

Posudek na diplomovou práci

Recentně aktivní L1 a B1 retrotransposony v myším genomu

Diplomant: Václav Janoušek

Oponent: Josef Bryja

Diplomová práce na 70 stranách textu s vmezeřenými tabulkami a grafy a na 27 stranách příloh se zabývá recentně aktivními retrotransposony v genomu myši domácí s hlavním důrazem na nalezení a popis inzerce ze dvou skupin tzv. non-LTR retroelementů: L1 z LINE a B1 ze SINE. Asi nejdůležitějším výsledkem práce je nalezení a detailní charakteristika 49 B1 inzerce a 116 L1 inzerce. Detailní analýza sekvencí tak umožnila pochopit vývoj těchto inzerce v myším genomu a upřesnila některé nesrovnalosti v jejich klasifikaci. Srovnání výskytu jednotlivých inzerce v několika inbredních kmenech odvozených z různých druhů a poddruhů rodu *Mus* zase vedlo k vytíování vhodných markerů pro studium populačně-genetické struktury a případně hybridizace modelové skupiny domácích myší. Student zvládl celou řadu bioinformatických i molekulárně-genetických přístupů a sepsal rozhodně významné dílo, jehož alespoň některé části si zaslouží publikování v odborném časopise typu *Mammalian Genome*. Popsané znaky odlišující jednotlivé poddruhy myší si zase jistě najdou cestu k analýzám introgrese v projektu studia speciace v oblasti hybridní zóny těchto poddruhů.

Celá práce je psána stručným srozumitelným stylem s minimem věcných chyb. Většina provedených analýz je zcela logická a jejich výsledky jsou smysluplně interpretovány. Abych jako recenzent mohl práci alespoň něco vytknout, tak jsem se musel zaměřit na některé formální prohřešky, které většinou ve svých posudcích vůbec neuvádím (viz níže). Zároveň bych měl několik dotazů na věci, které z textu práce nevyplývají úplně jednoznačně, ale silně podněcují mou zvědavost:

Nejprve tedy dotazy:

- 1) Proč byl pro hlavní část práce použit genom *M. m. molossinus*? Vzhledem k návaznosti na modelové studium evropské hybridní zóny bych očekával použití genomu odvozeného z *M. m. musculus*. Nebo takový není k dispozici?
- 2) Na str. 12 (např. Obr. 3, a posléze i ve Výsledcích a Diskuzi) jsou charakterizovány klasifikace retroelementů založené na substitucích v určitých diagnostických pozicích. Existuje u těchto inzerce možnost zpětné záměny na diagnostickém místě, a kam by se pak takovýto transposon řadil?
- 3) V práci bylo objeveno 49 B1 inzerce a 116 L1 inzerce, kterými se lišily genomy dvou poddruhů myší (oddělených před cca 500 000 let). Podle Tabulky 1 se celkem v genomu myší vyskytuje 564 000 kopií B1, což je dvojnásobně více než má potkan. Za cca 12 miliónů let (tj. rozdělení linie *Mus/Rattus*) tedy v myším genomu vzniklo 282 000 inzerce (tj. rychlost 23500 inzerce/1 milión let), přičemž za posledních 500 tisíc let (oddělení *M. m. domesticus* a *M. m. molossinus*) pouze 49 (tj. rychlost 98 inzerce/1 milión let). Kde je v mém výpočtu chyba, případně co jsem nepochopil správně?
- 4) Jak by se dalo jednoznačně zjistit, který ze scénářů na Obr. 12 je pravdivý? Šlo by to nějakým větším samplingem, tj. srovnáním více osekvenovaných genomů?

5) V databázi RepBase je 10 B1 podrodin a v diplomové práci byly nalezeny pouze 4 z nich. Dá se tedy jednoznačně říci, že 6 nenalezených podrodin tvoří inserce, které za posledních cca 500 000 let nevyprodukovaly žádnou kopii?

69) Ad str. 61, ř. 1-2: Ví se něco o aktivitě LTR retroelementů v myším genomu? Jsou v současnosti aktivní nebo ne?

Formální připomínky (prosím odpovídat jen na tučně zvýrazněné)

- pro délky jednotlivých fragmentů DNA je uváděna buď zkratka "bp" nebo "nt". Mělo by to asi být v celé práci jednotně.

- Tabulka 1 obsahuje několik nepřesností:

- proč je v ní uvedeno "Alu", když to je transposon, který se u myši nevyskytuje
- poslední sloupec by bylo přehlednější uvést ve formě % z předposledního sloupce. Tj. např. u B1 by bylo uvedeno v posledním sloupci 100%. V textu se pak na tato procenta odkazuje (např. druhý řádek nad tabulkou)
- str. 13, 1. řádek - číslo 660 000 kopií odpovídá všem LINEs, nikoliv L1
- u SINEs chybí kategorie MIR/MIR3, které se u myši taky vyskytují (viz. str. 9, 3. odst.)

- str. 19, 1. ř. - opět jsou uvedeny jenom 4 skupiny SINE u myši - jak to tedy je s MIR/MIR3 insercemi - vyskytují se u myši nebo ne?

- str. 20, Obr. 7 - jednotlivé zkratky nejsou vysvětleny

- str. 20, 5. ř. - jména linií (např. Hystriognathni ...) buď psát počestně nebo odborným názvem

- často se objevují anglikanismy - většinou mně nevadí, ale některé zde přímo bijí do očí (varíruje, pattern, defoltní, biasem)

- v práci se často objevují nevhodné odkazy typu "..., jak jsem již uváděl v první kapitole,..." které se do vědeckého textu příliš nehodí

- str. 23, obr. 8 - u obrázku jsou dva odkazy a každý má jiné datování

- str. 25, 4. ř. od konce - není mně jasný mechanismus excise inserce pomocí TSD - ví se o tom mechanismu něco více?

- str. 26: Cíle práce mohly být definovány přesněji a detailněji, tj. ke každému z těchto stručných cílů mohl být ještě krátký odstavček o tom, proč je to důležité

- str. 27 - na prvním řádku je definován "cíl", který není mezi "Cíle práce" na předchozí straně

- str. 41 - co znamená, že "Většina nalezených recentních L1 je kratších než 1/10 z délky konsensu."

- str. 45, Obr. 17 - co nám říká tento obrázek navíc oproti Network analýze? Tj. proč byla prováděna "mismatch distribution", která se v klasické fylogeografii používá na testování demografických modelů?

- Tabulka 1 v Příloze 4: z důvodu přehlednosti by bylo dobré uvést záhlaví tabulky na každé straně

Závěr: Diplomant odvedl při zpracování zadaného tématu obrovské množství práce a své výsledky dokázal převést do formy kvalitní diplomové práce. Ta téměř neobsahuje věcné chyby a jen minimum chyb formálních. Z tohoto důvodu ji **jednoznačně doporučuji k obhajobě.**

Ve Studenci 17.9.2010

Doc. Mgr. et Mgr. Josef Bryja, Ph.D.