

Oponentský posudek na magisterskou diplomovou práci:
"Phylogeny of the *Imaginal Disc Growth Factors (Idgf)* gene family of *Drosophila melanogaster*" autorky Lucie Kučerové

Tato útlá magisterská práce (11 stran textu bez příloh a seznamu literatury) je sepsána v anglickém jazyce a je strukturována jako rukopis vědeckého článku připravený k odeslání do vědeckého časopisu. Z tohoto důvodu jsem se snažil ve svém posudku postupovat jako recenzent časopisu. Bohužel není v práci upřesněno, do jakého časopisu bude (nebo už byl?) rukopis odeslán, což poněkud práci posuzovatele ztěžuje.

Jak již název spisu naznačuje, jedná se především o práci fylogenetickou. Cíle práce jsou pak definovány takto: zhodnotit možné scénáře evoluce genové rodiny *Idgf*, určit fylogenetické vztahy uvnitř rodiny, vysledovat genové duplikace v historii této genové rodiny, a odhadnout čas těchto duplikací. Práce je členěna do 6 kapitol (Introduction, Objectives, Materials and Methods, Results, Discussion, and References) a je doplněna 16 přílohami.

V úvodu (Introduction) autorka stručně seznamuje čtenáře s genovou rodinou *Idgf*, současnými poznatky o funkci a původu těchto genů.

V další kapitole (Materials and Methods) pak popisuje způsob získání sekvencí jak z dostupných databází, tak vlastní sekvenací cílových úseků genomové DNA či cDNA. Tento popis je však, zejména v případě bioinformatického vyhledávání poněkud strohý a nedostatečný. Není tak jasné, zda autorka pracovala pouze s anotovanými geny, či zda též extrahovala data z needitovaných, hrubých sekvencí, popřípadě EST databází. Řada sekvencí není z neznámých důvodů uvedena v tabulce *Idgf* genů (Appendix 2), přestože jsou v genové bance k dispozici (např. *Drosophila pseudoobscura*, v genové bance od srpna 2005; *Bombyx mori*, od března 2001; *Anopheles gambiae*, od listopadu 2004; *Aedes aegypti*, od prosince 2004) a jsou patrně použity i ve vlastní fylogenetické analýze. U sekvencí získaných z FlyBase databáze pak jakékoli identifikační číslo či kód zcela chybí a konkrétní použité sekvence jsou tak naprosto nedohledatelné. Popis laboratorního postupu při amplifikaci a sekvenování genů z genomové DNA a cDNA je pak naopak zbytečně podrobný, a jsou v něm použity poněkud nestandardní formulace. Některé údaje uvedené v metodickém popisu jsou pak bezobsažné (např. množství DNA templátu v PCR reakci je uvedeno v ul bez udání koncentrace). Autorka též uvádí v popisu řadu primerů bez udání literárního zdroje, i když je

zřejmě, že je nenavrhovala sama. Fylogenetická analýza, zejména pak sestavení alignmentů, je odbyta řadou obecných proklamací nemajících valné vypovídající hodnoty. Není vůbec jasné, zda autorka použila k tvorbě alignmentu nukleotidové, či aminokyselinové sekvence, není jasné jaké sekvence byly pak použity k fylogenetické analýze, a čtenář musí tyto informace vysлідit v kontextu, někdy marně. Použité metody fylogenetické rekonstrukce jsou pak, dle mého názoru, nepříliš vhodné a popis jejich parametrů je zcela ignorován. Konstatování: "We tested different models and parameters settings." je poněkud vágní, zvláště pak, když se ve výsledcích o těchto imaginárních testech nedozvíme ničeho.

Výsledky vlastní laboratorní práce jsou nedostatečně popsány. V Appendix 3 je sice patrné, jaké druhy hmyzu byly sekvenovány, jak byla tato sekvenace úspěšná a co bylo vlastně sekvenováno (genomová DNA nebo cDNA) už však nikoliv. V popisu výsledků je možné dohledat počty *Idgf* získaných sekvencí homologů u druhů sekvenovaných autorkou (*Scaptodrosophila lebanonensis*, *Scaptomyza palmae*, *Chymomyza costata*, *Rhagoletis cerasi*, *Neobellieria bullata*), délka sekvencí, stejně jako typ konkrétního templátu (genomová DNA, cDNA) zůstávají však čtenáři skryty. Sekvence získané diplomantkou pak nejsou nikterak označeny ve fylogenetických stromech a je tak tedy obtížné vůbec určit, nakolik byla tato laboratorní část práce náročná a úspěšná. Hodnocení výsledků práce je navíc ztíženo tím, že žádný z prezentovaných fylogenetických stromů není editován do publikovatelné formy. Prakticky to znamená, že ve stromě nejsou uvedena jména jednotlivých druhů, nýbrž pouze jejich blíže nevysvětlené zkratkovité kódy, což značně ztěžuje orientaci čtenáře a obírá oponenta o čas. V některých případech (Appendix 9) jsou pak stromy jen velmi těžce čitelné, protože čísla statistických podpor (bayesian posterior probabilities) se překrývají. Taxon sampling stromů konstruovaných pomocí různých metod vzájemně neodpovídá (Appendix 8 a 9 versus Appendix 10 a 11), což poněkud ztěžuje jejich porovnání a interpretaci. Popis výsledků v textu pak čtenáři bohužel příliš nepomůže. Autorka například píše, že stromy sestavené z kompletních sekvencí potvrzují blízkou příbuznost *Idgf4* a *Chit* na jedné straně a *Idgf* 1-3 a *Idgf5* na straně druhé. Dále se pak dovíme, že stejná topologie byla získána analýzou kódující oblasti. Není mi zcela jasné, co je touto formulací myšleno. Jsou snad ostatní uvedené stromy konstruovány na základě nekódujících oblastí? Jakých? Fylogenetické stromy jsou kořeněny dalšími hmyzími *Idgf* geny, patrně díky předpokladu monofylie *Idgf* genů z dipter. Takový předpoklad však nemusí být nutně pravdivý! Kořen stromu by bylo daleko vhodnější nalézt společnou analýzou s příbuznými chitinázami. Autorka též ve výsledcích tvrdí, že se pokusila o analýzu časování *Idgf* genových duplikací, přestože v

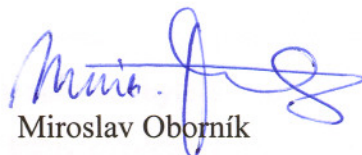
popisu fylogenetických metod takovou analýzu vůbec nezmiňuje. Z důvodů, které jsem nepochopil pak pro "time divergence estimate analysis" použila pouze třetí pozici v kodónu, přičemž konstatuje, že nízká podpora bazálních větvení ve stromě (Appendix 14) je důsledkem nízkého informačního obsahu třetí pozice v kodónu. Troufám si tvrdit, že je to ale důsledek faktu, že tato třetí pozice je saturována. Mohla by autorka tento aspekt blíže vysvětlit? Jde saturace třetí pozice v kodónu ověřit? Jakým způsobem?

Chápu, že interpretace podobných výsledků a následná diskuse není zcela jednoduchá. Proto budu velmi rád, když nám autorka ozřejmí své závěry během obhajoby práce. Autorka například tvrdí, že duplikace *Idgf4* - Chit je tou nejrecentnější, přestože ve všech ostatních stromcích vychází jako bazální. Autorka v diskusi tvrdí, že všechny analyzované geny jsou funkční. Jaká byla úspěšnost amplifikace těchto genů z cDNA? Mají tyto geny EST podporu?

Souhrn:

Předložená práce je velmi stručná, metody, zejména způsob získání sekvencí z databází, konstrukce alignmentu, a fylogenetická analýza jsou nedostatečně popsány. Popis použitých sekvencí je nekompletní (Appendix 2). Zdroje některých informací (sekvence primerů) jsou nejasné, stejně jako původ obrázku v Appendix 4 (není uveden literární zdroj, ani gen, na základě kterého byl strom sestaven). Výběr prezentovaných fylogenetických stromů je chaotický, stromy jsou v needitované a tudíž nepublikovatelné podobě, jejich popis je nedostatečný (chybí počet znaků, u MP stromů počet informativních znaků; zda jsou stromy sestaveny na základě NT nebo AA dat je patrné jen z kontextu; u žádné analýzy nejsou uvedeny parametry). Provedení "time divergence analysis" je nejasné, a není popsáno v metodice. Diskuse kolem časování genových duplikací se mi jeví ukvapená a nepodložená. Pokud bych hodnotil tuto magisterskou práci jako rukopis článku, zcela jistě bych jej v této podobě nedoporučil k publikaci. I přesto, že se autorka snažila pečlivě skrýt výsledky své nemalé laboratorní práce před oponentem, doporučuji její magisterskou práci k obhajobě a hodnotím ji známkou **dobře**.

V Českých Budějovicích dne 19.1. 2006



Miroslav Oborník

Parazitologický ústav AV ČR
Braníšovská 31
České Budějovice

Oponentský posudek

na magisterskou diplomovou práci **Lucie Kučerové**

Phylogeny of the Imaginal Disc Growth Factors (*Idgf*) gene family of *Drosophila melanogaster*

Práce se zabývá molekulární evolucí multigenové rodiny růstových faktorů imaginálních terčků, původně objevených u *D. melanogaster*, u nejpočetnější a nejvyspělejší vývojové větve hmyzu – skupiny Endopterygota. Tato problematika je v současné době velmi aktuální a vědecky atraktivní, neboť výsledky studia evoluce multigenových rodin mohou významně přispět k poznání jednoho ze základních mechanismů evoluce genomů – genových duplikací.

Práce je psána anglicky, podotýkám velmi dobrou a srozumitelnou angličtinou s minimem drobných chyb. Autorka zvolila pro diplomové práce netradiční formu publikace, což je sice v souladu s platným studijním a zkušebním řádem BF JU, avšak přineslo to s sebou i některé problémy – zejména přílišnou stručnost.

Úvod práce je stručný, jasný a výstižný, zcela odpovídající formě publikace. Osobně bych ale dal v diplomové práci přednost obsáhlejšímu úvodu formou literární rešerše. Cíle práce jsou jasně definovány.

Kapitola Materiál a metody (str. 4, přílohy 2 a 3) asi nejvíc trpí přílišnou stručností, troufám si tvrdit, že některé části jsou příliš stručné i pro potřeby publikace. Autorce se v této kapitole umělo podařit skrýt značné množství vlastní provedené práce. Během ústní obhajoby doporučuji objasnit, které sekvence byly získány z Genbank a které byly osekvenovány v rámci této práce, což z textu této kapitoly není jednoznačně vidět. Taky mi není zcela jasný text podkapitoly "Database searches". Byly homologní sekvence hledány v genomech 15 v textu uvedených druhů s použitím sekvencí 7 jiných druhů, uvedených v příloze 2? Je to tak? Jsou druhy v příloze 3 ty, u kterých byly sekvence získány experimentálně? Pokud jsem dobře porozuměl, degenerované primery byly navrženy autorkou – tudíž bych uvítal, kdyby jejich seznam včetně sekvencí byl uveden v příloze.

Kapitola Výsledky je rozdělena do tří částí – v první z nich jsou prezentovány výsledky hledání homologů *Idgf* genů u hmyzích taxonů spolu s výsledky izolace homologů vlastní experimentální prací (příloha 5). Druhá část obsahuje srovnávací analýzu všech studovaných genů na základě mutačních rychlostí (přílohy 5-7). Ve třetí nejrozsáhlejší části jsou pak prezentovány výsledky fylogenetické analýzy (přílohy 8-14). Tato kapitola jednoznačně dokazuje, že autorka získala velmi zajímavé, původní a kvalitní poznatky, z nichž vyvodila odpovídající a dobře podložené závěry. Dle mého názoru tak získala materiál, který bude po dopracování a odstranění některých nedostatků vhodný pro publikaci v některém prestižním genetickém časopise. Výsledky již prezentovala na mezinárodní konferenci (viz příložená kopie velmi hezky připraveného posteru). Ke zpracování výsledků mám následující připomínky a komentáře.

- 1 - Na str. 6 v podkapitole „Hledání homologů *Idgf* u hmyzu“ jsou uvedena Hemiptera, ale v kladogramu přílohy 4 není tato skupina označena u křísa *Oncometopia nigricans* ("sharpshooter"; Cicadellidae). Podotýkám, že dle současného taxonomického pojetí jsou křísi řazeni do taxonu Auchenorrhyncha.

Autorka správně v kladogramu uvádí, že druhy, u kterých nebyl osekvenován celý genom, nemusí být počet nalezených homologů konečný. Proč nejsou v kladogramu naznačeny fylogenetické vztahy mezi hmyzími řády, přestože jsou k dispozici publikované údaje, ze kterých lze vycházet? Jinými slovy řečeno, kladogram v levé části neodpovídá současným znalostem o fylogenezi řádů Diptera, Lepidoptera, Hymenoptera, Coleoptera a vzdálenější větve Hemiptera.

Dle informace od Davida Heckel, se kterým jsem se setkal začátkem ledna na konferenci, bylo právě dokončeno sekvenování genomu potemníka *Tribolium castaneum* (viz <http://www.bioinformatics.ksu.edu/BeetleBase/>). Doporučuji zopakovat hledání a doplnit do připravované publikace.

-2- Na str. 6 v podkapitole „Analýza rychlosti substitucí mezi *Idgf* geny“ autorka uvádí, že poměr synonymních ku nesynonymním substitucím (Ka/Ks) je menší než 1, což je typické pro kódující oblasti. Ve skutečnosti je výrazně menší a dosahuje hodnot 0,20- 0,36, jak je vidět z údajů v příloze 6; to je třeba zdůraznit.

U přílohy 7 je uvedeno, že u *Pieris rapae* nebyly hodnoty Ks počítatelné u všech *Idgf* genů kromě *Idgf3*, což je logické, protože ten jediný byl u běláška nalezen – přesto by to mělo být uvedeno v legendě přílohy, aby čtenář nemusel trápit mozkové závity.

-3- Na str. 7 v podkapitole „Fylogenetická analýza“ mě zaujalo, že výsledky překvapivě dobře podporují současné pojetí příbuzenských vztahů mezi zkoumanými taxony a přesně odhalily i vztahy mezi hmyzími řády (to je to, co jsem postrádal u kladogramu v příloze 4). Dále ukázaly vzájemnou příbuznost jednotlivých genů. Dle mého názoru jsou tyto výsledky velice dobře interpretovány; oceňuji i přiměřenou kritičnost autorky při interpretaci výsledků. O to větší škoda, že v přílohách k této podkapitole je řada formálních, ale docela podstatných nedostatků:

- u kladogramů v přílohách 8-12 postrádám vysvětlivky ke zkratkám druhů;
- kladogram v příloze 9 je prakticky nerozluštitelný; pokud to není jen v mnou posuzované kopii, doporučuji předělat a vlepit do práce;
- nejsem fylogenetik a tudíž přesně neznám, jak zachází software s "outgroup" (v tomto případě kříš *Oncometopia nigricans*), ale domnívám se, že by ve správně zakořeněných kladogramech v přílohách 8 a 9 měl kříš tvořit sesterskou větev k dalším dvěma větvím, tj. větvi Diptera a větvi ostatní Endopterygota, a nikoliv tvořit s oběma větvemi rovnocennou větev; podotýkám, že kříši patří do skupiny Paraneoptera, jež je sesterskou skupinou celé skupiny Endopterygota;
- u rekonstrukce ancestrální *Idgf* sekvence (přílohy 11-13) mi z popisu v textu ani v legendách není jasné, jakým způsobem došla autorka k této sekvenci.

Diskusi považuji za nejlepší část práce. Zvlášť mě zaujala pasáž o "concerted evolution" (str. 9). Souhlasím se závěrem, že tento evoluční mechanismus nebude příliš důležitý pro evoluci rodiny *Idgf* genů, jen bych doplnil, že mechanismy, které ho podmiňují jako např. genová konverze a nerovnoměrný crossing over, se týkají především tandémových duplikací. U *D. melanogaster* jsou v těsné vazbě pouze geny *Idgf1*, *Idgf2* a *Idgf3*, ostatní *Idgf* geny lokalizovány na jiných částech genomu. S tím souvisí i můj dotaz: Diptera jsou známa výraznou evolucí karyotypu přestavbou celých ramen chromosomů a četnými inverzemi. U *D. melanogaster* odpovídá stupeň podobnosti *Idgf* genů jejich umístění v genomu s jedinou výjimkou: geny *Idgf4* a *Chit*, které jsou dle výsledků nejmladší duplikací, jsou na různých chromosomech, *Idgf4* na X a *Chit* na pravém rameni chromosomu 2, což je poněkud v rozporu s výsledky fylogenetické analýzy. To by ale mohla osvětlit pozice těchto genů u jiných druhů rodu *Drosophila*. Je známa lokalizace jednotlivých genů u jiných druhů?

Doplňkové dotazy k diskusi:

- Má autorka nějakou hypotézu nebo možné vysvětlení, proč právě u řádu Diptera nedošlo k degeneraci duplikovaných genů rodiny *Idgf* ve srovnání s jinými taxony hmyzu?
- Je evoluce genů *Idgf* srovnatelná s evolucí jiných multigenových rodin u rodu *Drosophila* (jako např. genů pro cecropiny, histony, lectiny, amylázy, rDNA, atd.) nebo se výrazně odlišuje?

Z á v ě r

Po přečtení a zhodnocení předložené diplomové práce nemám sebemenší pochyby o tom, že autorka zcela splnila zadané cíle práce a že získala kvalitní původní výsledky. Způsobem interpretace a diskutováním těchto výsledků i vyvozenými závěry prokázala, že velmi dobře pronikla do problematiky molekulární fylogeneze a že má všechny předpoklady pro další působení v základním výzkumu. Velká škoda, že dobrý dojem z výsledků a jejich interpretace pokazila řadou formálních nedostatků, zejména nedostatečným popisem příloh a odstrašujícím kladogramem v příloze 9.

Práci jednoznačně doporučuji k obhajobě, ale vzhledem ke zmíněným nedostatkům navrhuji její hodnocení klasifikačním stupněm **velmi dobře**.

V Českých Budějovicích, dne 22.1.2006



František Marec