

Posudek oponenta na diplomovou práci

oponentský posudek

Jméno posuzovatele:
RNDr. Michal Vinkler, Ph.D.

Datum:
26.01.2019

Autor:

Bc. Kalifa Samaké

Název práce:

Polymorfismus transkripčního faktoru NF- κ B a Toll-like receptoru 2 u produkční populace skotu (*Bos taurus* L.)

Cíle práce

Autor si za cíle práce stanovil (i) vytvoření knihovny genomické DNA ze sledované populace skotu, (ii) přípravu postup pro resekvenování vybraných genů přirozené imunity skotu (*NFKB1*, *NFKB2* a *TLR2*), (iii) zdokumentování diverzity ve vybraných imunitních genech pro sledovanou populaci skotu pomocí resekvenování, (iv) interpretaci nalezené genové varianty a predikci haplotypů, (v) navržení a validaci analytické reakce pro vybrané varianty, (vi) stanovení distribuce genových variant ve sledované populaci pomocí navržených genotypovacích analýz a (vii) interpretaci výsledků ve vztahu k infekční odolnosti skotu.

Z cílů práce nevyplývá, proč se autor ve své metodice (Tab.1) zabývá také dalšími geny, které již v práci dále diskutovány nejsou – doporučoval bych vysvětlit.

Také mi přijde, že některé body jsou spíše metodické povahy, a proto bych doporučoval je jako samostatné cíle nepovažovat (např. i, ii, příp. v, kde se však už může jednat o tvůrčí výsledek). Podobně interpretace výsledků je ve vědecké práci samozřejmostí (body iv, a vii). Osobně bych doporučoval raději méně dobře vymezených a vysvětlených cílů. Ty by však měly být zřejmé také z abstraktu práce, který by měl stručně shrnout i jejich dosažení (některé nejsou v abstraktu zmíněny).

Struktura (členění) práce, odpovídá požadovanému? ANO

Rozsah práce (počet stran): 101, velkou část rozsahu práce tvoří souhrnné tabulky, z nichž některé by bylo lépe vložit do přílohy práce.

Je uveden anglický abstrakt a klíčová slova? ANO

Je uveden seznam zkratk? ANO, ale ne všechny zkratky použité v textu jsou v seznamu uvedeny (např. IRF-3, IFN- β , SMRTbell atp.), což se ale v takto rozsáhlém textu stává.

Literární přehled:

Odpovídá tématu? ANO, ale jen částečně (viz níže)

Je napsán srozumitelně? ANO

Použil(a) autor(ka) v rešerši relevantní údaje z literárních zdrojů? ANO, ale ne všechny citace se specificky vztahují k tématu (viz níže) a některá důležitá tvrzení nejsou opatřena referencí (např. kapitola „NGS“ je prakticky bez citací).

Jsou použité literární zdroje dostatečné a jsou v práci správně citovány? ANO, ale použita

literatura by mohla být rozsáhlejší.

Úvod a literární přehled nejsou příliš uspokojivé - jsou psány jen velmi obecné a neposkytují pro čtenáře dobré vodítko k pochopení specifických cílů práce v kontextu imunity a imunogenetické variability skotu, domácích zvířat či obratlovců (např. pasáž o karyotypu skotu, obecný přehled imunity obratlovců nebo obecná kapitola o sekvenování nové generace jsou spíše nadbytečné). Tyto pasáže by měly být psány více specificky vzhledem k zaměření práce (přehled by měl např. zmiňovat dříve zjištěné poznatky týkající se evoluce a variability vybraných genů u tura, srovnání s jinými savci či obratlovci, funkční význam variability, diverzita ligandů atp.).

Některé informace se v po sobě jdoucích odstavcích opakují, aniž to přispívá k lepšímu vysvětlení souvislostí.

V úvodu autor zmiňuje „nečekaný“ objev TLRs, ale existence těchto receptorů byla dříve predikována (Janeway 1989) – z mého pohledu tedy nebyly objeveny nečekaně a jedná se o jeden z nejvýznamnějších dedukcí moderní imunologie.

NGS je spíše přístup než metoda.

Materiál a metody:

Odpovídají použité metody experimentální kapitole? ANO

Kolik metod bylo použito? Této otázce úplně nerozumím – pokud se NGS a analýza dat počítají jako jedna metoda, pak dvě (NGS a SNaPShot), pokud je každý úkon v rámci postupu práce hodnocen zvlášť (tak, jak jsou popsány v metodice), pak mnoho.

Jsou metody srozumitelně popsány? ANO, ale ne ve všech případech je legitimita zvoleného postupu dostatečně vysvětlena.

Dílní připomínky:

Materiál a metodika by měly začínat rekapitulací materiálu využitého pro danou studii (kolik jedinců, kolik plemen, kde byly získány, jak byly skladovány atp.).

Kap. 4.1., konec prvního odstavce. Pokud docházelo k chybám v automatickém pipetování, jak byl tento problém řešen? (autor neuvádí). V tomto kontextu je překvapivý závěr třetího odstavce, který uzavírá, že plně automatizovaný přístup zaručil vysoce reprodukovatelné výsledky.

Metodika izolace DNA je velice komplikovaná, ale autor nikde nezmiňuje proč – jaké měl daný přístup výhody. Autor dokonce neuvádí informace o výsledné kvalitě extrahované DNA (koncentrace, délka fragmentů), přičemž tyto informace jsou klíčové pro porovnání s jinými přístupy a vysvětlení výhod zvoleného přístupu. Není jasné, proč je srážení krve považováno za znehodnocení vzorku (str. 28).

Proč byla ke krvi přidávána EDTA jako antikoagulant, když je známo, že EDTA funguje jako inhibitor PCR (např. Huggett et al. 2008, doi:10.1186/1756-0500-1-70) – není tento přístup kontraproduktivní? Jaký význam má z hlediska molekulárně genetického využití vzorků bránit srážení krve?

Proč bylo nutné homogenizovat krev pískem před vlastní lýzou? Krev je typický příklad tkáně, která enzymaticky lyzuje velice dobře. Písek byl dekontaminován autoklávováním, ale autoklávování není dobrý způsob dekontaminace DNA – při vysokých teplotách DNA

denaturuje, ale nedegraduje a po ochlazení opět renaturuje. Např. při běžných podmínkách mikrobiologické sterilizace (110°C, 2 bar) zůstává zachováno cca 80% DNA (např. Karni et al. 2013, DOI: 10.1089/dna.2013.2056). Kontaminace tedy velice pravděpodobně přetrvávala, pokud byl stejný materiál používán opakovaně.

Metodika neobsahuje některé důležité informace, které by umožňovaly opakování postupu – např. v jakém přístroji byla provedena fluorescenční kvantifikace DNA popsaná na str. 29? (jak byly snímky digitalizovány – přístroj, nastavení)

Jelikož primery uvedené v Tab. 1 jsou podle kap. 4.4 převzaty z literatury, bylo by správné uvést příslušný zdroj pro jednotlivé primery i v tabulce. Tabulka zahrnuje i primery pro geny mimo tuto práci, což by bylo dobré vysvětlit. Určitý náznak vysvětlení se objevuje na str. 58, ale to je skryto v textu a navíc velice daleko. Krom toho takto rozsáhlou tabulku by bylo lepší uvést formou přílohy.

Poněkud zarážející pro mne je skutečnost, že sekvenované fragmenty (amplikony uvedené v Tab. 1) mají délky pouze 400-1200 bp, což je významně méně než je standardní délka sekvence získatelná při sekvenování na platformě PacBio. Nebyly tedy využity možnosti této platformy. V kap. 5.6 se sice hovoří o „dlouhých čtených úsecích“, ale je potřeba si uvědomit, že čtené úseky mohou být jen tak dlouhé, jak jsou dlouhé vstupní amplikony. Proč se autor omezil jen na takto krátké fragmenty a nepokusil se pomocí long-range PCR získat sekvence celé CDS v jednom čtení? Z tohoto paradoxu pak vyplývá i určitá kontroverze při stanovování sekvence haplotypů (viz dále). Může autor svůj přístup vysvětlit?

Autor uvádí, že primery byly navrženy tak, aby pokrývaly celé exony. To však z Tab. 1 nevyplývá (tabulka neuvádí o které exony se jedná) – schématické zobrazení nasedání primerů na sekvenci genu by bylo výhodou.

Na str. 38 autor zmiňuje kapilární sekvenování, ale o tom v textu výše zatím nebyla řeč.

Popis PCR zmiňuje na str. 38 extenzi při teplotě 72°C o délce 2 minuty, což je hodně s ohledem na relativně krátkou délku většiny produktů.

V postupu přípravy sekvenační knihovny mne překvapilo, že nebyly použity žádné indexy (tagy) pro označení ampliconů jednotlivých jedinců. Proč tomu tak bylo? Důsledkem je značné snížení informativní hodnoty získaných dat. Využití indexů by umožnilo získat sekvence přiřaditelné k jednotlivým vzorkům.

Pokud jsou kroky uvedené v metodice shodné s doporučením výrobce daného kitu, není potřeba jednotlivé kroky popisovat. Je ale naprosto zásadní uvést, jaký kit byl použit – jaký sekvenační kit byl použit?

U úkonů, jakými je např. vyhodnocení sekvencí v programu Geneious je potřeba uvést nastavení programu.

Co přesně je myšleno větou „Pro další vyhodnocení a návrh genotypovacích reakcí byly použity v prvé řadě polymorfismy, jejichž existence byla potvrzena nezávisle ze dvou zdrojů.“? Metodika by měla obsahovat informaci o počtu polymorfních míst validovaných jinými metodami a způsobu validace (poslední odstavec kap. 4.6)

Statistická reprodukce haplotypů zmíněná v kap. 4.7 bývá nepřesná, pokud každá varianta není v datasetu potvrzena např. klonováním. Informace o použitém softwaru a jeho nastavení je v tomto případě zásadní. Na str. 84 (ve výsledcích) se čtenář dozvídá, že byl využit algoritmus PHASE – ten však nelze spolehlivě použít, pokud není v datasetu zastoupena velká část alel v homozygotním stavu, popř. pokud do analýzy nejsou zahrnuty rozřešené sekvence získané na základě klonování.

Chybí velmi důležitá informace o tom, jak velký vzorek byl genotypován přes NGS a jak velký vzorek pak pomocí dalších genotypovacích přístupů.

Kap. 4.9 – nehodnotily se fyzikálně-chemické vlastnosti záměn a jejich konzervativita?

Podle mého názoru by bylo vhodnější velkou část textu zahrnutého v části 5 výsledky přesunout do metodiky, kam logicky patří (např. 5.1, Tab. 3, Tab. 4, 5.3, Tab. 5, Tab. 6, větší část 5.4) – nejedná se o výsledky, tj. data, která by umožňovala vytvořit nějaký náhled na vytčené otázky, ale o aplikaci metod, které k těmto datům vedou. Není nic špatného na tom, mít dlouhou metodickou část a krátké výsledky (pokud je potřeba použít složité metody, aby se člověk dostal k relevantním datům).

Překvapuje mne použití BioEditu pokud byl zároveň používán program Geneious, který plní stejnou funkci, ale je výrazně novější.

Kapitola 5.2 popisuje poměrně složitý proces stanovování hranic exonů v *NFKB1* a *NFKB2* – byly zjištěné hranice exonů jiné než pokud by byly přímo staženy hranice exonů z volně dostupné databáze NCBI nebo Ensembl? Pokud ano a to byl důvod pro takto složitý postup, čekal bych o tom v textu nějakou zmínku.

Jsou v mRNA genů *NFKB1* a *NFKB2* UTR oblasti? Pokud ano, liší se mezi geny?

Legenda k Obr. 7 mluví o optimalizaci Ta, ale z obrázku gelu mi není jasné, v čem přesně tato optimalizace spočívá.

V kapitole 5.4 je zmíněno poolování alikvotů extrahované DNA, ale podle mých zkušeností je lepší poolovat až amplikony – PCR může pro jednotlivé sekvence probíhat s různou efektivitou a výsledné zastoupení variant tak nemusí být stejné.

Metodika jasně neříká, jak se autor vypořádal s artefakty chimérických sekvencí.

Experimentální část:

Je vysvětlen cíl experimentů? NE – práce není ze své povahy experimentální

Je dokumentace výsledků dostačující? NE - viz poznámky níže k výsledkům

Postačuje množství experimentů k získání odpovědí na zadané otázky?

Nejedná se o experimentální výzkum.

Otázky a komentáře k výsledkům:

Na str. 64 mi není jasné, co jsou „strukturní varianty“ zmíněné v první větě.

V Tab. 9 je zarážející hodnota signifikance „0,001>“.

Na str. 72 bych poprosil o vysvětlení tvrzení: „Vzhledem k výskytu rekombinantních haplotypů v amplikonu 2 a 20 nebyl diagnostický soubor zjednodušen. Na druhé straně absolutní a silná vazba markerů v amplikonech 4 a 20 umožnila jednotlivé reakce vynechat.“.

Na str. 77 uvádíte: „Koncová dvojice 1044 a 1047 tvoří v amplikonech 2 a 3 třídy s odlišnou frekvencí. Toto pozorování je obtížné interpretovat a není vyloučena metodická chyba při zpracování souborů s multisekvencemi BAM a FASTA.“ Jak si má čtenář takovou chybu představit? Nehrozí podobná chyba i u vyhodnocení genů *NFKB*? Nemůže mít vliv skutečnost, že různé PCR mají různou efektivitu a tak frekvence detekovaných SNP ve směsi neodpovídá frekvenci alel v populaci? Osobně si myslím, že frekvence SNP získané

sekvenováním ampliconů ze směsných vzorků DNA neodrážejí spolehlivě frekvence alel v populaci. Byl tento přístup předem testován?

Co znamenají prázdná pole v Tab. 15?

Některé závěry týkající se výsledků mi přijdou lezce zavádějící – např. na str. 85 autor zmiňuje, že byla „pozorována dobrá shoda výsledků dvou odlišných metod“, která se však týká pouze dvou pozic v rámci genu *TLR2* a není úplně jasné, jak je definována „dobrá shoda“.

Na str. 87 autor uvádí, že rozsah modelů *NFKB1* a *NFKB2* je omezen stávajícími templáty. Bylo by však možné pro přibližné stanovení celé struktury použít nástroj s funkcí de-novo modelování – např. Phyre2.

Diskuze:

Je opravdu diskuzí, nejde jen o konstatování vlastních výsledků? ANO, ale diskuse je relativně velmi krátká (pouze 6 stran)

Jsou výsledky porovnávány s literaturou? ANO, ale výčet diskutovaných a srovnávaných prací je relativně omezený (pokud počítám dobře, pak do 15 prací, z nichž některé jsou zmíněny spíše povrchně)

Jsou uvedeny nějaké hypotézy či návrhy na další řešení problematiky? NE

Poznámky k diskusi:

Některé diskusní části se vyskytují již v kapitole výsledků (porovnání výsledků v rámci této práce).

Některá tvrzení v diskusi se zdají být v kontextu výsledků práce poněkud nadhodnocená (např. na str. 90: „Byl získán přehled o polymorfismu těchto genů v populaci českého strakatého skotu a byly zhodnoceny možnosti pro využití tohoto polymorfismu pro účely zvýšení rezistence k infekcím.“ – z práce přímo nevyplývá, která polymorfní místa jsou využitelná pro účely zvýšení rezistence k infekcím – k tomu je zdá se ještě celkem daleká cesta ověřování funkčního významu zjištěného polymorfismu).

V práci je opakovaně zmiňována vysoká chybovost sekvenování na platformě PacBio, ale už není vysvětleno, jak byla tato skutečnost v práci zohledněna/ošetřena (autor formuluje konstatování obvykle jako závěr “i přes to bylo možné metodu použít...“ – ano, metodu lze použít i tak, ale není jasné, jaký vliv má tato chybovost na kvalitu získaných výsledků, jak časté byly artefakty – mohly mít např. vliv na počet nově zjištěných polymorfních míst zmiňovaný na str. 91 jako „neobvyklé“ zjištění – mohlo se jednat o artefakty?).

V práci mi pro srovnání s jinými studiemi (a také jinými geny) chybí údaje o nukleotidové, resp. haplotypové diverzitě.

Autor na str. 92 konstatuje, že není známo, že by byly plošně používány plemenářské programy pro zvyšování heterozygotnosti pracující s genetickými markery a proto je překvapivá zvýšená heterozygotnost studovaných jedinců v genu *TLR2*. Nemohou ale stejnou roli pro zvyšování heterozygotnosti plnit plemenné knihy, pokud existují?

Haplotypovou diverzitu by bylo vhodné hodnotit na úrovni alel celého genu, resp. alespoň CDS – kratší úseky jsou voleny obvykle nestandardně a je těžké je porovnat mezi studiemi, geny atp.

Na str. 93, kap. 6.5 autor uvádí pozice pod pozitivní selekcí zjištěné v práci Jann et al. 2008, ale již neuvádí, zda byl polymorfismus na těchto pozicích zjištěn také v této práci. Poznatky zjištěné v této práci je potřeba dát do vztahu s informacemi z literatury, protože jiná se jedná o dvojici mimoběžných informací, ze kterých nelze udělat společný závěr.

Překvapivá je informace uvedená na str. 94 „Shodu publikované mutace v NFKB2 s některým ze SNP nalezených v populaci ČESTR se nepodařilo potvrdit, nicméně není vyloučeno shodu s pozorovanou variantou 7545C>T, pokud bude určena příčina posunu referenčních sekvencí.“ Shodu lze přeci na základě alignmentu určit – anebo tomu něco brání?

Závěry (Souhrn) :

Jsou výstižné? NE - závěr mi přijde poněkud stručný a spíše povrchní – nevyzdvihuje hlavní zjištění práce, spíše obecně konstatuje možný přínos.

Formální úroveň práce (obrazová dokumentace, grafika, text, jazyková úroveň):

Obrazová dokumentace neobsahuje dostatečně vysvětlující legendy – místy není zřejmé, co přesně by měl čtenář na obrázku vidět, co je na něm to důležité. Některé brázky a tabulky by bylo lepší přesunout do přílohy práce.

Překlepů není mnoho, ale v práci jsou (což je ale běžné).

Termín „polymorfismus“ nelze (v češtině ani angličtině) užívat jako počítatelné označení jednotlivých variabilních pozic = „polymorfních míst“ (vyskytuje se na různých místech práce, např. v abstraktu)

Po formální stránce bych v práci doporučoval sjednotit značená polymorfních míst (jednopísmenné vs. trojpísmenné zkratky aminokyselin).

Splnění cílů práce a celkové hodnocení:

Přes všechny dílčí nedostatky uvedené v tomto posudku po vědecké stránce splnila práce svůj účel. Data byla shromážděna, vyhodnocena a jakžtakž diskutována. Ačkoliv si myslím, že měl autor lépe promyslet některé metodické aspekty své diplomové práce než na ní začal pracovat a i poté minimálně mohl věnovat větší pozornost diskusi výsledků, lze konstatovat, že ve všech ohledech hodnocených u diplomových prací potřebné úsilí vynaložil (místy patrně nemalé) a výsledků dosáhl.

Otázky a připomínky oponenta:

Z výše uvedených výtek prosím při obhajobě především o vysvětlení:

1. Jak byl řešen problém chyb v automatickém pipetování (Kap. 4.1)?
2. Proč byl zvolen komplikovaný postup extrakce DNA (vč. použití EDTA jako antikoagulantu) a proč byly ve výsledku sekvenovány pouze krátké fragmenty zvolených genů, když platforma PacBio umožňuje sekvenaci 10-20 000bp i více? Měl zvolený přístup nějakou výhodu? Proč nebyly použity indexy pro individuální značení ampliconů?
3. Jak by mohla vypadat metodická chyba při zpracování souborů s multisekvencemi BAM a FASTA zmíněná pro *TLR2* na str. 77? Nehrozí podobná chyba i u vyhodnocení genů *NFKB*? Nemůže mít vliv skutečnost, že různé PCR mají různou efektivitu a tak frekvence detekovaných SNP ve směsi neodpovídá frekvenci alel v populaci?
4. Jak byla v práci zohledněna vysoká chybovost sekvenování na platformě PacBio?

5. Nakolik uspokojivě mohou být využity plemenné knihy pro zvyšování heterozygotnosti v porovnání s přístupem využívajícím data ke konkrétním genům (genome-wide heterozygosity vs heterozygotnost na specifickém lokusu)?

Návrh hodnocení oponenta (známka nebude součástí zveřejněných informací)

výborně velmi dobře dobře nevyhověl(a)

Podpis oponenta: