

Oponentský posudek na magisterskou práci Bc. Jany Smrčkové – **Meta-analysis of methodological artifacts of the phylogenetic imbalance**

Předložená práce má rozsah 27 stran včetně 2 tabulek a 4 obrázků. Další 4 tabulky a 8 obrázků nalezneme v příloze. Seznam použité literatury čítá 54 položek.

O zpracování tohoto posudku jsem se aktivně hlásil, neb mě téma předložené práce zaujalo již na seminářích. Studenti, kteří před pobíháním s dalekohledem či mačkáním mikropipety dávají přednost přemýšlení nad výdobytky jiných učenců jsou na této katedře (a zřejmě nejen na ní) poměrně vzácným jevem, což následně platí i o tématech, která zpracovávají. Je ovšem otázka zda mé usilování bylo skutkem bohumilým, neb moje způsobilost posuzovat předloženou práci může být nepochybně zpochybněna. Nezabývám se výrobou fylogenetických stromečků a na příslušníky bratrstva statistiků hledím se směsicí posvátné úcty a nenávisti (přičemž to druhé převažuje). Chtěl bych se proto předem omluvit za případné nehoráznosti, jichž se v posudku následkem svojí nevědomosti snad dopouštím.

Téma předložené práce se mi jeví mimořádně zajímavé. Jak jsem předeslal nejsem sice výrobcem stromečků, konzumuji je však v množství větším než malém a, jako zřejmě každou takto postiženou osobu, mě již mnohokrát napadlo, co se skrývá za více či méně nápadnou asymetrií, jíž se většina jejich topologií vyznačuje. Její zdroje mohou být (a zřejmě i jsou) dvojího druhu: „metodické“ (vzniklé při výrobě stromu) a „věcné“ (způsobené rozdíly v rychlosti vymírání a/nebo speciace v jednotlivých větvích). Splnění cíle předložené práce, stanovit (a posléze snad odfiltrovat) „metodický podíl“ asymetrie, by před interpretátory stromečků otevřelo nesmírně lákavou možnost pátrat po příčinách její „věcné složky“ a následně významně zepičtět vyprávění o makroevoluci celé bioty respektive jejích složek. Po přečtení předložené práce musím bohužel konstatovat, že považuji tento cíl spíše za nesplněný. Nejsem si ovšem jist ani tím, nakolik vůbec splnitelný je (alespoň při stávající úrovni znalosti fylogeneze živých organismů).

Z mého pohledu nejzásadnější výtka se totiž týká východisek celé práce. Autorka se téměř nezabývá vlivem neúplného samplingu. V úvodu a hlavně v metodice jsem o něm nenalezl ani slovo, pouze z jedné poněkud záhadné věty v diskuzi (str. 19, druhý odstavec, třetí řádka) usuzuji, že pracovala převážně či výlučně s nekompletními stromečky, tedy takovými, které neobsahují všechny známé taxony. To by nemuselo příliš vadit, kdyby „nekompletnost“ byla distribuována ve stromečcích náhodně, jak ostatně ukázaly simulace citované v témže odstavci diskuse. Nikdo mě ovšem nepřinutí ani na okamžik věřit, že tomu tak opravdu je. Pominu-li nenáhodnost záměrnou, kdy autora zajímají některé zahrnuté skupiny více než jiné, nesčetné jsou i zdroje nenáhodnosti neúmyslné. Je nepochybné, že pro některé skupiny se shánějí data lépe zatímco pro jiné hůře. Příčiny mohou být politické (válčení, kriminalita, absence infrastruktury), sociální (vazby autora na dodavatele materiálu) ale i biologické (např. obývaný biotop, způsob života, obecná (ne)hojnost). Je navíc spíše pravidlem, že sběrné megataxony (hlavně rody) bývají samplovány „řidčeji“ než taxony malé neřkuli monotypické. Tradiční rodová taxonomie přitom velmi často nemá se skutečnými fylogenetickými liniemi nic společného a to ani u tak solidně probádaných skupin jako jsou vyšší obratlovci. K tomu všemu přistupují díry v alfa taxonomii. Ty ovšem ohrožují i stromečky obsahující údajně veškeré známé taxony. Nejedná se přitom jen o problém exotických skupin či území. Podívejme se, co provedli „pipetáři“ se západopalearktickými myšicemi či hrouzky. Stranou již ponechávám poněkud rozpliznutou definici druhu a to nejen třeba u kytek či některých ryb, s níž se autorka pokusila bojovat vyřazením stromů se subspeciemi. Mám-li vše shrnout, za nejzávažnější překážku získání smysluplné informace z asymetrie topologie fylogenetických stromečků považuji nedostatečnou znalost fylogeneze :-).

Z mé předchozí výtky se bezprostředně odvíjejí výhrady k pojetí provedené meta analýzy. Autorka do ní zahrнула velké množství stromečků lišících se zřejmě vším možným i nemožným. Pro vytčený cíl – hledání metodických zdrojů asymetrie to nepovažuji za příliš šťastné. Jediným ziskem byla

možnost testovat rozdíly v asymetrii mezi velkými fylogenetickými liniemi (což mimochodem ze zvoleného tématu poněkud vybočuje), zaplacená ovšem nemožností odlišit skutečný vliv testovaných (a netestovaných zvláště pak „věcných“) faktorů. Osobně bych volil práci s menším množstvím (nejvýše několika desítkami) stromečků s (řádoby)kompletním samplingem, k nimž by byla k dispozici také primární data (opět pokud možno co nejshodnější). To by umožnilo odděleně testovat nejen vliv metod užitých pro výrobu stromečků ale (pomocí manipulací s daty) i dopady kvality respektive úplnosti dat, mezer v samplingu a možná i leccíhos jiného.

Výše uvedené výtky, které mají spíše polemický charakter, nijak nesnižují mé hodnocení předložené práce. To bohužel neplatí o mých dalších výhradách směřujících vůči prezentaci a interpretaci vlastních analýz, tedy především vůči kapitolám výsledky a částečně i diskuse. Ve výsledcích mě popudila především jejich nevalná formální úroveň. Analýzy ať již uváděné v textu nebo v tabulkách a grafech jsou zcela nedostatečně popsány. K zuřivosti mě uvádělo zvláště obtížné dohledávání zkratk, které jsou i v metodice dovedně ukryty. Komentář působí při prvním pročtení chaotickým dojmem, čtenáři (alespoň mě) uniká, čím si vypíchnuté výsledky zasloužily mimořádnou pozornost, jež jim je věnována. Musím přiznat, že jsem na soustavné hodnocení rezignoval. Vybral jsem nicméně jeden detail, které mě udivil o něco více než jiné. Na str. 11, sedmá řádka odspodu a další, se píše, že závislost Collesova indexu na velikosti stromu byla úspěšně potlačena jakým si „LOESS smoother“ s $p = 0.9$ (v metodice o tomto postupu chybí zmínka) a následuje odkaz na obr.1. Ten ovšem žádné odstranění této závislosti neukazuje. Navíc mě překvapilo, že nejnižší hodnoty indexu vykazují velké stromy. Z metodického hlediska mám, vzhledem k nepochybné heterogenitě datového souboru, velké pochybnosti o informativní hodnotě provedených jednoduchých analýz (shrnutých v tab. 2). V diskusi postrádám soustavnější interpretaci získaných výsledků. To je mimochodem poměrně neobvyklé, většinou mají studenti problémy s jejich zasazením do kontextu stávajících poznatků. Vše ještě zhoršuje to, že se autorka stydí za slovo „má“ či „mé“ a píše o „těchto“ či „sesbíraných“ datech (tak jsem to alespoň pochopil). Zeptal bych se proto alespoň dodatečně: 1) Jak si autorka vysvětluje to, že FC medián nebyl testovanými faktory ovlivněn, zatímco FC „kvartilový rozptyl“ povětšinou ano? 2) Proč nebyl „kvartilový rozptyl“ ovlivněn počtem parsimonních stromů ani počtem znaků připadajících na taxon, zatímco konsistenčním a retenčním indexem ano? 3) Jak by dopadly analýzy v tab. 2, kdyby bylo místo se „stromovou“ počítáno s „nodální“ asymetrií? S nedostatečnou analýzou získaných výsledků kontrastuje suverénní konstatování na str. 21 dole, že tyto ukazují malý vliv metodických zkreslení na symetrii testovaných stromů.

Abych jen nehaněl, velmi se mi líbí autorčina práce s literaturou a to jak v úvodu, tak i v diskusi. Obě kapitoly jsou si také velmi podobné, což ovšem úplně správné není. Právě nezpochybnitelně prokázané proniknutí autorky do problematiky mě vede k tomu, že nemám (více méně) pochybnosti o úspěšnosti obhajoby předložené práce. Hodnocení bych ale ponechal na jejím vývoji.

Č.B. 21.1. 2011

Roman Fuchs



**Oponentský posudek na magisterskou diplomovou práci Jany Smrčkové
„Meta-analysis of methodological artifacts of the phylogenetic imbalance“**

Práce Jany Smrčkové se týká možných artefaktů vzniklých při analýzách imbalance fylogenetických stromů, tedy při analýzách asymetrie jejich větvení. Je psána anglicky a vypadá jako rukopis opravdického článku – proto taky tento posudek píšu tak, jako kdybych dělal opravdickou recenzi. Všechna zjištění jsou dobře podpořená a poměrně důležitá; autorka totiž shromáždila mnohem rozsáhlejší materiál než všichni badatelé, kteří se podobným úkolem zabývali před ní, a proto jsou její výsledky výrazně robustnější. Práce je napsána srozumitelně a až na pár neobratností také dobrou angličtinou. Jedinou vážnější výhradu mám k Diskusi, která mi přijde trochu zmatená, místy opakuje to, co bylo (lépe) řečeno v úvodu, a hlavně z ní zas tak moc neplyne, co přinesla práce opravdu nového – jde víc o shrnutí původních zjištění, než že by tato zjištění byla jasně kontrastována s novými poznatky (zvláště zmatený mi přijde konec prvního odstavce na str. 20; tam mi totiž není jasné, co je ten skutečný stav a co artefakt podle jmenovaných autorů). Doporučoval bych proto celou sekci před odesláním do časopisu (předpokládám, že to je v plánu) přepsat. Další připomínky a dotazy už jsou spíše minoritní:

Str. 4 uprostřed – ten „Fusco-Cronk“ index mi přijde dost nešikovně vysvětlený. Možná by bylo lepší jej nejdřív napsat a pak vysvětlit, co jednotlivé proměnné znamenají – a nakonec třeba zmínit tu interpretaci (která už je jasná).

Str. 5, 10. řádek odspodu – nemá v tom vzorečku být čtyřka místo dvojky? Jde přece o celkový počet větví.

Str. 6 dole – proč vlastně byly využity stromečky jen pro zelené rostliny, živočichy a houby? Je to tím, že stromečků pro protisty je málo, nebo je málo dobrých stromů, anebo je to proto, že se vzhledem k problematičnosti druhového konceptu nebo k mezerám ve znalostech diverzity protistů nedá stanovit kompletnost těch stromů?

Str. 7 nahoře – proč byly „prořezány“ všechny poddruhy jen když tvořily monofyletický taxon? Mám to chápat tak, že když jej netvořily, šlo vlastně ve skutečnosti o nezávislé druhy?

Str. 9, konec druhého odstavce – proč byla imbalance stanovená ve třech „reprezentativních“ stromečkách; k čemu to sloužilo?

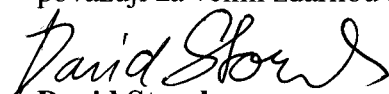
Str. 11, Tabulka 1 – proč je zajímavé frekvenční rozložení imbalance, respektive co plyne z toho, když zjistíme, že je nebo není normální? S touto informací se snad dále nepracuje, nebo snad ano?

Str. 16 dole – bylo by dobré uvést, v jakém směru se liší jednotlivé taxony, tj. u kterých je vyšší a kde nižší imbalance (tedy nejen, že se významně lišily).

Str. 21 uprostřed – potřeboval bych vysvětlit, jaktože náhodně generované stromy jsou často „imbalanced“. Měl jsem za to, že ta imbalance by měla vyjadřovat odchylku právě oproti „náhodnému“ stromečku. Jak to přesně je?

Str. 36-38 – tabulky musí mít legendu, i když jsou v appendixu.

Všechny tyto komentáře píšu spíš jako náměty před odesláním do redakce nějakého časopisu, práci považuji za velmi zdařilou a doporučuji její kladné ohodnocení.



David Storch

Centrum pro teoretická studia a Katedra ekologie PřF UK, Univerzita Karlova