

I. IDENTIFIKAČNÍ ÚDAJE

Název práce:	Detekce selekčních tlaků na lidské endogenní retroviry
Jméno autora:	Tuan Anh Ho
Typ práce:	bakalářská
Fakulta/ústav:	Fakulta elektrotechnická (FEL)
Katedra/ústav:	počítačů
Oponent práce:	Ing. Anh Vu Le
Pracoviště oponenta práce:	ČVUT FEL, Katedra počítačů

II. HODNOCENÍ JEDNOTLIVÝCH KRITÉRIÍ

Zadání	náročnější
<i>Hodnocení náročnosti zadání závěrečné práce.</i>	
Zadáním je detekce selekčních tlaků na lidské endogenní retroviry. Z infromatického hlediska jde o výpočet statistických testů, což je samo o sobě relativně přímočarý postup. Co dělá ale práci netriviální je pochopit biologickou stránku problematiky a sémantiku za těmito testy. Jde o velice specifický problém, který vyžadoval důkladnou rešerši a z tohoto důvodu zadání práce hodnotím proto jako náročnější	

Splnění zadání	splněno
<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>	
Cílem práce pak bylo seznámit se s 5 metodami odhadu selekčních tlaků na lidské endogenní retroviry. Implementovat kalkulace těchto odhadů a aplikovat je na odpovídající části lidského genomu. Posledním bodem zadání bylo porovnat tyto odhady se známými hodnotami z databáze <i>PopHuman</i> . Všechny tyto body student splnil, navíc našel pravděpodobnou chybu v datech referenční databáze.	

Zvolený postup řešení	vynikající
<i>Posuďte, zda student zvolil správný postup nebo metody řešení.</i>	
Student použil relevantní datový zdroj a data korektně předzpracoval, aby byly výsledky porovnatelné s referenčními postupy. Vypočítané odhady pak porovnal s hodnotami referenčními a ověřil správnost implementace.	

Odborná úroveň	B - velmi dobře
<i>Posuďte úroveň odbornosti závěrečné práce, využití znalostí získaných studiem a z odborné literatury, využití podkladů a dat získaných z praxe.</i>	
U odborné stránky oceňuji, že se studentovi podařilo zvládnout obtížný průnik oborů. V úvodu a teoretické části dokázal velmi strukturovaně a důkladně představit biologickou stránku problému. Bylo znát, že problematice rozumí i z jiné než programátorské stránky. V praktické části pak naimplementoval zadané odhady a byl schopný ověřit jejich korektnost. Velkým plusem je také spolupráce s externím výzkumným střediskem a reálné nasazení výsledků práce. Velká škoda ale je, že jsou všechny výsledky a postup práce shrnuty v pouhých 3 stránkách a dvou tabulkách. Ukazuje se na nich porovnání s referenční databází, což je relevantní ale mohlo být prezentováno více. Určitě by bylo zajímavé vidět např. distribuci hodnot selekčních tlaků napříč lidským genomem – kde jsou vysoké hodnoty, kde nízké – a popřípadě biologická interpretace těchto významných hodnot. Také by stálo za to porovnat zmíněné odhady – v jakých případech se od sebe odchyľují a proč. Z těchto důvodů hodnotím za B – velmi dobře.	

Formální a jazyková úroveň, rozsah práce	A - výborně
<i>Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku.</i>	
Z hlediska jazykové úrovně nemám výhrad. Práce byla psaná v angličtině a její úroveň byla bezpochyby hodná i větších prací, než je ta bakalářská. Formulace byly jasné, text srozumitelně strukturovaný, slovní zásoba vysoce nadstandartní.	

Výběr zdrojů, korektnost citací

A - výborně

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.

Práce se zdroji je rovněž bez výhrad. Student k pochopení problematiky a sepsání práce použil široké spektrum literatury a provedl důkladnou rešerši stávajících řešení. Zdroje odpovídajícím způsobem citoval.

Další komentáře a hodnocení

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

Vložte komentář (nepovinné hodnocení).

III. CELKOVÉ HODNOCENÍ, OTÁZKY K OBHAJOBĚ, NÁVRH KLASIFIKACE

Shrňte aspekty závěrečné práce, které nejvíce ovlivnily Vaše celkové hodnocení. Uveďte případné otázky, které by měl student zodpovědět při obhajobě závěrečné práce před komisí.

Práce je velmi kvalitní a má mnoho kladných bodů – úctyhodná je jak podrobná teoretická část, tak praktické nasazení a vysoká úroveň textu. Vysokých standardů teoretické části se však nepodařilo dostát ve stručné praktické části a závěru. Vzhledem k tomu, že by tato část měla tvořit polovinu práce, nemohu i přes všechny klady hodnotit nejvyšším stupněm.

Předloženou závěrečnou práci hodnotím klasifikačním stupněm **B - velmi dobře**.

Otázky:

1. V práci máte tabulku s ukázkami hodnot testů. Je vidět, že některé oblasti genomu mají hodnotu větší než 2 (resp. menší než -2). Je možné dané hodnoty interpretovat stejně jako v t-testech jako významné? Pokud ano, jaká je interpretace významnosti?
2. U popisu metod odhadu uvádíte, že se jednotlivé metody liší v tom, na co jsou v datech citlivé. V appendixu ve srovnání dvou z nich můžeme vidět, že se někdy shodují, někdy se liší až o řád. Jak tato pozorování interpretujete?
3. Váš nástroj předpokládá bi-alelický model. Uvádíte, že je to zjednodušení, které ve skutečnosti příliš neplatí. Jak moc tento předpoklad limituje analýzu genomu Vaším nástrojem?

Datum: 2.6.2020

Podpis:

