

Posudek na magisterskou práci Bc. Daniely Kotalové; Sestavení genomu cichlidy *Crenicichla semifasciata* kombinací sekvencí z různých sekvenačních platform.

Úvodní část diplomové práce Bc. Daniely Kotalové dobře shrnuje pozadí výzkumu a důvody, proč se laboratoř Dr. Piálka v tomto kontextu zabývá právě skládáním genomu. Část je věnována vývoji sekvenačních technologií a rozdíly mezi nimi stejně tak jako úvodu do bioinformatických nuancí sestavování genomů.

Cíle práce jsou jasně a stručně definované, dosažitelné a dávají smysl vzhledem k výzkumnému zaměření laboratoře. Oceňuji i to, že tato práce je přímým pokračováním předchozích projektů teamu.

Metody primárně popisují skripty, použité při bioinformatické analýze. Popis je to velmi detailní a usnadňuje případnou reprodukovatelnost dosažených výsledků, což oceňuji. Pokud bych měl poťouchle hledat co vytknout, tak bych snad poukázal na to, že kód by se lépe četl, pokud by byl formátován neproporcionálním písmem. Taktéž od stránky 22 (sekce 3.4.3. - Mapování krátkých sekvencí na sestavu Flye) se v kódu začíná objevovat proměnná "assembly", která není v textu ani v poznámkách vysvětlena. Oba tyto body však nepovažuji za příliš závažné a pro reprodukovatelnost analýzy by jistě nebyly velkou překážkou. Za nutné považuji vyzdvihnout aktuálnost použitého software a pečlivě zvolený přístup s jakým autorka přihlíží k rozdílnosti vstupních dat. Několik iterací assembly a jejich následné srovnání je velmi chvályhodné a dobře ukazuje, že autorka ví co dělá. Použité metodice tak vcelku není co vytknout.

Výsledky a diskuse jsou velmi deskriptivní a jasné a z mého pohledu jim nic nechybí. Porovnání s předchozí assembly ze stejných vzorků je skvělým příkladem toho, proč jako sekvenační středisko nutíme své zákazníky zálohovat surová data z Nanopore. S novější verzí programů, které tato data zpracovávají můžete dostat výrazně lepší výsledky analýzy. Takto konkrétní důkaz osobně považuji za velmi přínosný.

Celkové hodnocení:

Diplomová práce Bc. Daniely Kotalové spojuje rešeršní a praktickou část do vhodného celku. Množství činnosti bioinformatické i spisovatelské, dle mého názoru, dobře odpovídá (nebo přesahuje) to, co se od magisterského studia čeká a nemám zde sebemenších výhrad. Závěry vyvozované z výsledků nejsou nikterak přeceněné ani podceněné, což vítám. Osobně si též vážím (stejně jako v předchozích pracích stejně laboratoře) kvalitní dokumentace bioinformatických skriptů. Na čem by se do budoucna zapracovat dalo, je stylistická úprava a strukturní ucelenost práce. Jako příklad bych uvedl jistou redundanci v popisu basecallingu jak v úvodu, metodách, tak ve výsledcích, kterou jsem při čtení považoval za redundantní. Nepochybuji však, že jde o návyky, které si studentka rychle osvojí s větší praxí.

Stanovené cíle práce se podařilo více než obstojně splnit. Výsledné assembly bezesporu budou velmi důležitým nástrojem nejen pro laboratoř Dr. Piálka. Skládání genomu nemodelových organismů rozhodně nepatří mezi lehké úkoly, a i s pokrokem a dostupností rozmanitých sekvenačních metod a pokročilých informatických postupů nejde o triviální problém. Je tedy třeba velmi ocenit, že se studentka tohoto počinu zhostila a dokázala ho dotáhnout do úspěšného konce.

Práci jednoznačně doporučuji k obhajobě s hodnocením "výborně".

Dotazy:

- 1) Při popisu Oxford Nanopore Technologies je několikrát zmíněno, že má větší chybovost než Illumina. Může, prosím, autorka zmínit o jaký typ chyb se jedná, jestli jsou náhodné a jak se liší oproti chybám technologie Illumina?
- 2) Při vyhodnocení biologické kompletnosti assembly byl použit program BUSCO. Zajímalo by mě, jak na tom byla kupříkladu sestava "Hybridní_3000_1" s počtem duplikovaných BUSCO genů?
- 3) Bonusová otázka, možná spíše pro školitele: Plánujete zveřejnění složeného genomu či jeho další opracování? Pokud ano, myslíte, že assembly jako takové ještě něco chybí?

V Heidelbergu 08. 01. 2023

RNDr. Jan Provazník