

Posudek oponenta bakalářské práce

Jméno a příjmení uchazeče/ky : Magdaléna Turínská

Název práce: Vizualizace funkcí podobných proteinů

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 nekoresponduje 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 četnějšími drobnými závadami
	N - nevyhovující, s hrubými chybami

3. Uvedení použitých literárních a j. zdrojů	
	A - bez připomínek, všechny převzaté údaje s citací zdroje, celkový počet citací odpovídá charakteru práce
x	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	
	A - výborný, práce je napsána čtivě a srozumitelně, bez závažnějších gramatických n. pravopisných chyb
x	B - velmi dobrý, ojedinělé stylistické neobratnosti, gramatické n. pravopisné chyby
	C - uspokojivý, četnější 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	
	A - výborná, bez překlepů a chyb ve formátování
x	B - velmi dobrá, ojedinělé chyby formátu citací, překlepy, chybějící zkratky apod.
	C - uspokojivá, s ojedinělými většími (např. vynechání stránky) nebo četnějšími drobnými chybami
	N - nevyhovující, s četnými hrubými chybami

Případný slovní komentář k bodům 1. až 5.:

Magdaléna Turínská předkládá obsáhlou bakalářskou práci na velmi aktuální téma predikce funkce protein. Práce se vyznačuje podrobnými teoretickými úvody k problematice (včetně seznamu 20 základních amnokyselin) a někdy až banálními komentář, že 30 hodin je více než jeden, aniž by však zdůraznila že Needleman-Wunsch algoritmus narušil od heuristických metod zaručuje nalezení optimálního alignment. Líbilo se mi však, že z textu je znát, že autorka poctivě přemýšlela o základních předpokladech a principech používaných metod. Naopak bych více ocenil pokud by např. v oblasti metod pro predikci funkce šla do většího detailu a třeba zmínila silné a slabé stránky jednotlivých přístupů, když už ne metod – jako čtenář jsem pochopil, že metod je hodně, ale ne moc více. Chyběla mi tam jakákoliv zmínka o kvalitě daných nástrojů a případného srovnání jejich kvality, z které by třeba vyplývala potřeba vzniku nového nástroje. V této oblasti funguje už dlouho iniciativa CAFA, kterou vede Iddo Friedberg, a která je inspirována soutěží CASP, kterou autorka zmiňuje. Domnívám se také, že obsáhlé tabulky s webovými adresami jsou extrémně užitečné a oceňuji, že autorka zkontrolovala současnou dostupnost všech metod – webové adresy však mohly být spíše v přílohách. Místo nich se mohl v hlavním textu objevit alespoň jeden příklad výstupu autorčina programu, který je teď skrytý v přílohách. Citovaných prací je velké množství, i když v konkrétních příkladech jsem byl překvapen jakou práci text odkazuje – např. věta: *Prohledáváme databázi D neredundantních sekvencí, které v dubnu 2024 obsahovala 722 278 205 sekvencí*“ vede k citaci o programu Blast. Formát citací není 100 % konzistentní – v seznamu se objevuje Bioinformatics i BIOINFORMATICS apod. Neobvyklé je, že už při více než třech autorech není uveden plný seznam autorů, i když to není chyba. Autorka přidává rozsáhlý poznámkový aparát poznámek pod čarou s různými funkcemi, který by však šel někdy dobře vyřešit klasickou citací a pro lidi se slabším zrakem jako jsem já bylo snadné odkaz na poznámku pod čarou v textu identifikovat, a pak jsem je musel dohledávat. Jazyk práce je velmi dobrý ale některé pasáže jsem nepovažoval za podstatné pro prezentované téma (třeba specifika PDB formátu nebo příklad FASTA formátu sekvence). Líbilo se mi tabulky a obrázek 1.2. Chybělo mi také nějaké ucelené shrnutí kvality a limitů dosud známých metod ve všech oblastech, které autorka představuje.

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? ano

Je množství experiment adekvátní k cílům? ano

Je dokumentace výsledků dostačující? S výhradami

Jsou výsledky diskutovány a zasazeny do kontextu existující literatury?

ne

Případný další slovní komentář k výsledkům autora:

Moc se mi líbil samostatný vědecký projekt autarky – práce integruje několik různých typů dat a poskytuje zajímavou vizualizaci výsledků. Mám pár komentářů - Zdá se mi, že k práci existuje minimálně tato relevantní citace od Martina Steineggera a Pedra Beltra, která systematicky mapuje výsledky klastrování Foldseek hledání na GO anotace (<https://www.nature.com/articles/s41586-023-06510-w>). Ocenil bych kdyby autorka uvedla více detailů k metodice a na závěr blíže představila svou představu, jak by bylo správně naložit s výstupy z jejího programu a interpretovat tyto výsledky. Osobně bych se pokusil ukázat program na sekvenci, kterou mohu zveřejnit. Chybí mi diskuse k praktické části.

B. Obhajoba

Dotazy k obhajobě (povinná část posudku)

Je již program k dispozici na Githubu či jako webový server?

Jaké je nastavení programu BLAST? O jakou verzi se jedná, jaká skórovací tabulka je použita, jaké gap penalties?

Není 90% sekvenční identita příliš vysoká – nebudou mít velmi podobné anotace i sekvence s 80% sekvenční identitou? Zkoušeli jste jinou hranici sekvenční identity? Zkoušeli jste použít více než 10 sekvencí?

Jak má uživatel číst výstup z Vašich vizualizací?

3Lo by automatizovat analýzu těchto vizualizací, aby šlo dělat predikce např. pro celý genom?

C. Celkový návrh

Práci **doporučuji** k přijetí k dalšímu řízení: **ANO mimo vší pochybnost a s radostí.**

Navrhovaná celková klasifikace: velmi dobře

Datum vypracování posudku: 6.9. 2024

Jméno a příjmení, podpis oponenta (SIS):

Marian Novotný

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.