

Oponentský posudok diplomovej práce Bc. Anety Pilíkovéj

Aneta Pilíková sa vo svojej diplomovej práci venuje hľadaniu samičích molekulárnych markerov na spoľahlivé určenie pohlavia u priadzovca mukyňového (česky předivka brslenová), a to konkrétne unikátnych nerepetitívnych oblastí na chromozóme W, ktorý sa nachádza len u samíc. Autorka úspešne použila moderné aj klasické metódy molekulárnej biológie, a tiež netriviálne bioinformatické postupy, a vedecky tak určite spĺňa podmienky pre diplomovú prácu.

Práca samotná je napísaná čtivým štýlom, má 53 stránok, pričom bioinformatické kódy sú výhradne na internete, a po formálnej stránke obsahuje všetko, čo by mala obsahovať. Celkovo sa jedná o peknú prácu a je z nej vidieť, že autorka rozumie tomu, čo robila, a vie to zároveň vhodne podať čitateľovi. V práci som si nevšimla jediný preklep a je napísaná primerane vedeckým jazykom a vhodným slohovým štýlom. Jediné, čo pôsobí trochu rušivo, je, že autorka občas nazýva svoj modelový organizmus českým názvom, inokedy českým aj latinským, inokedy zas len latinským a niekedy zas len skratkou latinského názvu. Bolo by lepšie na začiatku práce pri prvom použití názvu druhu uviesť oba názvy a potom konzistentne už len jeden po zvyšok textu práce. A pri číslovaní kapitol autorka chybné uvádza bodku aj za poslednou číslicou, pričom tam by už bodka nemala byť, napr. kapitola 3.4 Názov kapitoly (za číslicou 4 už bodka nemá byť).

Teoretický úvod má 11 stránok, a primerane zoznamuje čitateľa s problematikou výskumu pohlavných chromozómov motýľov, dôležitosťou schopnosti určiť pohlavie u raných štádií motýľov, bioinformatickými postupmi a s priadzovcom mukyňovým ako modelovým druhom tejto práce. Tým, že práca je napísaná v češtine, autorka asi občas trochu bojovala s hľadaním správneho českého vedeckého výrazu – napríklad „pohlavní rozmnožování umožňuje promíchání genetického materiálu“ (ale samej mi nenapadá, ako to povedať lepšie), „skupina, u které se vyvinul externí příjem potravy“ (tento pojem by bolo lepšie aj popísať pár slovami), a tiež definovanie holocentrických chromozómov ako chromozómov s chýbajúcou primárnou konstrikciou v oblasti centroméry, čo vyvoláva dojem, že holocentrické chromozómy majú jednu centroméru, len v nej nie je konstriKCia.

Ciele práce sú výstižne popísané. Metódy sú napísané prevažne jasne, u sekcie „Spracovanie výsledkov YGS“ by prospel nejaký diagram prehľadne zobrazujúci jednotlivé filtrovacie kroky a použité parametre, ale dalo sa to pochopiť aj bez toho. Niektoré použité parametre nie sú jasne definované, napr. na strane 20 na konci sekcie 3.3 je uvedené, že po mapovaní v programe Geneious boli vyradené sekvencie, u ktorých bola nájdená zhoda. Nie je ale zrejmé, aká veľká zhoda to musela byť. V kapitole 3.5 o PCR overovaní kandidátnych markerov je napísané, že boli testované dve polymerázy, a rôzne vstupné koncentrácie DNA, a že boli vybrané tie najlepšie. Nie je jasné, na základe akých parametrov, a či tak boli testované len nejaké testovacie vzorky, alebo skutočné, u ktorých sa markery overovali, a ktoré primery takto boli testované (toto je relevantné kvôli tomu, že autorka vo výsledkoch ukazuje, že ani jedna z reakcií overovania markerov vlastne nefungovala optimálne, a čitateľ tak nemôže dostatočne posúdiť, či sa nemôže jednať aj o problémy technické). Ďalej je uvedené, že pre DNA na PCR bola získaná z desiatok jedincov oboch pohlaví, ale nie je na 100% zrejmé, či jedna vzorka v jednej PCR reakcii bola z jedného jedinca, alebo zmes viacerých jedincov. A ani v metódach ani vo výsledkoch nie je uvedené, ktorá vzorka bola použitá v ktorej reakcii – t.j. či v niektorých prípadoch boli použité rôzne primery u tých istých vzoriek, alebo sa jednalo zakaždým o inú vzorku.

Výsledky sú popísané primerane, ale v očakávaní, že čitateľ si presne pamätá kapitolu metód. Aspoň podľa mňa (ale možno to je len subjektívny názor) by pomohlo, keby na začiatku každej

podkapitoly autorka stručne (napr. jednou až tromi vetami) uviedla, čo v danej kapitole robila, ako to robila a prečo to robila. Bioinformatická časť je inak popísaná pekne, len v kapitole 4.1 o assembly genómu by názornosti pomohli grafy, napr. histogram dĺžky readov, či grafické výsledky Nanoplotu. Tiež sa mi zdá, že obr. 5 má asi chybu v popise osy X – nemá tam byť percento k-mérov namiesto percenta fragmentov. Na stránke 31 autorka píše, že *contig_24514* má 2542 bp, hoci v tabuľke IX uvádza dĺžku 5086 bp. Sekcia 4.4 o PCR overovaní identifikovaných samičích markerov svojou kvalitou bohužiaľ nedosahuje zvyšok práce – čitateľ z nej ale dokáže precítiť autorkino nešťastie z toho, že vlastne nič nevyšlo, dokonca ani kontrolné primery nefungovali, ako by mali. Takže na jednu stranu chápem, že autorke sa nechcelo príliš pipláť s množstvom negatívnych výsledkov, na druhú stranu aj negatívne výsledky sa dajú prezentovať vedecky správnejšie. Táto časť by sa dala vylepšiť napr. takto: (I) Bol by uvedený zoznam použitých vzoriek, u každej vzorky spôsob izolácie DNA, kvalita a koncentrácia vyizolovanej DNA (napr. Nanodropom), a výsledok použitia kontrolných primerov (aj keď povedzme robia dva produkty namiesto jedného), ktorý by ukazoval, že DNA tam je a funguje na nej PCR. Napr. v obrázku 7 vznikajú po použití primerov *Ycag-1* produkty len u samca – nevieme, či u samice nevznikajú preto, že tam daná sekvencia nie je, alebo preto, že DNA je nedostatočnej kvality. (II) U každej reakcie by bola uvedená teplota hybridizácie primerov a vysvetlené, či je to najlepšia možná teplota. Napr. na obrázku 8 je ukázané, že kontrolné primery *Ace2* tvoria pri teplote 57°C nešpecifické produkty, ale nevieme už, či pri teplote 58°C a viac tam už nie je produkt žiadny (t.j. či si je autorka istá, že pri vyššej teplote by nešpecifické produkty nezmizli). (III) Ak by to bolo možné, autorka mohla urobiť PCR zo vzorkami, ktoré boli použité na bioinformatickú analýzu (samčie aj samičie) na potvrdenie, či tam daná údajná samičia sekvencia je, alebo nie je.

Diskusia je dlhá a bohatá, je to rozhodne silná časť práce, a ukazuje, že autorka dokáže kriticky rozmyšľať nad svojimi výsledkami, a orientuje sa v tematike. To dokladá aj fakt, že v práci je citovaných viac než 100 zdrojov, pričom sa jedná väčšinou o pôvodné články. Chýba mi tam len zamyslenie sa nad otázkou, či je stále nádej, že skúmaný druh motýľa má nejaké unikátne sekvencie špecifické pre samice.

Celkovo sa jedná o vydarenú prácu, vymenované nedostatky sú skôr len drobnosti, aby autorka vedela, ako to prípadne nabudúce môže napísať ešte lepšie. Prácu doporučujem k obhajobe a mám tieto otázky:

1. Na str. 3 je uvedené, že na rozdiel od iných organizmov so ZZ/ZW určením pohlavia, u motýľov sú bežné fúzy autozómov a pohlavných chromozómov. Je nejaký dôvod, prečo sú u motýľov tak bežné?
2. Na str. 5 je uvedený predpoklad, že expresia pohlavne antagonistických génov je u raných embryí rovnaká medzi samcami a samicami, ale líši sa u dospelých. Autorka uvádza, že dlhodobým cieľom je analyzovať pohlavne antagonistickú selekciu u motýľov, a preto je treba poznať pohlavie motýľov už u raných embryí, aby sa napr. potom mohla porovnávať expresia génov u raných embryí a dospelých. Jedná sa podľa autorky teda o realistický predpoklad, že toto je naozaj charakteristika pohlavne antagonistických génov (a nie napr. génov, ktoré len u druhého pohlavia nie sú potrebné), a zaoberal sa touto témou ešte niekto okrem jedného článku z r. 2009 (Mank a Ellegren), ktorý autorka cituje?
3. Ako by dopadli podľa názoru autorky PCR s primermi *Ycag-1* až -8 pri použití tých istých vzoriek, ako boli použité na bioinformatickú analýzu (najmä u samcov)?
4. Prečo autorka použila program DANTE až v diskusii, kde ukázala, že údajne unikátne samičie sekvencie sú v skutočnosti repetitívne elementy, a nie počas filtrovania fragmentov s cieľom identifikovať najpravdepodobnejšie kandidátne samičie unikátne fragmenty?

5. Je stále nádej, že u tohto druhu sú nejaké unikátne samičie sekvencie na chromozóme W, ktoré by mohli byť využité pri určení pohlavia?

6. V diskusii sa uvádza, že v priebehu jej diplomovej práce bol v rámci projektu Darwin Tree of Life publikovaný genóm tohto druhu o veľkosti 575 Mbp, ktorý na základe BUSCO analýzy autorka vyhodnotila ako lepši a kompletnejši než ten jej o veľkosti 800 Mbp či z publikácie Voleníková et al. o veľkosti 626 Mbp, kde bola veľkosť stanovená prietokovou cytometriou na cca 710 Mbp, pričom ale BUSCO analýza u oboch týchto genómov naznačuje väčšie percento duplikovaných regiónov. Ako si to autorka vysvetľuje? Je skutočná veľkosť okolo 700 Mbp, ale v genóme Tree of Life tých chýbajúcich 125 Mbp nenesie v podstate žiadne ortológne gény?

V Českých Budejoviciach, 14.5.2023

Mgr. Lenka Gahurová, Ph.D.