

Oponentský posudek na bakalářskou práci Filipa Husníka: Molecular phylogeny of intracellular symbiotic Gammaproteobacteria in insects.

Předložená práce je klasicky členěná, rozsahem i obsahem pak výrazně přesahuje požadavky JU na bakalářské práce. Úvodní kapitola vyčerpávajícím způsobem shrnuje poznatky o bakteriálních symbiontech hmyzu včetně charakteristik jejich redukovaných genomů a jejich taxonomické distribuci. Velkou pozornost věnoval diplomant metodám fylogenetické analýzy, použití fylogenomiky při studiu evolučních vztahů mezi symbiotickými bakteriemi hmyzu, možným fylogenetickým artefaktům a též způsobům, jak se s nimi vypořádat. V kapitole věnované HGT postrádám definici HGT, autor zcela opomněl čtenáři říci, co to vlastně HGT je a jaké typy lze v přírodě najít (endosymbiotický HGT, ne-endosymbiotický HGT). V případě některých bakteriálních symbiontů hmyzu, kteří se strukturou a velikostí genomu přibližují organelám, by takový přehled nebyl na škodu. V souvislosti s tím mám následující otázku: **Vzhledem k minimální velikosti genomu některých bakteriálních symbiontů hmyzu se nabízí otázka, zda již nejde o organely. Jak se liší symbiont od organely?** Zároveň mi v kapitole chybí obecný pohled na vliv HGT na evoluci bakterií; podrobně se této problematice věnoval např. James McInerney (např. McInerney JO et al., The prokaryotic tree of life: past, recent and future? Trends in Ecology and Evolution 23, 276-281.). Autor soudí, že použití fylogenomických přístupů, například odvození evolučních vztahů z pořadí genů v genomu, pomůže vyřešit fylogenezi bakteriálních symbiontů. **Může být pořadí genů u vysoce redukovaných bakteriálních genomů ovlivněno konvergencí?** Podobná situace nastane při použití katenovaných datasetů. Byl bych trochu skeptický k navrhovanému postupu analyzovat bud katenované geny nebo individuální geny. Analýza katenovaných genů bez souběžné analýzy jednotlivých genů mi přijde velmi riziková. Právě vzhledem k obecně vysoké míře HGT u bakterií by měl být nejdříve vyloučen HGT pomocí separátních analýz jednotlivých genů.

Metodika je v předložené práci zpracována pečlivě, ale některé části jsou trochu matoucí. Komplexní metodický postup od bioinformatické analýzy (design primerů) přes PCR, TGGE, až po velmi solidně popsanou fylogenetickou analýzu ale mohu jen ocenit. Z vlastní zkušenosti s datasety s různým aminokyselinovým složením bych ještě doporučil použití programu PhyloBayes či HN_PhyloBayes s pokročilými modely, jako jsou např. CAT či CATBP modely. Spolu s AsaturA s LG modelem mi daly nejlepší výsledky. Popis fylogenetické analýzy pomocí LogDet paralign distances je nedostatečný. **Popište přesně krok po kroku, jak jste v případě analýzy pomocí LogDet distancí postupoval.** Postrádám též informaci o sekvencích získaných z genové banky. Není také zcela jasné, zda byly datasety pro slow-fast získány pomocí programu Slowfaster nebo manuálně. Zároveň bych v případě slow-fast viděl raději stromy ze všech datasetů, zvláště pak v diplomové práci, která není nikterak omezena rozsahem případných příloh.

Výsledky jasně ukazují, kolik Filip Husník udělal práce. Z předložené práce ale není jasné, jestli se diplomant pokoušel získat jmenované proteinové a rRNA geny ze symbiontů ze všech studovaných hmyzích druhů, či jen z těch, ze kterých je nakonec dostal. Poněkud matoucí je tabulka 7 sumarizující použité alignmenty, protože obsahuje geny, které diplomant nesekvencoval (FusA, GapA, GidA, GyrB, PurB, RpoB) a v metodice nezmiňuje, že by byly sestaveny na základě sekvencí získaných z banky. V případě analýzy nukleotidového datasetu (Figure 3, page 26) interpretuje diplomant výslednou topologii jako ovlivněnou LBA. **Odhalení artefaktu je vždy ošemetná záležitost, protože tato topologie může být i pravdivá a extrémně odlišné sekvence mohou být klidně společného původu. Jak diplomant přišel na to, že se v tomto případě jedná o artefakt? Konkatenovaný dataset**

(Figure 5, page 29) zjevně LBA nevyřešil, ale diplomant jej jako ovlivněný LBA nekomentuje. Proč?

Diskuse je na bakalářského studenta solidní. Diplomant však diskutuje i výsledky, které neukázal, jako např. LogDet analýzu, či použití nehomogenního modelu u metody Maximum likelihood (Galtier and Gouy, 1998). Je to škoda, protože když hovoří o tom, že LogDet LBA nevyřešil (to, ostatně ani nemůže, protože je designován pro řešení nukleotidového biasu), mohu si jen představovat, jakou topologii asi dostal. Totéž se týká ostatních utajených stromů. Slow-fast analýza se ukázala jako nejlépe použitelná, a pokud mohu dle stromů soudit, na ní postavená alternativní hypotéza k monofýlii hmyzích symbiontů se mi jeví jako opodstatněná.

Ocenil bych dostupnost datasetů na CD. Jen těžko se hodnotí správnost provedených analýz nemáte-li k dispozici výchozí data. Zároveň postrádám seznam analyzovaných sekvencí a jejich accession numbers.

Shrnutí

Předložená bakalářská práce jednoznačně převyšuje požadavky na podobný typ prací na JU. Je napsaná srozumitelnou angličtinou s minimem překlepů a úctyhodným seznamem použité literatury. Přes jisté nedostatky v metodice a v popisu výsledků však musím tuto práci jednoznačně hodnotit jako výbornou a doporučit ji k obhajobě.

Miroslav Oborník
Biologické centrum AVČR
Parazitologický ústav
Branišovská 31
37005 České Budějovice

