

I. IDENTIFIKAČNÍ ÚDAJE

Název práce:	Explainable and transferable fungal intron models
Jméno autora:	Barbora Mašková
Typ práce:	diplomová
Fakulta/ústav:	Fakulta elektrotechnická (FEL)
Katedra/ústav:	Katedra počítačů
Oponent práce:	RNDr. Martin Pospíšek, Ph.D.
Pracoviště oponenta práce:	Přírodovědecká fakulta Univerzity Karlovy, Kat. genetiky a mikrobiologie

II. HODNOCENÍ JEDNOTLIVÝCH KRITÉRIÍ

Zadání	náročnější
<i>Hodnocení náročnosti zadání závěrečné práce.</i>	
Předem musím zmínit, že nejsem informatik. Mojí odborností je molekulární biologie a zaměřím se více na hodnocení práce z hlediska jejího využití v biologickém výzkumu. Cílem bylo analyzovat stávající neuronovou síť navrženou pro detekci mRNA intronů v metagenomech hub a pokusit se najít minimální počet modelů potřebných pro hledání intronů v druhově rozmanité říši hub.	

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>	
Podle mého soudu studentka zadání splnila. Během své práce však zjistila, že zvolená metoda nevede k vytvoření univerzálního modelu pro všechny houby. Nakonec v rámci této diplomové práce zvolila jako optimální 4 modely pro 5' konce intronů a 5 modelů pro 3' konce intronů. Sama však podtrhuje v závěru, že práce zejména ukazuje cestu a nabízí potenciál k dalšímu výzkumu včetně případné přípravy univerzálnějších modelů pro detekci intronů v houbových genomech.	

Zvolený postup řešení	správný
<i>Posuďte, zda student zvolil správný postup nebo metody řešení.</i>	
Pokud jsem byl schopen posoudit, jeví se mi zvolený postup jako správný. Jednotlivé úseky práce jsou dobře a logicky vysvětleny včetně provázání informatické části a biologických východisek. Práce vyžaduje od studenta kromě znalostí informatiky a statistiky též určitou orientaci v taxonomii hub a jistou úroveň pochopení složitého procesu vystřihování intronů z eukaryotní pre-mRNA.	

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>	
Vzhledem k mé odbornosti bylo pro mne čtení předložené práce poněkud náročné, nicméně oceňuji, jakým způsobem autorka vysvětlila informatickou část práce i její relativně úspěšnou snahu vyrovnat se s popisem biologické části. Jak píše výše, práce do určité míry závisí na pochopení taxonomie hub a vhodném rozřídění genomů do jednotlivých taxonomických kategorií. Z těchto důvodů bych doporučil použití modernější taxonomie hub než té, která byla použita v práci. Například dle Miguel A. Naranjo-Ortiz, Toni Gabaldón et al., Fungal evolution: diversity, taxonomy and phylogeny of the Fungi. 2019. Biological Reviews, https://doi.org/10.1111/brv.12550 .	

Formální a jazyková úroveň, rozsah práce	C - dobře
<i>Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku.</i>	
Práce je napsána anglicky, je úpravná a graficky přehledná. Při orientaci v práci mi scházeli legendy u obrázků a tabulek. Orientace čtenáře v grafické části práce včetně příloh je tedy náročnější. Výsledková část je napsána velmi úsporně. Vědecká diskuse autorčiných výsledků a jejich konfrontace s odbornou literaturou včetně širšího formulování dalšího možného vývoje tohoto výzkumu a případných hypotéz vyplývajících ze získaných výsledků v podstatě chybí. Přítomná diskuse je svým obsahem spíše pokračováním výsledkové části. To je zřejmé i z toho, že diskuse neobsahuje žádné odkazy na odbornou	

literaturu.

Výběr zdrojů, korektnost citací

C - dobře

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 obsahuje 57 citací. Přesto si myslím, že vzhledem k tomu, že práce kombinuje vzdálené obory jako je informatika, taxonomie hub a molekulární biologie, mohlo by být citací více. Autorka s citacemi šetří. Například v diskusi se žádné nevyskytují. Vzhledem k tomu, že řazení jednotlivých genomů do taxonomických jednotek následně ovlivňuje trénování modelů a získané výsledky, doporučil bych zabývat se tímto tématem hlouběji a využít modernější zdroje, než jaké byly v práci použity. Autorka používá číselné odkazy na literaturu v textu práce, které však nejsou vhodné, pokud jsou citace v seznamu literatury řazeny abecedně. Při abecedním řazení seznamu literatury je nutné začínat odkaz příjmením, dle kterého je seznam řazen, a nikoliv křestním jménem. Některé zdroje nejsou dostatečně citovány, např. [7] Suzanne Clancy, 2008.

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.

Přes drobné výtky považuji práci za zdařilou a její výsledky za zajímavé. Další komentáře uvádím níže.

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.

Předloženou práci studentky Barbory Maškové považuji za zdařilou a výsledky za zajímavé.

Doporučil bych se více věnovat taxonomickému zařazení použitých genomů, protože může výrazně ovlivnit získané výsledky a jejich interpretaci. Rád bych se zeptal, v čem byl problém při vytváření datové sady pro *Cryptomycota*? Jedná se o poměrně obskurní a spíše „sběrnou“ taxonomickou jednotku. Její vyřazení z důvodu malého počtu vzorků možná nebylo třeba, protože někteří autoři řadí *Cryptomycota* spolu s další heterogenní skupinou *Microsporidia* do (nad)kmene *Opisthosporidia*. Možná stačilo pouze přiřadit jednu analyzovanou kryptomycetu ke skupině 8 analyzovaných mikrosporidií.

Velmi se mi líbí použití log pro analýzu sekvencí. Bohužel ne vždy se mi podařilo se v nich vyznat. Nemyslím si, že by tento přístup byl nějak přínosný pro vylepšení detekce intronů, ale může přinést zajímavé poznatky o stavbě intronů a sestřihového aparátu v jednotlivých taxonech hub a z tohoto hlediska bych doporučil v této práci v budoucnu v rozšířené formě pokračovat. Bylo by možné ukázat loga pro všechny zástupce jednotlivých modelů s grafickým vyznačením 5' a 3' sestřihového místa? Bylo by možné filtrovat vystřižené introny tak, aby pro tvorbu loga byly využity je ty, které poskytují v přilehlém exonu dostatečně dlouhou kódující oblast nepřerušovanou STOP kodónem? Pokud by to bylo možné, myslíte si, že by tato úprava vedla k výraznější změně log u některých zástupců použitých modelů?

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

Datum: 24.1.2024

Podpis: