BIOLOGY CENTRE CAS
Institute of Parasitology
address: Branišovská 1160/31, 370 05 České Budějovice, Czech Republic
IBAN – CZ24 0800 0000 0000 0606 3942 | SWIFT CODE – GIBACZPX | VAT No.: CZ60077344
phone: +420 387 775 403 | fax: +420 385 310 388 | www.paru.cas.cz | e-mail: paru@paru.cas.cz

České Budějovice, 30. 8. 2020

Oponentský posudek

Diplomová práce: Assembly a anotace genomu včely *Ceratina nigrolabiata* s obourodičovskou péčí

Magisterská práce Bc. Terezy Fraňkové je dobře a povětšinou jasně napsána a zabývá se velice zajímavou tématikou. Bylo mi tak potěšením vypracovat tento oponentský posudek. Předkládaná diplomová práce má celkovou délku 50 stran a odkazuje na více jak 200 referencí. V textu jsem nenašel žádné závažnější nedostatky či chyby.

Práce se zabývá sekvenováním, assembly, anotací a analýzou genomu biparentální včely *Ceratina nigrolabiata.* Jedním z cílů autorky je, jak sama uvádí: „Understanding of programs and software needed for successful analysis of the genome of *Ceratina nigrolabiata"*, z čehož usuzuji, že s bioinformatikou a genomikou neměla před započetím práce rozsáhlejší zkušenosti. Vzhledem k tomu, že *de-novo* assembly genomů je velmi náročná práce i pro bioinformatika pracujícího ve specializované „genomické" laboratoři, je tak odvedená práce o to impozantnější. Výsledky a diskuze práce obsahuje dvě hlavní složky – vlastní assembly a analýzu genomu, na kterou navazuje analýza specifických genů hrajících roli v regulaci genové expresse, popřípadě roli v parentálním nebo dokonce eusociálním chování hmyzu.

Jako většina vědeckých prací, i tato má jisté nedostatky. Je ovšem třeba říci, že tyto nedostatky jsou spíše literárního charakteru a nejedné se přímo o faktické či metodické chyby. Některé části metodiky a výsledků by nicméně mohli být vysvětleny jasněji, především pak postup při genové predikci a anotaci, který jsem musel projít několikrát. Při porovnání genomů ze dvou různých druhů *Ceratina* autorka uvádí, že nejdelší scaffold *C. calcarata* je 52 351 bp a N50 bylo 73 643 bp což však technicky není možné a s největší pravděpodobností se tak u jednoho z uvedených čísel jedná o překlep.

Celkové zpracování a výsledky této diplomové práce jsou nesporným přínosem pro naši znalost včel a získaná data mají nemalou vědeckou hodnotu. Navíc, jak z textu vyplývá, předkládaným textem práce nekončí a autorka se chystá ve studii genomiky a transkriptomiky biparentálních včel pokračovat. Je také důležité uvést, že výše uvedené nedostatky v žádném případě nic neubírají na tom, že předkládaná práce je velmi vysoké úrovně. Autorka, dle mého názoru, zcela splnila požadavky na Diplomovou práci na Přírodovědecké Fakultě Univerzity Karlovy, a proto ji s potěšením doporučuji k obhajobě s hodnocením výborná.

Mgr. Martin Kolisko, PhD
Laboratoř Genomiky a Diverzity Prvoků
Parazitologický Ústav, Biologické Centrum, AVČR
Branišovská 31, České Budějovice, 370 05
kolisko@paru.cas.cz
+420387775478

Otázky a komentáře pro autorku:

1. Není mi zcela jasné, jaký je vztah mezi genovou predikcí pomocí programu Braker a Maker. Sloužily výsledky obdržené programem Braker jako jeden ze vstupů pro program Maker?

2. Pokud jsem správně pochopil metodiku, tak autorka použila do jisté míry „meta-assembly"', kdy contigy z assembleru Mira byly použity jako vstupní data pro assembler SPAdes? Mohlo by takto dojít ke vzniku miss-assembly? A pokud ano, byla proveden kontrola?

3. Transkriptomická data ze studovaného druhu nebyla k dispozici a tak autorka použila transkriptomická data blízce příbuzného druhu. Jaká je odhadem genetická vzdálenost mezi těmito druhy. Mohl případný rozdíl mezi těmito druhy mít negativní impakt na „downstream" analýzu?

4. Na stránce 29 autorka uvádí: „We found 15 out of 23 unique orthogroups previously discovered in *C. calcarata*. We presume our genome contains more shared unique orthogroups, however, they were not identified in our analyses." Jak byl v tomto případě definován termín "orthogroup"?

5. Jak by autorka dále postupovala při pokračování výzkumu?