

Posudek dizertační práce

Autorka dizertační práce:

Ing. Helena Škutková

Název práce:

Vyhodnocení příbuznosti organismů pomocí číslíkového zpracování genomických dat

Předložená dizertační práce se zabývá číslíkovým zpracováním genomických dat pro účely komparativní genomiky. Autorka v práci představuje originální přístup k popisu evoluce v sekvenci DNA spočívající jednak ve využití signálové reprezentace sekvencí DNA a dále v použití zarovnávacích metod a vzdálenostních metrik pro genomické signály za účelem jejich klasifikace či fylogenetické analýzy.

Téma práce je podnětné a vědecky aktuální – s ohledem na současný rapidní rozvoj oborů bioinformatika a číslíkové zpracování genomických signálů (GSP – Genomic Signal Processing). Práce zasahuje do řady vědních oblastí od molekulární biologie přes bioinformatiku až po zpracování a analýzu číslíkových signálů, a vykazuje tedy vysokou míru interdisciplinarity.

Autorka ke zpracovávanému tématu přistupuje systematicky a problematiku jednotlivých kapitol popisuje velmi podrobně a srozumitelně. Úvodní části vysvětlují teoretické pozadí analýzy sekvencí DNA v jejich běžné, tedy znakové podobě, a to včetně představení základních principů existujících algoritmů pro zarovnávání sekvencí a pro konstrukci fylogenetických stromů. Ve druhé kapitole je podán přehled vybraných existujících metod pro konverzi znakové reprezentace sekvence DNA na reprezentace numerické a signálové, přičemž pro vlastní výzkum si autorka vybrala na základě předem stanovených požadavků metodu rozbalené fáze. Dále je v této kapitole podán přehled metod pro klasifikaci sekvencí DNA převedených do různých reprezentací a zároveň jsou tyto metody porovnány s využitím jednotného souboru dat.

Po třetí kapitole, která sumarizuje cíle disertační práce, se autorka věnuje výhradně vlastním návrhům a validacím metod pro zpracování genomických signálů – v kapitole 4 – a pro jejich analýzu – v kapitole 5. Na pole zpracování genomických signálů přichází autorka s originální ideou využít metodu dynamického borcení časové osy (DTW), obvyklou v oblasti zpracování a analýzy řečových signálů. Metodu DTW různými způsoby modifikuje a využívá pro zarovnávání signálových reprezentací sekvencí DNA tak, aby byla možná následná klasifikace genomických signálů i při jejich různých délkách způsobených mutacemi typu inserce-delece. Na pole fylogenetické analýzy signálových reprezentací sekvencí DNA pak autorka přináší návrh metriky nazvané proporcionální vzdálenost genomických signálů, s jejíž pomocí pak rozlišuje mezi signály obsahujícími bodové změny, a to lépe než v případě použití Euklidovy vzdálenosti. V závěru práce jsou pak demonstrovány výsledky fylogenetických analýz provedených nad sekvencemi obvyklých genových markerů z veřejně dostupné databáze a také ovšem nad velmi dlouhými sekvencemi celých genomů různých organismů.

Práce je napsaná v českém jazyce, její rozsah je 118 stran včetně všech formálních náležitostí. Samotný text práce na 98 stranách je rozdělen do 5 kapitol s přehlednou a logickou hierarchií podkapitol. Práce obsahuje 73 obrázků, 13 tabulek a odkazuje na 185 literárních pramenů. Po formální stránce nelze autorce vytknout nic závažného – pouze několik nevýznamných gramatických

či stylistických chyb, kterých je v úvodních kapitolách až nadbytek (str. 9: „formy, které se z nich odvinuli“; str. 9: „tyto základní kritéria“; str. 20 obr. 1.7: chybí slovo „vzdálenosti“). Tyto drobné nedostatky však nesnižují celkově velmi pozitivní dojem z formálního zpracování práce i z jejího samotného obsahu.

Zadání k vypracování oponentského posudku vyžaduje, aby se oponent vyjádřil k následujícím bodům:

1) Odpovídá námět práce oboru disertace a je aktuální z hlediska současného stavu vědy?

Ano. S odkazem na druhý odstavec mého posudku – dizertační práce Ing. Heleny Škutkové dobře zapadá do oboru Biomedicínská elektronika a biokybernetika.

2) Vykazuje práce původní přínosné části? Konkretizujte prosím, v čem spatřujete originální přínos

Ano, práce vykazuje původní přínosné části.

Hlavní vědecký přínos této práce spatřuji v originálním přístupu ke klasifikaci sekvencí DNA s využitím signálových reprezentací a dynamického borcení časové osy. Autorka se svým přístupem k analýze genomických dat dosahuje oproti běžně používaným postupům založeným na symbolických reprezentacích DNA vyšší přesnosti i podstatně nižší výpočetní složitosti. Vyšší přesnost je dosažena jednak v části zarovnávání genomických signálů, a to použitím techniky dynamického borcení časové osy s výpočtem korelace v klouzavém okně. Dále je pak vyšší přesnost dosažena i při samotné fylogenetické či fylogenomické analýze, neboť nízká výpočetní náročnost dosahovaná díky téměř bezztrátové decimaci signálů umožňuje pro klasifikace využít sekvence či signály s řádově větší délkou, a tedy potencionálně použít pro analýzu více diskriminační informace.

3) Bylo jádro disertační práce na potřebné úrovni publikováno?

Z přehledu tvůrčích aktivit doktorandky a z dodaných reprintů článků vyplývá, že své výsledky publikovala v řadě článků na konferencích i v časopisech. Je potřeba zejména vyzdvihnout prvoautorskou publikaci v časopisu *Journal of Theoretical Biology* (2015, IF = 2.12), prvoautorskou publikaci v suplementu časopisu *BMC Bioinformatics* (2013, IF = 2.67) a patrně podstatné spoluautorství publikace v časopisu *Computers in Biology and Medicine* (2016, IF = 1.24). Autorka má výraznou stopu v rejstřících Scopus (23 výsledků, 224 citací, h-index: 7) a WebOfScience (21 výsledků, 208 citací, h-index: 7).

4) Vyplývá ze seznamu vědecké činnosti uchazečky, že se jedná o pracovníka s vědeckou erudicí?

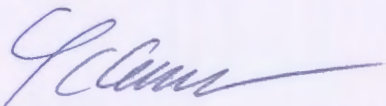
Ano, uchazečka publikovala řadu prací v časopisech a dále na domácích a zahraničních konferencích. Vytvořila nebo se podílela na tvorbě řady softwarových produktů (14 ×). Z pozice hlavní řešitelky získala a vyřešila FRVŠ projekt inovace výuky v předmětech zaměřených na genomiku a proteomiku na FEKT VUT v Brně.

V závěru svého oponentského posudku konstatuji, že předložená práce Ing. Heleny Škutkové splňuje požadavky obvykle kladené na úroveň dizertačních prací. Dizertační práce i přehled tvůrčích aktivit doktorandky vyhovuje nad rámec obecně uznávaným požadavkům k udělení titulu Ph.D.

Při obhajobě dizertační práce bych rád předložil následující dotazy:

- 1) Na Vašem pracovišti byla nedávno (2015) obhájena doktorská práce zaměřená na molekulární identifikaci organismů s využitím reprezentace sekvencí DNA pomocí nukleotidových denzitních vektorů a Euklidovy vzdálenosti (dr. Maděránková). Můžete srovnat hlavní rozdíly Vašeho přístupu s výše zmíněným?
- 2) Použitím decimace signálů může dojít – jak sama v práci na mnoha místech uvádíte – ke ztrátě užitečné informace pro následující klasifikaci. Při volbě decimálního faktoru se opíráte o kritérium zachování 99.5% energie signálu. Mohou signály odvozené ze sekvencí DNA Vámi zvolenou metodou rozbalené fáze nést nějakou užitečnou diskriminační informaci v zanedbaných složkách? Souvisí nějak energie signálových složek s jejich diskriminační schopností?
- 3) Signálové reprezentace spolu s Vaším přístupem pro jejich zpracování a analýzu využíváte pro hodnocení příbuznosti organismů – tedy pro obecně biologické téma. Domnívám se, že díky dosažené nízké výpočetní složitosti se nabízí praktická využitelnost i na poli klinické medicíny. Vidíte pro tuto oblast nějaký konkrétní potenciál Vašich metod? Nabídnete Váš software např. klinickým pracovištím zaměřeným na lékařskou genetiku?

V Brně dne 3.3.2016



Daniél Schwarz