

**I. IDENTIFIKAČNÍ ÚDAJE**

<b>Název práce:</b>	Nástroj pro objevování sekvenčních motivů v RNA-Seq datech
<b>Jméno autora:</b>	Petr Schimperk
<b>Typ práce:</b>	diplomová
<b>Fakulta/ústav:</b>	Fakulta elektrotechnická (FEL)
<b>Katedra/ústav:</b>	Katedra počítačů
<b>Vedoucí práce:</b>	Martin Pospíšek
<b>Pracoviště vedoucího práce:</b>	Katedra genetiky a mikrobiologie, PŘF UK

**II. HODNOCENÍ JEDNOTLIVÝCH KRITÉRIÍ**

<b>Zadání</b>	<b>náročnější</b>
<i>Hodnocení náročnosti zadání závěrečné práce.</i>	
Jsem molekulární biolog, proto je pro mne těžké zhodnotit náročnost zadané práce. Vycházím z toho, že podobnou práci jsem žádal bez úspěchu před časem po profesionálních bioinformaticích, kteří ji však vyhodnotili jako příliš náročnou na čas a úsilí.	

<b>Splnění zadání</b>	<b>splněno s menšími výhradami</b>
<i>Posuďte, zda předložená závěrečná práce splňuje zadání. V komentáři případně uveďte body zadání, které nebyly zcela splněny, nebo zda je práce oproti zadání rozšířena. Nebylo-li zadání zcela splněno, pokuste se posoudit závažnost, dopady a případně i příčiny jednotlivých nedostatků.</i>	
Jak to často bývá, problém se v průběhu práce košatil a řešení se ubíralo i neočekávanými směry. V některých případech pak zbylo málo času na dokončení původně plánovaných úloh. Mezi úlohy, které zůstaly nedodělané a neobjevily se tak v závěrečné verzi práce, patří zejména navržení zpracování konkrétních sekvencí z RNA-Seq dat. Navržený software využívá názvy genů získaných některou z analýz RNA-Seq dat, ale následně zpracovává sekvence stažené k jednotlivým genům z databáze Ensembl pomocí nástroje BioMart. Je nutno podotknout, že pro naprostou většinu aplikací je tento způsob hledání motivů dostačující a v současné době pravděpodobně převažuje. Na druhou stranu se Petr Schimperk pustil do přípravy rozsáhlého grafického rozhraní, které by uživatelsky přívětivým způsobem umožnilo biologům analyzovat jejich data. Jedná se významné rozšíření cílů práce. V tomto úsilí byl Petr Schimperk úspěšný a vytvořil nástroj, který může obsluhovat bez problémů běžný uživatel a já bych mu chtěl touto cestou za vytvořený software poděkovat.	

<b>Aktivita a samostatnost při zpracování práce</b>	<b>A - výborně</b>
<i>Posuďte, zda byl student během řešení aktivní, zda dodržoval dohodnuté termíny, jestli své řešení průběžně konzultoval a zda byl na konzultace dostatečně připraven. Posuďte schopnost studenta samostatně tvůrčí práce.</i>	
Petr Schimperk si nechal na práci málo času, což se mu vymstilo tím, že nestihl udělat vše, co si předsevzal a samotné psaní práce pro něj bylo velmi stresující. Při tvorbě softwaru i při psaní práce byl velmi samostatný, což bylo i nezbytné, protože s vlastním programováním jsem mu nemohl nijak poradit. Konzultace byly pravidelné a pan Schimperk byl na ně vždy perfektně připraven.	

<b>Odborná úroveň</b>	<b>B - velmi dobře</b>
<i>Posuďte úroveň odbornosti závěrečné práce, využití znalostí získaných studiem a z odborné literatury, využití podkladů a dat získaných z praxe.</i>	
Text práce, který vznikl hekticky během velmi krátké doby a velmi samostatně, je podle mého soudu překvapivě dobrý. Pan Schimperk dokázal nastudovat celou problematiku hledání sekvenčních motivů v RNA a zpracovat ji ve své práci přehledně, i když poněkud úsporně. Rozvržení textu je spíše netradiční. Testování dat bych si představoval uvedené raději před diskusí, jako součást výsledků. Velmi mne však potěšil výsledek testování, který kromě potvrzení dobré funkčnosti softwarového balíku ukázal též na důležitost současného používání různých přístupů pro vyhledávání motivů v rámci jednoho souboru dat. Zároveň student našel i nové nepublikované motivy, respektive rozšíření známých motivů. Výsledný software je použitelný obecně pro různá data na rozdíl od výchozího jednoúčelového nástroje vytvořeného Janem Holčákem.	

<b>Formální a jazyková úroveň, rozsah práce</b>	<b>C - dobře</b>
---	------------------

*Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku.*

Ačkoliv byla práce sepsána v rekordním čase, obsahuje relativně malé množství překlepů a jazykových neobratností. Dovedl bych si však představit větší rozsah výsledkové části, více testování a rozsáhlejší úvod a zejména diskusi, která by šířeji rozebrala srovnání autorova řešení s výchozím řešením Jana Holčáka a s řešeními publikovanými v odborné literatuře.

**Výběr zdrojů, korektnost citací**

**D - uspokojivě**

*Vyjádřete se k aktivitě studenta při získávání a využívání studijních materiálů k řešení závěrečné práce. Charakterizujte výběr pramenů. Posuďte, zda student využil všechny relevantní zdroje. Ověřte, zda jsou všechny převzaté prvky řádně odlišeny od vlastních výsledků a úvah, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami.*

Pokud jsem schopen posoudit, jsou literární zdroje voleny dobře. Nicméně autor často cituje až po delším úseku textu. Bylo by vhodnější citovat častěji a vkládat relevantní citace hned za jednotlivá tvrzení. Doporučoval bych se též více zaměřit na původní zdroje a méně citovat tzv. přehledové články. Počet citací by byl v takovém případě výrazně vyšší. Je třeba počítat s tím, že při psaní přehledových článků často dochází k určitému posunu významu a některé jednotlivé informace mohou být interpretovány jinak, než zamýšlel původní autorský kolektiv.

**Další komentáře a hodnocení**

*Vyjádřete se k úrovni dosažených hlavních výsledků závěrečné práce, např. k úrovni teoretických výsledků, nebo k úrovni a funkčnosti technického nebo programového vytvořeného řešení, publikačním výstupům, experimentální zručnosti apod.*

Bez dalších komentářů.

### **III. CELKOVÉ HODNOCENÍ A NÁVRH KLASIFIKACE**

*Shrňte aspekty závěrečné práce, které nejvíce ovlivnily Vaše celkové hodnocení.*

Nehledě na dílčí kritiku v jednotlivých bodech hodnocení bych rád vyzdvihнул aktivní přístup studenta k řešenému projektu a celkový praktický přínos předložené práce. Vznikl software, který bude bezpochyby rutinně používán v naší laboratoři a laboratořích spolupracujících. Někteří kolegové se již na něj těší. Pokud by se podařilo učesat některé nedodělky a případně dokončit i část pracující přímo se sekvencemi z RNA-Seq, bylo by možné řešení šířeji zveřejnit v odborné komunitě. Chtěl bych vyzdvihnout navržené zohlednění počtu čtení v RNA-Seq datech při hledání motivů. Pokud je mi známo, tento nápad není v současné době v analýzách uplatňován, ačkoliv zohlednění počtu čtení jednotlivých motivů může celkovou interpretaci výsledku hledání motivů např. v RNA-IP-Seq datech výrazně pozměnit a pravděpodobně i více přiblížit realitě.

Předloženou závěrečnou práci hodnotím klasifikačním stupněm **C - dobře**.

Datum: 15.6.2024

Podpis: