



Oponentský posudok bakalárskej práce: Pavel Hronek

Cieľom práce Pavla Hronka bolo zostaviť genóm motýľa *Lycorea halia*, identifikovať pohlavne viazané sekvencie (s očakávaním, že bude identifikovaný neo-Z chromozóm) a testovať hypotézu o spoločnom evolučnom pôvode neo-Z chromozómu v rámci širšej skupiny motýľov. Jedná sa o bioinformatickú prácu s použitím štandardných programov, a tiež skriptov vytvorených inými členmi tímu. Celá práca má 30 strán, pričom 23 je samotná práca a zvyšok sú formálne časti a zoznam referencií. Členenie práce je štandardné a obsahuje všetky potrebné časti. Zoznam referencií obsahuje reviews aj pôvodné články, a tiež články popisujúce bioinformatické programy použité v práci.

Práca je po formálnej stránke v poriadku a splňuje požiadavky na bakalársku prácu. Celková úroveň je porovnateľná s inými prácami, nevyčnieva pozitívne ani negatívne. Chcela by som vyzdvihnúť veľkú časť metód (ohľadne skladania genómu), kde sa autor snaží detailne ale zároveň zrozumiteľne vysvetliť fungovanie a zmysel jednotlivých programov. Z vyslovene formálnych nedostatkov som si všimla, že medzi odstavcami sú rôzne veľké medzery. V práci sa striedajú časti napísané dobre, s časťami, ktoré sú nedostatočne vysvetlené až vôbec nevysvetlené. Zároveň práca obsahuje niekoľko anglických výrazov, ktoré podľa mňa mohli byť preložené do češtiny: pattern, tribe. Chcela som tam zaradiť aj slovo pík, ale školiteľ práce tvrdí, že sa jedná o české slovo. K jednotlivým častiam mám nasledovné pripomienky.

Úvod:

1. Z celkových 7,5 strán teoretického úvodu je 3,5 strany venované popisu motýľa monarcha sťahovavého, a to konkrétne podrobným informáciám ohľadne jeho migračných stratégií, orientácie počas migrácie atď – bez zjavnej relevancie pre prácu a bez prechádzajúceho vysvetlenia, prečo to v práci je. Neskôr som z ďalších sekcií práce pochopila, že sa jedná o motýľa príbuzného druhu študovaného autorom, ale aj tak mi details o migrácii prídu nerelevantné. Naopak by bolo lepšie viac rozviesť ďalšie podkapitoly ohľadne určenia pohlavia u motýľov, o vzniku a funkciách pohlavných a neopohlavných chromozómov a pod.

2. V texte úvodu je niekoľko nedostatočne vysvetlených alebo nepresných informácií:

i.) str.5: „Karyotyp u organismů bývá často konzervovaný a napříč řádem Lepidoptera je to ve většině případů pravda.“ – chýba informácia, na akej úrovni autor očakáva konzervovanosť sekvencie – druh, rod?

ii.) str. 7-8 – celá časť o C lokuse je nejasná, a nie je jasné, k čomu sa vzťahujú genotypy CC, Cc a cc.

Ciele práce:

3. Sú definované jasne, ale nie je jasné, prečo bol zvolený práve ten konkrétny druh motýľa, a v akom je vzťahu k motýľovi *Melitaea cinxia*, s ktorého genómom bude porovnávaný. Resp. čitateľovi to tak nejak dôjde po prečítaní celej práce, ale bolo by lepšie, keby to bolo jasne napísané napr. niekde v úvode, prípadne priamo v cieľoch.



Metódy:

4. Časti venujúce sa jednotlivým krokom zloženia genómu sú napísané zrozumiteľne, so všetkými potrebnými informáciami a použitými parametrami, a tiež vysvetleniami, čo dané programy robia a prečo je potrebné ich použiť. Ďalšie dve analýzy, ktoré už využívajú zostavený genóm, v kapitolách 3.2.7 a 3.2.8 už obsahujú výrazne menej informácií. U kapitoly 3.2.7 chýbajú prípadné bližšie informácie k použitiu Bedtools a Samtools, a stálo by zváženie tam pridať aj použitý R skript (aj keď bol vytvorený niekým iným). U sekcie 3.2.8 chýbajú akékoľvek informácie o použitom postupe.

Výsledky:

5. Výsledky sú napísané dosť stručne. Bolo by lepšie, keby sa autor viac rozpísal a pred samotný výsledok vždy napísal aj čo bolo robené, prečo to bolo robené, a prípadne stručne ako to bolo robené. Aj konkrétne výsledky (hlavne pri odkazovaní sa na graf – týka sa to hlavne grafov 1 a 4, do menšej miery dva) by mohli byť popísané zrozumiteľnejšie – pochopenie, prečo daná informácia vyplýva z grafu vyžaduje od čitateľa značnú dávku fantázie, intelektu a ideálne expertízy v odbore.

6. Graf 2 má v legende príliš malé nečitateľné písmo.

7. Tabuľka 3 – bolo by dobré vysvetliť jednotlivé skratky vo výsledkoch programu BUSCO.

8. Str. 16: Autor tvrdí, že použitie programu Ratatosk malo značný efekt na vylepšenie kvality, z tabuľky 1 to ale podľa mňa nevyplýva. Ako je možné, že po filtrácii programom Ratatosk je celkový počet bází v genóme vyšší, než bez úprav?

9. V texte výsledkov sú uvedené čísla, u ktorých nie je zrejmé, čoho sa týkajú a k čomu sa vzťahujú. Napr. na str. 16, že po odstránení 29 sekvencií zvýšilo 1114 sekvencií, alebo na str. 18 je uvedených 6684 sekvencií. Ani v jednom prípade nie je zrejmé, odkiaľ autor získal dané počty sekvencií, ak som dobre hľadala, nikde v texte ani v tabuľkách predtým nie sú. A keby aj boli, v texte by malo byť jasne uvedené, odkiaľ autor vzal dané hodnoty.

10. Graf 3 a text k nemu: Autor by mal uviesť, či cut-off 0.5 na rozlíšenie autozomálnych a pohlavne viazaných fragmentov je štandardne používaný, alebo k nemu došiel nejakou úvahou. A tiež by to mal spomenúť v texte, nie len v obrázku. Ďalej nie je úplne zrejmé, čo autor myslel sekvenciami okolo hodnoty 0,7 – všetky s hodnotami medzi 0,5 a 1? Nešlo by to pomenovať nejak lepšie?

11. Graf 4 a sekcia 4.3 – nie je vysvetlené, čo presne je na grafe a ako autor došiel k tomuto obrázku (to je presne tá časť, kde chýbajú informácie aj v metódach).

Diskusia:

12. Diskusia je na dve stránky, ale cca polovica z nej patrí asi skôr do teoretického úvodu. Autor mohol viac diskutovať technickú stránku práce (ako by išlo genóm vylepšiť a či vôbec, či použité bioinformatické programy fungujú ideálne a ak nie, aké problémy so sebou prinášajú atď), a prípadne navrhnúť možnosti pokračovania jeho projektu, čo už na úrovni bioinformatickej alebo experimentálnej.

Hoci by prácu išlo rozhodne vylepšiť, ako som už písala, spĺňa požiadavky na bakalársku prácu a doporučujem ju k obhajobe. K hodnoteniu sa vyjadrím po obhajobe.



Mám pre autora tieto otázky:

1. Autor v úvode práce uvádza, že motýle sú jedinečná skupina pre štúdium neopohlavných chromozómov kedy heterogametické pohlavie je samičie. Prečo v súvislosti v neopohlavnými chromozómami záleží na tom, či je heterogametické pohlavie samičie alebo samčie? Na čo to má vplyv? U akých skupín okrem motýľov, resp. hmyzu, ešte existujú neopohlavné chromozómy?
2. Mohol by autor vysvetliť ten lokus C a jeho súvislosť s neopohlavnými chromozómami, spomenutý v úvode na str. 7-8?
3. Mohol by autor objasniť komentáre č. 8 a 9?

„8. Str. 16: Autor tvrdí, že použitie programu Ratatosk malo značný efekt na vylepšenie kvality, z tabuľky 1 to ale podľa mňa nevyplýva. Ako je možné, že po filtrácii programom Ratatosk je celkový počet bází v genóme vyšší, než bez úprav?

9. V texte výsledkov sú uvedené čísla, u ktorých nie je zrejmé, čoho sa týkajú a k čomu sa vzťahujú. Napr. na str. 16, že po odstránení 29 sekvencií zvýšilo 1114 sekvencií, alebo na str. 18 je uvedených 6684 sekvencií. Ani v jednom prípade nie je zrejmé, odkiaľ autor získal dané počty sekvencií, ak som dobre hľadala, nikde v texte ani v tabuľkách predtým nie sú. A keby aj boli, v texte by malo byť jasne uvedené, odkiaľ autor vzal dané hodnoty. „
4. Mohol by autor detailnejšie objasniť, ako získal graf 4, čo na ňom je za hodnoty a ako si ho interpretovať?