

Oponentský posudek na diplomovou práci Natálie Čížkové – Evoluce množství jaderné DNA rodu *Mallomonas* a její souvislost s ekofyziologií

Vzhledem k celkové deficienci podobných studií zabývajících se problematikou velikosti genomu u protist je jakákoliv zacílená práce, která je metodicky dobře ukotvena a využívá „best practice“ metodický aparát u této obecně obtížné skupiny organismů, velmi přínosná a vítaná. Nejen, že nám značně rozšiřuje poznání o samotné variabilitě velikosti genomu, ale zároveň poukazuje na dosud netušenou míru diverzity mezi jednotlivými kmeny, které jsou řazeny k těmž druhu. Zde studovaný rod *Mallomonas* nebyl dosud z tohoto hlediska zevrubně zkoumán, a práce tak přináší řadu nových poznatků, které mají značný vědecký přínos. Jako jednoznačný klad práce vnímám i to, že byla získaná data o variabilitě ve velikosti genomu analyzována v kontextu dalších fyziologicko-ekologických vlastností jednotlivých druhů, včetně např. rychlosti pohybu. Nejde tedy jen o deskriptivní studii, která by sama o sobě byla také přínosná, ale o dobře strukturovanou práci, kde se zjištěné poznatky dávají do širšího kontextu a je zde jasná snaha o jejich vysvětlení ve vzájemných interakcích. Nespornou devízou práce je i fakt, že bylo možno využít poznatků z paralelně řešeného projektu na zájmové skupině a získaná data analyzovat např. ve fylogenetickém kontextu.

Na druhou stranu nemohu nezmínit i určité nedostatky práce, resp. nepoukázat na nejasnosti v některých analýzách. Při korelační analýze velikosti genomu a ekologických podmínek se pracuje s environmentálními daty, které nejsou pro sladkovodní organismy asi úplně ideální. Takový postup je pochopitelný, protože jsou data snadno získatelná, nicméně vzbuzuje otázku, zda analýza vůbec může něco ukázat. Mimoto se pro mě dost nepochopitelně pracuje jen s průměry, a to jak ve velikosti genomu (navzdory vnitrodruhové variabilitě), tak v charakteristikách prostředí. Tento „druhový“ pohled je validní a možný, ale chtěl bych vedle něj vidět i analýzu, kde se pracuje s konkrétními kmeny separátně („populační“ přístup), a který může lépe škálovat gradienty v datech.

S tím úzce souvisí fakt, že vnitrodruhová variabilita velikosti genomu je u mnohých taxonů velmi vysoká a průměrná hodnota může být velmi zavádějící. Zároveň je při interpretaci dat a hodnocení jejich relevance nutné brát v potaz, že měření velikosti genomu v rámci jednotlivých kmenů je mnohde velmi variabilní, a není výjimkou, že opakovaná měření ukazují mezivzorkovou fluktuaci i kolem 20%. U vzorku *M. hexareticulata* (IE140) dokonce došlo k přehlédnutí faktu, že se podařilo pravděpodobně zachytit haploidně-diploidní fázi, neb rozdíl v jednom měření dosahuje téměř 100% (pokud jsou tedy data v Příloze 6 zaznamenána správně).

Při analýze velikosti genomu ve fylogenetickém kontextu postrádám snahu vysvětlit nejpravděpodobnější model evoluce – místo toho jsou používány metody, které apriorně předpokládají určitý evoluční model (*contMap* funkce v R balíku *phytools* implementuje Brownův model; PGLS defaultně implementuje Pagelův λ model). Proto bych nejraději viděl nejprve test nejvhodnějšího modelu (např. pomocí jednoduchého srovnání pomocí *fitContinuous* funkce z R balíku *geiger*, která umožňuje testovat různé evoluční modely, a navíc dovoluje pracovat s chybovostí vstupních dat), a teprve na základě výsledků testu zvolit další postup analýzy. V interpretační části práce mi přijdou některé teorie o vnitrokmenové variabilitě velikosti genomu až moc divoké – viz moje první otázka.

Dílčí nedostatky:

V kapitole 1.2.1.1 je snaha vysvětlit princip průtokové cytometrie. Navzdory tomu, že se u rostlin využívá nejjednodušší možná konfigurace cytometru zpravidla s jedním, nebo maximálně dvěma zdroji záření (ale vždy používanými odděleně), neobešla se tato pasáž bez chyb, zejména pak v popisu „forward“ a „side scatter“ – u nich není principiálně možné pracovat s intenzitou fluorescence (tedy emitovaným světlem), ale naopak s excitačním světlem.

V kapitole 3.2.2 jsou popisovány standardy pro FCM s jejich velikostmi genomu. V práci se pracuje hned se třemi standardy, ale to vyžaduje jejich „re-kalibraci“ vůči primárnímu standardu tak, aby

velikosti genomu měřené s jakýmkoliv standardem byly porovnatelné. Taková korekce velikostí standardů tu však není uvedena, což může být v interpretaci dat zavádějící.

Kapitoly o zpracování statistických dat zmiňují využití programu R-studio ver. 4.2.1 – R-studio je ale pouhá grafická nadstavba vlastního R softwaru, jehož verze je zde uvedena, ale ve spojitosti s jiným programem.

Určité nedostatky vidím i ve způsobu prezentace dat – např. nerozumím tomu, proč je pro každou dílčí analýzu velikost genomu v jiných jednotkách (v evolučních analýzách logaritmus velikostí v Mbp, v korelačních analýzách logaritmus velikostí v pg). Všechny prezentované stromy (Obr. 15–17) mají nečitelná jména taxonů – zjevně jde o nesoulad fontů v R a v textovém editoru, nebo v programu pro úpravu obrázků.

Otázky:

1) V kapitole 5.2 je zevrubně diskutována zjištěná variabilita v množství DNA v rámci jednoho vzorku po re-izolaci kultury z jediné buňky. Pominu-li nejprozaičtější zdroj zjištěné variability, tedy kontaminaci, které byla snaha zabránit právě pěstováním kultury z jediné původní buňky, tak mne překvapuje, že zde vůbec nebyla diskutována možnost indukce variability ve velikostech genomu pomocí akumulace či ztráty transponovatelných elementů (TE). A to navzdory tomu, že je právě variabilita v TEs v práci uvedena jako jeden z hlavních mechanismů (vedle polyploidizace) ovlivňujících velikost genomu. Co vedlo k tomu, že taková možnost byla a priori vyloučena, a místo toho je zde zmíněna krajně nepravděpodobná teorie o částečné endoreduplikaci, která je navíc specifickou variantou endoreduplikace, tedy procesu výhradně spojeného s mnohobuněčnými organismy? Je natolik nepředstavitelná představa, že se v rámci kultury, sice vzniklé z jediné buňky, ale po x-cyklech dělení, najde vnitro-kmenová variabilita podmíněná akumulací či ztrátou TEs? Pokud je to „nemožné“, tak jaké hlavní překážky v tom vidíte? Jaká je pak obecná představa o tom, že vůbec nějaká variabilita ve velikosti genomu u protist vzniká, a zejména v jaké fázi jejich vývoje?

2) Kapitola 5.3.1 je věnována vnitrodruhové variabilitě ve velikosti genomu a významu polyploidizace. U vybraných druhů s variabilitou, která neodpovídá polyploidizaci, je snaha vysvětlit její existenci oddělováním evolučních linií a tedy speciací. Jako příklad je uveden i druh *M. lelymene*, kde je odlišná velikost genomu italského kmene IT02 vysvětlována zvláštním postavením ve fylogenetickém stromě. Nicméně tři kmeny z Česka jsou fylogeneticky neoddělitelné (Obrázek 16) a přitom mají každý úplně jinou velikost genomu (jeden kmen totožnou s velikostí italského vzorku). Můžete prosím vysvětlit tento zjevný interpretační rozpor? Čemu přisuzujete rozdílnost českých kmenů, když nejde o speciaci?

3) V diskuzi na straně 51 popisujete vztahy mezi velikostmi genomu a velikostmi buněk, což je zcela v pořádku, ale není mi moc jasné, co má vyjadřovat veličina „koeficient variace“, kterou zde používáte. Ve skutečnosti jde o veličinu popisující rozptyl v datech, konkrétně o směrodatnou odchylku dělenou průměrem. Tím neříkám, že by nebylo možné vztahy mezi různými velikostmi organismů v jednotlivých trofických kategoriích takto popisovat, jen nevím, co nám řekne o jedné skupině $CV = 56\%$, a u jiné skupiny $CV \sim 100\%$. Nešlo ve skutečnosti o úplně jinou metriku, např. koeficient determinace? Ať už je to v práci popsáno správně, nebo ne, tak co tím mělo být řečeno o rozdílech mezi heterotrofními a fotoautotrofními chrysomonádami? Co to o nich vypovídá a co z toho vyplývá, resp. jaký vztah je očekáván v kontextu různých způsobů obživy?

4) Na straně 54 je diskutován Obrázek 28, tedy závislost velikosti genomu a četnosti výskytu jednotlivých druhů, který poukazuje na fakt, že druhy s větším genomem jsou nalézány častěji než druhy s menším genomem. Potřeboval bych vysvětlit, co si z tohoto výstupu mám odnést. Znamená četnost nálezů jednotlivých druhů (resp. jejich šupin) automaticky i větší areál výskytu (tedy schopnost kolonizace na větší vzdálenost)? Může být vůbec prezentovaný graf fakticky správně, vezmeme-li v potaz velkou míru vnitrodruhové variability velikosti genomu? Například

nejpočetnější druh *M. crassisquama* vykazuje obrovskou variabilitu přisuzovanou polyploidizací (strana 50), přesto jsou všechny nálezy druhu asociovány s jedinou, střední hodnotou (Příloha 4). Co Vás vedlo k této interpretaci? Jak by bylo možné docílit toho, že prezentovaná závislost bude věrohodná?

Závěr:

I když v práci vidím určité nedostatky, nejde o žádné systémové chyby, které by byly nenapravitelné. Naopak velmi pozitivně vnímám fakt, že jsou primární data sebrána a analyzována asi nejlepším možným způsobem, který lze pro danou skupinu vymyslet, a jsou tudíž velmi dobře použitelná v budoucí publikaci, která doufám z DP vznikne. Některé dílčí analýzy by naopak bylo vhodné doplnit, jiné trochu přepracovat, a některé interpretační závěry více propracovat. Nicméně práci celkově hodnotím jako velmi kvalitní a doporučuji ji k obhajobě. Prozatím ji hodnotím stupněm velmi dobře.

V Průhonicích, 20. 6. 2024

Pavel Trávníček

