

Posudek oponenta bakalářské práce

Jméno a příjmení uchazeče/ky : Lujza Milotová

Název práce: Bioinformatické přístupy k vyhodnocení frekvenčních populačních dat založených na sekvenčních polymorfismech DNA

A. Bodové hodnocení jednotlivých aspektů práce (označte právě jednu z možností)

1. Rozsah BP a její členění	
x	A - přiměřené, odpovídají charakteru BP a významu jednotlivých částí
	B - nevyrovnané, členění není logické n. rozsah jednotlivých částí nekorresponduje s jejich významem
	C - uspokojivé, rozsah některých částí nedostačuje
	N - nedostatečné

2. Odborná správnost	
	A - výborná, bez závažnějších připomínek
x	B - velmi dobrá, s ojedinělými drobnými závadami (nejasnost výkladu, chyby ve vzorcích nebo chemických názvech, nedokonalý popis metod nebo výsledků)
	C - uspokojivá, s čtenějšími drobnými závadami
	N - nevyhovující, s hrubými chybami

3. Uvedení použitých literárních a j. zdrojů	
x	A - bez připomínek, všechny převzaté údaje s citací zdroje, celkový počet citací odpovídá charakteru práce
	B - uspokojivé, s občasnými neobratnostmi zejm. v umístění odkazů, nebo s celkově nižším počtem citací
	C - s vážnějšími závadami, např. převažují "nestandardní" odkazy na učebnice, přednášky, webové stránky, nebo se ojediněle vyskytuje opominutí odkazu na zdroj převzatých dat
	N - nevyhovující, velmi málo citací, ev. rysy plagiátu (časté opomíjení odkazu na zdroj převzatých dat, popř. opsání velkých částí textu)

4. Jazyk práce	
x	A - výborný, práce je napsána čtivě a srozumitelně, bez závažnějších gramatických n. pravopisných chyb
	B - velmi dobrý, ojedinělé stylistické neobratnosti, gramatické n. pravopisné chyby
	C - upokojivý, čtenější slohové neobratnosti, gramatické n. pravopisné chyby, ojediněle se vyskytují obtížně srozumitelné n. nejednoznačné formulace
	N - nevyhovující, s četnými hrubými chybami

5. Formální a grafická úroveň práce	
x	A - výborná, bez překlepů a chyb ve formátování
	B - velmi dobrá, ojedinělé chyby formátu citací, překlepy, chybějící zkratky apod.
	C - uspokojivá, s ojedinělými většími (např. vynechání stránky) nebo čtenějšími drobnými chybami
	N - nevyhovující, s četnými hrubými chybami

Případný slovní komentář k bodům 1. až 5.:

Práce obsahuje některé nepřesnosti. Pro samotnou práci nejsou podstatné, ale považuji za dobré je zmínit.

"Až přibližně 99,9 % z celkových 3 miliard bazových párov v lidském genomu je pre všetkých ľudí spoločných."

Citovaný zdroj (PMID:11148583) je odhad staršího data a nezahrnuje větší strukturní variace. Skutečná variabilita je přibližně pětikrát větší, viz například (PMCID:PMC1964779), doporučuji tedy použít novější odhad 99.5%.

"Priemerne sa vyskytujú každých 300 nukleotidov, čo znamená, že v ľudskom genomu je približne 10 miliónov SNP."

Toto tvrzení je ve své jednoduchosti ošidné, záleží na studované populaci a definici. Tak jak jsou SNP v této práci uvedené, tedy všechny jednobázové rozdíly bez ohledu na jejich frekvenci v populaci, je číslo značně podceněné (přinejmenším 85 mil). Na druhou stranu, pokud byl záměr nastínit variabilitu typického lidského genomu, pak je číslo přeceněné (<5 mil), viz např. PMID:26432245. Ať už je uvedené jakékoliv číslo, mělo být blíže vysvětleno.

"V súčasnosti nastupujú sekvenačné metódy vysokej priepustnosti druhej generácie (angl. high-throughput - next generation sequencing, HT-NGS), ktoré umožňujú sekvenovanie jedinej molekuly."

Toto není zcela správně. Nešťastně zvolený termín "next-generation sequencing, NGS" se začal používat pro masivně paralelní metody druhé generace vyvinuté na přelomu století. Díky nim bylo možné sekvenovat ve velkých objemech a začít studovat genom jako celek. Nejúspěšnější z nich je Illumina sequencing. Nanopore a PacBio patří mezi metody třetí generace, které přinášejí možnost číst větší úseky DNA v celku (long read sequencing).

"Ultimátnou sekvenačnou platformou sa stáva Oxfordské nanopórové sekvenovanie"

Oxford Nanopore má malý podíl, menší než technologie PacBio. Zdaleka nejvíce sekvencí DNA se dnes celosvětově čte na přístrojích Illumina.

Obsahuje-li práce i vlastní výsledky uchazeče/ky (nejsou povinnou součástí práce), pak prosíme o Vaše stanovisko k následujícímu:

Jsou řádně stanoveny a vysvětleny cíle experiment?

Je množství experiment adekvátní k cílům?

Je dokumentace výsledků dostačující?

Jsou výsledky diskutovány a zasazeny do kontextu existující literatury?

Případný další slovní komentář k výsledkům autora:

B. Obhajoba

Dotazy k obhajobě (povinná část posudku)

Výhrady:

- github repository by mělo obsahovat README či INSTALL soubor popisující, jakým způsobem nainstalovat potřebné závislosti. Soubor by měl rovněž popsat, jak spustit testovací scripty
- scripty by měly být psané tak, aby se při chybě jednoho z příkazů přerušil celý script, to se dá zajistit např. pomocí `set -e`
- opětovné spuštění přerušeno scriptu by nemělo vést k chybě, např. pokud stahovaný VCF soubor již existuje. Ideálně by script měl zajistit, aby se velký soubor nestahoval znovu, pokud již existuje

C. Celkový návrh

Práci **doporučuji** k přijetí k dalšímu řízení: **ANO** / **NE**

Navrhovaná celková klasifikace: 2

Datum vypracování posudku: 10.6.2024

Jméno a příjmení, podpis oponenta (SIS): Petr Daněček

Instrukce pro vypracování a odevzdání posudku:

- Pro vypracování posudku bakalářské práce použijte tento formulář.
- Posudek můžete sami vložit do SIS, anebo s předstihem zaslat v elektronické podobě na adresu: marian@natur.cuni.cz, a dále zajistit dodání podepsaného originálu (v 1 výtisku, jako součást protokolu o obhajobě) na sekretariát Katedry buněčné biologie PŘF UK (p. Růžičková), Viničná 7, 128 44 Praha 2. Podepsaný originál posudku musí být dodán před vlastní obhajobou, bez něho nesmí být obhajoba zahájena!

- Student by měl být s posudkem seznámem nejméně tři dny před obhajobou – posudek mu můžete poslat Vy nebo Váš přepošleme.