

Posudek na magisterskou diplomovou práci Evy Novákové „*Arsenophonus*, an emerging clade of intracellular symbionts with a broad host distribution“

Ve své magisterské diplomové práci se autorka zabývá fylogenezí bakteriálních symbiontů hmyzu rodu *Arsenophonus*. Tyto γ -proteobakterie jsou díky širokému hostitelskému spektru a veliké diverzitě vztahů k hostiteli zřejmě na nejlepší cestě stát se jednou z modelových skupin ve výzkumu symbióz hmyzu. Práce je psána srozumitelnou angličtinou a občasné překlepy či gramatické chyby (např. několik různých podob slova Enterobacteriaceae, *Pseudomonas aeroginosa* místo *aeruginosa*, nebo desetinné čárky místo teček) nenarušily můj celkově dobrý dojem z textu.

Práce je sepsána ve formě článku připraveného k zaslání do tisku. Z toho nutně vyplývá jistá stručnost oproti „klasické“ formě diplomových prací. To jsem si uvědomoval především v sekci Úvod (3 strany) shrnující dosavadní znalosti o rodu *Arsenophonus*, kde bych jako nespecialista v daném oboru uvítal podrobnější informace i o problematice hmyzo-bakteriálních symbióz obecně.

Metodická část zahrnuje 4 strany. Autorka si zjevně osvojila základní metody molekulární biologie a fylogenetiky, jako jsou izolace, klonování a sekvenace DNA, tvorba alignmentu a práce s ním, tvorba fylogenetických stromů. Zvláště oceňuji její schopnost neztratit hlavu u analýzy rozsáhlých datasetů. K metodice mám několik dotazů:

1. V textu postrádám informaci o tom, zda byla DNA izolována z celého hmyzího jedince či pouze z některých jeho částí. Především v jakých orgánech se *Arsenophonus* vyskytuje?
2. Byl GC obsah vypočítáván z celých sekvencí nebo z ořezaných alignmentů?
3. Jakým způsobem a proč byl zvolen substituční model GTR + Γ + I?
4. Proč nebyla provedena bootstrapová analýza? Bylo to pouze z důvodu nadměrné velikosti datasetů? A proč nejsou ani v textu ani v obrázcích uvedeny hodnoty Bayesian posterior probability? Je to z důvodu, že jim autorka nedůvěřuje?
5. Nebylo by lepší označovat symbionty jinak než pouze jmény jejich hostitelských druhů? Takto se v jednom stromu vedle sebe objevují *Pseudomonas aeruginosa* a *Lipoptena cervi*.
6. Jednalo se u rodu *Lipoptena* skutečně o druh *L. cervi*? Pokud si dobře vzpomínám, Dr. J. Votýpka nedávno zjistil, že naprostá většina zástupců tohoto rodu v jeho sbírce patří do druhu *L. fortisetosa*.

Sekce Výsledky se skládá ze dvou stran textu a řady fylogenetických stromů. Autorka získala 34 sekvencí SSU rDNA rodu *Arsenophonus* z cca 13 druhů klošů, muchulí, mravenců a mšic. Ze získaných sekvencí a z příbuzných sekvencí z GenBanku vytvořila 6 alignmentů lišících se ve složení sekvencí a v míře „ořezání“. Tyto datasety poté analyzovala.

Mám následující dotaz:

7. Jaká byla prevalence *Arsenophonus* u studovaného hmyzu? Jaký byl celkový soubor zkoumaného hmyzu?

Diskuse (7 stran) je čtivá a srozumitelná. Autorka pomocí výsledků svých fylogenetických analýz věcně diskutuje tak rozmanité problémy, jako je význam „taxon samplingu“, míru koevoluce zástupců rodu *Arsenophonus* s jejich hostiteli, míru laterálního genového transferu genu pro SSU rRNA a význam intragenového polymorfismu v tomto genu. K diskusi mám následující dotazy:

8. Autorka konstatuje, že po odstranění nejednoznačně alignovatelných pozic z alignmentu dojde ke ztrátě podstatné části fylogenetické informace. Dochází k logickému závěru, že tyto pozice jsou v analýze důležité. Uvážila však možnost, že tyto pozice mohou vinou své „nezalignovatelnosti“ vnášet artefakty do fylogenetické analýzy?
9. Na str. 12 je uvedeno, že v případě sekvence z *Trichobius yunkerii* jde nejspíše o chiméru. Jedná se spíše o artefakt vzniklý během PCR nebo klonování, anebo o výsledek rekombinace mezi dvěma nepříbuznými molekulami?
10. Zjišťovala autorka, co se stane s topologií fylogenetického stromu, odstraní-li z alignmentu 41-bázový úsek, jehož evoluce je odlišná od evoluce zbytku sekvence?

K práci jako celku mám následující dotazy:

11. Je známo něco více o zástupcích rodu *Arsenophonus* z rostlin (např. jestli tam mohli být zaneseny hmyzem nebo nejedná-li se o krátkodobé infekce)?
12. *Riesia* a *Phlomobacter* jsou nejspíše synonyma *Arsenophonus*. Na základě čeho byly tyto rody popsány? Pouze na základě sekvencí nebo i na základě nějakých jiných znaků?
13. Co si autorka myslí o možnostech reálného odhadu doby divergence mezi rody *Escherichia* a *Salmonella* (anebo o odhadu doby divergence mezi jakýmkoli bakteriemi, kdy nemůžeme použít ani fosilní, ani fylogeografická data, ani kospeciace s hostiteli se známou dobou divergence)?

Na závěr konstatuji, že předloženou práci Evy Novákové považuji za kvalitní a že podle mého úsudku splňuje formální podmínky kladené na magisterské diplomové práce na PřF JU. Předpokládám, že po nevýznamných úpravách bude možno práci odeslat do některého mezinárodního mikrobiologického periodika. Hodnotící komisi tedy doporučuji, aby byla obhájena s kladným hodnocením.

RNDr. Ivan Čepička, Ph.D.
Katedra zoologie PřF UK
Viničná 7
128 44 Praha 2

Posudek na magisterskou práci Evy Novákové „*Arsenophonus*, an emerging clade of intracellular symbionts with a broad host distribution“

Vypracoval Mgr. Oldřich Říčan, PhD

Magisterská práce Evy Novákové je sepsána velmi dobrou angličtinou a je i formou a obsahem evidentně směřována k publikaci, což je nejvyšším cílem magisterské práce.

Práce přináší 34 nových sekvencí symbiontů z 15 hmyzích taxonů a tato nová data kombinuje s rozsáhlým datovým souborem studovaného taxonu z publikovaných prací v kombinaci s potenciálními outgroupovými taxony.

Otázka / komentář: Přijde mi zvláštní (v úvodu) uvádět, že *Arsenophonus* definují 4 velmi odlišné vlastnosti (jejich kombinace), které nejsou přítomny u všech zástupců. Definují tedy asi jen jednotlivé podskupiny? Nebo jak to přesně je s distribucí oněch 4 znaků?

Pozitivní je celý přístup ke studovanému problému, od formy spisu, přes analytický aparát po diskuzi výsledků. Analytický postup použitý autorkou je ve všech směrech velmi příkladný a krásně umožňuje oponentovi nahlédnout pod pokličku (což je chvályhodné), ale předpokládám, že pro publikaci by bylo vhodné ho jistým způsobem usměrnit, jinak lze očekávat více všetečných otázek než je žádoucích (viz níže).

Velmi vhodně byl zvolen přístup k alignmentu, který je pro fylogenetické studie zcela zásadní. Bylo vytvořeno několik různých matic, z nichž *Basic matrix* měla všechny pozice a celý taxon sampling. Alignment zjevně nebyl jednoduchý, když délkové rozdíly PCR-amplifikovaných fragmentů jsou velmi výrazné (632 – 1198 bp).

Otázka / komentář: Proč obr. 2 (největší strom, *Basic matrix*) a i další stromy postrádají podpory uzlů? Přitom větev vedoucí k monofyletickému *Arsenophonus* ve poměrně dlouhá, tj. pravděpodobně robustní. Je příčinou chybění podpor uzlů velikost matice a tím pádem výpočetní náročnost? Nebo autorka podpory nezahrnula poté co podpora pro *Arsenophonus* nebyla přesvědčivá? Nebývá totiž příliš zvykem publikovat ML stromy bez podpor (u MP se to dělá, ale také pouze u některých frakcí jako např. v časopise *Cladistics*).

Conservative matrix opět obsahovala celý taxon sampling, ale obtížně alignovatelné pozice byly vyloučeny (ne od oka, ale opakovatelnou metodou). Tento přístup je z mé zkušenosti skoro vždy velmi radikální, a i zde vedl k redukci na pouhých 284 bp.

Uváděný velmi vysoký stupeň polytomie obr. 3 je tedy jasný.

Otázka / komentář: Je podle mého názoru zavádějící srovnávat různé alignmenty (obr. 2 a 3), když jsou analyzovány jinými fylogenetickými metodami (ML u obr.2 a MP u obr. 3). ML je alespoň z mé zkušenosti metodou dávající vyšší rozlišení oproti MP, čímž by se efekt uváděný v práci ještě dále prohluboval. Zde bych pro publikaci viděl problém, který asi označí většina oponentů (nemonofylie *Arsenophonus* v obr. 3 může

být tímtež efektem, stromy ze stejných matic se v jednotlivých metodách zpracování často rozcházejí).

Velmi pozitivní je analýza několika (5ti) fylogenetických matic s redukováným taxon samplingem za účelem zjištění vlivu taxon samplingu na robustnost fylogenetických hypotéz. Podle očekávání (v souhrě s relativně nízkým počtem znaků) se i efekt redukováných matic co do počtu a skladby taxonů projevilo sníženým fylogenetickým rozlišením. Autorka v diskuzi správně uvádí, že se velmi pravděpodobně jedná o metodologický artefakt, navíc odhaluje i artefakt LBA a v souladu s předpoklady ho nachází mezi topologiemi MP analýz. Autorka uvádí, že účelem těchto taxonomicky okleštěných analýz byla simulace neúplného samplingu (který ani teoreticky nemůže být kompletní), což je velmi vhodný přístup. V této práci bohužel tato analýza koliduje s efektem malého počtu znaků a je podle mého názoru těžko interpretovatelná.

V metodice jsou popsány parametry Bayesovské analýzy a i to, že nebylo výpočetně snadné je provést. Ale ve výsledcích ani v dalších částech práce se stromy z MrBayese vůbec nevyskytují. Bayesovské analýzy se pro velké matice dají běhat na clusterech (např. volně přístupný Cornell).

Otázka / komentář: Proč se výsledky Bayesovských analýz v práci nevyskytují? Výběr kladogramů mi mimo obr. 2 přijde spíše náhodný. Nebo je v tom záměr (ukazují co se hodilo?) Nejsou v nich zobrazeny ani stromy z MrBayese, ani z metody T92 a převažují MP analýzy, které jsou asi pro molekulární analýzy nejméně vhodné.

Clock matrix má podle mého názoru a vzhledem k vysoké diverzitě studované skupiny příliš malý taxon sampling. Kalibrace nejsou popsány příliš detailně a neumožňují mi nahlédnout příčinu vysoké disproporce v časových odhadech (navíc jsou tu další komplikace, viz. níže).

V diskuzi mi něco chybí, přijde mi malinko neučesaná a ne zcela kulminující, a zároveň jsem na pochybách zda to není jen mojí neznalostí.

V této souvislosti by mě zajímalo, jaká je korespondence mezi délkou sekvenovaných fragmentů a délkou větví v kladogramu (obr. 2). Mou otázkou tedy je zda jsou dlouhovětevné taxony a krátkovětevné taxony odlišené mutační rychlostí, nebo delecemi (kdy by tedy platilo, že krátké sekvence jsou ty deleční). Jak jsem se dočetl, jedná se v případě dlouho- a krátkovětevných taxonů o to čemu se říká P- a S-symbionti, kteří se mají lišit způsobem svého života co se koevoluce týče. Krátkovětevné taxony jsou tedy ty, kde není hostitelská specifita a koevoluce (což jasně plyne z obr. 2, kde jsou všechny koevoluce pouze mezi dlouhovětevnými taxony – předpokládám, že se jedná o P-symbionty)? Dovedl bych si představit jak s délkou větví (tedy jejich homogenizací) souvisí přenos genových fragmentů (LGT), což je v práci také diskutováno a byla pro to nalezena podpora v alignmentu. Je to tedy tak, že všechny krátkovětevné taxony jsou co se týče délky větví takto homogenizovány? Nebo je to několik odlišných procesů (příčin) majících tentýž efekt? Na druhou stranu se ale v úvodu píše, že P-symbionti, tedy ti koevoluční, jsou těmi u koho dochází k degeneraci genomu (a poklesu obsahu GC) (což by byla obrácená interpretace než ta v předchozí situaci).

Autorka v této souvislosti uvádí, že nespecifické taxony jsou ty, které přeskakují mezi hostiteli. Nemyslím si ale, že tato tautologie vysvětluje odlišné délky větví ani to, že krátkovětevné taxony jsou ty, které nejsou hostitelsky specifické. Klastrování dlouhých linií do monofyletické linie je vždy podezřelé. Toto v kombinaci s problémem paralogie, který dále může situaci rozostřovat a s přičtením horizontálních přesunů genových fragmentů (což je oboje v práci diskutováno a podpořeno), se nelze divit, že jak časový odhad rámce fylogeneze tak biologické interpretace jsou zatím jen velmi těžce studovatelné a interpretovatelné. (Jelikož nejsem v problematice symbiontů a koevolucí příliš zběhlý, může většina předchozího odstavce být pouze mým nedorozuměním).

Otázka/komentář: Jaké je autorčino rozuzlení korespondence délky větví, délky alignmentů a hostitelské specifity (koevolucí)?

Ve shodě s autorkou se domnívám, že pouze mnohem větší fylogeneze jak po stránce taxonomické tak znakové má potenciál dobrat se k biologickému smyslu věci. Klady práce ale jasně převažují nad neúplností biologických interpretací. Kvalitu předkládané práce hodnotím velmi vysoko, s vědomím toho, že se jedná v podstatě o explorační fázi celého velkého a velmi složitého biologického fenoménu. Autorka do práce musela vložit velké úsilí doprovázené jistě notnou dávkou štěstí. Navrhuji hodnotit práci stupněm „výborně“.

V Českých Budějovicích
21. 5. 2008
Oldřich Říčan

