

- V Tab. č. 3 ideálně doplnit také taxon a případně délku genu.
- V Tab. 5 by bylo vhodné vztáhnout délku sekvence ke konkrétnímu taxonu (sekvenci z GenBank).
- Pro seznam obrázků a tabulek není nutné generovat popis legendy do obsahu.
- V textu se objevuje velké množství anglicismů – někdy přeložených doslovně, jindy nikoliv. Bylo by vhodné je uvést např. pomocí uvozovek, kurzívy či se je pokusit česky popsat, kde to lze apod. („Homologické sekvence dvou geneticky vzdálených organismů si nemusí být dostatečně podobné a mohou obsahovat značné množství indelů“ → lépe např. insercí a delecí; „V alignmentech se po jejich rafinaci v refineAlignment už nevyskytovaly žádné posuny čtecího rámce“ apod..). Obdobně si nejsem jistý, jestli pozicové modely a modely větví-pozicí jsou v češtině ustálené pojmy.
- Byť je většina konstant z modelů vysvětlena v metodice, legendy v tabulkách by měly být samo vysvětlující.
- Pozor na některá zjednodušující tvrzení, např. „Synonymní substituce nepředstavují žádnou funkční změnu a vyvíjí se většinou neutrálně, stejně tak je tomu v případě nesynonymních substitucí, které neovlivní funkci proteinu (Salemi et al., 2009).”

I synonymní mutace mohou ovlivnit dostupnost kodonu (codon bias) a následně genovou expresi přes rychlost translace či stabilitu RNA.

Otázky:

1. Může autorka graficky či pomocí tabulky jednoduše shrnout hlavní výsledky pro jednotlivé markery z různých metod?

Např. síla selekce (síla dN/dS, počet pozic pod selekcí apod.), rozdíl mezi branch-site vs. site modely, vhodnost/nevhodnost markeru pro danou skupinu?

2. Mohla by autorka na základě své práce a jiných studií navrhnout doporučení, jak přistupovat při fylogenetické analýze s ohledem na možné zkreslení způsobené selekcí?

Např. pro analýzu 30 druhů hlodavců – jak vybrat vhodné markery, jak silná musí být selekce (pozitivní i negativní), aby marker vyloučila či provedla tzv. konkatenaci stromu? Je větším problémem silná selekce odhalená pomocí tzv. branch-site metod nebo jen site-modelu?

3. Autorka píše, že se zaměřila přednostně na protein-kódující geny kvůli použití dN/dS metod.

Lze však otestovat sílu selekce i na nekódujících úsecích, které se často také využívají ve fylogenezi – např. introny či ultrakonzervativní elementy? A pokud ano, jakými metodami?

4. Kapitola 3.3: Jak bylo pracováno s transkriptomickými variantami?
Obvykle se pro analýzu selekce berou v úvahu nejdelší či nejčastější varianty – znamená to, že šlo vždy o varianty 1?

5. Autorka v kapitole 3.4 uvádí:
„Většina alignmentů neobsahovala žádné stop kodony uvnitř sekvence. Výjimkou byly alignmenty genů APOB a RBP3, které obsahovaly 11 a 23 stop kodonů. Sekvence, kde se stop kodony vyskytovaly, jsem vizuálně zkontrolovala a vyloučila možnost, že by se mohlo jednat o chybu v alignmentu."

Pracovalo se tedy se sekvencemi se stop kodony, které byly následně prostřednictvím programu MASCE maskovány, nebo byly tyto sekvence obsahující stop kodony před analýzou vyloučeny?

6. V kapitole 2.1 autorka uvádí, že podmínkou pro zařazení řádu do selekční analýzy bylo, aby obsahoval alespoň čtyři zástupce.
Není to příliš málo pro selekční analýzu? Obvykle se doporučuje alespoň 10–15 sekvencí/taxonů pro robustní selekční test.

Závěr:

I přes výše zmíněné kritické připomínky práci hodnotím velmi pozitivně, plně doporučuji k obhajobě a těším se na diskusi během obhajoby.

Navržené hodnocení:

prospěl	prospěl výborně; 40-33	prospěl; 32-16	neprospěl <16

V Brně dne: 16.06.2025

RNDr. Martin Těšický, Ph.D.