

Posudek na diplomovou práci Barbory Chalupové

Haemoproteus u pěvců: prevalence a dynamika infekce

Diplomová práce je založena na úctyhodném vzorku více než tisíce krevních vzorků pěvců. Je potěšitelné, že na značné části odběrů vzorků se autorka sama podílela. Studentka sama detekovala krevní parazity, což při tomto počtu vzorků, použití nested PCR a poctivého vyřezávání z gelů musela být značná řehole. Metodika je pečlivě a podrobně sepsána. Získané sekvence byly srovnány s již dostupnými sekvencemi známých linií. Oceňuji, že se autorka nespokojila jen s identifikací do linie, ale snaží se k liniím přiřadit i známé morfotypy. Vlastní data pak vhodně kombinuje s údaji z krevních roztěrů, která získala školitelka. Autorka také skvěle předem uvádí, které charakteristiky hostitelských druhů by mohly mít vliv na přítomnost parazitů. Tyto charakteristiky pak autorka pěkně uplatnila v diskusi, kde příkladně probírá biologické vlastnosti hostitelů a parazitů. V diskusi jsou také zdařilé úvahy o lokálních přenosech konkrétních linií, což považuji za nejcennější příspěvek celé práce. Obecně lze na základě podrobné a dlouhé diskuse říci, že se autorka v biologii sledovaných druhů a jejich parazitů velmi dobře orientuje. Oceňuji také, že abstrakt práce je napsán velmi dobře srozumitelnou angličtinou.

K práci mám však i řadu kritických připomínek. V několika ohledech mě práce překvapila poněkud nemilým způsobem. Například fakt, že v práci není použit jediný statistický test, mi doslova vyrazil dech. Získaná data přitom přímo žadoní o výpočet. Ortogonální povahu vysvětlujících proměnných a jednoduché rozložení dat mohou autoři jiných studií jen tiše závidět. Jednoduchý model z prvních lekcí statistiky by to zřejmě celé vyřešil na jediném krátkém řádku v R. Na některá srovnání by snad stačil prostý chí kvadrát. Data se zdají být připravena pro test, který se však ke značnému překvapení čtenáře neobjeví. Výsledky tak zůstávají v podobě jakéhosi podivného polotovaru. Vlastně nevíme, jak to celé dopadlo, co se liší a co ne. Místo výsledků tak vidíme jen hrubá data. Je třeba zdůraznit, že všechny údaje se proto týkají jen chyceného vzorku ptáků. Bez jakékoli statistiky těžko říci, co to vlastně vypovídá o zkoumaných populacích.

Vzhledem k fascinujícímu potenciálu vzorků jsou nemile překvapivé také cíle práce. Z cílů práce se zdá, že si autorka neklade žádnou otázku, nechce řešit žádný problém ani vyvracet či podpořit nějakou hypotézu. Cíle jsou pohříchu popisné a upřímně řečeno strašně nezajímavé a nudné. S minimální znalostí literatury by šlo rovnou ze židle ve Viničné odhadnout dopředu, že v Milovickém lese budou mít ptáci krevní parazity, asi tam bude pár nových linií, ptáci budou často pozitivní při opakovaných odchycích a že počty infikovaných nebudou přesně stejné, pokud se použijí různé metody. Proč tam tedy vůbec jezdit? Přitom na takto pěkném datasetu by šlo otestovat celou řadu ohromně zajímavých otázek. Například v jedné nedávné práci na obdobně velkém datasetu na stejném modelu bylo najednou otestováno deset různých hypotéz o parazitech a hostitelích. Bylo by úchvatné se podívat, co z toho vyjde stejně. Těch deset navíc rozhodně není nějaký finální počet. Hypotézy se týkaly jen určitého vybraného typu interakcí parazitů a hostitelů. Absence otázek samozřejmě může být alespoň částečně spíše chybou zadání práce. I v takovém případě bych ale čekal, že si studentka nějakou otázku vymyslí, jak i trochu naznačovaly některé části diskuse. Moc tomu nerozumím. Řešit jen ty uvedené cíle musela být strašná otrava. Pro jistotu jsem ověřil zadání v SIS. Upřímně řečeno, není o mnoho lepší. Je zde navíc jen cosi o „biodiverzitě“. Nevím, co se tím přesně myslí. Každopádně by diversita společenstva počítat šla a bylo by to zajímavé. Metodika pro výpočet alfa a beta diversity, dominance, ekvitability nebo detekce klíčových hráčů ve společenstvu je důkladně propracovaná. Vždyť se tím ekologové zabývají již děsně dlouho. Nic z toho v práci není, ale vzhledem ke stručnosti zadání nemohu posoudit, co se tou „biodiverzitou“ vlastně myslelo. Nejsm si

také jist, zda došlo ve shodě mezi studentkou a školitelkou v pochopení toho, co má být dynamika. V zadání se objevuje nezávisle dynamika a perzistence. Perzistence se v práci řeší (a říká se jí kupodivu často dynamika). S dynamikou si nejsem jist. Spíše bych si pod tím představil třeba změny infekční intenzity. To se ale pomocí zvolené metody měřit nedá, což muselo být dopředu jasné.

Výčet připomínek k práci tímto bohužel nekončí. Proberu zde postupně, kde v práci shledávám chyby a nedostatky nebo kde bych jen doporučil jiné řešení.

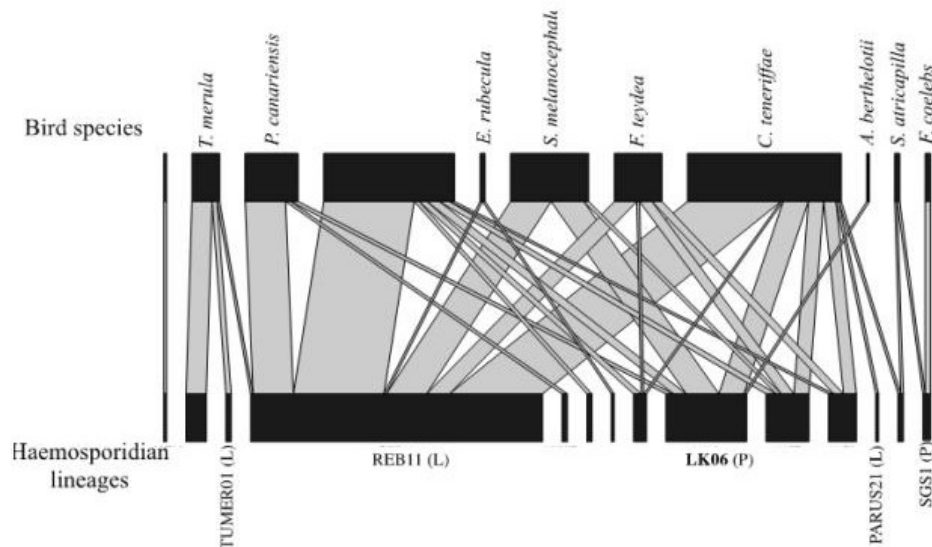
Název práce neodpovídá obsahu. Práce celkem spravedlivě řeší dva rody parazitů. Druhý se však v názvu nevyskytuje. Vzhledem k metodice muselo být jasné, že se budou oba rody detekovat společně. Nechápu, proč se tedy název neupravil. *Na stránce s názvem je taky špatně rok.*

Diplomová práce má mít, stejně jako jiné vědecké práce, pořádný úvod. V úvodu se mám dozvědět, co se na dané úzké problematice dělalo a kde jsou mezery našeho poznání. Z úvodu mám pochopit, proč se to celé vůbec dělalo, čím je to nové a co překvapivého by se mohlo zjistit. Z úvodu předložené diplomové práce se to nedozvím. Úvod je zde na necelou stránku a je zcela obecný. Za úvodem následují kapitoly literárního přehledu. Smysl sepisování této části diplomové práce mi již leta zcela uniká. Většinou to skončí přesně jako zde. Literární přehled je tak široký (v podobném rozsahu jako je jedna celá tlustá kniha), že se pak o problematice diplomky dozvím málo. Rozhodně to však není výjimečný problém jen této diplomové práce. Napsat dobrý úvod je těžké a povede se to v diplomové práci zřídka. Konkrétně se zde třeba píše o ostrovních efektech (*mimoходом špatně, protože Loiseau et al. zjistili ve skutečnosti něco jiného*) nebo o detailech životního cyklu parazitů. To se však v diplomce vůbec neřeší. Na vlastní problematiku diplomky tu pak chybí místo a energie, ač autorce nemohu upřít značnou snahu zde o studované problematice aspoň něco napsat. Často se zde však citují výsledky jednotlivých prací, aniž by se pak z toho učinil nějaký závěr nebo aspoň shrnutí. Obecné přesahy zcela chybí. Jakési shrnutí nalezneme snad jen v části o diagnostických metodách. Bohužel to není shrnutí úplně šťastné. Autorka tak trochu pozapomněla, že PCR může být taky kvantitativní a proto velmi vhodná pro určení intenzity infekce. Řada autorů to využila. Chybí mi tu také uvedení využití metod masivního sekvenování nebo jemné rozlišení linií pomocí alternativních markerů (třeba mikrosatelitů).

Již jsem zmínil, že metodika detekce parazitů je řádně popsána. Docela však překvapí, že byl pro sekvenování použit vždy jen jeden primer. Myslím, že u nových linií by osekvenování z obou stran mělo být standardem. Upřímně řečeno, problémy mám i se samotnou detekcí pozitivních vzorků a zarážející je také počet vzorků, u kterých nebyla určena konkrétní linie. Víc jak 20 procent fakt není málo. Nested PCR je nesmírně citlivá metoda, což je vyváženo rizikem falešně pozitivních nebo negativních signálů. Koneckonců i „normální“ PCR při optimálních podmínkách někdy prostě nevyjde a napodruhé se zjeví krásné pruhy na elektroforéze. Bývá proto zvykem při detekci parazitů PCR opakovat a také zahrnout do setu vzorků negativní či pozitivní kontroly. **Byla detekce založena opravdu na jediné nested PCR reakci a byly použity nějaké kontroly?** Některé linie se mohou lišit právě v místě nasedání primeru. S tím se ale dá alespoň částečně bojovat sekvenováním nebo i amplifikací jinými primery. I výskyt dvojitých peaků na elektroforetogramech lze řešit bioinformaticky, klonováním nebo specifickými primery pro běžné linie (po potvrzení se běžná linie „odečte“ v místě dvojitých peaků a můžeme přečíst druhou linii smíšené infekce). Bohužel se nic z uvedeného zřejmě nestalo. **Může autorka okomentovat, jaký mohou mít uvedené skutečnosti vliv na zjištěné prevalence?** V metodice chybí popis výpočtů. Mnoho se v práci sice vskutku nepočítalo, ale i výpočet prevalence vlastně není triviální. Není tak vůbec jasné, z čeho se prevalence počítala (viz níže). *Z drobných chyb zmíním, že u složek reakčních směsí je třeba udávat koncentrace. (Například údaj 0,5 µl primeru neříká vůbec nic, protože nevíme, jak koncentrovaný byl původní roztok primeru.) Záznam kroků PCR je poněkud nešťastný. Není úplně jasné, které kroky se cyklicky opakují.*

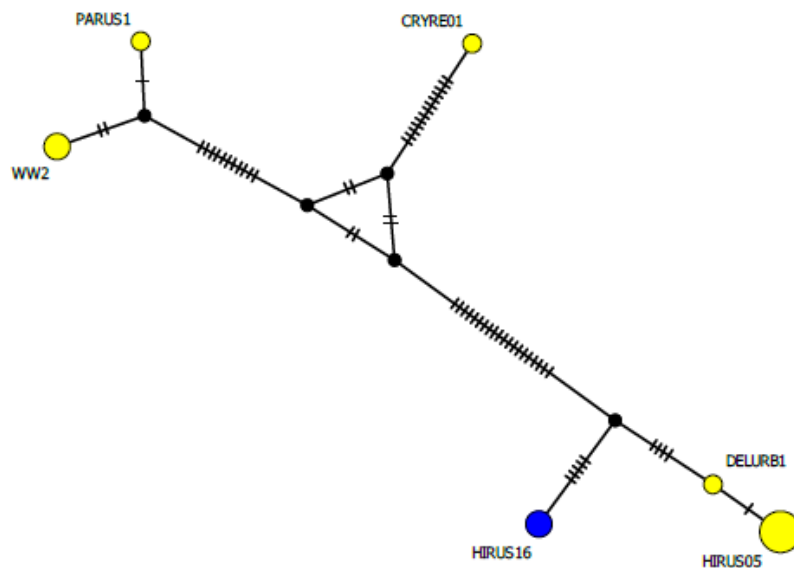
Mimochodem cyklus tvoří jen tři z vedených pěti kroků. Celé se to jmenuje třeba program pro PCR nebo kroky (steps) PCR. Na gelu nevidíme dvojpeaky (ty mohou být a asi i byly na elektroforetogramu) ale spíš více proužků (bands). Myslím, že lepší by bylo identifikovat sekvence podle MalAvi než přes genbank. Některé linie jsou v genbank špatně. MalAvi je specializovaná a průběžně kontrolovaná databáze. Je to tedy daleko spolehlivější zdroj.

Kapitola s výsledky je dost nepřehledná. Většina tabulek by patřila do příloh diplomové práce. Zde spíš překáží. Tabulky jsou částečně duplicitní s textem a i v textu se některé údaje opakují. Například není potřeba v jednom odstavci několikrát opakovat, kolik bylo celkem vyšetřeno vzorků dospělých nebo mladých jedinců, což se v práci mnohokrát děje. Naopak není jasné, co vlastně znamenají údaje o prevalenci. V abstraktu se píše o 1092 vzorcích, v metodice o 1092 jedincích. **Kolik tedy bylo vzorků a kolik jedinců? Byla prevalence počítána ze všech vzorků, tedy i z více vzorků odebraných stejnému jedinci? Co pak tedy údaj o prevalenci vlastně říká?** Říká se, že dobrý obrázek vydá za víc než tisíc slov. Platí to i pro tuto práci. Dohledat v tabulkách a textu interakce hostitelů a parazitů a udělat si tak celkový obrázek o specificitě systému je skoro nadlidský úkol. Obrázek by to ale ukázal snadno a nahradil by tak možná několik stran textu. Nejlepší by byl obrázek sítě interakcí. Je víc možností, jak takovou síť vytvořit, ale na vše jsou již hotové programy nebo skripty. Já mám rád síť z balíku bipartite v R. Na první pohled vidím početnost linií a specificitu vztahů hostitelů a parazitů. Ten balík také rovnou vypočítá, nakolik jsou obecně interakce specifické a dá se to porovnat s jinými pracemi. Není těžké to vytvořit. Vypadá to nějak takhle:



Ohromně by také porozumění autorčiných objevů usnadnilo nějak zobrazit podobnost (a tedy předpokládanou příbuznost) sekvencí parazitů. Hodně by to pomohlo u nově objevených linií. Jednou z možností by byl nějaký fylogenetický strom. Vůbec bych se nebál použít klidně i nějakou jednoduchou metodu založenou jen na distancích a v něčem jednoduchém (MEGA třeba). Stejně to

bude mít mizerné podpory. Já bych spíš ale preferoval haplotypové sítě. Třeba v programu PopArt se jednoduše nakreslí velmi efektní obrázky v publikační kvalitě. Vypadá to nějak takhle:



S příbuzností parazitů a bohužel i hostitelů se v práci skoro nic neřeší. To je velká škoda, zvláště když jsou od parazitů získány sekvence a o příbuznosti hostitelů celkem dost víme z literatury. **Mohla by autorka alespoň u obhajoby okomentovat vzájemnou příbuznost nejběžnějších hostitelů a zvážit, jaký vliv by mohly mít příbuzenské vztahy na výskyt parazitů?**

Hlavní problém výsledků jsem již zmínil. Vlastně tu nejsou. Autorka sice slibuje různá porovnání, nikdy však na ně nedojde. Čísla v tabulce nebo sloupcové grafy nejsou žádné porovnání. Říkají totiž něco jen o našem vzorku (tedy třeba o konkrétních odchycených dlascích nebo modřinkách). Zda to platí pro celé populace (tedy třeba všechny dlasčky a modřinky v Milovickém lese) odhalí až statistický test, na který ale nikdy nedošlo. Bez statistického testu se o rozdílech, shodě, změně či trendu prostě vůbec nedá mluvit. Abychom mohli něco spočítat, musíme si navíc být jisti, že máme náhodný vzorek z populace. U některých druhů hostitelů by to asi šlo ověřit. **Mohla by autorka říci důvody, proč její vzorek nemusí být náhodným vzorkem populace a jak by to mohla na svých datech otestovat?**

Již jsem zmínil, že autorka diskusi zjevně věnovala mnoho času, úsilí a péče. Problém však je, že bez korektně vypočítaných výsledků je těžké něco diskutovat nebo dokonce vyvozovat něco obecně platného. Také víme, že i výpočty prevalencí jsou zatíženy chybou. Sama autorka se navíc snaží ukázat, že se prevalence může hodně změnit z roku na rok. Srovnání prevalencí s jinými pracemi, kde navíc byla použita jiná metoda detekce, mi tedy přijde poněkud problematické. Se změnami prevalence je to ještě horší. Usuzovat ze tří let na nějaký dlouhodobý trend změn prevalencí, navíc bez statistického testu, je mírně řečeno poněkud nešťastné. I pokud by tu nějaký trend byl, nemá asi valného smyslu ho srovnávat se studii úplně jinde, jinak a jindy. Přes zmíněné limitace však v diskusi autorka odvádí obdivuhodný výkon. Ukazuje tu dokonalou znalost systému i prima biologický úsudek a pro svá tvrzení se pečlivě snaží najít podporu v literatuře. Bohužel se ale děsivě zaměřuje jen na svůj studovaný systém. Nijak se nesnaží z vlastních výsledků vytáhnout něco obecnějšího (ano, vím, šlo by to těžko) a konfrontovat to s jinými modely. Naprostá většina citací se tak týká jen modelu autorky.

Celkově se mi diplomová práce hodnotí strašně špatně. Autorka odvedla ohromné množství práce v laboratoři i při základní práci se sekvencemi a zřejmě i v terénu. Je jasné, že se v problematice výborně orientuje, dokáže pracovat s literaturou a konfrontovat ji se svými výsledky. Z diskuse usuzuji, že jí diplomka snad i musela bavit. Sepsání diskuse a metodiky jistě autorka věnovala mnoho času a příkladně pilného úsilí. V práci je řada menších nedostatků, ale to se týká skoro každé diplomové práce. Pořádně sepsaný úvod se v diplomkách také vyskytuje spíše sporadicky. Nejsem schopen rozlišit, zda absence smysluplných otázek a hypotéz je problém studentky nebo zadání práce. Nejsem si však jist, zda mají naši fakultu úspěšně absolvovat studenti, kteří nejsou schopni provést jediný, byť třeba triviální, statistický test a neumí přehledně a smysluplně ukázat své výsledky. Obávám se, že úroveň zpracování dat zde neodpovídá nejen posledním trendům ale snad ani tomuto tisíciletí. Nevím skoro nic o struktuře povinných a doporučených předmětů na parazitologii. Pokud však studentku minuly všechny přednášky ze statistiky, mohla si základní znalosti osvojit z učebnic. Existují dobré učebnice statistiky i v češtině. Ani dostupnost statistického programu dnes není limitací. Stejně se skoro všechno nejlépe spočítá v R, což je zadarmo.

V Přebudově 8.8.2020

Pavel Munclinger