

Oponentský posudek na diplomovou práci Bc. Jany Skopalíkové: „Využití metody Hyb-Seq pro rekonstrukci retikulární vnitrorodové fylogeneze: příklad z polyploidního rodu *Curcuma* L. (Zingiberaceae)“

Diplomová práce Bc. Jany Skopalíkové je zaměřena na rekonstrukci fylogenetických vztahů v rodě *Curcuma*. Tento rod představuje taxonomicky složitou polyploidní skupinu s vysokou frekvencí mezidruhové hybridizace. Evoluční historie této skupiny již byla předmětem dvou předcházejících studií, v rámci kterých byly identifikovány její hlavní vývojové větve, avšak vztahy v rámci těchto linií jsou stále nevyjasněné. Předložená diplomová práce využívá ke studiu této problematiky velice moderní metodický přístup, založený na analýze sekvencí několika stovek jaderných low-copy genů. Takto rozsáhlý soubor nezávislých markerů má velký potenciál objasnit fylogenetické vztahy studované skupiny objektivně, přesně a robustně. Jenže použití jakkoliv slibného souboru dat předpokládá zvládnutí mnohdy náročných analytických postupů. Není proto nijak překvapivé, že hlavním cílem práce bylo spíše otestovat možnosti využití takto rozsáhlého souboru dat a způsobů jejich analýzy než vlastní rekonstrukce vztahů nad rámec již předtím objasněných.

Práce studuje vztahy mezi 27 druhy rodu *Curcuma* (každý reprezentovaný jedním jedincem), u kterých bylo pomocí metody Hyb-Seq sekvenováno 1154 low-copy jaderných genů, téměř celý chloroplastový genom a jaderná ribosomální DNA. Pro práci byly zvoleny druhy, které reprezentují již potvrzené základní evoluční linie rodu. Zároveň se jedná o druhy spíše bezproblémové, charakterizované nižšími ploidiemi a nižší mírou intraindividuálního polymorfismu ITS sekvencí, na základě čeho lze očekávat spíše jednodušší (i když u takto složité skupiny je potřeba tento pojem brát s rezervou!) pattern fylogenetických vztahů. Součástí výběru jsou i 4 taxony, u kterých byl v předchozích studiích prokázán hybridní původ.

Ve svém jádru se práce zabývá dvěma základními cíli. Porovnáním vhodnosti různých výpočetních metod a datasetů (low-copy geny, nrDNA, cpDNA) pro rekonstrukci fylogenetických vztahů a možností využití sekvencí takto velkého počtu low-copy genů k identifikaci hybridů. Za tímto účelem bylo použito několik rozdílných výpočetních přístupů a bylo provedeno velké množství různých analýz, umožňujících dobře prozkoumat tento komplexní soubor dat. I přes velkou snahu kandidátky prezentovat výsledky práce podrobně a zároveň přehledně mám k práci několik připomínek:

Incongruence mezi výsledky fylogenetických analýz založených na různých markerech (cpDNA, low-copy geny a nrDNA) je demonstrována na celém souboru dat, kterého součástí jsou i druhy, které jsou prokazatelně hybridního původu. Tyto druhy mají vliv na topologie porovnávaných stromů a podporu některých skupin. Bylo by zajímavé vidět, do jaké míry jsou rozpory mezi jednotlivými markery způsobeny jevy, které nejsou tak markantní a proto prezentovat i výsledky analýz bez těchto hybridů.

Výsledky identifikace hybridů pomocí lineage movement analýzy a testů monofylie nejsou prezentovány dostatečně jasně (viz dotaz níže).

Práci považuji za velice kvalitní, doporučuji k obhajobě a hodnotím známkou 1.

V Praze 5.9.2017

Karol Krak

Otázky k obhajobě:

1. Výsledky analýz z 811 (resp. 1154) exonů low-copy genů se až tak výrazně neliší od výsledků práce Záveská et al. 2016, která je postavena na analýze 4 low-copy genů včetně jejich nekódujících částí. V čem spatřujete výhody použití Vašeho přístupu oproti použití sice výrazně menšího množství markerů, kterou jsou ale vysoce variabilní (introny low-copy genů)?
2. Většina analyzovaných jedinců byla polyploidní. Vykazovali jednotlivé úseky v rámci jedinců po mapování readů na pseudoreferenci polymorfismy? V následných analýzách byl každý jedinec zastoupen pro každý úsek jenom jednou sekvencí. Z popisu metodiky není zcela jasné, jakým způsobem byly tyto polymorfismy ošetřeny před fylogenetickými analýzami. Bylo by možné na základě těchto dat fázovat jednotlivé alely? Zkoušela jste odhadnout počty alel pro jednotlivé úseky u různých jedinců?
3. Podle práce Záveská et al. 2012 se vyznačují zástupci rodu *Curcuma* relativně vysokou mírou intraindividuální variability sekvencí nrDNA. Jak vysoká byla tato variabilita ve Vašich vzorcích?
4. Není mi úplně jasné, proč považujete na základě výsledků lineage movement analýzy a testu monofylie čtyři taxony za jednoznačné hybridy a ostatní druhy za taxony nehybridogenního původu. Přitom podle hodnot v Tab. 10 a 11 se od ostatních taxonů výrazně neliší (např. *C. candida* a *C. aeruginosa*)? Zkoušela jste použít i testy, které byly použity v práci Záveská et al. 2016 (STEM-hy a JML)?
5. Jaký je potenciál využití dané metody pro studium vztahů v rámci jednotlivých evolučních linií? Které otázky/hypotézy byste chtěla pomocí této metody dále testovat?