

V. Brožová: Revision of Juncaceae and Cyperaceae phylogeny based on cpDNA and nDNA

Diplomová práce V. Brožové byla pravděpodobně koncipována jako rukopis vědeckého článku; je psána v angličtině a její těžiště spočívá v podrobné diskusi výsledků, zatímco úvodní rešeršní část a metodické kapitoly jsou poměrově stručnější než ve standardní DP. Většina vlastních výsledků fylogenetických analýz je k dispozici pouze v elektronické formě na přiloženém CD, stejně jako soupis použitých sekvencí (bohužel i zde bez bližších údajů o původu dokladových položek). Při podrobnějším ponoření se do textu ovšem brzy vyjde najevo, že text je zatím ve fázi spíše přípravného materiálu ke kýženému rukopisu, který bude muset být docela významně zestručněn a zprehledněn, aby nově zjištěné výsledky vůbec vynikly, o jazykové revizi nemluvě.

Autorka se ovšem pustila do opravdu neobvykle ambiciózního projektu – revize fylogenetických vztahů obrovské, byť zřejmě fylogeneticky koherentní skupiny zahrnující šáchorovité a sítinovité, tedy skupiny o zhruba 6 000 druzích. Z textu je zřejmé, že zde navazuje na předchozí práce své školitelky, respektive spolupracuje s ní na současných projektech včetně tohoto, konkrétní podíl však v práci vymezen není. Práce byla provedena na základě fylogenetické analýzy sekvenčních dat ze tří lokusů (2 chloroplastové + jaderná ITS), přičemž hrubá data byla ze značné části kompilována z veřejně dostupných zdrojů (GenBank) a doplněna o nově získané sekvence z necelých 50 druhů. Kompilaci z externích zdrojů samozřejmě není možné se v rozsáhlejších fylogenetických pracích vyhnout, navíc lze předpokládat, že řada sekvencí, které vystupují v soupisu jako externí zdroje, pochází z předchozích prací školitelky a kolektivu [bylo by asi vhodné tato data v datasetu označit], a je tedy možné ručit za jejich spolehlivost. Přesto mám obecně pocit, že víra autorky „As the volume of freely accessible molecular data is still increasing, analysis of these data is easy [sic] obtainable and is promising to resolve relationships...” je možná až příliš optimistická. Bližší pohled na dataset (Suppl. 1) totiž odhaluje, že data jsou dosti děravá: z asi 1800 sekvencí patřících 1168 taxonům (některé na infraspecifické úrovni) jich pouze 130 má data od všech lokusů. K tomu je potřeba připočíst skutečnost, že ani u druhů, u nichž jsou k dispozici data ze všech úseků není jasné, zda se jedná o izoláty z jedné položky, nebo zda je provenience vzorků různá; to může vést ke zkreslujícím informacím. Velmi bych se přimlouval za to, aby ve výsledném rukopisu informace o provenienci položek k jednotlivým sekvencím nechyběla. V případě ITS navíc není jasné, v kterých případech se jedná o celý úsek, a kdy pouze o dílčí úseky ITS1, resp. 2, popřípadě jestli dílčí úseky opět pocházejí z jednoho izolátu či byly zkombinovány různé. Paradoxně nejpodrobnější informaci dostáváme o vzorcích, z nichž se sekvence nepodařilo získat (Tab. 3).

Bohužel, vlastní výsledkovou část je možné komentovat jen velmi omezeně: jsou prezentovány pouze fylogenetické stromy (a to ještě pouze v příloze), a na těchto stromech nejsou čitelné podpory (zřejmě omyl při volbě jednotného černého pozadí). Popisy se ve značné míře koncentrují na rozdíly mezi jednotlivými typy analýz na datasetech, místo aby přehledně sumarizovaly rozdíly získané z různých datasetů. Diskuse je velmi rozsáhlá a pečlivě porovnává zjištěné vztahy s předchozími pracemi, ale k rozporným zjištěním, ať už mezi vlastními dílčími analýzami nebo s výsledky jiných autorů se příliš analyticky nevyjadřuje. Čtenář nepříliš orientovaný v konkrétní problematice tak hledá jádro novinek jen velmi obtížně. Chybí mi úvahy nad reprezentativností zvoleného výběru a jeho přínosu oproti předchozím studiím, nad kombinací datasetů (domnívám se například, že chloroplastová data z jediného lokusu jsou obecně povětšinou nedostatečná a prezentované úvahy nad vhodností konkrétního lokusu k určitému typu analýzy mi přijdou spíše irelevantní) a zhodnocení rozdílů ve fylogenetickém signálu z chloroplastových a jaderných dat a jejich příčin. Návrhy na alternativní klasifikaci nejsou dotaženy do reálných taxonomických návrhů s výjimkou alternativního tribálního členění šáchorovitých a rovněž nejsou důkladněji diskutovány.

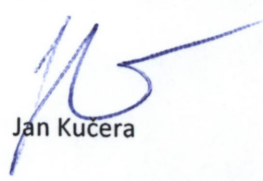
Přesto jsem přesvědčený, že autorka odvedla obrovský kus dobré práce, a po dotažení věřím, že svá zjištění spolu se školitelkou dotáhne k slušné a citované publikaci. Dílčích připomínek a otázek k diskusi by bylo možné uvést o dost víc, než je časový rámec obhajoby, proto položím jen některé:

- Proč v případě nově získaných sekvencí byl amplifikován jen jediný lokus (zpravidla trnL-F)?
- Podle jakého klíče byly „příliš krátké sekvence“ z datasetů odstraněny (osobně mi přijdou 269 bp dlouhé sekvence kompletní ITS stále velmi krátké, podobně tak 68bp pro ITS1 nebo 120bp pro ITS2)?
- Proč nebyla spojena data z chloroplastových úseků pro celkové porovnání s jaderným signálem?
- Jak byla zkonstruována matice všech úseků o 592 položkách, když žádný z dílčích datasetů jich tolik neobsahoval (kolik chybějících dat bylo tolerováno a z kterých úseků)? Neovlivnily chybějící úseky výsledek?

Z obecných připomínek bych asi zmínil, že v podobné práci by bylo vhodné uvádět autorské zkratky taxonů při první zmínce; v úvodní části nejsou uvedeny popisy některých tribů; popisy nastavení ML a BI analýz nejsou dostatečné, není uvedeno, jak byly definovány ‚partitions‘; pozor na nevysvětlené používání zkratk v textu, např. SHC, LBA; mnohé obrázky v textu jsou nečitelné...

Práci i přes uvedené výtky jednoznačně doporučuji k obhajobě s hodnocením v závislosti na průběhu obhajoby mezi stupni výborně a velmi dobře.

V Českých Budějovicích, 13.5.2018



Jan Kučera

Oponentský posudek na diplomovou práci Viktorie Brožové "Revision of Juncaceae and Cyperaceae phylogeny based on cpDNA and nrDNA"

Předložená diplomová práce v rozsahu necelých 50 stran (bez citací a příloh) se zabývá rekonstrukcí molekulární fylogeneze dvou sesterských čeledí. Autorka shromáždila velké množství sekvencí ze dvou chloroplastových (*rbcL*, *trnL-trnF*) a jednoho jaderného (nrDNA ITS) úseku DNA. Většina sekvencí pocházela z databáze GenBank, 69 sekvencí bylo vlastních. Na základě alignmentů byly provedeny rekonstrukce fylogenetických stromů pomocí tří různých metod – maximální parsimonie, maximum likelihood a Bayesovské analýzy. Jednotlivé úseky byly analyzovány nejprve zvlášť a poté i dohromady.

V úvodu se autorka věnuje relativně hodně morfologickým znakům, které definují jednotlivé skupiny. Chybí ale nějaké obecné představení morfologie studovaných čeledí (včetně obrázků), pojednání o tom, které znaky jsou taxonomicky důležité a srovnání stavů vybraných znaků v popisovaných čeledích (formou tabulky nebo i obrázků).

Na konci úvodu jsou celkem vágně a částečně i nejasně formulovány otázky, ze kterých vyplývá, že hlavním cílem práce je rekonstrukce molekulární fylogeneze čeledi a návrh možných taxonomických změn tak, aby výsledné taxony byly monofyletické.

Práce je jednoznačně rozčleněna na základní kapitoly obvyklé ve vědeckém textu, na konci nechybí stručné shrnutí. K práci mám následující připomínky a dotazy:

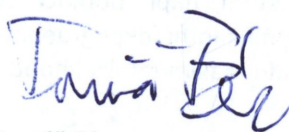
1. Celá práce je psána anglicky a celkem srozumitelně, což je třeba ocenit. Jen občas je text příliš komplikovaný a místy jsou používány konstrukce nebo slova, která znesnadňují pochopení textů. Například je opakovaně zaměňováno slovo „basis“ a „base“. Nesprávně je používáno slovní spojení „on the basis of the tree“, kdy autorka evidentně myslí „na bázi stromu“, tj. „at the base of the tree“.
2. Přehled starších i recentních fylogenetických klasifikací obou studovaných čeledí je představen v Tabulce 1, která je však do značné míry chaotická když si buňky (tribus) v jednotlivých řádcích neodpovídají.
3. Alignment – text popisující jednotlivé matice, počty vzorků a jejich délky by bylo lépe shrnout do tabulky. Také zde chybí vysvětlení, co jsou to tzv. „extended“ matice, byť je tento pojem dále často používán. Dále je uvedeno, že sekvence, které byly příliš krátké, byly z alignmentu odstraněny; není však napsáno, jak byly „krátké“ sekvence definovány. Při pohledu na některé alignmenty je zřejmé, že některé sekvence jsou kratší, tj. v alignmentu je určité množství chybějících dat. Je někde uvedeno procento chybějících dat pro jednotlivé taxony/matice? Nemůžou tato chybějící data ovlivnit analýzy, resp. nemají právě některé „problematicky umístěné“ taxony větší množství chybějících dat?
4. Redukce alignmentu – při pohledu na alignment nekódující DNA (tj. *trnL-trnF* a ITS) s velkým množstvím „gapů“ se nabízí otázka, zda by nebylo vhodné nejednoznačně alignované části automaticky odstranit, např. pomocí algoritmu Gblocks (<http://www.esi.umontreal.ca/Gblocks/>). Přemýšlela autorka o tom, jak výrazně může nejednoznačný alignment ovlivnit výsledné analýzy?
5. Analýza ITS – jaký je důvod pro oddělenou analýzu ITS1 a ITS2?
6. Analýza spojených matic – bylo nějakým způsobem testováno, zda je signál mezi jednotlivými maticemi (obzvláště mezi cpDNA a nrDNA) kongruentní, např. pomocí ILD testu? Co nám takový test může říct a jak postupovat v případě signifikantní inkongruence? Bylo provedeno rozdělení matic do jednotlivých partitions (v případě maximum likelihood a Bayesovské analýzy)?
7. Fylogenetické analýzy – některé části metodiky nejsou úplně jednoznačně a detailně popsány. Jaký je rozdíl mezi provedeným bootstrappingem a jackknifingem? Jackknifing čeho

byl prováděn? Jsou ještě nějaké další metody pro definování statistické podpory větví? Kolik jednotlivých hledání (ML searches) bylo provedeno při analýze v programu RAxML? Byl nějaký důvod k použití starší verze programu (8.2.4.)?

8. Bayesian inference – jak byla testována konvergence MCMC řetězců, pomocí nějaké statistiky?
9. Výsledky analýz pro jednotlivé matice – opět by zde pomohlo shrnutí do tabulky (počet sekvencí, délka stromu/likelihood, variabilní a parsimonicky informativní místa atd.)
10. Sekce Results zcela postrádá obrázky, tj. výsledné fylogenetické stromy, čtenář je odkazován do Supplementu, který ovšem existuje pouze v elektronické podobě. Není tedy možné jednoduše sledovat to, co je popisováno. Naopak Diskuze obsahuje množství částí fylogenetických stromů (což je do určité míry ilustrativní a mělo být už součástí Výsledků), nicméně u většiny z nich jsou popisky jednotlivých taxonů špatně čitelné (malé písmo) a pokud jsou některé části stromu obarvené, není v popisku uvedeno, o co konkrétně se jedná. Diskuze také někdy popisuje další vztahy na nižší úrovni, které už na obrázcích nejsou. Bohužel v tomto případě chybí jednoznačný odkaz na strom/stromy, ke kterému se text vztahuje. Použití jedné barvy na celý strom není příliš vhodné, spíše by se hodilo označit různými barvami jednotlivé skupiny, rody nebo druhy, které jsou dále diskutovány v textu. V textu (ve Výsledcích) by měl být alespoň jeden výsledný fylogenetický strom přímo mezi textem (stejně jako v článku také nebude vše jen v Supplementu).
11. Statistická podpora větví – text práce při popisu a interpretaci zjištěných vztahů téměř zcela ignoruje statistické podpory větví ve fylogenetických stromech. Podpora je místy komentována pouze slovně (jako např. *strong*, *well*, *high*, *very strong*, *significant* nebo naopak *poor*, *low*, *unsupported*, *insignificant*, *no support*), aniž by někde bylo uvedeno, co se těmito „termíny“ myslí. Jaká je (obecně přijímaná) interpretace statistické podpory větví fylogenetických stromů? Jaký je rozdíl mezi bootstrapovými podporami a *posterior probabilities* z Bayesovské analýzy?
12. Používání zkratk – určitě by bylo vhodné udělat někde na počátku práce seznam nejčastěji používaných zkratk. Občas jsou dokonce použity zkratky, které nejsou nikde vysvětleny, např. SHC.
13. V rámci diskuze o ITS je zmíněn termín *incomplete lineage sorting*? Co tento termín znamená a jak souvisí právě s ITS.
14. Jedním z výsledků práce je také návrh taxonomické klasifikace, např. v čeledi Juncaceae jsou navržena dvě řešení pro rody z jižní polokoule: buď budou zahrnuty do rodu *Juncus*, nebo se rod *Juncus* rozpadne na více rodů. Které řešení má větší podporu v morfologických datech? Jsou zde nějaké argumenty pro zachování parafyletických taxonů?

Celkově práce působí kompaktně, studentka provedla řadu fylogenetických analýz, jejichž výsledky popsala a diskutovala v kontextu většiny recentních prací. Přes výše uvedené výhrady práci hodnotím kladně a navrhuji klasifikaci stupněm velmi dobře.

V Kolíně dne 14.5.2018



Tomáš Fér

Katedra botaniky, PřF UK, Praha