

## Oponentský posudek na diplomovou práci

**Autor:** Bc. Jan Röslein

**Název:** Mutační a substituční tempo u sexuálních a klonálních forem: možný klíč k vysvětlení persistence sexu u modelové skupiny sekavců

Předložená diplomová práce se zabývá analýzou „next-generation“ sekvenčních dat z jater a oocytů několika druhů sekavců rodu *Cobitis* a jejich asexuálních hybridních linií. Data jsou využita k testování třech evolučně genetických otázek, kterými jsou (1) identifikace molekulárních mechanismů souvisejících s iniciací asexuality, (2) testování přítomnosti genomového imprintingu u hybridních linií a (3) porovnání rychlosti akumulace nesynonymních substitucí v genomu asexuálních hybridních linií a jejich sexuálních rodičovských druhů. Za tímto účelem byla provedena *de novo* assembly transkriptomu jater a oocytů, analýza genetického polymorfismu mezi zkoumanými jedinci a analýza diferenciální genové exprese mezi vybranými skupinami jedinců.

Práce je velmi rozsáhlá (zahrnuje celkem 119 stran) a zabývá se zajímavými a aktuálními tématy. Z výsledků je patrné, že autor provedl veliký kus práce a velmi kladně oceňuji, jak se dokázal ponořit do poměrně složitých bioinformatických procedur, které vyžaduje analýza „next-generation“ sekvenčních dat. Přesto mám k práci několik kritických výhrad.

Nemůžu jinak než souhlasit s autorovou sebekritickou poznámkou na samém konci souhrnu práce: „Myslím si, že moje práce patří spíše ke komplexnějším genomickým studiím, které na asexuálních organismech byly podniknuty; tomu také odpovídá její rozsah, v některých částech však možná na úkor její kvality“. Svou tématickou šíří i rozsahem provedených analýz by studie klidně vystačila i na dizertační práci. Tato šíře s sebou bohužel nese i to, že ne všechny otázky, které jsou v práci řešeny, jsou v úvodu probrány do dostatečné hloubky. Např. otázce genomického imprintingu je v úvodu věnován pouze jeden kratší odstavec. Úvod práce je tak trochu nevyvážený a kromě toho by vzhledem k šíři řešených témat asi stálo za to pozměnit samotný název práce. Současný název se týká prakticky jen třetí otázky.

Co se týče metod a výsledků práce. Velmi oceňuji práci, kterou si autor dal se základním zpracováním dat zahrnujícím assembly transkriptomu, anotaci získaných kontigů a mapování readů na assembly za účelem zjištění genetických polymorfismů mezi analyzovanými jedinci. Tomu odpovídá i poměrně rozsáhlý a detailní popis těchto metod v metodách. Také oceňuji, že si dal autor práci s validací získaných výsledků diferenciální exprese pomocí RT-qPCR. Nicméně, mám určité výhrady či připomínky k prezentaci některých dalších výsledků. Asi největší výhrady se týkají prezentace, jak sám autor píše, „ústředních výsledků této práce“, tj. otázky rychlosti akumulace nesynonymních (škodlivých) mutací v genomu asexuálních hybridních linií a jejich sexuálních rodičovských druhů. K prezentaci těchto stěžejních výsledků se autor dostal až na straně 89 a trochu se bohužel zdá, že mu již došel čas či energie k popsání těchto výsledků, neboť jsou uvedeny jen velice stručně (na jedné straně) a navíc se mi zdá, že prezentované výsledky neodpovídají tomu, co autor píše v abstraktu či diskuzi. V abstraktu i diskuzi práce autor píše, že „výsledky evoluce hybridních linií poukazují na zvýšenou akumulaci nesynonymních mutací v genomu hybridních linií“. Když se ovšem podívám na graf č. 48, který by měl ukazovat poměr dN/dS u rodičovských sexuálních druhů a hybridních asexuálních linií, zdá se mi, že dN/dS je naopak u většiny asexuálních linií nižší ve srovnání se sexuálními druhy. Mimochodem, u grafu chybí popis osy y a měřítko je trochu podivné. Tabulku č. 14 se statistikou porovnávající dN/dS mezi různými páry sexuálních druhů a asexuálních linií bych doporučila k přezkoumání, neboť je trochu podezřelá, že p-value vyjde mezi některými páry

přesně 1 a u zbylých je  $<2e-16$ . Ve srovnání s poměrně nízkou kvalitou závěru výsledků bych chtěla pochválit velmi pěknou diskuzi, ze které mimo jiné vyplývá, že si je autor většiny nedostatků práce plně vědom, což je určitě pozitivní.

Po formální stránce je práce na slušné úrovni, i když se v ní občas vyskytují překlepy, gramatické chyby, chyby ve formátu citací (ne však ve velké míře). Občas jsou popisky k obrázkům umístěny na další straně než samotný obrázek, což nepůsobí úplně dobře. Také názvy a popisky k delším skriptům a k tabulkám bych doporučila dát nad spíše než pod tabulku, jak je to v literatuře běžné. V textu se také občas vyskytují anglické termíny, které by stálo za to přeložit, neboť existují dobře zavedené české ekvivalenty. Asi nejvíce do očí bijící je používání termínu „loci“ (např. „s cílem najít loci, které mohou souviset s...“), který by asi bylo vhodnější přeložit jako lokusy.

Celkově, i přes uvedené výtky, považuji práci za velice kvalitní a zajímavou a věřím, že po dotažení některých analýz z práce vznikne hned několik pěkných publikací. Níže uvádím nějaké další komentáře a otázky k práci. Jelikož jich je celkem hodně, postačí, když autor při obhajobě bude reagovat jen ty tučně označené.

1. Autor by si měl ujasnit, co znamená termín fitness. Nelze říci, že „fitness klonální reprodukce je 2x větší nežli u sexuální reprodukce“ (str. 11), či „mutace mají nezávislou fitness“ (str. 12).
2. Str. 11. „možnost využití nadpočetné kopie DNA k reparaci HEJ, NHEJ.“ Není vůbec jasné, co je HEJ, NHEJ.
3. **Autor uvádí vedle známé hypotézy Mullerovy rohatky také méně známou hypotézu Kondrashovovy sekery (str. 12). Může nám autor vysvětlit, co přesně Kondrashovova sekera znamená?**
4. **Na str. 12 autor uvádí: „Asexuální linie mohou díky snížení efektivní velikosti populace (selekce na pozadí) a driftu přežít opravdu dlouhou dobu“. Tomuto tvrzení příliš nerozumím. Může nám autor vysvětlit, jak to myslí?**
5. **Na str. 13 autor píše, že „sexuální reprodukci můžeme označit za evoluční past, protože vedla k vývoji genomického imprintingu mezi pohlavími...“. Může autor říci, zda se genomový imprinting (myšleno jako přítomnost genové exprese pouze z jedné maternální či paternální alely) opravdu vyskytuje u všech sexuálně se množících organismů včetně ryb, kterými se zabývá? Existují i jiné formy genomového imprintingu než ten popsany výše, který se vysvětluje odlišnými zájmy mateřského a otcovského genomu? Mám pocit, že je v této problematice v práci trochu zmatek.**
6. Str. 17. Kapitola 1.3 je bohužel poněkud nesrozumitelná.
7. Str. 18-21. Je otázka, zda se kapitola o metodách studia transkriptomu do úvodu hodí. Osobně bych tuto problematiku vynechala nebo zmínila v metodách a v úvodu se soustředila jen na evoluční otázky. Stejně tak cíle práce (str. 22) bych omezila na tři hlavní evoluční otázky (3,4,5) a technické cíle (1,2) bych mezi ně nezahrnovala.
8. **Na str. 42 autor popisuje srovnání heterozygotnosti (tj. přítomnosti dvou nukleotidů v jedné pozici) u různých jedinců. Někteří z jedinců jsou ale triploidi. Předpokládám, že většina programů na detekci heterozygotních pozic pracuje s předpokladem, že studujeme diploidní organismy. Jak se autor vypořádal s tímto problémem u sekavců a jak detekoval heterozygotní pozice u triploidů?**
9. Překvapivý je velmi vysoký počet zjištěných stop kodonů v ORF. Nemůže to vysvětlit přítomnost sekvenačních chyb vedoucích k indel polymorfismům, které jsou časté zejména v 454 datech?

10. Na str. 64 autor píše, že dN/dS počítal pro všechny páry sekvencí. Nebylo by lepší využít nějaký vzdálenější outgroup, který by umožnil identifikaci více SNP?
- 11. Na str. 71 autor píše, že při srovnání diferenciální exprese mezi sexuálními a asexuálními sekavci našel 473 diferenciálně exprimovaných genů v oocytech a 821 v játrech. Jak si autor vysvětluje, že více rozdílů v expresi je v játrech než v oocytech, když oocyty by měly být asexualitou postiženy více?**
12. Grafy 32-35. Není mi příliš jasné, proč vzájemně klastrují jednou sexuální vs. asexuální druhy (linie) a podruhé diploidní vs. triploidní druhy (linie), podle toho mezi kterými skupinami byla komparativní analýza provedena. Nemělo by to vycházet podobně nezávisle na tom, které skupiny kontrastujeme? Navíc rodičovské druhy ee a tt spolu v Heatplotech klastrují většinou dohromady, přesto, že autor píše, že mezi nimi je nejvíce diferenciálně exprimovaných genů. Může to autor vysvětlit? Možná jsem jen ale nepochopila co Heatploty znázorňují.

RNDr. Radka Reifová, Ph.D.  
Katedra zoologie PřF UK  
E-mail: radka.reifova@natur.cuni.cz

23. 8. 2016, Bělá pod Bezdězem