

Posudek oponenta na diplomovou práci

Autor práce: Bc. Kateřina Jurásková

Název práce: Bioinformatická analýza distribuce transpozonů na pohlavních chromozomech rostlin

Oponent práce: Eva Hřibová, Ph.D.

Poř. číslo	Kritérium hodnocení	Body (0-5)
1	Ucelenost a aktuálnost rešeršní části práce	3
2	Kvalita úvodní části práce (množství použitých původních pramenných zdrojů, vhodnost výběru)	4
3	Naplnění cílů práce	5
4	Logika postupu při vlastní rešeršní nebo experimentální práci	4
5	Úplnost popisu používaných metodik a postupů	3
6	Úroveň zpracování výsledků (vhodné používání grafů a tabulek atd.)	4
7	Adekvátnost interpretace získaných výsledků a jejich diskuse	4
8	Výstižnost souhrnů práce v českém a anglickém jazyce	4
9	Grafická úprava textu a obrázků	4
10	Jazyková a stylistická úroveň, respektování platného názvosloví	4
11	Správnost a úplnost legend u obrázků a tabulek (srozumitelnost bez zřetele k ostatnímu textu, vysvětlení značek, jednotky uváděných veličin)	3
12	Správnost používání citačních odkazů (přítomnost necitovaných údajů, dodržování jednotného stylu citací, používání oficiálních zkratk časopisů)	4
Celkem bodů		46

max
60

Konkrétní připomínky a dotazy (možno připojit samostatný list)

Diplomová práce Bc. Kateřiny Juráskové nazvaná Bioinformatická analýza distribuce transpozonů na pohlavních chromozomech rostlin, si kladla za cíl využít illumina sekvenční data pro analýzu repetitivních sekvencí DNA, analýzu složení samčího genomu a pohlavních chromozomů *Rumex acetosa* a využití získaných výsledků pro potvrzení modelu parentálně specifického šíření transpozonů na pohlavních chromozomech. Vlastní práce je rozdělena na teoretický úvod (současný stav řešené problematiky) a experimentální část popisující použité bioinformatické nástroje a dosažené výsledky.

V teoretickém úvodu studentka podává takřka vyčerpávající popis jednotlivých typů transponovatelných elementů, věnuje se také regulaci jejich aktivity v genomu a nakonec zmiňuje výskyt transponovatelných elementů u rostlinných druhů s pohlavními chromozomy. Vzhledem k tomu, že jedním z hlavních cílů práce byla de novo identifikace a charakterizace repetitivních DNA sekvencí, spojená se srovnávací analýzou přítomnosti jednotlivých typů

repetic a jejich zastoupení na autozomech a pohlavních chromozomech druhu *R. acetosa*, literární rešerše by si dle mého názoru zasloužila začlenění kapitoly poskytující náhled na nejpoužívanější bioinformatické nástroje, které se v dnešní době využívají pro identifikaci a charakterizaci repetitivních DNA sekvencí a zmínění výhod/nevýhod daných nástrojů. Předpokládám, že z této kapitoly by velmi jednoznačně vyplynuly výhody programu Repeat Explorer, který byl zvolen pro analýzu repetitivní DNA v předkládané diplomové práci.

První tři kapitoly Experimentální části se věnují obecnému popisu již zmíněného Repeat Exploreru, shlukové analýzy a programovacímu jazyku R. Popis samotného programového balíku, či shlukové analýzy, na kterém je sestavování repetitivních elementů založeno, není po formální a jazykové stránce pečlivě vypracován, a zasloužil by si větší pozornost. Studentka zde také používá jazykově nepřesné až zavádějící výrazy, které významně ubírají na kvalitě předložené práce, např. popis na str. 31 a 32: „Transpozony mohou být roztrhané na menší části (Obr. 8A), které se mohou spojit na základě podobnosti s jinými roztrhanými transpozony (Obr. 8B) a vytvořit tak nový klast (tvořený čteními z rozdílných rodin TE).“ „LTR transpozony, které mají na svých koncích repetici, by se v grafu měly spojit do „tenkého kruhu“ se zdvojenou LTR částí.“ „...RepeatExplorer dlouhé transpozony roztrhá na více menších klastů, které se zobrazují jako „čára“...“, „Další způsob jak určit, o jaký transpozon se jedná, jsou superklasty. Superklast je skupina klastů, které mají hodně spojení (překryvů) přes párová čtení. Párová čtení jsou fyzicky blízko u sebe. Pokud jsou blízko sebe párová čtení z různých klastů, může to znamenat, že pochází ze stejného elementu.“

Bohužel i kapitola Metody a Výsledky by si zasloužila větší pozornost. Použitá metodika není mnohdy dostatečně popsána a čtenář tak tápe, jaká přesně data a jaké parametry programů byly použity pro dané analýzy. Některá místa v metodice jsou dosti zmatečná, např. str. 33, kdy jsou zmíněny parametry trimování programem Trimmomatic („... crop 215,...“), ale v následující větě je popsáno: „Získali jsme párová čtení o minimální délce 100 bází a maximální délce 251 bází.“, či tvrzení „Výsledný soubor připravený pro klastrovou analýzu byl po úpravách a vzorkování zmenšen na 500 000 párových čtení“, což je jistě důležitá informace, ale dále v textu tyto úpravy nejsou podrobněji popsány. Další velmi zmatečná část se vyskytuje na str. 35., kdy studentka popisuje „Velikost samčího genomu je přibližně 7335 Mbp.“, ale ve vzorci níže používá velikost genomu 14,95, tedy pokud to chápu správně, protože zmíněná rovnice není vůbec popsána.

Neúplný popis Metodiky tak má i vliv na jednoznačné pochopení získaných a popsaných výsledků. Nepochybuji, že studentka v rámci diplomové práce získala řadu původních výsledků, avšak množství jazykových nepřesností a nižší stylistické úrovně významně ubírá na kvalitě předložené práce. Přes veškeré jmenované výhrady diplomovou práci doporučuji k obhajobě.

V souvislosti s obhajobou diplomové práce, bych chtěla autorce položit následující doplňující dotazy:

- 1) Mohla byste objasnit, jaký samčí genotyp *R. acetosa* byl analyzován v rámci vaší práce? Jakou měl velikost genomu, jak byl zjištěn obsah DNA jednotlivých pohlavních chromozomů a autozomů?
- 2) Kolik sekvenčních dat bylo získáno illumina sekvenováním tří skupin chromozomů a jaké procento dat bylo získáno po trimování? Jaké pokrytí (tzv. coverage) představovaly datasety sekvencí specifických pro A, X a Y chromozomy? Zkoušela jste použít větší dataset pro komparativní analýzu repetit na A, X a Y chromozomech pomocí Repeat Exploreru?
- 3) Jaké „úpravy“ souboru trimovaných illumina sekvencí byly použity před samotnou analýzou programem Repeat Explorer (viz tvrzení str. 33)? Týkaly se tyto úpravy skutečnosti, že párová čtení byla trimována na velmi odlišné délky (nejdelší čtení byla 2,5 x delší než nejkratší illumina čtení)? Pokud byla cílem komparativní analýza a poměrné zastoupení

jednotlivých typů repetice na A, X a Y chromozomech, nebylo by tedy výhodnější trimovat Illumina data na stejně dlouhé sekvence?

4) Proč bylo pro vyjádření rozdílů ve složení pohlavních chromozomů pracováno se dvěma soubory - „před a po očištění od satelitů“? A kdy k tomuto „očištění“ došlo? Před samotným klastrováním pomocí programu Repeat Explorer?

5) Potvrdila analýza repetitivních částí tříděných chromozomů přítomnost (nahromadění) clDNA sekvencí na Y chromozomu *R. acetosa*, jak bylo zjištěno v práci Šteflová et al. (2014)?

6) Zkoušela jste použít větší soubor dat pro komparativní analýzu repetice na A, X a Y chromozomech pomocí Repeat Exploreru?

7) Na str. 40 popisujete: „Zajímalo nás tedy, kolik procent DNA pohlavních chromozomů je tvořeno sdílenými, stejnými, repetitivními elementy a kolik procent DNA pohlavních chromozomů je tvořeno unikátními kopiemi repetitivních elementů, tedy elementy, které jsou na chromozomu X navíc proti chromozomu Y a naopak.“

Mohla byste definovat tyto skupiny repetitivních elementů? Jakou homologii sdílí „stejně repetitivní elementy“ a naopak, jaká variabilita na úrovni DNA byla potřebná pro charakterizaci „unikátních kopií“ repetitivních elementů? Je vůbec možné na základě analýzy sekvencí reprezentujících tak malé procento genomu identifikovat unikátní kopie repetitivních elementů?

Chyby, které je nutno opravit

Závěr: práci doporučuji k obhajobě.

V Olomouci dne: 2. 6. 2020

Podpis

