

Posudek na diplomovou práci Daniely Kotalové „Sestavení genomu cichlidy *Crenicichla semifasciata* kombinací sekvencí z různých sekvenačních platforem“

Posuzovaná diplomová práce je metodické povahy. Studentka pracovala se surovými sekvenčními daty a cílem její práce bylo sestavit z těchto dat různými metodami genom příslušného druhu cichlidy a porovnat vlastnosti takto sestavených genomů. Zvláštní pozornost přitom byla věnována předpokládaným přínosům kombinování dat získaných pomocí odlišných sekvenačních technologií.

Samotnou bioinformatickou práci studentka úspěšně provedla, ale kromě toho si na jejím textu cením především dvou skutečností. Za prvé jsou v něm cíle práce nejenom jasně definovány, ale navíc uvedeny textem, který čtenáři objasňuje, proč je zajímavé a důležité se právě tomuto tématu věnovat. Předem je vysvětleno v čem se liší surová data získaná pomocí odlišných technologií, v čem jsou komplementární a na čem tedy staví předpoklad, že jejich kombinování bude pro přesnost sestaveného genomu prospěšné. Za druhé práce pokrývá všechny důležité stránky studovaného problému, a to v logickém pořádku a s určitými výhradami i v přiměřeném rozsahu. Autorka čtenáři poskytne přehled sekvenačních technologií, jednotlivých kroků bioinformatických analýz a nezapomene ani na seznámení s biologickým problémem, k jehož řešení mají tato konkrétní data přispět. Moje připomínky jsou proto spíše dílčího rázu.

Je v pořádku nezabíhat do přílišných detailů, ale na některých místech by textu přece jen prospěl odstavec navíc. Například popis sekvenačních technologií (části 1.3 a 1.4) je sice nejspíš technicky v pořádku, ale je zestručněný tak, že čtenář už nezíská konkrétní představu, jak to celé vlastně funguje. Podobně v popisech algoritmů (části 1.5.2.1 a 1.5.2.2) by mělo být jasné nejen to, jaký typ grafu pro sebe vytváří, ale také jak pohyb v tomto grafu sestavuje genom. Konstatování „je nutné každou hranu [De Bruijnova grafu] projít podle Eulerovského tahu“ (část 1.5.2.2) by mohl nahradit odstavec s obrázkem, za kterých by bylo jasné, co toto konstatování znamená a proč je právě tento Eulerovský tah klíčem k sestavení genomu. Také z popisu výpočtu N50 (1.5.4) jsem získal jen přibližnou představu, co toto číslo vyjadřuje. Toto činí práci poněkud nepřesvědčivou, neboť z textu není jisté, nakolik studentka rozuměla operacím, které na datech provádí. Je docela dobře možné, že jim rozuměla bezvadně, ale z textu nejde vždy s jistotou poznat, zda se při práci nedržela pouze instrukcí v návodu („použijte tento příkaz s těmito argumenty“) a zda text není souhrnem napůl pochopených informací z literatury.

Také diskuse by měla být v některých bodech konkrétnější. Pokud se například kompletnost genomu zvýší z 96.6 % na 98.1 %, je to hodně nebo málo? V práci se sice píše, že „V takto vysokých hodnotách je významné každé zlepšení.“, ale nejsem si jistý, jak je to míněno. Znamená to jen, že je každého dalšího zlepšení těžké dosáhnout? Nebo že každé takové zlepšení zpravidla ukáže něco zajímavého? Bylo by také lépe citovat i obdobné metodické práce provedené na jiných organismech. Zde kontext vytváří jen bakalářská práce provedená na stejných datech.

Z hlediska stylistického má práce jednoznačně charakter vědeckého textu. Tento dojem narušují jen občasné neobratné formulace, jako například „změna morfologie těla proběhla na základě různé potravy“ nebo „z těchto důvodů se vymýšlely další postupy“.

Na závěr položím otázky, které sice úzce nesouvisí s diplomovou prací samotnou, ale vzbudily moji zvědavost a myslím, že biologicky zaměřený čtenář by uvítal jejich objasnění i v textu samotném. Sekvenování genomů bylo motivováno snahou porozumět evoluci druhových hejn v rodu *Crenicichla*. V úvodu je zmíněno, že „Tradiční mitochondriální markery (cytochrom b a ND2) bohužel neposkytly věrohodnou evoluční hypotézu.“ a „genotypování pomocí techniky ddRAD [...] práci významně posunulo (Burriss et al., 2022)“. Co tedy bylo v citované práci zjištěno? Jedná se o případ mitonukleární neshody (discordance)? Jaká je její povaha a jakou evoluční hypotézu lze tedy pro tato druhová hejna formulovat?

Celkově hodnotím práci jako úspěšně zvládnutou – splňuje vytyčené cíle, přináší vědecky hodnotné údaje a má povahu dobře rozvrženého vědeckého textu, byť na jeho konkrétní podobě by podle mého názoru bylo co zlepšovat. Jednoznačně ji však doporučuji k obhajobě a navrhuji hodnocení „velmi dobře“.

Ve Studenci, 4. 1. 2023



Ondřej Mikula