

**Univerzita Karlova v Praze**  
**Přírodovědecká fakulta**

Studijní program: Biologie  
Studijní obor: Botanika



Bc. Alžběta Böhmová

**Evolvece velikosti genomu v čeledi *Costaceae***  
**Genome size evolution in *Costaceae***

Diplomová práce

Školitel: Mgr. Tomáš Fér, Ph.D.

Praha 2016

Prohlášení:

Prohlašuji, že jsem závěrečnou práci zpracovala samostatně a že jsem uvedla všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

V Praze, 02. 05. 2016

Podpis

## **Poděkování**

Ráda bych poděkovala svému školiteli Mgr. Tomáši Férovi Ph.D. za odborné vedení a cenné rady při zpracovávání dat do mé diplomové práce. Dále bych ráda poděkovala vedoucí laborantce Ing. Lence Flaškové za pomoc v DNA laboratoři a užitečné rady při optimalizaci PCR reakcí. Mgr. Monice Pospíšilové za odvedenou laboratorní práci při přípravě knihoven a rady při psaní diplomové práce. Spolužákům Mgr. Kristýně Hlavaté, Mgr. Elišce Záveské Ph.D. a Mgr. Evě Kondrysové za rady a oporu při zpracovávání vzorků do DP. Dr. Janě Leong-Škorníčkové za všechnen poskytnutý rostlinný materiál a cenné rady ohledně taxonomie. Děkuji botanickým zahradám v Singapuru, Troji, Edinburghu a Washingtonu, Johnovi Moodovi za poskytnutý rostlinný materiál. A v neposlední řadě mé rodině za morální podporu a kontrolu formální stránky DP.

## Abstrakt

Variabilita ve velikosti genomu dosahuje u rostlin velkého rozpětí; u krytosemenných rostlin může být toto rozpětí až 2400-násobné. Proto častým předmětem studia je snaha vysvětlit tyto rozdíly v mnoha kontextech. Jedním z nich je i fylogenetické hledisko, kdy se studuje vztah velikosti genomu k jednotlivým taxonomickým skupinám. Doposud bylo publikováno mnoho prací zabývajících se rostlinami temperátního pásu. Prací, které reflektují velikost genomu u tropických rostlin, je však velmi málo. Proto byla předmětem této práce analýza velikosti genomu u pantropicky rozšířené čeledi *Costaceae* Nakai.

Tato čeleď v současnosti zahrnuje sedm rodů: *Monocostus* K.Schum., *Dimerocostus* Kuntze, *Costus* L., *Paracostus* C.D.Specht, *Chamaecostus* C.D.Specht & D.W.Stev., *Hellenia* Retz. a *Tapeinochilos* Miq., které dohromady čítají přibližně 300 druhů. Čeleď je sesterská k čeledi *Zingiberaceae*, která je díky svému ekonomickému významu výrazně lépe prozkoumána.

Cílem této práce byla částečná revize fylogeneze, kterou publikovala Specht (2006). K analýze fylogenetických vztahů byly použity molekulární metody, a to jednak klasické sekvenování úseku *trnL-F*, jednak next-generation sequencing – konkrétně metoda HybSeq, kdy byly použity předem designované proby pro získání cílových genů. Dalším z cílů této práce bylo srovnat absolutní velikosti genomu čeledi *Costaceae* se získanými fylogenetickými daty. Byla použita metoda průtokové cytometrie a z ní získané hodnoty byly poté namapovány na fylogenetické stromy. Z výsledku této analýzy se ukázalo, že v této čeledi absolutní velikost genomu blízce koreluje se zde navrženými fylogenetickými vztahy; tyto vztahy byly většinou rodově specifické.

**Klíčová slova:** Velikost genomu, *Costaceae*, fylogeneze, next-generation sequencing, průtoková cytometrie.

## Abstract

Genome size in plants varies greatly; in angiosperms this variability can make a 2400-fold difference. This is why the subject of many studies is the attempt to explain these differences in various contexts. One of these is the phylogenetic viewpoint, which studies the relationships between genome size and different taxonomic groups. Many studies have been published so far which explore the plants of the temperate zone; there are however very few studies investigating genome size in tropical plants. For this reason, the subject of this thesis is the analysis of genome size in the pantropical family *Costaceae* Nakai.

This family presently includes seven genera: *Monocostus* K.Schum., *Dimerocostus* Kuntze, *Costus* L., *Paracostus* C.D.Specht, *Chamaecostus* C.D.Specht & D.W.Stev, *Hellenia* Retz. and *Tapeinochilos* Miq., counting altogether about 300 species. The family is sister to the family *Zingiberaceae*, which has been notably better explored due to its economic importance.

One of the aims of this thesis was a partial revision of the phylogeny published by Specht (2006). Molecular methods were used to analyse the phylogenetic relationships; these methods included classical sequencing of the trnL-F region and next-generation sequencing, where probes designed in advance were used to obtain the targeted genes. Another aim of this thesis was to compare the genome size in *Costaceae* with obtained phylogenetic data. Flow cytometry was used to measure the absolute genome size. The acquired values were then mapped onto phylogenetic trees, showing that in this family the absolute genome size correlates closely with the phylogenetic relationships suggested herein; most of these relationships were genus-specific.

**Keywords:** Genome size, *Costaceae*, phylogeny, next-generation sequencing, flow cytometry.

# OBSAH

ÚVOD	1
1.1 KLDENÉ OTÁZKY	2
1.2 VELIKOST GENOMU	3
1.2.1 Evoluce velikosti genomu u rostlin	4
1.2.2 Vlastnosti rostlin spojené s velikostí genomu	7
1.2.3 Velikost genomu tropických rostlin	8
1.3 REKONSTRUKCE FYLOGENEZE	8
1.3.1 Maximální parsimonie	9
1.3.2 Maximum likelihood	10
1.3.3 Bayesovská analýza	11
1.3.4 Vybrané úseky DNA	11
1.3.5 NGS Next-Generation Sequencing	13
1.3.5.1 Hyb-Seq	15
1.3.5.2 Chybějící data	16
1.3.6 Gene trees versus Species trees	16
1.3.6.1 Konsenzuální metody	18
1.3.6.2 Konkatenace	18
1.3.6.3 Koalescenční metody	18
1.3.6.3.1 Summary methods	19
1.3.6.3.1.1 ASTRAL, ASTRAL-II: Accurate Species TRee ALgorithm	19
1.3.6.3.1.2 ASTRID: Accurate Species TREes from Internode Distances	19
1.3.6.3.1.3 MDC: Minimizing deep coalescence	19
1.3.6.3.1.4 MP-EST: Maximum Pseudo-likelihood approach for Estimating Species Trees under the coalescent model	20
1.3.6.3.1.5 NJst	20
1.3.6.3.1.6 STAR: Species Tree estimation using Average Ranks of coalescences	20
1.3.6.3.2 Co-estimation methods	21
1.3.6.3.2.1 BEST Bayesian Estimation of Species Trees	21
1.3.6.3.2.2 *BEAST: Bayesian Evolutionary Analysis by Sampling Trees	21
1.4 COSTACEAE NAKAI	22
1.4.1 Charakteristika	22
1.4.2 Rody	27
1.4.2.1 Chamaecostus C.D.Specht & D.W.Stev.	27
1.4.2.2 Costus L.	28

1.4.2.3	Dimerocostus Kuntze	28
1.4.2.4	Hellenia Retz.	29
1.4.2.5	Monocostus K.Schum.	30
1.4.2.6	Paracostus C.D.Specht	30
1.4.2.7	Tapeinochilos Miq.	31
1.4.3	Fylogeneze čeledi Costaceae	32
1.4.4	Chromozomové počty	34
<b>2.</b>	<b>MATERIÁL A METODY</b>	<b>37</b>
2.1	MATERIÁL	37
2.2	METODY	37
2.2.1	Průtoková cytometrie	37
2.2.2	Molekulární analýzy	38
2.2.2.1	Izolace DNA	38
2.2.2.2	Sekvenování úseku ITS a trnL-F a jejich zpracování	39
2.2.2.2.1	PCR	39
2.2.2.2.2	Přečištění	41
2.2.2.2.3	Sekvenování DNA	41
2.2.2.2.4	Práce se sekvencemi	42
2.2.2.2.5	Tvorba Datasetu	42
2.2.2.2.6	Fylogenetické stromy	43
2.2.2.2.6.1	Maximální parsimonie	43
2.2.2.2.6.2	Maximum likelihood	43
2.2.2.2.6.3	Bayesovská analýza	44
2.2.2.3	Příprava knihovny, získání NGS dat a jejich zpracování	45
2.2.2.3.1	Příprava knihovny pro Hyb-Seq	45
2.2.2.3.2	Obohacení (Enrichment)	47
2.2.2.3.3	Zpracování NGS dat	50
2.2.2.3.3.1	Testy monofylie	51
2.2.3	Vyhodnocení Velikosti genomu a fylogeneze	53
2.2.3.1	Korelace fylogenetické hypotézy s velikostí genomu	53
2.2.3.2	Rekonstrukce ancestrální velikosti genomu	54
<b>3.</b>	<b>VÝSLEDKY</b>	<b>55</b>
3.1	MĚŘENÍ VELIKOSTI GENOMU	55
3.2	DOPLNĚNÍ DRUHŮ A REKONSTRUKCE FYLOGENEZE	60
3.2.1	Fylogenetické analýzy z klasického sekvenování	61

3.2.1.1	Maximální parsimonie	61
3.2.1.2	Maximum likelihood	61
3.2.1.3	Bayesovská analýza	61
3.2.2	Fylogenetické analýzy z NGS-HybSeq	62
3.2.2.1	Odhady species trees	62
3.2.2.2	Testy monofylie	62
3.3	KORELACE VELIKOSTI GENOMU S FYLOGENEZÍ	63
3.3.1	Evoluce velikosti genomu	63
3.3.2	Odhady škálovacích parametrů	63
<b>4.</b>	<b>DISKUZE</b>	<b>65</b>
4.1	NOMENKLATURA A TAXONOMIE	65
4.2	SEKVENCE DNA	65
4.3	REKOSTRUKCE FYLOGENEZE	66
4.4	VELIKOST GENOMU	68
4.5	VELIKOST GENOMU A FYLOGENEZE	69
4.6	VELIKOST GENOMU A POČET CHROMOZÓMŮ	69
	<b>ZÁVĚR</b>	<b>71</b>
	<b>SEZNAM SYMBOLŮ A ZKRATEK</b>	<b>72</b>
	<b>POUŽITÁ LITERATURA</b>	<b>73</b>
	<b>PŘÍLOHY</b>	<b>74</b>

# ÚVOD

Studium rostlinné velikosti genomu je jedním z aktuálních botanických témat. Ať už se jedná o studie mezidruhové, zabývající korelacemi velikostí genomu s různými parametry či její evolucí. Nedílnou součástí jsou i studie na vnitrodruhové úrovni, které se zabývají přítomností různých cytotypů v rámci jednoho druhu.

Obecně je variabilita ve velikosti genomu (genome size) velmi široká, kdy dosahuje více než 24 000 násobného rozdílu s faktem, že velikost genomu často nekoreluje s mírou komplexity a evoluční odvozeností jednotlivých organismů. Hlavními zdroji takto široké velikostní variability jsou především transponovatelné elementy, které svým zmnožováním dokáží výrazně ovlivnit výslednou genomovou velikost. U některých skupin transponovatelné elementy mohou tvořit až 75 % celkového genomu. Dalším důležitým zdrojem variability jsou polyploidizační události, které dokáží výrazně měnit chemismus a komplexní chování organismů. Dalšími původci variability ve velikostech genomu jsou nerovnoměrná rekombinace u homologních chromozomů a ilegální rekombinace, jež vedou k celkovému snížení obsahu DNA v jádře. Ze znalosti těchto obecných mechanismů ovlivňujících velikost genomu vyplývá otázka, zda lze pozorovat určité evoluční trendy u jednotlivých rostlinných skupin. Souvislost mezi velikostí genomu a fylogenetickými hypotézami z doposud publikovaných prací příliš nevyplývá, nicméně v některých skupinách byly určité trendy pozorovány, a to například v rodech *Festuca* a *Allium* nebo čeledích *Liliaceae* či *Zingiberaceae*. Pro pochopení a správnou interpretaci evoluce velikosti genomu je nezbytné rozšířit studium co největším počtem taxonomických skupin pokrývajících dosavadní diverzitu rostlin (Leitch *et al.*, 1998; Bennetzen, 2000; Soltis *et al.*, 2003; Puttick *et al.*, 2015; Wendel *et al.*, 2016).

Doposud publikované práce se zabývaly především rostlinami temperátního pásu, naopak těch, které se věnují rostlinám tropickým, je velmi málo. Byly naznačeny některé trendy typické pro tropické rostliny, jako je obecná tendence k větším velikostem genomu a délce chromozomů, než u temperátních rostlin. Avšak jedná se o velmi kusé informace. Proto je náplní této diplomové práce rozšířit znalosti o velikosti genomu u tropických rostlin.

Konkrétně se tato práce zabývá analýzou velikosti genomu pantropicky rozšířené čeledi *Costaceae* Nakai., která je sesterskou skupinou výše zmíněné čeledi *Zingiberaceae*. Tato čeleď má centrum druhové diverzity v Jižní Americe na rozdíl od *Zingiberaceae* s nejvyšší diverzitou v Jihovýchodní Asii. *Costaceae* v současnosti čítá něco mezi 150 až 300 druhy

náležících do sedmi rodů: *Monocostus*, *Dimerocostus*, *Chamaecostus* rozšířené v Jižní Americe, *Costus*, *Hellenia*, *Paracostus* rozšířené na více kontinentech a *Tapeinochilos* s omezenou distribucí v Jihovýchodní Asii. Tento taxonomický koncept vyplývá z v minulosti publikovaných prací Chelsie Specht (2001 -2006), kdy byly pro analýzu fylogenetických vztahů použity molekulární i morfologické metody. Výsledky těchto prací sice objasnily některé vztahy v této čeledi, nicméně připouští i jistou míru nejasností. Proto je součástí této diplomové práce i částečná fylogenetická revize. Jsou zde použity molekulární metody konkrétně DNA sekvenování. Byl použit jak klasický přístup sekvenování Sangerovou metodou několika lokusů, tak sekvenování nové generace v podobě HybSequ, který využívá specificky vyvinutých prob pro sekvenování low-copy genů (Specht *et al.*, 2001; Specht, 2006a; Specht & Stevenson, 2006; Cronn *et al.*, 2012; Weitemier *et al.*, 2014).

Díky zařazení NGS je možné nejen získat fylogenetické hypotézy založené na mnoha genech, ale je též možné tento přístup porovnat s doposud používanými metodami pro objasnění fylogeneze v této čeledi.

## 1.1 Kladené otázky

- 1) Jaké jsou fylogenetické vztahy v čeledi *Costaceae*, jsou srovnatelné s již publikovanou prací nebo lze nalézt i alternativní hypotézu?
- 2) Koreluje velikost genomu s fylogenezí čeledi?
- 3) Mají různé geografické skupiny i odlišnou velikost genomu, a pokud ano, jsou tyto skupiny monofyletické?
- 4) Je možné pozorovat jednoznačný trend v evoluci velikosti genomu, nebo se genom opakovaně zvětšuje a zmenšuje?
- 5) Jakou roli hraje v této čeledi polyploidizace?

## 1.2 Velikost genomu

Velikost genomu je dnes velmi studovanou charakteristikou druhu, nejčastěji bývá popsána jako C-hodnota neboli množství DNA obsažené v jádře s nereplikovanou haploidní sadou chromozómů (Swift, 1950). S tím souvisí některé dřívější nesrovnalosti v termínech, a proto byly zavedeny pojmy holoploidní velikost genomu (C-hodnota), která udává množství nereplikované DNA v jádře bez ohledu na ploidní stupeň, oproti tomu monoploidní velikost genomu (Cx-hodnota) vyjadřuje obsah DNA v jedné chromozomové sadě, výpočet Cx-hodnoty lze provést vydělením 2C-hodnoty stupněm ploidie. Platí že, v případě diploidního organismu se 1C-hodnota rovná 1Cx-hodnotě, oproti tomu u polyploidního organismu je 1Cx-hodnota většinou menší než 1C-hodnota (Greilhuber, 2005; Suda *et al.*, 2015).

Velikost genomu bývá nejčastěji vyjádřena v pikogramech (pg), ale je možné provést přepočítání na počet párů bází. Přepočítání lze provést vynásobením velikosti genomu v (pg)  $0,978 \times 10^9$  (Dolezel *et al.*, 2003). V průběhu studia velikosti genomu bylo prokázáno, že obsah jaderné DNA neodpovídá komplexitě jeho nositele, tento fakt byl označen jako paradox C-hodnoty (Thomas, 1971). Tento paradox lze vysvětlit faktem, že převážnou většinu obsahu jaderné DNA tvoří nekódující úseky, tady ty co nemají na první pohled zjevný vliv na svého nositele a naopak bylo zjištěno, že počet genů napříč rostlinami je v celku konzistentní (Bennetzen, 2000).

U rostlin bylo zjištěno, že v rámci druhu je velikost genomu zpravidla konstantní, ale mezi druhy, i v rámci rodu, se často může výrazně lišit (Bennett *et al.*, 2000). Z doposud publikovaných prací má nejmenší velikost genomu *Genlisea aurea* s velikostí 60 Mbp, zatímco největší velikost byla naměřena u druhu *Paris japonica* (>150 Gbp), což odpovídá více jak 2000 násobku (Wendel *et al.*, 2016). Obecně proti extrémním hodnotám má velká část krytosemenných rostlin genom malý. Pro exaktní představu škály velikostí bylo navrženo třídění na genom velmi malý, což odpovídá 1C hodnotě  $\leq 1,4$  pg, malý s 1C hodnotami  $\leq 3,5$  pg, intermediární 1C 3,51 – 13,99 pg, velký 1C  $> 14,0$  pg a velmi velký kdy 1C hodnota je  $> 35,0$  pg (Leitch *et al.*, 1998; Soltis *et al.*, 2003).

Byly hledány mechanismy, které vysvětlují takto enormní rozrůzněnost v genomových velikostech nezávislých na komplexitě jejich nositele. Výsledkem bylo zjištění, že velikost genomu se může v evoluci jak zvětšovat, tak zmenšovat nikoli, pouze nabývat na objemu, jak předpokládala teorie genomické obesity (Bennetzen & Kellogg, 1997). Za nárůst celkového

obsahu DNA na molekulární úrovni mohou jednak transponovatelné elementy. U rostlin jsou to především LTR-retrotranspozony, což jsou sekvence DNA, které jsou schopné přesouvat se z místa na místo prostřednictvím RNA intermediátu, který je opětovně vložen do DNA reverzní transkripcí. Tato skupina elementů je specifická tím, že nese na obou svých koncích dlouhou repetici. Za zvětšení velikosti genomu je též zodpovědná polyploidizace, kdy dochází k celogenomové duplikaci (WGD – whole genome duplication), tento proces je velmi častý u mnoha rodů i čeledí, uvažuje se o polyploidním původu 30 – 70 % kvetoucích rostlin. Představuje tak důležitý proces v diverzifikaci krytosemenných rostlin. Naopak ke snížení velikosti genomu vede nerovnoměrná rekombinace u homologních chromozomů a ilegální rekombinace, jež je sice méně efektivní co do objemu ztráty DNA, za to je častější (Masterson, 1994; Devos *et al.*, 2002).

Za celkovou kontrolu a stabilitu obsahu jaderné DNA jsou zodpovědné některé epigenetické faktory, jako jsou methylace či různé rodiny malých RNA, které zabraňují v nekontrolovatelné proliferaci transponovatelných elementů tím, že v místě jejich výskytu udržují heterochromatin. Na základě toho je usuzováno, že transponovatelné elementy, které dokáží tak mocně zvětšit obsah jaderné DNA, mohou volně fungovat pouze v případě, že vnitřní regulační mechanismy jsou dočasně odstraněny, což může být zapříčiněno například vnějším stresem či mezidruhovou hybridizací. Nutno dodat, že toto platí pro krytosemenné rostliny, pro ostatní linie cévnatých rostlin nejsou doposud epigenetické procesy příliš známé (Leitch & Leitch, 2012; Wendel *et al.*, 2016).

Dalšími faktory ovlivňující velikost genomu jsou: selekce, genetický drift a efektivní velikost populace (Grover & Wendel, 2010).

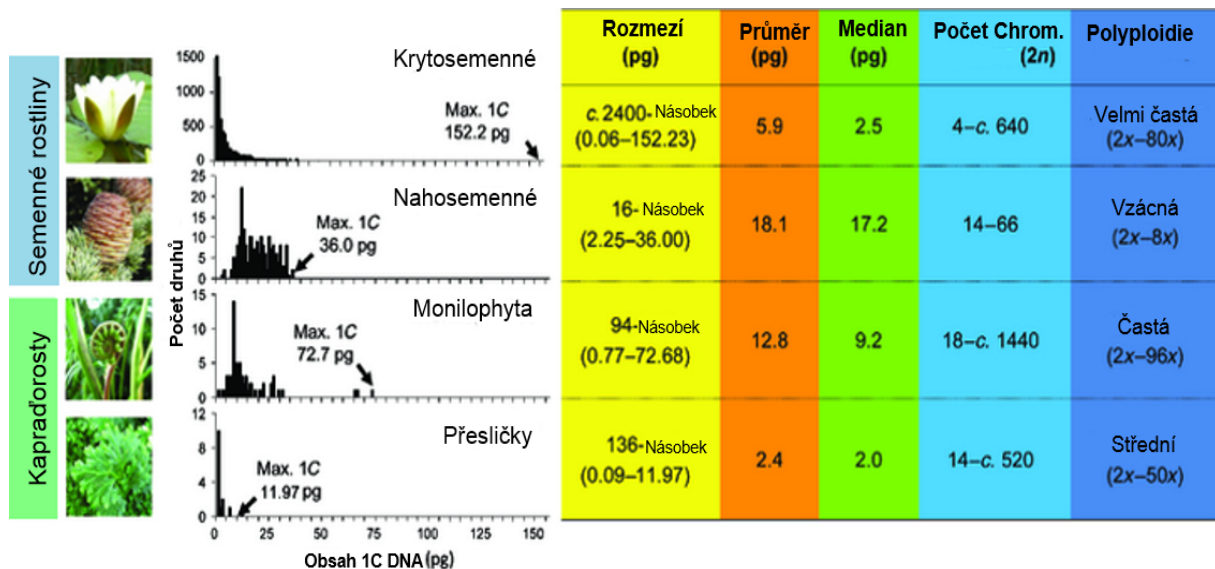
## 1.2.1 Evoluce velikosti genomu u rostlin

Jedním z hledisek, jak je nahlíženo na velikost genomu, je jeho postavení v evolučním kontextu. Zde jsou stručně shrnuty zásadní poznatky. Byly zjištěny zásadní rozdíly mezi jednotlivými „odděleními“. V případě plavuní jsou velikosti genomu převážně malé, rozpětí velikostí se pohybuje mezi 1C 0,09 – 11,97 pg, tedy v rozpětí 136 násobku. V případě Monilophyt, kam dnes náleží kapradiny, přesličky a Psilophyta, má nejvíce druhů malý až střední velikost genomu, ačkoli jsou zde zastoupeny druhy s výrazně velkým genomem, a to konkrétně u rodu *Psilotum* a *Ophioglossum*, kdy se velikost genomu pohybuje mezi 1C 65 – 72,7 pg, u zbytku Monilophyt se pohybují 1C hodnoty od 0,77 do 30 pg. Celkově mají 1C

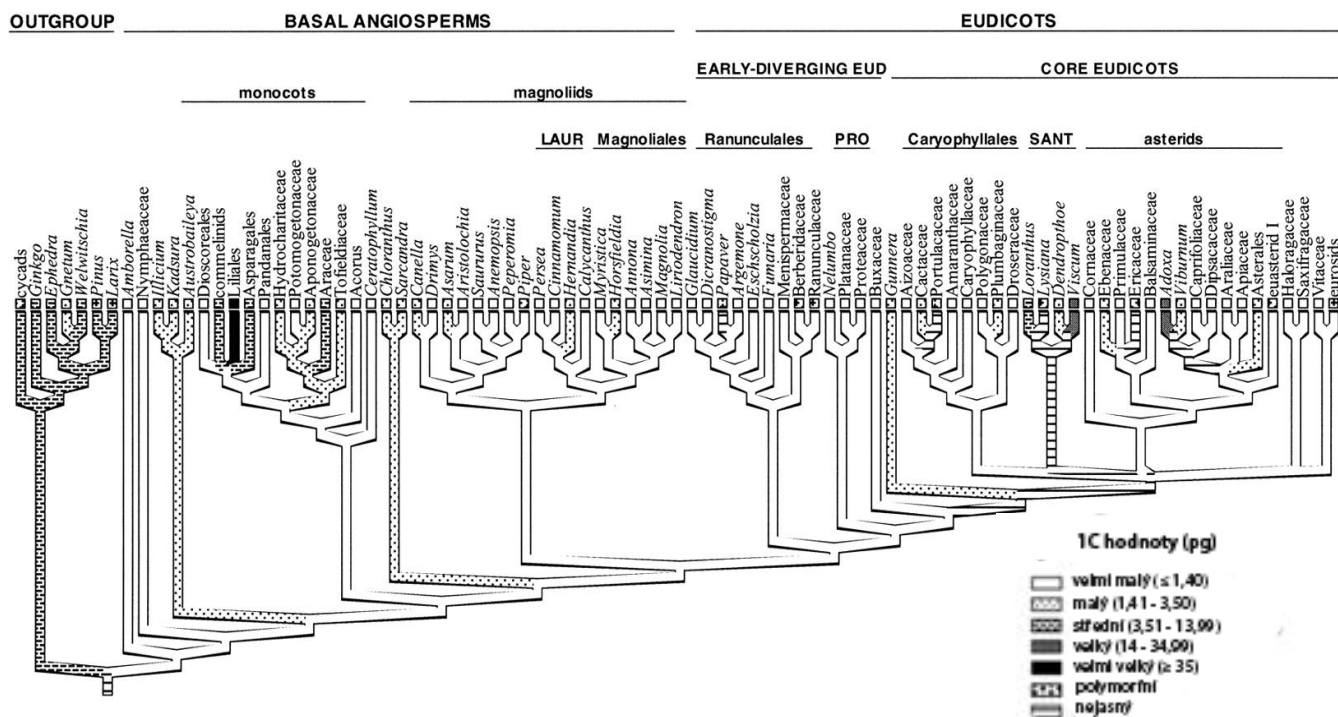
hodnoty rozpětí 94 násobné. V případě nahosemenných rostlin byl překvapivě nalezen velmi malý rozdíl ve velikostech genomu mezi jednotlivými taxony, celkově odpovídající pouze 16 násobku. Oproti tomu krytosemenné rostliny se od sebe liší více jak 2400 násobně v rozmezí 1C 0,06 – 152,26 pg viz Obr. 1, kde jsou navíc uvedeny průměrné hodnoty, škála počtu chromozomů a četnost polyploidizace (Leitch & Leitch, 2012; Suda *et al.*, 2015).

Co se konkrétně krytosemenných rostlin týká Soltis *et al.* (2003) vynesli pomocí parsimonické analýzy velikosti genomu na fylogenetický strom krytosemenných rostlin (Obr. č. 2), kde je patrné, že většina bazálních skupin má velmi malý genom. Konkrétně se to týká bazálních krytosemenných, bazálních jednoděložných, bazálních dvouděložných a některých z hlavních cladů odvozených dvouděložných např.: Caryophyllales nebo Asteridy. Naopak v některých vrcholných liniích jsou zastoupeny taxony s velmi velkými genomy např.: v řádu Liliales, rod *Viscum* náležící do řádu Santalales a rod *Adoxa* z řádu Ericales.

Zřejmý rozdíl ve variabilitě velikosti genomu a celkově odlišné evoluční úspěšnosti vedl k podrobnému zkoumání příčin odlišností mezi nahosemennými a krytosemennými rostlinami. Jedním z důvodů je, že nahosemenné rostliny představují dřeviny, které mají obecně pomalejší rychlost evoluce (Smith & Donoghue, 2008). To souvisí s nízkou rychlostí oplození, pomalým růstem embrya, dlouhověkostí a častým opílováním větrem. Tyto vlastnosti vedou k tvorbě velkých populací, kde je malý prostor pro genetický drift, přičemž selekce působí pomalu. Dále nahosemenné rostliny mají nízkou frekvenci polyploidizace a genetickou diverzitu. Dohromady tyto faktory vedou k pomalé evoluci genomu u nahosemenných. V případě krytosemenných rostlin je situace opačná, tedy je u nich častá polyploidizace, vysoký podíl transponovatelných elementů, mezidruhové hybridizace, rychlý růst embrya, tvoří krátkověké byliny s častou entomofilií, tedy dochází k tvorbě malých fragmentovaných populací, kde je prostor pro rychlou selekci a genetický drift, které umožňují rychlou evoluční odpověď (Leitch & Leitch, 2012). V současnosti se jeví jako zásadní vysoký podíl polyploidních událostí u krytosemenných rostlin, kdy po WGD následuje zmenšení genomu, v důsledku fragmentace genomu a ztráty velké části duplikovaných genů. Je tedy volný prostor pro utváření nových funkcí. Ukazuje se, že u krytosemenných rostlin rychlost změny velikosti genomu úzce koreluje s rychlostí evoluce, viz Obr. 2 (Puttick *et al.*, 2015; Wendel *et al.*, 2016).



**Obrázek 1:** Rozložení velikostí genomu (1C-hodnot) cévnatých rostlin se shrnutím průměrných hodnot, mediánů a rozmezí 1C hodnot, frekvencí polyploidizace a škálou počtu chromozomů (Suda *et al.*, 2015).



**Obrázek 2:** Rekonstrukce velikosti genomu kryptosemenných rostlin pomocí maximální parsimonie. Jsou uvedeny 1C-hodnoty v pikogramech (Soltis *et al.*, 2003).

## 1.2.2 Vlastnosti rostlin spojené s velikostí genomu

Velikost genomu lze také chápat jako fenotypový znak, který je korelován s mnoha proměnnými ať už pozitivně, či negativě, silně, nebo méně patrně. Pozitivní vztah k velikosti genomu má velikost buňky a rychlost buněčného dělení, potažmo rychlost růstu tkání. Už pouhé tyto vztahy naznačují, že velikost genomu je jedním ze základních vlastností, která ovlivňuje rostlinou fitness (Bennett, 1987; Knight *et al.*, 2005; Knight & Beaulieu, 2008; Hodgson *et al.*, 2010; Šimová & Herben, 2012). Dále je velikost genomu korelována se specifickou listovou plochou (SLA – specific leaf area), ačkoli je síla tohoto vztahu rozdílná mezi různými rostlinnými skupinami např.: U nahosemenných rostlin, jedná se o korelaci převážně negativní oproti krytosemenným rostlinám, kde je většinou pozorována korelace pozitivní (Grotkopp *et al.*, 2004; Knight *et al.*, 2005; Morgan & Westoby, 2005; Beaulieu *et al.*, 2008).

Dále byl empiricky pozorován negativní vztah velikosti genomu s relativní růstovou rychlostí semenáčků, oproti tomu byla pozorována pozitivní korelace mezi minimální generační dobou. To vysvětluje fakt, že rostliny s většími genomy jsou typicky vytrvalé. Navíc je velikost genomu korelována s hmotností semen pozitivně a jejich počtem negativně (Thompson, 1990; Grotkopp *et al.*, 2004; Knight *et al.*, 2005; Leitch & Bennett, 2007; Herben *et al.*, 2012). Dalším s velikostí genomu negativně korelovaným znakem je hustota umístění průduchů, naproti tomu jejich velikost je s velikostí genomu korelována pozitivně (Beaulieu *et al.*, 2008b; Hodgson *et al.*, 2010).

Bylo analyzováno mnoho dalších proměnných např. Vinogradov (2003) pozoroval, že ohrožené druhy rostlin mívají větší genomy než rostliny nechráněné. Z toho je vyvozován závěr, že rostliny s většími genomy snáze podléhají extinkci. Dále byl analyzován vztah velikosti genomu a regionální abundance, kdy rostliny s většími genomy mají nižší abundanci než rostliny s malým genomem. Tento vztah vyšel signifikantnější u jednoletých rostlin, oproti trvalkám (Šimová & Herben, 2012). Nakonec jedním z aktuálních témat je studium invaznosti rostlin a není divu, že do tohoto studia byla zahrnuta i velikost genomu, kdy se ukazuje, že znaky spojené s velkými genomy jsou neslučitelné s úspěšnou invází (Rejmánek, 1996, 2000; Suda *et al.*, 2015).

Obecně lze říci, že rostliny s malým genomem mohou zaujímat širší ekologickou niku a mohou mít rozličné životní strategie, oproti tomu rostliny s velkým genomem jsou obligátně vytrvalé. Navíc rostliny s velkým genomem mohou být selektovány v na živiny chudých

stanovištích, kde je nedostatek fosforu a dusíku, jelikož jsou nedostatkem živin limitovány při syntéze DNA. Obecně rostliny s velkým genomem nejsou schopny zaujímat extrémní stanoviště, kde jsou limitujícími faktory např. sucho, nízká nebo vysoká teplota (Knight & Ackerly, 2002; Beaulieu *et al.*, 2007; Šmarda *et al.*, 2013; Leitch *et al.*, 2014).

### 1.2.3 Velikost genomu tropických rostlin

Ohriho práce (2005) se zabývala tématem, jak je velikost genomu ovlivněná klimatickými podmínkami a životní formou, kdy byly použity C4 hodnoty získané z databáze C hodnot pro krytosemenné rostliny (Bennett *et al.*, 2000). Zde jsou stručně shrnuty poznatky týkající se tropických rostlin. Škála použitých C4 hodnot se pohybovala v rozmezí 356-ti násobku, kdy nejnižší použitá hodnota byla 0,66 pg druhu *Amoreuxia wrightii* (Cochlospemaceae) a nejvyšší 235,4 pg pro *Haemanthus katherinae* (Amaryllidaceae). Ukázalo se, že celkově tropické rostliny mají signifikantně menší velikost genomu  $4C = 17,35$  pg než rostliny temperátní  $4C = 29,2$  pg. Stejný pattern bylo pozorováno při vzájemném srovnávání samostatně jednoděložných a dvouděložných rostlin, dále při samostatném porovnání bylin a dřevin.

V práci Levin & Funderburg (1979) byla též porovnána  $4C$  hodnota tropických a temperátních rostlin, kde navíc byla zkoumána i průměrná délka chromozómu. Výsledek byl obdobný. Tropické rostliny měly jak menší velikost genomu, tak průměrnou délku chromozómu. U tropických rostlin byla průměrná délka chromozómu 3,1  $\mu\text{m}$  oproti temperátním rostlinám s průměrnou délkou 5,6  $\mu\text{m}$ .

## 1.3 Rekonstrukce fylogeneze

Rekonstrukce fylogeneze neboli analýza příbuzenských vztahů mezi taxony je jeden z přístupů jak taxonomicky řadit organismy. Výsledkem těchto analýz jsou fylogenetické stromy, které mohou být zkonstruovány řadou metod za použití rozličných typů dat. Data používaná ke konstrukci fylogenetických stromů mohou být ve formě vlastních znaků např. přítomnost určitého nukleotidu na konkrétním místě DNA sekvence, nebo ve formě distancí neboli vzdáleností např. procentuální rozdílnost taxonů ve sledovaných znacích. Používané znaky mohou být jak kvalitativní tak kvantitativní. Důležitou podmínkou pro volbu znaků je

jejich vzájemná nezávislost, a pro porovnání konkrétního znaku, jejich homologie. Pro rekonstrukci fylogeneze mohou být použity např. molekulární znaky, kdy se může jednat o pořadí nukleotidů v sekvenci DNA, nebo mikrosatelity (krátké tandemové repetice rozmístěny napříč celým genomem) a další. Do fylogenetických analýz lze zapojit i morfologické znaky, avšak je nezbytné, aby byly homologické, což v případě morfologických znaků nemusí být na první pohled zřejmé. ([http://www.zoologie.frasma.cz/fylogeneze/fylogeneze\\_C.html](http://www.zoologie.frasma.cz/fylogeneze/fylogeneze_C.html); <http://www.protistologie.cz/files/MolTax/Molekularni%20taxonomie1-text.pdf>).

Pro tvorbu fylogenetických hypotéz můžeme zvolit dva hlavní přístupy. Prvním přístupem je definice algoritmu, který vede ke konstrukci stromu. Sem spadají všechny metody shlukové analýzy např. metoda Neighbour-joining (Nejbližšího souseda) či některé z ordinačních metod jako PCA nebo PCoA. Pro všechny tyto metody platí, že je vždy vytvořena určitá matice vzdáleností a na základě ní je pak odvozen dendrogram. Druhý přístup se snaží definovat kritéria k porovnávání alternativních hypotéz a rozhodnout, která hypotéza je nejoptimálnější. Sem patří metody Maximální parsimonie (MP), Maximální věrohodnosti (ML), Bayesovská analýza (BI), Fitchova-Margoliashova metoda a metoda Minimální evoluce (ME) (Marhold & Suda, 2001; <https://web.natur.cuni.cz/~muncling/FYLO.DOC>).

### **1.3.1 Maximální parsimonie**

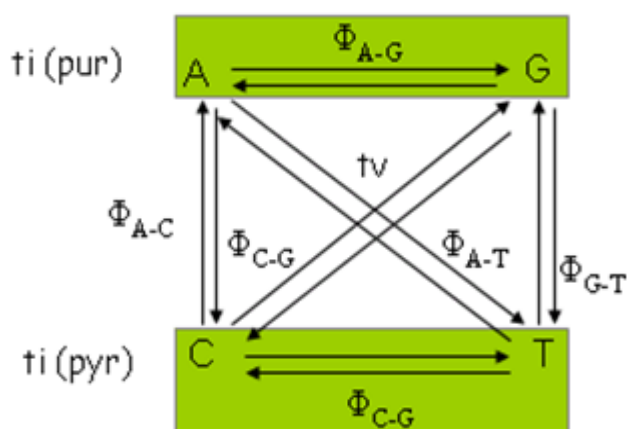
Metoda Maximální parsimonie je založena na principu hledání takových stromů, které budou tvořeny co nejmenším počtem evolučních kroků neboli budou co nejkratší. Tato metoda je založena na principu Occamovy břitvy, která spočívá v tom, že nejjednodušší možná hypotéza, vysvětlující daný problém, je ta nejsprávnější. Tedy ta hypotéza, která vyžaduje nejméně evolučních změn. Při hledání nejparsimoničtějších stromů je možnost prohledat všechny možné stromy a z nich vybrat ten nejparsimoničtější (toto lze uplatnit u omezeného množství taxonů zahrnutých do analýzy, jedná se o rozmezí mezi 8 – 20 taxony, záleží na výkonosti výpočetní techniky) nebo lze využít tzv. heuristických algoritmů, které umožní prohledávat prostor všech stromů pomocí hledání lokálních optim. Například při použití heuristické metody postupného přidávání (Stepwise addition) je na počátek stromu tvořen třemi taxony, poté je na strom přidán další taxon, pro který je spočteno, která topologie je nejkratší, takto se přidává po jednotlivých taxonech, dokud nejsou zařazeny všechny. Při

použití heuristických algoritmů je výsledek analýz vždy přibližný, a proto často následuje krok hodnocení spolehlivosti stromu např. pomocí Bootstrapu nebo Jackknifu.

### 1.3.2 Maximum likelihood

Statistická metoda Maximum likelihood (Maximální věrohodnosti) se snaží zodpovědět otázku, jaká je pravděpodobnost, že budou pozorována vstupní data za použití konkrétního modelu evoluce DNA. Substituční model určuje pravděpodobnost jednotlivých evolučních změn s tím, že je brána v potaz topologie stromu i délka větví, které se též podílí na určení pravděpodobnosti modelu. Jsou generovány stromy a na základě vstupních dat jsou skórovány podle zvoleného modelu. Za nejlepší strom je pak považován ten s největším likelihood skóre (Huelsenbeck & Crandall, 1997).

V této diplomové práci je použit model GTR+GAMMA. Gamma popisuje kontinuální model, který popisuje míru heterogenity závislou na rychlosti změny znaku. K popisu je používán parametr  $\alpha$ , jehož hodnota je rovna reciproké hodnotě gamma a jeho vzrůstající hodnota svědčí o snižující se míře heterogenity. V případě, že se  $\alpha$  blíží nekonečnu, platí, že se všechna místa na stromě vyvíjí stejnou rychlostí. GTR (Generalised time-reversible) je jeden z modelů týkající se evoluce sekvence DNA, kdy jsou určeny parametry, za jakých okolností může být jeden nukleotid nahrazen za druhý. V případě GTR je pro čtyři znaky možných šest různých parametrů popisujících substitučních rychlost.



**Obrázek 3:** Nákres substitučního modelu GTR. ti = tranzice, tv = transverze, pur (purinové báze adenin a guanin) pyr (pyrimidinové báze cytosin a thymin) GTR model má pro každou změnu unikátní substituční rychlost ([http://www.ccg.unam.mx/~vinaesa/Model\\_fitting\\_in\\_phylogenetics.html](http://www.ccg.unam.mx/~vinaesa/Model_fitting_in_phylogenetics.html)).

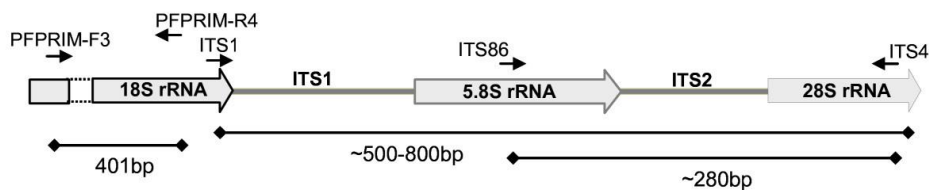
### 1.3.3 Bayesovská analýza

Bayesovská analýza je založena na Bayesovském teoremu o podmíněné pravděpodobnosti, tedy že lze spočítat posteriorní pravděpodobnost stromu z pravděpodobnosti apriorní, což jsou náhodně sestavené stromy ze vstupních dat před jakýmkoli testováním, kdy jsou během analýzy apriorní stromy testovány, s jakou pravděpodobností vzniknou za použití konkrétních dat. Výstupem je soubor stromů, které jsou nejpravděpodobnější při použití daného modelu a dat. S ohledem na velké množství různých možných topologií stromu při použití většího počtu taxonů, je pro prohledávání krajiny stromů použita Metropolisova-Hastingsova metoda výběru, založená na metodě Monte Carlo a vycházející z předpokladů Markovových řetězců (Markov chain Monte Carlo metod, MCMC). Analýza probíhá následovně: Na začátku je vybrán jeden strom, a to buď náhodně, nebo je předem definován, poté je modifikací délek větví nebo přemístěním jedné z větví navržen nový stav řetězce a vypočítána jeho pravděpodobnost. Strom je akceptován v případě, že je pravděpodobnost větší než původní stav. V opačném případě zůstává stav topologie stromu původní. Takto je prohledána celá krajina stromů s tím, že řetězce si mezi sebou vyměňují informace, aby se zamezilo dosahu lokálního maxima. Takto analýza pokračuje, dokud nedojde k tzv. konvergenci stromů neboli stabilní hodnotě pravděpodobnosti (<https://web.natur.cuni.cz/~muncling/FYLO.DOC>).

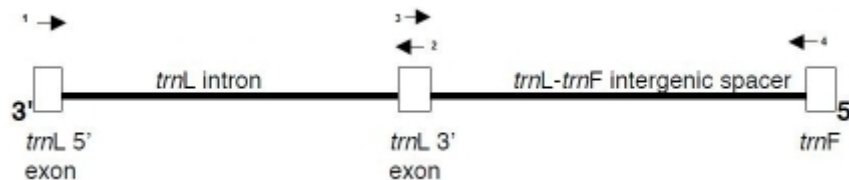
### 1.3.4 Vybrané úseky DNA

V této diplomové práci byly sekvenovány dva lokusy - jaderný ITS a chloroplastový *trnL-F*, které byly použity jako zdroj dat pro fylogenetické analýzy. ITS (Internal transcribed spacer) neboli vnitřní přepisovaný spacer je součástí genů kódujících ribosomální DNA. Tento úsek se nejčastěji skládá z části 18S rRNA, vnitřní přepisované oblasti (ITS1), celého 5,8S rRNA, vnitřní přepisované oblasti (ITS2) a části 25S rRNA. Tento úsek se v genomu vyskytuje v tandemových repetících. Tedy je tento úsek v genomu přítomen ve stovkách až tisících kopií. Z toho plyne jedna z výhod využití tohoto úseku, a to že stačí malé množství DNA k úspěšné amplifikaci tohoto úseku. Druhou výhodou je relativně velká variabilita, díky níž je možné tento úsek využít pro objasnění vztahů na různých taxonomických úrovních. S tímto úsekem je spojen další fenomén, a to concerted evolution. Jelikož se jedná o jaderný

úsek, je zřejmé, že se dědí kopie tohoto úseku od obou rodičů, nicméně tento úsek podléhá rychle intramolekulární rekombinaci (homogenizaci). Díky tomuto jevu je možné tento úsek využít pro fylogenetické analýzy rovnou přímým sekvenováním bez klonování, na druhou stranu tento jev může představovat riziko, v případě, že studovaní jedinci vznikli dávnou hybridizací nebo allopoliploidizací. V tomto případě můžeme dostat zkreslenou informaci díky přítomnosti částí DNA z jiného druhu. Nicméně tento jev neplatí univerzálně a bylo již publikováno mnoho prací, kde to zdaleka neplatí např. (Mayol & Rosselló, 2001; Nieto Feliner & Rosselló, 2007). V případě, že úsek není homogenizován, a nebo je jen z části homogenizován, je nezbytné pro získání sekvencí tohoto úseku klonovat. Na druhou stranu lze takto získané sekvence využít pro zjištění rodičů. Druhým úsekem je chloroplastový *trnL-F*, který zahrnuje geny kódující tRNA. Jehož přibližná délka je 1000 bp. Tento úsek je na kruhové molekule cpDNA umístěn v LSC (long single copy) regionu a zahrnuje *trnL* intron, 3'*trnL* exon a *trnL-F* intergenový spacer. *TrnL-F* je vhodný pro systematické studie blízké příbuzných druhů, avšak může být problematické alignovat jednotlivé sekvence zvláště u příbuzensky vzdálenějších taxonů.



**Obrázek 4:** Schématický náčrt sekvenovaného úseku ITS pro houby s vyznačenými místy pro vazbu primerů. (Rostlinný úsek je složen místo z 28S rRNA z 25S rRNA) (<http://www.biomedcentral.com/1471-2415/8/7/figure/F1?highres=y>)



**Obrázek 5:** Schématický náčrt sekvenovaného úseku *trnL-F* ([http://openi.nlm.nih.gov/detailedresult.php?img=3055156\\_97320630005405F1&req=4](http://openi.nlm.nih.gov/detailedresult.php?img=3055156_97320630005405F1&req=4))

### 1.3.5 NGS Next-Generation Sequencing

Na rozdíl od tradičního Sangerova sekvenování, které probíhá v kapilárách a je založeno na principu zaznamenávání fluorescenčně značených dideoxinukleotidů (ddNTP) vážících se na DNA místo běžných nukleotidů, je NGS (next-generation sequencing) kvantitativní metodou. To znamená, že poskytuje jak primární nukleotidové sekvence, tak kvantitativní informaci kolikrát byla daná báze zaznamenána tzv. sequencing depth (Shendure & Ji, 2008; Straub *et al.*, 2012). Dále je schopno poskytnout několika násobně větší objem dat, v řádu stovek až miliard bází za jeden běh. V případě NGS je možné sekvenovat celé genomy, nebo použít nějaký z přístupů cíleného obohacení (target enrichment), který cíleně sníží množství DNA určené k sekvenování, což je velmi praktické, obzvláště v případech rostlin s velkými genomy. K těmto přístupům patří PCR-enrichment (PCR-oboahcení), kdy jsou vyvíjeny speciální primery pro cílené obohacení, dále hybridization-based enrichment (na hybridizaci založené obohacení), jenž funguje na principu vývoje specifických prob, které hybridizují s DNA, dále restriction enzyme-based enrichment (oboahcení založené na štěpení pomocí restrikčních enzymů) a izolace mRNA, kdy je tento přístup soustředěn na část genomu, která je transkribována tzv. transkriptom. Transkriptom se stává z relativně malé části celkového genomu v rozmezí 25 – 1 % u krytosemenných rostlin (Cronn *et al.*, 2012).

Pro lepší orientaci jsou nejpoužívanější NGS technologie shrnuty v Tab. č. 1, kde jsou zároveň stručně uvedeny výhody a nevýhody jednotlivých metod.

**Tabulka 1:** Porovnání next-generation sequencing platform převzato z: (Metzker, 2010).

Platforma	Knihovna/příprava templátu	Princip sekvenování	Délka readů (báze)	Doba běhu (dny)	Gb na běh	Cena přístroje (US\$)	Klady	Zápory
Roche/454's GS FLX Titanium	Frag, MP/emulzníPCR	Pyrosekvenování	330*	0,35	0,45	500.000	Dlouhé ready, které zlepšují mapování na repetitivní regiony; rychlá doba běhu	Vysoká cena chemie, vysoký error rate u homopolymerů
Illumina/Solexa's GAII	Frag, MP/solid-fáze	Reverzibilní terminace	75 či 100	4 <sup>‡</sup> , 9 <sup>§</sup>	18 <sup>‡</sup> , 35 <sup>§</sup>	540.000	V současnosti nejpoužívanější platforma	Nízká schopnost multiplexování vzorků
Life/APG's SOLiD 3	Frag,MP/emulzníPCR	Ligace	50	7 <sup>‡</sup> , 14 <sup>§</sup>	30 <sup>‡</sup> , 50 <sup>§</sup>	595.000	Two-base kódování provádí přirozené korekce chyb	Dlouhá doba běhu
Polonator G.007	Pouze MP /emulzníPCR	Ligace, štěpení prob	26	5 <sup>§</sup>	12 <sup>§</sup>	170.000	Nejlevnější platforma; možnost volného použití alternativní NGS chemie	Povinnost udržovat a kontrolovat kvalitu činidel; nejkratší délka readů ze všech
Helicos BioSciences HeliScope	Frag, MP /jednomolekulové	Reverzibilní terminace	32*	8 <sup>‡</sup>	37 <sup>‡</sup>	999.000	Není třeba předchozí PRC a je tak zamezeno chybám při amplifikaci	Vysoké error rates oproti RT chemi
Pacific Biosciences (target release: 2010)	Pouze fragmenty /jednomolekulové	Sekvenování v reálném čase	964*	N/A	N/A	N/A	Má největší potenciál vytvářet ready větší než 1 kb	Vysoké error rates v porovnání s ostatními NGS chemiemi

\*Průměrná délka readu.  
<sup>‡</sup>Fragmenty.  
<sup>§</sup>Mate-pair ready. Frag, fragment; GA, Genome Analyzer; GS, Genome Sequencer; MP, mate-pair; N/A, není k dispozici; NGS, next-generation sequencing; RT, reversible terminator; SOLiD, support oligonucleotide ligation detection.

### 1.3.5.1 Hyb-Seq

Hyb-Seq neboli Hybridization based sequencing je jednou z NGS metod, která kombinuje cílené obohacení (target enrichment) a mělké sekvenování (genome skimming) pro vývoj prob, kdy jejich použití vede ke snížení celkové genomové komplexity. Výsledkem je získání stovek až tisíců low-copy i high-copy jaderných exonů. Mimo jiné jsou získány i malé části intronů, vyskytující se bezprostředně před a za samotným exonem tzv. splash zones, které je možné využít pro řešení vztahů na nižší taxonomické úrovni. Tento přístup poskytuje rozsáhlý objem dat, který lze použít pro fylogenomické studie (Weitemier *et al.*, 2014).

Celý proces HybSeq metody se skládá ze tří částí. První je design prob, druhý je obohacení o cílené úseky a třetí částí je úprava získaných readů a analýza dat. První je tedy design prob, kdy jako vstupní materiál slouží genom získaný genome skimmingem a transkriptom od jednoho taxonu. Proby se skládají z RNA oligonukleotidů připojených vodíkovými můstky ke komplementární sekvenci (RNA-DNA hybridizace), jsou tedy vysoce specifické. Princip designu prob spočívá ve srovnání genomu a transkriptomu, kdy jsou hledány single nebo low copy úseky, které vykazují vhodnou variabilitu. Tedy ne příliš malou, aby bylo možné rekonstruovat fylogenetické vztahy a ne příliš velkou, aby se zamezilo získání paralogů. Též je požadavek na jejich určitou délku, a to je v případě exonu minimálně 120 bp a v případě genu sestaveného z těchto exonů o minimální délce 960 bp. Během tohoto procesu je odfiltrována většina repetitivních úseků, cpDNA a mtDNA.

Samotné obohacení může být provedeno dvěma způsoby jednak on-array neboli na pevném podkladu, který tvoří nylonový filtr, sklíčko, nebo microarray. V tomto případě je testovaná DNA nalita na destičku, kde požadované úseky hybridizují s probami. Poté se odstraní nenahybridizované úseky DNA. Nakonec je obohacená DNA namnožena pomocí PCR, takto jsou vzorky připraveny pro sekvenování. Druhý přístup je obdobný s tím rozdílem, že místo pevného podkladu probíhá obohacení ve vodném roztoku (solution hybridization), proby jsou umístěny přímo v roztoku a jsou většinou opatřeny biotinem. Zbytek procesu je stejný jako v případě on-array enrichmentu. Jak bylo výše zmíněno, součástí HybSeq protokolu je i zpracování sekvenováním získaných dat. V principu jsou nejdříve odstraněny duplikáty, ořezány adaptory a vyfiltrovány kvalitní ready. Poté jsou z readů vytvořeny contigy, které jsou identifikovány pomocí mapování na pseudoreferenci. Pseudoreference jsou sestavené sekvence všech exonů, mezi kterými jsou vložena N (any), bývá jich kolem 200-400. Je to z důvodu, aby se jednotlivé mapované contigy nemohly

překrývat. Následně jsou namapované cotigy alignovány a zkonkatenovány tak, aby tvořily jednotlivé exony. Pak už lze konstruovat species tree hypotézy na základě rozličných přístupů viz dále (Cronn *et al.*, 2012; Weitemier *et al.*, 2014).

### 1.3.5.2 Chybějící data

V této diplomové práci jsou použity pro rekonstrukci fylogeneze proby primárně designované pro čeleď Zingiberaceae, což znamená, že po namapování sekvencí na pseudoreferenci, bude generovat velké množství chybějících dat, to nutně nemusí být na překážku, jak bylo prokázáno v minulých letech v pracích např. (Philippe *et al.*, 2004; Thomson & Shaffer, 2010). Pro rekonstrukci musí být splněny určité podmínky, a to že dat musí být relativní dostatek. Čím více vzorků, tím lépe, i když to znamená větší množství chybějících dat. V mnoha případech se jedná o více jak 90 % chybějících dat. Obecně přítomnost chybějících dat může být zapříčiněna několika faktory. Prvním je nedostatečná hloubka sekvenování při NGS experimentech, dále degradovaná vstupní DNA nebo RNA, zejména se to týká materiálu získaného ze starých herbářových položek. Také se to týká skupin s velkou substituční rychlostí, kde je přítomná velmi vysoká variabilita v sekvencích. Výsledkem je pak nefunkčnost univerzálních primerů či prob, zvláště pokud jsou uplatňovány na široké škále taxonů.

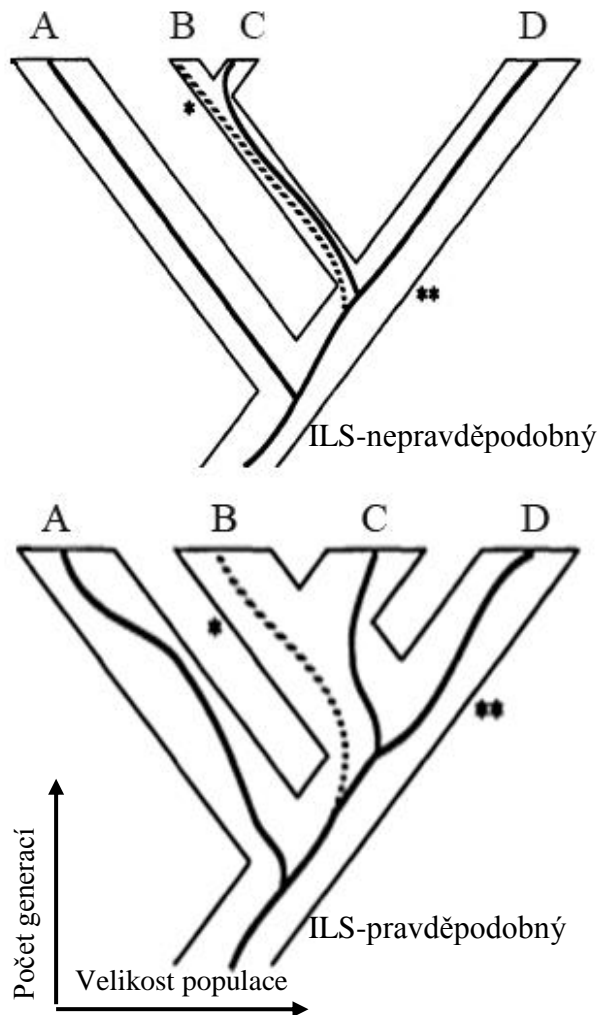
Chybějící data mohou být jak náhodně rozmístěná, tak mohou tvořit i určitá pattern, a to jak napříč geny, tak i napříč studovanými taxony (Xi *et al.*, 2015).

### 1.3.6 Gene trees versus Species trees

Jednou ze základních otázek při rekonstrukci fylogeneze je, zda genová genealogie neboli strom vytvořený na základě jednoho či více genů (gene trees) skutečně odpovídá fylogenezi druhů, které studujeme (species tree). Ukazuje se, že ne vždy jsou gene trees totožné se species trees, jelikož každý gene tree reflektuje unikátní historii, která je nepochybně spjatá s historií druhu, ale často se také v mnohém liší (Maddison, 1997; Szöllösi & Daubin, 2012). Odlišnosti mezi gene trees a species trees mohou být zapříčiněny mnoha rozličnými procesy. Patří sem horizontální genový transfer, mezidruhová hybridizace, neúplné sortování linií (ILS – incomplete lineage sorting), genová duplikace potažmo

polyploidizace či extinkce, rekombinace a selekce (Maddison, 1997; Degnan & Rosenberg, 2009).

Horizontální genový transfer je spíše problémem prací, zabývajících se bakteriemi či kvasinkami, u kterých je běžné vyměňování si genetické informace jiného jedince, ačkoli není jeho potomkem. Daleko větší problém u rostlin představuje mezidruhová hybridizace ať už



**Obrázek 6:** Species trees s naznačeným incomplete lineage sortingem (ILS) pro gene tree taxonu B, C a D. V případě (a) je ILS méně pravděpodobný, z důvodu nízké velikosti efektivní populace a přítomnosti dlouhých větví, (b) vyšší pravděpodobnost ILS z důvodu velké efektivní populace s přítomností krátkých větví. \* speciální událost pro taxona B a C, \*\* speciální událost, kdy se odělely taxon D a předek B a C. Převzato z: (Maddison, 1997).

recentní nebo historická. Výsledek je obdobný, kdy se můžou vzdálenější druhy jevit příbuznější právě v důsledku sdílení genetické informace po hybridizační události. Oproti tomu incomplete lineage sorting se týká všech skupin. Dochází k němu v případě, že je zachován ancestrální genový polymorfismus tzn., existuje více jak jedna alela na jeden lokus, které přetrvávají během druhové divergence. Výsledkem tohoto procesu mohou být mylné podobnosti mezi méně příbuznými taxony vedoucí k nesprávnému species tree viz Obr. č. 6. Dalším problémem je genová duplikace, kdy bezprostředně po duplikaci vznikají koexistující linie genů, to činí problém v případě, že při fylogenetické analýze nedojde k zachycení pouze ortologních genů. V případě, že není znám skutečný species tree a jsou pro rekonstrukci fylogeneze použity ortologní i paralogní geny, které prošly samostatnou evolucí (extinkce částí linií), může dojít k vytvoření nesprávné hypotézy o příbuznosti jednotlivých taxonů a řešitel na to nemusí vůbec přijít (Maddison, 1997; Degnan & Rosenberg, 2009).

Rozpory mezi jednotlivými gene trees obsahují informace o rysech species

tree, jako je jeho topologie, doba divergence a velikostech populace. Proto je třeba použít takových metod, které nepracují s předpokladem gene tree = species tree, jsou statisticky konzistentní a jejich algoritmus je přiměřeně výpočetně složitý. Řadí se sem: konsenzuální metody, konkatenace a koalescenční metody (Degnan & Rosenberg, 2009). Některé z nejpoužívanějších metod pro odhady species trees jsou níže stručně popsány.

#### **1.3.6.1 Konsenzuální metody**

Konsenzuální metody zkonstruují strom, který shrnuje topologii jednotlivých vstupních stromů. Na základě vstupních stromů je vytvořena matice, která je pak analyzována. Jednotlivé vstupní stromy nemusí vždy obsahovat stejný počet taxonů. V tomto případě se pak hovoří o tzv. supertree metodách. Příkladem je supertree (MRP) Matrix Representation with Parsimony, který s ILS nepočítá, navzdory tomu vykazuje celkem konzistentní výsledky za krátký výpočetní čas, dále pak (MRL) Matrix Representation with Likelihood, tato metoda oproti MRP zahrnuje maximum likelihood algoritmus, kdy dosahuje lepších výsledků než MRP (Bayzid & Warnow, 2012; Nguyen *et al.*, 2012).

#### **1.3.6.2 Konkatenace**

Další z přístupů je konkatenace, kdy jsou všechny získané geny (sekvence) pro každý taxon spojeny dohromady a následně vyhodnoceny. Tato metoda ignoruje odlišnosti v evoluční historii pro různé lokusy, které teoreticky vedou k nadhodnocení podpory pro nesprávné species tree (Degnan & Rosenberg, 2009). Konkrétním příkladem je RAxML (randomized accelerated maximum likelihood), kde je možné použít velký dataset za použití maximum likelihood (Stamatakis, 2014).

#### **1.3.6.3 Koalescenční metody**

Koalescenční metody se snaží ve svých species tree odhadech zohlednit přítomnost deep coalescent neboli incomplete lineage sortingu. Přitom je kladen důraz na jejich statistickou konzistentnost a schopnost zahrnout co největší možný počet taxonů a znaků do

analýzy. Jako vstupní data mohou sloužit buď přímo sekvence DNA, nebo jednotlivé gene trees. V případě použití koalescenčních metod je třeba dobře zvážit, zda je jejich použití vhodné, jelikož se ukazuje, že v případě velkého počtu chybějících dat s nízkou mírou ILS je vhodnější použít metodu konkatence (Xi *et al.*, 2015). Tyto metody je možné rozdělit na tzv. summary methods a co-estimation methods. Pro summary methods slouží jako vstupní data zkonstruované gene trees, které jsou danou metodou vyhodnocovány, nakonec je odhadnut species tree. Co-estimation methods používají jako vstupní data alignment, kdy jsou současně konstruovány jednotlivé gene trees a species tree.

### **1.3.6.3.1 Summary methods**

#### **1.3.6.3.1.1 ASTRAL, ASTRAL-II: Accurate Species TRee ALgorithm**

Tato metoda je statisticky konzistentní a je velmi rychlá. Dokáže zpracovat i velké množství taxonů, což v případě \*BEASTu není výpočetně možné. Jako vstupní data slouží nezakořeněné gene trees. Poté je v případě velkého objemu dat použito heuristické vyhledávání a jako species tree je vybrán ten, který souhlasí s největším počtem kvartetových stromů, které byly vytvořeny ze vstupních gene trees. ASTRAL-II je vylepšen o další parametry a zvládá tak ještě větší objem dat v kratším čase (Mirarab *et al.*, 2014; Mirarab & Warnow, 2015).

#### **1.3.6.3.1.2 ASTRID: Accurate Species TRees from Internode Distances**

Jedná se o metodu založenou na NJst vzdálenostech, ačkoli je výpočetně rychlejší a lze u ní použít chybějící data. Výpočet se skládá ze dvou kroků. V prvním je vypočtena matice vzdáleností a v druhém kroku proběhne velmi přesný odhad species tree. (Vachaspati & Warnow, 2015).

#### **1.3.6.3.1.3 MDC: Minimizing deep coalescence**

Tato metoda je založená na maximální parsimonii, kdy je heuristiky hledán species tree. Jsou porovnávány jednotlivé topologie gene trees a jako nejlepší species tree je vybrán ten, který

připouští nejmenší počet deep coalescent událostí a zároveň nejsou v rozporu s žádným gene tree (Huang *et al.*, 2010; Tonini *et al.*, 2015).

#### **1.3.6.3.1.4 MP-EST: Maximum Pseudo-likelihood approach for Estimating Species Trees under the coalescent model**

Zde je pro odhad species tree použita tzv. pseudo-likelihood funkce. Výsledkem této metody je topologie species tree s délkami větví. Tato metoda nepracuje s možným vlivem genového toku nebo horizontálního genového transferu, zato je velmi robustní čili přítomnost malé míry HGT nemusí nutně vadit. Tato metoda je vhodná pro velké objemy dat, avšak pouze se středním množstvím taxonů, navíc vyžaduje zakořeněné stromy, což může činit potíže v případě, že není znám skutečný outgroup (Liu *et al.*, 2010).

#### **1.3.6.3.1.5 NJst**

V případě NJst je species tree odhadován ze stromu vytvořeného pomocí neighbour joining (NJ) z matice vzdáleností, kde vzdálenost mezi dvěma druhy je definována jako průměrný počet internodií mezi dvěma druhy napříč gene trees = gene-tree internode distance (NJst). Tato metoda je statisticky konzistentní v případě, že jsou odhadovány nezakořeněné species trees. Tato metoda připouští chybějící data a díky jednoduchému výpočetnímu algoritmu je velmi rychlá, avšak nepřipouští chybějící taxony v jednotlivých gene trees (Liu & Yu, 2011).

#### **1.3.6.3.1.6 STAR: Species Tree estimation using Average Ranks of coalescences**

V této metodě jsou definované tzv. koalescenční ranky. Jedná se o deskriptivní statistiku koalescence prostřednictvím řazení očekávaných pozic koalescence mezi sekvencemi. Konkrétně je vždy na kořeni gene tree stejný počet koalescenčních ranků, jako počet použitých taxonů. Pozice se vždy sníží o 1 při pohybu směrem od kořene ke špičce stromu. Tato metoda je na čas velmi nenáročná, oproti odhadům Bayesovského modelování a lze tak do ní zahrnout velký objem dat (Liu *et al.*, 2009; Tonini *et al.*, 2015).

### 1.3.6.3.2 Co-estimation methods

#### 1.3.6.3.2.1 BEST Bayesian Estimation of Species Trees

Zde je implementován Bayesovský hierarchický model, kdy jsou společně odhadovány jednotlivé gene trees (samostatně) a species tree přímo ze sekvencí. Dále jsou odhadnuty časy divergence a velikosti populací. Tento model předpokládá, že rozpory mezi jednotlivými gene trees jsou zapříčiněny ILS za volné rekombinovatelnosti mezi geny, nikoli mezi nimi samotnými. Proto není tuto metodu vhodné používat v případě, že nesoulad mezi gene trees je jiného původu např. HGT. Tato metoda je výpočetně náročná a v případě použití většího množství taxonů to může představovat vážný problém (Liu, 2008; Tonini *et al.*, 2015).

#### 1.3.6.3.2.2 \*BEAST: Bayesian Evolutionary Analysis by Sampling Trees

Tato metoda je velmi podobná metodě BEST, používá též Bayesovský přístup za použití Markovových řetězců. Rozdílem oproti metodě BEST je, že jsou odhadovány gene trees a species tree společně najednou. Výsledkem jsou kvalitní gene trees spolu se species tree. Nicméně tento přístup je vysoce výpočetně náročný. Pro výpočet je limitující počet genů větší než 40 a počet taxonů větší než 50 (Heled & Drummond, 2010).

## 1.4 Costaceae Nakai

### 1.4.1 Charakteristika

Čeleď *Costaceae* náleží do řádu *Zingiberales* spolu s čeledí *Zingiberaceae*, která je k této čeledi sesterskou. V současné době zahrnuje sedm rodů *Monocostus* K.Schum., *Dimerocostus* Kuntze, *Costus* L., *Paracostus* C.D.Specht, *Chamaecostus* C.D.Specht & D.W.Stev, *Hellenia* Retz. a *Tapeinochilos* Miq. Celkový počet akceptovaných druhů je přibližně 140 s tím, že publikovaných jmen je kolem 300 (<http://www.theplantlist.org/1.1/browse/A/Costaceae/>). Jedná se o nearomatické vytrvalé rostliny tvořící oddenky, které jsou většinou silné a baňaté. Neotropické druhy jsou terestrické, na rozdíl od některých afrických zástupců, jež mohou růst epifyticky. Rostliny většinou dosahují do výšky 1 – 2 metrů až na výjimky, kdy mohou tvořit pouze listové růžice nebo přesahovat výšku 6 m. Stonky jsou specificky spirálně stočené zpravidla nevětvené. Na bázích jsou kryty listovými pochvami bez čepele. Listy vyrůstající výše od země a jsou spirálně uspořádaný podél stonku monostichiálně. Listová čepel bývá k listové pochvě připojena krátkým řapíkem. Listové pochvy jsou kruhového tvaru s vyčnívajícím jazýčkem. Květenství jsou většinou bohatá, kulovitého nebo válcovitého tvaru, směrem k vrcholu se zužující. Květenství může vyrůstat na vrcholcích stonku, popřípadě z paždí listu, nebo na samostatném neolistěném stonku. Neolistěný stonek může být dlouhý a vzpřímený, či velmi krátký popřípadě plazivý.



**Obrázek 7:** *Costaceae*. *Costus megalobracteata* K.Schum. (a) celkový vzhled rostliny, (b) Květenství, (c-f) *Costus lucanusianus*. (c) Květ, (d) Podélný řez semeníkem, (e) Petaloidní tyčinka s věnečkem chlupů, (f) Blizna, boční a přední pohled. (g, h) *Costus cuspidatus*. (g) Tobolka se semeny, (h) Semena k sobě připojena arilem. Převzato z: (Larsen, 1998)

V takovém případě leží květenství na povrchu nebo může být i částečně v zemi. Květenství se skládá z vytrvalých listenů, které na sobě mohou nést listovité přívěšky. U mnoha druhů jsou na listenech přítomna extraflorní nektaria, lákající mravence, kteří nejen rostlinu chrání, ale i šíří její semena (Schemske, 1980). V květenství bývá přítomen listenec, jenž může mít lodičkovitý tvar (boat-shaped), nebo kruhovitý s dvěma žebry (bicarinate). Květy jsou zygomorfni, rozdělené na kalich, květní trubku a korunu. Kalich je obvykle malý, laločnatý, srostlý ze tří většinou nestejně velkých kališních lístků. Koruna tvoří tři korunními lístky, které jsou na bázích srostlé. Zadní korunní list je výrazně větší než zbylé dva boční. Barva koruny varíruje od bílé přes žlutou a oranžovou po červenou. V květu je přítomna pouze jedna petaloidní tyčinka se dvěma prašnými pouzdry. Zbylé tyčinky jsou přeměněny na staminodia, vzájemným srústem tvoří labellum, které svým vzhledem připomíná korunu Obr. 7. Labellum nabývá dvou základních forem. V prvním případě je široce otevřené a jeho barva bývá ve světlých odstínech, tyto květy opyluje hmyz respektive včely, pro které široké labellum slouží jako přistávací plocha a směřuje opylovače ke zdroji nektaru. Druhá forma labella tvoří úzkou a krátkou trubku, která s pomocí kalicha a koruny tvoří pevný útvar. V tomto případě se jedná o adaptaci na opylování ptáky. U těchto druhů labellum a koruna mají červenou, oranžovou až tmavě žlutou barvu. Posun k opylování ptáky vznikl v této skupině několikrát nezávisle na sobě. Proto lze nalézt oba typy labella napříč celou čeledí (Specht *et al.*, 2012). Bliznu tvoří tři laloky, čnělka prorůstá mezi dvěma prašnými pouzdry a semeník je bi- nebo trilokulární. V semeníku jsou septální nektaria ústící do květní trubky. Plodem jsou nažky. Semena jsou hnědavé až černé barvy a mají bílý nebo žlutavý míšek (arilus). Hlavním zásobním pletivem semen je perisperm s bohatým obsahem škrobu, naopak endosperm je velmi slabě vyvinut bez přítomnosti škrobu.

Čeled *Costaceae* je charakteristická svou pantropickou distribucí. Největší druhová diverzita je koncentrována v Centrální a Jižní Americe. Většina druhů je součástí nížinného tropického deštného lesa, nicméně některé druhy se mohou nacházet v lesích do 2000 m n. m. Jedinci rostou buď v lesních světlínách, nebo naopak ve velmi vlhkých lokalitách, jako jsou břehy řek nebo močály.

Tato čeď se netěší přílišnému ekonomickému významu, na rozdíl od sesterské čeledi *Zingiberaceae*. Oddenky druhu *Hellenia speciosa* jsou v Asii lokálně využívány pro jejich zdroj škrobu a jako léčivá bylina (anthelmintikum). Trubkovité květy některých jihoamerických druhů se konzumují díky sladkému nektaru. Africký *Costus lucanusianus* se používá jako zelenina, léčivka i jako rostlina pro rituální účely. Nezanedbatelnou součástí

využití Costaceae je jejich dekorativní vzhled, díky němuž jsou v tropických oblastech hojně pěstovány jako okrasné rostliny (Larsen, 1998; Specht & Stevenson, 2006b)(Mass, 1972; Larsen, 1998; Specht and Stevenson, 2006).

*Costaceae*, dříve považované za podčeleď v rámci čeledi *Zingiberaceae*, byly rozeznány jako samostatná čeleď Nakaiem (1941), který definoval diagnostické znaky: Nepřítomnost aromatických látek ve vegetativních částech rostlin, spirálně stočená fylotaxe a přítomnost přívěsku na tyčince. Dalšími rozdílnými znaky jsou: Přítomnost nevětvených chlupů, dobře vyvinutá hypodermis, slabě vyvinuté průduchy přítomné pouze na svrchní straně listu a siliková tělíska hvězdicovitého tvaru (Tomlinson, 1956, 1962). Labellum je tvořeno pěti srostlými petaloidními tyčinkami na rozdíl od *Zingiberaceae*, kde je labellum srostlé pouze ze třech tyčinek (Kirchoff, 1988).

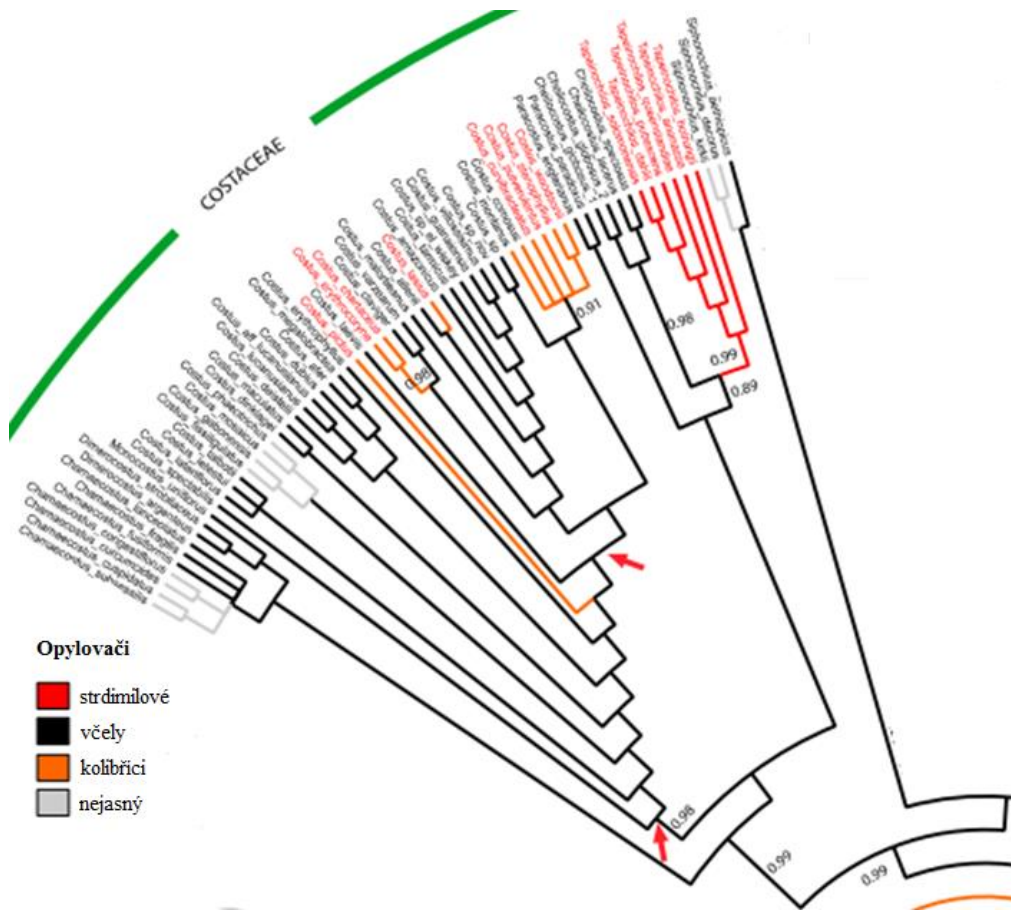


○ = Kalich   ● = Koruna   ● = Fertilní tyčinka   ● = Staminodium   \* = Neprítomná tyčinka

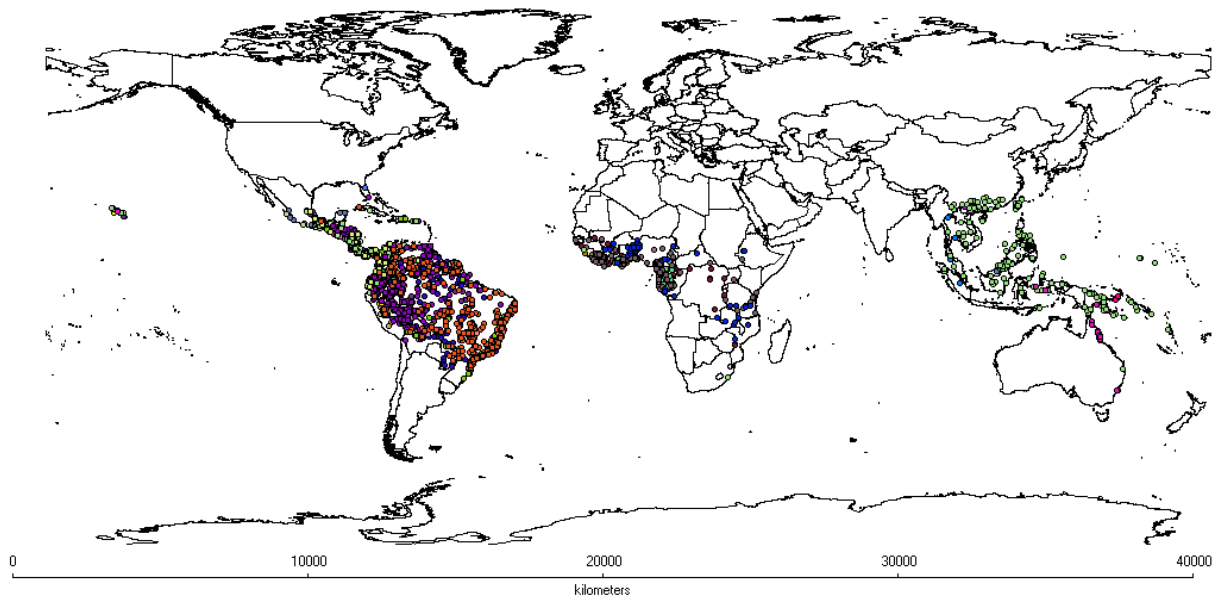
**Obrázek 8:** Květní diagram čeledi *Costaceae* (Kress, 1990).



**Obrázek 9:** *Costus villosissimus*, ukázka květní stavby čeledi *Costaceae*. V případě tyčinky je vidět zbytek květní trubky ve spodní části. Pestík je vidět se všemi částmi, blizna a čnělka ve vrchní části, semeník v části spodní, kde na něj nasedá kalich (vlastní foto).



**Obrázek 10:** Druhy opylovačů vstažené k fylogenetickým vztahům č. *Costaceae* (Specht *et al.*, 2012).



**Obrázek 11:** Rozšíření čeledi *Costaceae* podle databáze GBIF.

## 1.4.2 Rody

### 1.4.2.1 *Chamaecostus* C.D.Specht & D.W.Stev.

Rostliny náležící do rodu *Chamaecostus* jsou charakteristické svým nízkým vzrůstem, který nepřesahuje výšku 1 metru. Mívají velmi krátký stonek, popřípadě tvoří pouze listovou růžici. Listy jsou eliptické a na koncích zúžené. Tvoří hlávkovité květenství kryté listeny, jež mohou být plně bylinné, nebo připomínají pergamen, někdy mají listovité přívěšky. Barva listenů je zelená nebo žlutá. Každý květ kryje jeden listenec, v případě rodu



*Chamaecostus* je membranózní se dvěma žebry na adaxiální straně. Kalich je blanitý válcovitého tvaru. Koruna vyčnívá z kalichu a je tvořena velkými kopinatými laloky. Labellum je velké, široce otevřené. Kalich i koruna mívají stejnou barvu a to žlutou, oranžovou nebo bílou. Tyčinka je petaloidní.

**Obrázek 12:** *Chamaecostus cupsidatus* (<http://www.infojardin.com/foro/showthread.php?t=71079&page=19>).

Blizna je pohárovitá, semeník trilokulární. Plodem jsou pukavé dužnaté tobolky (Mass 1972, 1977; Specht and Stevenson 2006).

V současnosti je do tohoto rodu řazeno 8 druhů, které jsou svým výskytem omezeny na Jižní Ameriku (od Guianského štítu po amazonské nížiny v Bolívii a Brazílii). Druhy tohoto rodu mají tendenci růst v početných populacích na malém území, ale na celkovou rozlohu areálu jsou zastoupeny vzácně. (Specht and Stevenson 2006; André et. al. 2015)



**Obrázek 13:** *Chamaecostus cupsidatus*, listeny s vyvinutými listovými přívěšky (<http://www.gingersrus.com/images/ImageFrame.php?ImageID=13165>).

### 1.4.2.2 *Costus* L.

Zástupci rodu *Costus* L. jsou středně velké vytrvalé byliny dosahující výšky 2 – 5 m. Stonky jsou nevětvené na bázi kryté listovými pochvami. Listy jsou na lodyze spirálně uspořádané. Jazyček bývá velký a nápadný. Květenství vyrůstá buď terminálně na olistěné lodyze, nebo vyrůstá samostatně na neolistěném stonku přímo z oddenku. Listeny vyrůstají kónicky a navzájem se překrývají. Jsou kožovité, úzce či široce vejčité a mohou mít zelenou, žlutou, oranžovou nebo červenou barvu. Často bývají na koncích vyvinuty listové přívěsky. Listenec je buď trojhranný, nebo trubkovitý. Kalich je tvořen třemi laloky, zřídka svou délkou přesahující listeny. Koruna je bílá, žlutá, oranžová nebo červená. Labellum je malé nebo velké, bílé, žluté, oranžové či červené. Tyčinka je petaloidní s prašníkem umístěným ve prostřed. Blizna je pohárovitá nebo bilamelární. Semeník je trilokulární. Plodem je pukavá nebo nepukavá tobolka. Semena jsou černá nebo hnědá a opatřena bohatým míškem (Mass, 1972; Larsen, 1998; Specht & Stevenson, 2006b)

Tento rod je největším z celé čeledi a zahrnuje přibližně 90 druhů. Druhy rodu *Costus* L. se vyskytují v tropických deštných lesech Afriky a Jižní Ameriky s centrem diverzity v neotropis (Larsen, 1998; Specht & Stevenson, 2006b).

### 1.4.2.3 *Dimerocostus* Kuntze

Tento rod je typický svým vysokým vzrůstem přesahujícím do výšky 3 m. Masivní stonky se nevětví a rostou jen mírně spirálně. Listy vyrůstají až na koncích stonku. Samotné listy jsou protáhlé a na konci se zužují. Jazyček je velmi krátký nebo chybí. Květenství vyrůstá na vrcholcích stonků, je podlouhlé ve tvaru šišťice. Listeny jsou oválného až trojhranného tvaru, na jejichž konci jsou někdy vyvinuty listovité přívěsky. Barva je v odstínech zelené a žluté. Listenec je trubkovitý bikarinátní. Kalich je velký tvořený třemi nestejně velkými laloky obvykle přerůstající listeny. Koruna je zbarvena bíle nebo žlutě.



**Obrázek 14:** *Costus lasius* (Lyon Arboretum).

Labellum je nápadně velké a má buď žlutou, nebo bílou barvu. V případě, že má labellum bílou barvu, bývá veprostřed žlutá skvrna. Tyčinka je petaloidní na vrcholku zahnutá. Prašníky jsou vprostřed tyčinky. Pyl je velký dikolpátní nebo kolpo-porátní. Blizna je pohárovitá a semení bilokulární. Plodem je tobolka, někdy pukavá. Semena jsou leskle černá opatřena míškem (Mass 1972, Larsen 1998, Specht and Stevenson 2006).

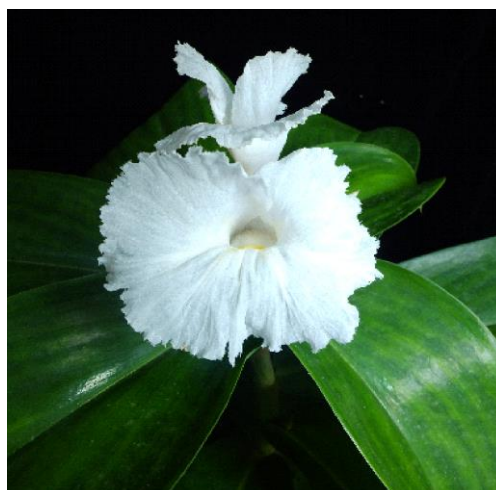


**Obrázek 15:** *Dimerocostus strobilaceus* (Lyon Arboretum).

Rod *Dimerocostus* se vyskytuje na území Střední a Jižní Ameriky od Nikaragui po severní Bolívii (Specht and Stevenson, 2006). Distribuce většinou kopíruje východní svah And. V současnosti jsou akceptovány tři druhy (*Dimerocostus argenteus*, *Dimerocostus cryptocalyx*, *Dimerocostus strobilaceus*).

#### 1.4.2.4 *Hellenia* Retz.

Zástupci rodu *Hellenia* dosahují výšky více než 1,5 m. Stonky jsou spirálně stočené a v terminálních částech se bohatě větví. Listy jsou k listové pochvě připojeny krátkým řapíkem a u některých druhů jsou sezóně opadavé. Květenství vyrůstá přímo na olistěné lodyze nebo na samostatném neolistěném stonku případně vyrůstá přímo z oddenku. Listeny jsou tuhé až dřevnaté, zelené, červené nebo hnědé bez listových přívěšků.



Listenec je bikarinátní. Kalich je tvořen třemi stejně velkými laloky. Květní trubka vyčnívá z kalicha.

**Obrázek 16:** *Hellenia speciosa* (vlastní foto).

Samotná koruna je tvořena membranózními laloky. Labellum je velké, široce otevřené s dlouhou trubkou. Může mít bílou, žlutou, oranžovou barvu. Tyčinka je trojhranná, petaloidní, kryjící vchod do květní trubky. Blizna je pohárovitá s přívěsky. Semeník je trilokulární s nektarií na vrcholu. Plodem je po stranách pukající tobolka. Semena jsou malá, černá opatřena bohatým arilem, který pojí jednotlivá semena k sobě, viz Obr. č. 7.

Tento rod zahrnuje čtyři druhy s omezeným výskytem v Jihovýchodní Asii, Malajsii a Nové Guinei (Specht & Stevenson, 2006b)

Současný název rodu (*Hellenia* Retz.) byl navržen jako nejstarší správné jméno pro tento rod (Dutta; Govaerts, 2013). Vzhledem k tomu, že toto jméno je potenciálně problematické (*Hellenia* Willd. byla, i když illegitimně, použito v příbuzné čeledi *Zingiberaceae* a je v potenciálním rozporu také s *Hellenia* Mill./*Hellenium* L. v čeledi *Asteraceae*), připravuje se návrh na konzervaci jména *Cheilocostus* (Leong-Škornickova & Šída, ústní sdělení).

#### 1.4.2.5 *Monocostus* K.Schum.

*Monocostus* je monotypickým rodem zahrnujícím druh *Monocostus uniflorus*. Jedná se o byliny velmi malého vzrůstu netvořící oddenky. Výška se pohybuje od 20-60 cm. Stonky se nevětví a jsou na bázi pochvaté. Listy jsou malé, po okrajích načervenalé. Jazyček je drobný nebo chybí. Květy vyrůstají jednotlivě v horní části stonku z paždí listu na krátké stopce. Listenec je dlouhý trubkovitého tvaru. Koruna je kratší než labellum a má světle žlutou barvu. Labellum je velké oválného tvaru žlutě zbarvené. Tyčinka je petaloidní na vrcholu ohnutá. Prašníky vyrůstají uprostřed délky tyčinky. Blizna je pohárovitá a semeník bilokulární. Plodem je po stranách pukavá tobolka. Semena jsou lesklá, černé barvy s přítomným slabě vyvinutým arilem.



**Obrázek 17:** *Monocostus uniflorus*, převzato z: Gingers are us.

Tento rod se vyskytuje pouze ve východním Peru v Jižní Americe. Je známo pouze málo lokalit výskytu (Mass 1972, Specht and Stevenson 2006).

#### 1.4.2.6 *Paracostus* C.D.Specht

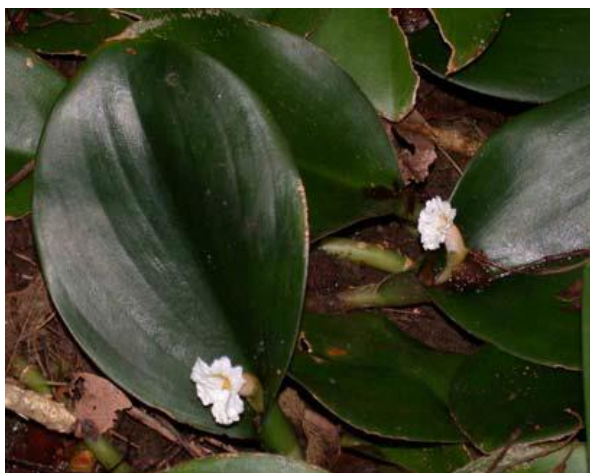
Rod *Paracostus* je formálně zastoupen dvěma druhy (*Paracostus englerianus*, *Paracostus paradoxus*). Jedná se o poléhavé oddenkaté byliny, které mohou tvořit vystoupavé popínavé lodyhy. Dorůstají do výšky mezi 10 - 50 cm. Na lodyze je přítomno pouze pár listů, které mohou sloužit jako opora pro axilární květenství.

Rostliny tvoří dlouhé plazivé oddenky. Na jedné lodyze jsou přítomny 1 - 4 lesklé listy s krátkými řapíky, jimiž jsou připojeny k listové pochvě, která vytváří pouze krátký jazýček. Rostliny jsou celé lysé. Květenstvím je zkrácený klas vyrůstající z úžlabí listu. Listeny jsou membranózní zelené barvy. Listence jsou trubkovité připomínající pergamen. Kalich je tvořen třemi laloky stejné velikosti. Korunní trubka vyčnívá z kalichu, tvoří blanité laloky úzce kopinatého tvaru. Labellum je velké, úzké a obvejčité. Na bázi je tvořeno úzkou a dlouhou trubkou, na vrcholu široce otevřenou. Barva labella je bílá, někdy se žlutou skvrnou uprostřed, naproti fertilní tyčince. Tyčinka je petaloidní a má trojúhelníkovitý tvar, jež kryje vstup do květní trubky. Tyčinka je na svém konci zahnutá směrem nahoru. Semeník je trilokulární. Blizna je pohárovitého tvaru bez jakýchkoli přívěšků. Plodem je tobolka. Semena jsou opatřena míškem (Specht & Stevenson, 2006).



**Obrázek 18:** *Paracostus paradoxus* (Lyon Arboretum).

Distribuce tohoto rodu je omezena na Afriku a ostrov Borneo v Jihovýchodní Asii.



Kdy v Africe se vyskytuje druh *Paracostus englerianus* a na Borneu *Paracostus paradoxus*. V současné době byly popsány další tři druhy z Bornea (*Costus bullatus*, *Costus eburneus* a *Costus muluensis*), ale jejich autoři neuznávají novou klasifikaci rodu *Paracostus*, a proto jsou prozatím formálně evidovány pod rodem *Costus* (<http://www.gingersrus.com/Paracostus.php>).

**Obrázek 19:** *Paracostus englerianus* (Lyon Arboretum).

#### 1.4.2.7 *Tapeinochilos* Miq.

Jedná se o byliny dosahující středního a vysokého vzrůstu se vzpřímenými nebo pouze mírně spirálně stočenými stonky. Rostliny se na vrcholcích bohatě větví a jsou hustě olistěné. Listy mají velmi krátký jazýček a listová čepel je eliptického nebo obvejčitého tvaru.

Květenství má šišticevitý tvar a vyrůstá buď přímo na olistěné lodyze, nebo samostatně na neolistěném stonku. V květenství jsou nejvýraznější složkou listeny, které mají kožovitou až dřevnatou strukturu a bývají zbarveny v odstínech červené. Listence většinou chybí. Květy svou velikostí stěží přesahují délku listenů. Mají žlutavou až hnědavou barvu. Kalich bývá kožovitý tvořící tři navzájem srostlé nestejně velké laloky. Koruna je na bázi srostlá a tvoří tři krátké laloky, kdy zadní je širší a delší než zbylé dva vyrůstající na bocích. Labellum je nenápadné a po obvodu zubaté. Tyčinka je petaloidní. Blizna je trojhraně zploštělá, semeník je bi- či trilokulární.



**Obrázek 20:** *Tapeinochilos* (Lyon Arboretum).

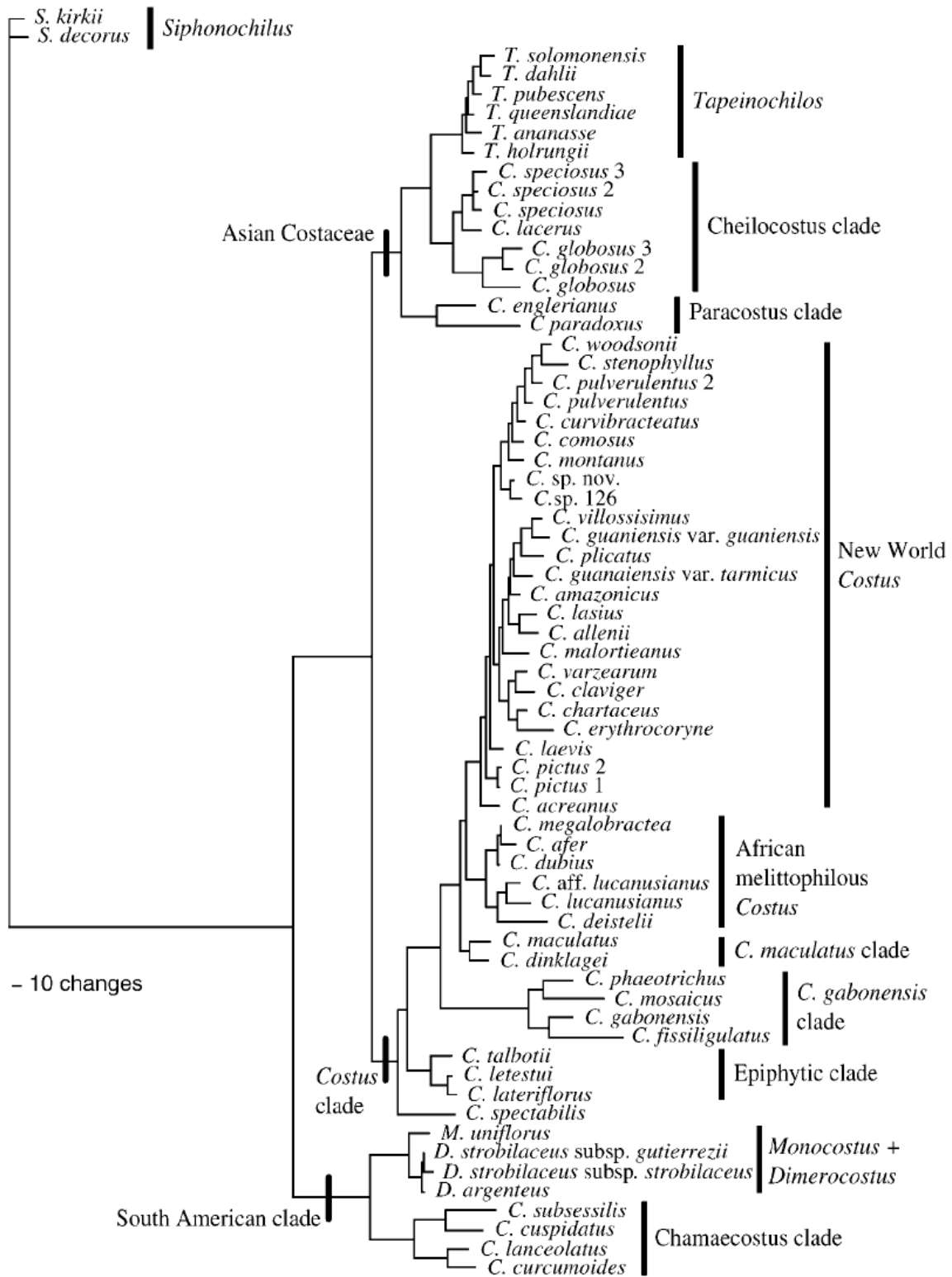
Rod *Tapeinochilos* vždy tvořil separátní rod díky své jedinečné morfologii a byl řazen do blízké příbuznosti s jihoamerickými druhy opylovanými kolibříky, nicméně fylogenetické analýzy prokázaly, že se jednalo o mylnou teorii a že je tento rod příbuzný asijskému rodu *Hellenia* (Specht & Stevenson, 2006).

### 1.4.3 Fylogeneze čeledi *Costaceae*

V letech 2001 a 2006 byly publikované práce Chelsea Specht, které se zabývaly fylogenezí čeledi *Costaceae*, vztahy byly postulovány na základě molekulárních a morfologických dat. Jako molekulární data posloužily dva chloroplastové lokusy *trnL-F*, *trnK* a jeden jaderný ITS. Pro morfometrickou analýzu bylo zahrnuto 73 znaků.

Z těchto dat byla vytvořena souhrnná matice o 67 taxonech a 5827 znacích. V programu PAUP\*4.0b10 byla vytvořena analýza maximální parsimone s binárně okódovanými gapy, dále byly provedeny ILD testy pro zjištění zda lze kombinovat data z jednotlivých úseků dohromady (Specht, 2006a). Pro měření podpory topologie stromu byl použit jackknife (jk). Výsledkem této analýzy bylo 12 nejparsimoničtějších stromů, z čehož jeden z nich je na Obr. č. 21. Výsledek fylogenetické analýzy ukázal, že dříve popsány rody *Tapeinochilos*, *Monocostus* a *Dimerocostus* jsou monofyletické a mohou být dále uznávány jako rody. Rod *Costus* se ukázal být parafyletickým, na stromě Obr. č. 21 jsou popsány jednotlivé clady (*Cheilocostus* clade, *Paracostus* clade, *Costus* clade, *Chamaecostus* clade).

Na základě výsledů této analýzy byly tyto clady popsány jako samostatné rody v práci (Specht & Stevenson, 2006).



**Obrázek 21:** Fylogram jednoho ze 12 nejparsimoničtějších stromů (Specht, 2006a).

#### 1.4.4 Chromozomové počty

Doposud publikované počty chromozómů jsou známy pro 59 druhů – čeledi *Costaceae* (Tab. č. 2),  $2n=18, 27, 28, 36, 72$ . To odpovídá minimálně čtyřem ploidiím s tím, že u většiny zástupců bylo spočítáno  $2n=18$ . Největší škála spočítaných chromozómů byla pro druh *Hellenia speciosa* ( $\equiv$  *Costus speciosus*, *Cheilocostus speciosus*). Základním chromozomovým číslem je  $x=9$ , v případě rodů *Monocostus* a *Dimerocostus* pak  $x=7$  nebo  $x=14$ . Pro rod *Costus* bylo nalezeno 48 záznamů s počty  $2n=18, 27, 28, 36$ . Pro rod *Paracostu* 1 druh  $2n=36$ , *Hellenia* 2 druhy s počty  $2n=13, 18, 27, 34, 36, 72$ , *Tapeinochilos* Miq. 1 druh  $2n=18$ , *Monocostus* K.Schum. 1 záznam  $2n=18$ , *Dimerocostus* Kuntze 2 druhy  $2n=28$  a pro rod *Chamaecostus* C.D.Specht & D.W.Stev. 2 druhy s  $2n=18$  (Mahanty, 1970; Mass, 1972, 1977; Gill, 1978; Subrahmanyam & Khoshoo, 1986; dos Santos Guerra, 1988; Vovides & Lascurain, 1995; Edeoga & Okoli, 2000).

Tyto získané údaje o počtech chromozómů jsou brána jako orientační s ohledem k faktu, že žádná z prací neposkytuje uspokojivou dokumentaci spočítaných chromozómů. V některých pracích není uveden žádný doklad, ve zbytku je doklad publikován ve formě namalovaného obrázku (Mahanty, 1970; Mass, 1972, 1977; Gill, 1978; Subrahmanyam & Khoshoo, 1986; dos Santos Guerra, 1988; Vovides & Lascurain, 1995; Edeoga & Okoli, 2000).

**Tabulka 2:** Soupis z literatury dohledaných chromozomových počtů seřazený podle příslušnosti k rodům. Akceptovaná jména (podle World checklist of selected plant families) jsou vtištěna silně, synonyma pod kterými byl počet uveden je v závorce). Tabulka byla sestavena na základě těchto prací: (Mahanty, 1970; Mass, 1972, 1977; Gill, 1978; Subrahmanyam & Khoshoo, 1986; dos Santos Guerra, 1988; Vovides & Lascurain, 1995; Edeoga & Okoli, 2000).

Druh	Počet chromozómů	
	2n	n
<b>Costus</b>		
Costus afer	18, 36	
Costus albus (Costus dubius)	36	
Costus bicolor	18	
Costus deistelii	18	
Costus dinklagei	18	
Costus dubius	36	
Costus englerianus (Paracostus englerianus)	36	

Costus giganteus	18	
Costus lucanusianus	18, 27, 36	
Costus megalobracteata	36	
Costus schlechteri	27	
Costus spectabilis	18, 27	
Costus talbotii	18	
Costus zechii (Costus dubius)	18	
Costus dhaninivatii	18	
Costus rumphianus	18	
<b>Hellenia</b>		
<i>Hellenia speciosa</i> (Costus speciosus, spec)	13, 18, 27, 34, 36,	
<i>Costus tonkinensis</i> [member of <i>Hellenia</i> , combination not yet	18	
<b>Tapeinochilos</b>		
Tapeinochilus ananassae	18	
Costus allenii	18	
Costus amazonicus	18	
Costus arabicus	18	
Costus brasiliensis (Costus arabicus)	18	
Costus claviger	18	
Costus comosus	18	
<b>Chamaecostus</b>		
Costus congestiflorus (Chamaecostus congestiflorus)	18	
Costus curvibracteatus	18	
Costus cylindricus (Costus spicatus)	18	
Costus dirzoi	27, 28	
Costus discolor (Costus arabicus)	18	
Costus elegans (Costus malortieanus)	18, 27	
Costus erythrocorone	18	
Costus fredrichsenii (Costus villosissimus)	18	
Costus glabratus (Costus arabicus)	36	
Costus glaucus	18	
Costus guanaiensis	18	
Costus igneus (Chamaecostus cuspidatus)	18	
Costus lacerus (Cheilocostus lacerus )	18	
Costus lima	18	
Costus macrostrobilus (Costus guanaiensis var. macrostrobilus)	18	
Costus malortieanus	18, 27, 36	
Costus niveopurpureus (Costus arabicus)	18	
Costus niveus (Costus arabicus)	18	
Costus pictus	18, 36	

Costus pulverulentus	18	
Costus scaber	18	
Costus spiralis	18	19
Costus stenophyllus	18	
Costus varzeorum	18	
Costus villosissimus	18	
Costus villosus (Costus spiralis var. villosus)	27	
Costus vinosus	18	
<b>Dimerocostus</b>		
<i>Dimerocostus aregenteus</i>	28	
<i>Dimerocostus strobilaceus</i>	28	
<b>Monocostus</b>		
Monocostus uniflorus	28	
Costus acreanus	18	
C. sp (asi cylindricus); (Costus spicatus)	16	

## 2. MATERIÁL A METODY

### 2.1 Materiál

Rostlinný materiál pochází z živých sbírek Botanické zahrady v Singapuru, Královské Botanické zahrady v Edinburghu, ze skleníku National Museum of Natural History, Botanické zahrady v Innsbrucku a sbírky Johna Mooda na Havaji. Rostlinný materiál byl dovezen v podobě samostatných listů uzavřených v zip-lockovém sáčku. Část vzorků pochází z přirozených lokalit přenesených do kultury a část vzorků pochází z kultury s neznámým původem. Do analýz byly preferovány vzorky se známou lokalitou a důležité druhy, které nejlépe pokrývají diverzitu čeledi *Costaceae*. Živý materiál byl zpracován do čtrnácti dnů od jeho pořízení.

### 2.2 Metody

#### 2.2.1 Průtoková cytometrie

Pro měření absolutní velikosti genomu byl použit průtokový cytometr Partec CyFlow® za použití postupu podle Doležela *et al.* (2007). Pro izolaci jader bylo použito rostlinné pletivo z listů. Ustřížený rostlinný materiál o přibližné velikosti 1 cm<sup>2</sup> byl vložen do plastové Petriho misky společně se standardem (*Bellis perennis* nebo *Solanum pseudocapsicum*). K materiálu bylo přidáno 500 µl vychlazeného pufru Otto I (0,1 M monohydrát kyseliny citrónové, 0,5% Tween 20; Otto 1990), poté byla směs důkladně rozdrovena žiletkou a pomocí pipety přefiltrovaná přes nylonový filtr o průměru oka 42 µm do zkumavky. Suspenze se nechala odstát v pokojové teplotě po dobu přibližně 15 minut.

Před samotným měřením byla suspenze barvena 15 minut barvicím roztokem složeným z 1 ml propidium jodidu (Sigma), 1 ml RNázy IIA (Sigma), 40 µl β-mercaptoethanolu (Sigma) a pufru Otto II (0,4 M Na<sub>2</sub>HPO<sub>4</sub>·12H<sub>2</sub>O; Otto, 1990).

Měření probíhalo na přístroji Partec CyFlow® ML se zeleným solid state laserem 532 nm a přidruženým softwarem Partec FloMax®. Z každého vzorku bylo změřeno 3500 jader s tím, že všechna měření byla prováděna pro každý měřený druh minimálně třikrát, pokaždé v jiný den. V případě, že se 2C-hodnoty nelišily o více než 2 %, bylo měření ukončeno. V případě, že se hodnoty odchylovaly o více než 2 % a rostlinný materiál byl ještě v

použitelném stavu, měřilo se počtvrté popř. popáté. Vzdálené hodnoty nebyly použity pro výpočet průměru z naměřených hodnot.

Jako ideální se ukázalo používat větší množství vzorku než standardu a celkově používat větší kusy listů a méně je sekát, dále pak vzorky nenechávat příliš dlouho odstát. V krajních případech docházelo ve zkumavce k tvorbě hustého slizu, který bylo nutné znovu přefiltrovat.

Někdy bylo třeba cytometr promýt vodou během měření, jelikož suspenze byla velmi hustá a docházelo k zanášení přístroje. Při měření byla nejčastěji použita rychlost 0,1–0,2  $\mu\text{l/s}$ , což odpovídalo měření do 20-ti jader za sekundu. Průměrná délka trvání jedné analýzy byla 15–20 minut. Jako primární standard bylo používáno *Solanum pseudocapsicum*. V případě, že nebylo možné odečíst hodnoty vzorku, byla použita *Bellis perennis*.

Hodnoty CV v cca 70 % nepřekročily 2, ve zbylých případech nepřekročily 4,5. Celkem bylo změřeno 146 vzorků. Měření bylo prováděno v Národním přírodovědném Muzeu v Horních Počernicích.

## **2.2.2 Molekulární analýzy**

Veškeré laboratorní práce související s prací s DNA byly provedeny na Přírodovědecké fakultě Univerzity Karlovy v Praze v DNA laboratoři (<http://botany.natur.cuni.cz/dna/>).

### **2.2.2.1 Izolace DNA**

DNA byla izolována ze sušeného rostlinného materiálu uchovávaného v silikagelu. Pro samotnou izolaci byl použit Invisorb Spin Plant Mini Kit (Invitex). Byl následován příložený protokol s několika úpravami týkajícími se přidaného množství pufru a doby eluce, viz podrobnější popis níže.

Do 2ml zkumavek byla přidána část listu o velikosti cca 1x3  $\text{cm}^2$ , v případě, že byl list pokryt velkým množstvím chlupů, bylo přidáno rostlinného materiálu více. Dále byly do zkumavek přidány 2 wolframkarbidové kuličky. Poté byly zkumavky umístěny do mlýnku Retsch Mixer Mill 200, kde bylo rostlinné pletivo drceno 3 minuty při 30ot./s.

K rozdrobenému materiálu bylo přidáno 400  $\mu$ l Lysis Bufferu P a 20  $\mu$ l Proteinase K, směs byla inkubována v Thermomixeru comfort (Eppendorf) po dobu 30 min při 1400 rpm. Poté byla směs centrifugována v centrifuze 5415 R (Eppendorf) 30s při otáčkách 13200 rpm. Vzniklý supernatant byl přepipetován do Prefilter kolonky a centrifugován 1 minutu při 12000 rpm. K filtrátu bylo přidáno 5  $\mu$  RNase A. Po 5 minutách bylo k odstátému filtrátu přidáno 200  $\mu$ l Binding Bufferu P a celá směs byla přepipetována do Receiver Tube s připravenou Spin Filter kolonkou. Směs byla centrifugována 1 minutu při 12000 rpm. Kolonky byly přemístěny do nových zkumavek, kde byl produkt promyt nejprve Wash Bufferem I poté Wash Bufferem II v objemu 550  $\mu$ l.

Kolonky byly přemístěny do nových zkumavek, kde bylo přidáno 70  $\mu$ l Elution Bufferu F a směs se nechala odstát 30 minut, poté byly vzorky centrifugovány 1 minutu při 10000 rpm. Následně byly kolonky přemístěny do nových zkumavek, do kterých bylo přidáno 50  $\mu$ l Elution Bufferu F, necháno odstát 45 minut, poté následovala centrifugace 1 minutu při 10000 rpm. Poté byla orientačně změřena koncentrace DNA pro ověření, zda izolace proběhla úspěšně. Izolovaná DNA byla označena a uchovávána v lednici při 5 °C.

## **2.2.2.2 Sekvenování úseku ITS a *trnL-F* a jejich zpracování**

### **2.2.2.2.1 PCR**

Pro namnožení požadovaného úseku DNA byla použita PCR (Polymerase Chain Reaction). Celkově byly amplifikovány 2 úseky. První byl použit jaderný úsek ITS (Internal transcribed spacer) a jako druhý chloroplastový úsek *trnL-F*, který zahrnuje *trnL* intron i exon a *trnL-F* intergenic spacer.

Tyto dva úseky byly zvoleny na základě publikované práce (Specht *et al.*, 2001), kdy byly použity totožné primery a dodržena teplota annealingu. V případě ITS byly použity dvě sady primerů. Jednak totožné z práce (Specht *et al.*, 2001) a zadruhé primery ITS4, ITS5 (White *et al.*, 1990). Pro úsek *trnL-F* byly použity totožné primery z práce (Specht *et al.*, 2001).

Pro oba úseky byla napipetována reakční směs, viz Tab. č. 3 o objemu 19  $\mu$ l, k ní byl přidán 1  $\mu$ l DNA vzorku z 2. eluce. Poté byl spuštěn reakční cyklus v termocykleru Mastercycler Pro (Eppendorf), popis cyklu je uveden v Tab. č. 4. Po doběhnutí cyklu byla testována úspěšnost PCR reakce pomocí horizontální agarosové elektroforézy. Byl připraven

0,8% gel smícháním příslušného množství Agarosy I s 1xTAE pufrem a 1 kapkou ethidiumbromidu. Na připravený gel v elektroforetické vaně s 1xTAE pufrem bylo nanášeno 1,5 µl žebříčku O'GeneRuler™ 100bp Ladder Plus (Fermentas) a 2-3 µl obarveného vzorku. V případě použití MyTaq polymerázy se vzorky nebarvily, v případě Immolasy byly vzorky smíchány s 1 µl barvicí směsí DNA Gel Loading Dye. Po skončení elektroforézy byl pro vizualizaci úseku DNA použit přístroj Gel Logic 100 (Kodak), kdy pomocí UV-transiluminátoru a přidruženého programového vybavení 1D Image Analysis Software (Kodak) byl výsledek vizualizován. Příklad získaných obrázků gelů, viz Obr. č. 22.

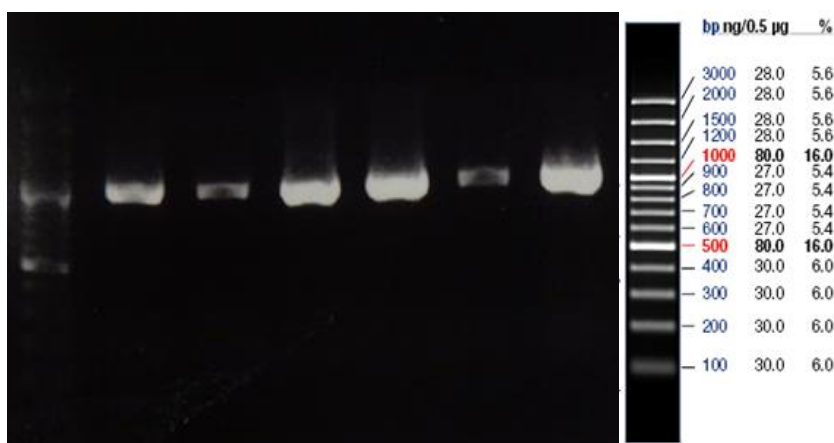
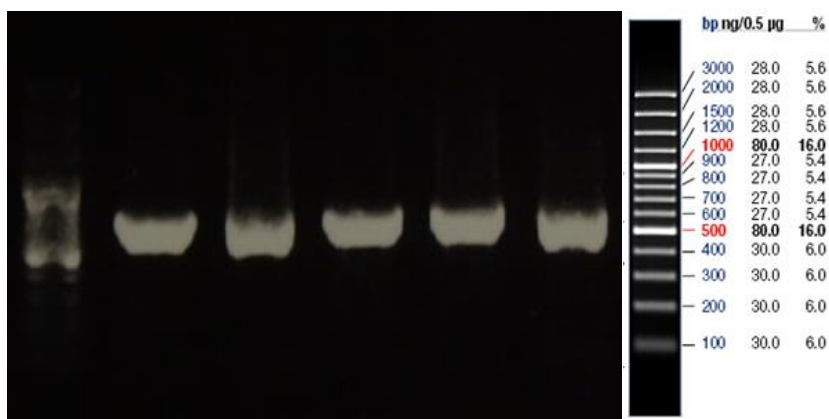
**Tabulka 3:** Složení směsí pro PCR reakce.

Reakce ITS	µl	Reakce <i>trnL-F</i>	µl
Sterilní Mili-Q H <sub>2</sub> O	13,8	Sterilní Mili-Q H <sub>2</sub> O	14,7
MyTaq pufr	4	IMMO pufr	2
Primer forward (ITS 5/ITS leu)	0,5	MgCl <sub>2</sub>	1
Primer reverse (ITS 4)	0,5	dNTP	0,5
MyTaq HS polymeráza (Bioline)	0,2	Primer forward <i>trnLc</i>	0,3
		Primer reverse <i>trnFf</i>	0,3
		Immolase polymeráza (Bioline)	0,2

**Tabulka 4:** Teplotní

Cyklus	ITS	Cyklus	<i>trnL-F</i>	
		95°C	7 min	
94°C	5 min	95°C	1 min	} 35x
94°C	30 sec	56°C	45 sec	
55°C	30 sec	68°C	2 min	
72°C	45 sec	72°C	7 min	
72°C	7 min	10°C	∞	
10°C	∞			

cykly použité při PCR reakcích.



**Obrázek 22:** Ukázka 0,8% gelu s namoženými úseky, (a) ITS a (b) *trnL-F* s po pravé straně připojeným obrázkem žebříčku O'GeneRuler™ 100bp Ladder Plus (<http://eshop.biogen.cz/generuler-100-bp-plus-dna-ladder#lightbox>).

#### 2.2.2.2.2 Přečištění

Pro odstranění nežádoucích nevyužitých prvků po PCR reakci byly PCR produkty přečištěny pomocí Gel/PCR DNA Fragments Extraction Kitu s úpravou, že místo Elution Bufferu byla použita sterilní Mili-Q H<sub>2</sub>O a doba eluce byla prodloužena.

K PCR produktu byl přidán DF Beffer v pětinasobku množství reakční směsi. Poté byla směs přenesena do kolonek a následně centrifugována 30 sekund při 13400 rpm. Filtrát byl odstraněn a PCR produkt byl promyt skrze kolonku 600 µl Wash Bufferu. Produkt byl eluován 15 min po přidání 20 µl sterilní Mili-Q H<sub>2</sub>O do kolonek. Takto přečištěný PCR produkt byl použit k přípravě vzorků na sekvenační reakci.

#### 2.2.2.2.3 Sekvenování DNA

Všechny úseky byly sekvenovány pouze z jedné strany s ohledem na délku úseku, která je v případě ITS přibližně 700 bp a u *trnL-F* 1000 bp dlouhá. Tedy je možno získat skoro celou sekvenci již po sekvenování z jedné strany.

Sekvenační reakce byla připravena do celkového objemu 8  $\mu$ l. Do zkumavky byl napipetován 1  $\mu$ l příslušného primeru a zbytek byl doplněn Sterilní Mili-Q vodou s amplifikovanou DNA v příslušné koncentraci, která se řídila podle rozpisu na 1000 bp koncentrace 100 ng/ $\mu$ l. Primer byl použit buď forward, nebo reverse pro příslušný úsek. Sekvenování probíhalo na šestnáctikapilárovém sekvenátoru 3130xl Genetic Analyzer (Applied Biosystems) v sekvenační laboratoři na Přírodovědecké fakultě Univerzity Karlovy v Praze, Viničná 7.

#### **2.2.2.2.4 Práce se sekvencemi**

Sekvence získané ve formátu \*.ab1 byly nejprve zkontrolovány a primárně editovány v programu FinchTV 1.4.0 (Geospiza, Inc.; Seattle, WA, USA; <http://www.geospiza.com>). Kontrolou je míněno zhlédnutí parametrů sekvence, konkrétně čistoty píků a relativní fluorescence. Úpravy sekvencí spočívaly v nahrazení bází, u kterých nebylo patrné, která má být zařazena za N (any), tedy za chybějící informaci. Takto upravené sekvence byly uloženy ve formátu FASTA (\*.fas) pro další manipulaci.

#### **2.2.2.2.5 Tvorba Datasetu**

Pro další analýzy byl vytvořen Dataset spojený ze získaných sekvencí úseku *trnL-F* a upravené datové matice použité v práci (Specht 2006), která byla získána z TreeBASE (A Database of Phylogenetic Knowledge, M1009). V datasetu C.D. Specht byl zahrnut jaderný úsek ITS, chloroplastový úsek *trnL-F*, *trnK* a výsledky morfometrické analýzy.

Pro tvorbu Datasetu byl C.D. Specht dataset převeden do formátu FASTA a rozdělen na jednotlivé partitions (oddíly) v programu Cygwin pomocí perlowského skriptu cutfasta.pl. Byly vytvořeny tři datové oddíly 1. pro ITS, 2. pro sekvence pocházející z chloroplastu a 3. pro morfologická data. Dále z 2. datového oddílu byla v programu Geneious 7.0.6 (Kearse *et al.*, 2012) vytvořena konsenzuální sekvence a k ní namapovány z GeneBanku stažené

sekvence pro *trnL-F* a *trnK* z práce (Specht, 2006). Tak bylo zjištěno, kde v 2. oddílu začínají a končí jednotlivé úseky. Pak v programu Cygwin byl tento oddíl rozdělen na jednotlivé lokusy. Nakonec byly vytvořeny oddíly čtyři (ITS, *trnL-F*, *trnK*, morfologická data).

K oddílu *trnL-F* byly připojeny získané sekvence *trnL-F* regionu a dále z tohoto souboru byl vytvořen alignment pomocí algoritmu MAFFT multiple sequence alignment software (Katoh & Standley, 2013) za použití webového rozhraní. Alignment byl zkontrolován a upraven v programu BioEdit (Hall, 1999).

Nakonec byl spojen upravený oddíl *trnL-F* s oddílem *trnK* a oddílem morfologie pomocí skriptu FASconCAT-G\_v1.02.pl (Küeck & Koenig, 2014). Takto vytvořený Dataset byl použit pro vytvoření fylogenetických stromů.

#### **2.2.2.2.6 Fylogenetické stromy**

Pro tvorbu fylogenetických stromů byly použity tři metody: Maximální parsimonie (Maximum parsimony, MP), Maximální věrohodnosti (Maximum likelihood, ML) a Bayesovská analýza (Bayesian inference, BI). Jako vstupní matice byl použit výše zmíněný Dataset.

##### **2.2.2.2.6.1 Maximální parsimonie**

Fylogenetický strom konstruovaný metodou Maximální parsimonie byl vytvořen v programu PAUP\* 4.0a146 (Swofford, 2003). Do programu PAUP\* byl importován Dataset ve formátu FASTA, kde byl převeden do formátu NEXUS. Pro konstrukci stromu byla použita parsimonická analýza s předem nastavenými parametry. Pro hledání nejkratšího stromu bylo použito heuristické vyhledávání metodou stepwise addition. Sekvence byly přidávány náhodně a jako substituční model byl zvolen GTR. Poté byl vypočítán bootstrap, který byl nastaven na 100 opakování. Výsledné podpory byly namapovány na jeden z nejkratších stromů. Stejný postup byl proveden pro soubor dat z práce Specht (2006) získaných z treeBASE.

##### **2.2.2.2.6.2 Maximum likelihood**

Pro konstrukci fylogenetického stromu pomocí metody Maximum likelihood byl použit program RAxML v8.2 (Stamatakis, 2014). Byl použit algoritmus rapid Bootstrap, který najde nejpravděpodobnější strom a současně vypočítá bootstrapové podpory. Substitučním modelem byl GTRGAMMA s nastavením 100 opakování. Použitý příkaz vypadal takto: `raxmlHPC -f a -m GTRGAMMA -p 12345 -x 12345 -# 100 -s DNA.phy -n rbs`. Výsledkem byl fylogenetický strom s namapovanými bootstrapovými hodnotami.

#### **2.2.2.2.6.3 Bayesovská analýza**

Pro výpočet byl použit program MrBayes v3.1.2 (Huelsenbeck & Ronquist, 2001). Jako substituční model byl zvolen GTR + I + gamma (I + gamma = gamma s proporcí invariabilních pozic). Pro dosažení vhodné změny pozic chladného řetězce tzv. swapování byla nastavena teplota 0,01. Počet generací byl nastaven od 5000000 po 10000000 s tím, že vždy prvních 25 % stromů bylo odstraněno (burn-in).

Byly provedeny předběžné analýzy variability jednotlivých úseků použitých v práci Specht (2006) pro *trnL-F*, *trnK* a ITS. Bylo nastaveno 5000000 generací a burn-in 12500 zbytek odpovídal Tab. č. 5. Pro výslednou analýzu Datasetu bylo nastaveno 10000000 generací a burn-in 25000 Tab. č. 5. Stejným způsobem byl vytvořen strom obsahující pouze druhy z treeBASE, ke kterým lze přiřadit velikost genomu.

**Tabulka 5:** Příkazy použité v programu MrBayes pro analýzu Datasetu.

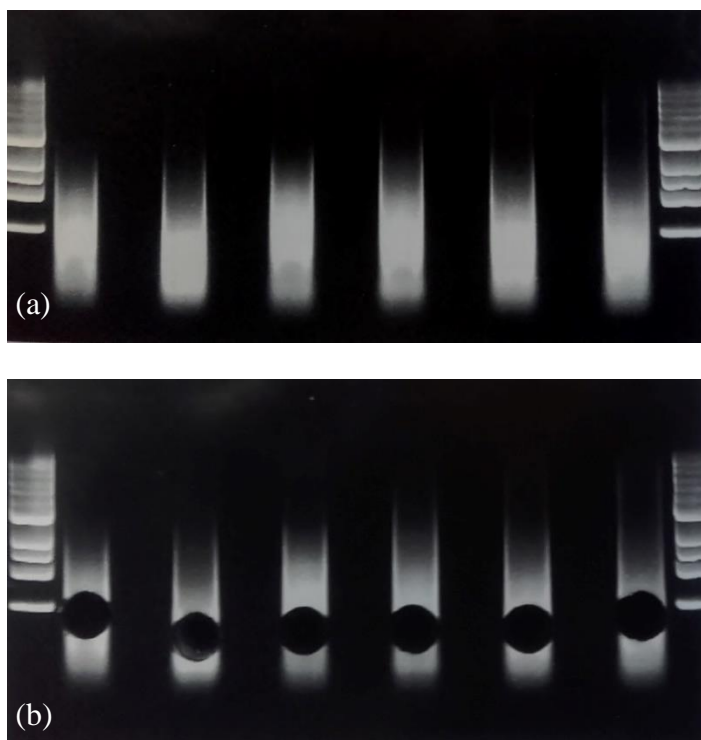
```
begin mrbayes;
log start filename=Dataset.log;
set autoclose=yes nowarn=yes;
help lset;
lset nst=6 rates=invgamma;
help mcmc;
mcmc ngen=10000000 Temp=0.01 samplefreq=100 printfreq=100 nchains=4
savebrlens=yes;
mcmc;
sump burnin=25000;
sumt burnin=25000;
end;
```

### 2.2.2.3 Příprava knihovny, získání NGS dat a jejich zpracování

#### 2.2.2.3.1 Příprava knihovny pro Hyb-Seq

Při přípravě samotné knihovny byly nejprve opraveny délky fragmentů a následně byl na jejich konce navázán adenin. Dále byly k fragmentům ligací navázány adaptory. Po ligaci byly produkty přečištěny pomocí kitu NEBNext Adaptor for Illumina (BioLabs). V dalším kroku byla provedena délková selekce na 1% agarózovém gelu a následné vyřezání fragmentů o přibližné délce 500 bp. Vzorky byly obarveny barvou orange DNA loading dye, vyřezávání proběhlo pod UV lampou pomocí obrácených 1ml špiček. Byl vyříznut terčík o průměru cca 0,5 cm. Vyřezané produkty byly promyty kitem QIAquick Gel Extraction Kit (Qiagen). Dále byly produkty namnoženy pomocí PCR, konkrétní cyklus a použitý mix jsou uvedeny v tabulce 6 a 7. Namnožené produkty byly dvakrát přečištěny pomocí Agencourt AMPure XP kitu (Beckman Coulter) a výsledná kvalita ověřena na 1% agarózovém gelu. Dále byla knihovna tzv. ekvimolárně zpoolována, kdy byly všechny připravované vzorky smíchány do jedné zkumavky tak, aby od každého vzorku bylo přítomno stejné množství fragmentů. Před samotným poolováním byla změřena koncentrace všech vzorků na fluorimetru Qubit 2.0 (Invitrogen), aby bylo možné napipetovat do celkového objemu každého vzorku stejné množství v ng. Výsledný objem zpoolovaných vzorků musel dosahovat 6  $\mu$ l, pro dosažení

požadovaného objemu byly vzorky zkoncentrovány na koncentrátoru IR Concentrator NB 503CIR (N-Biotek).



**Obrázek 23:** Vzorky nanesené po sonikaci na 0,8% agarózovém gelu. (a) vizualizace vzorků na gelu, (b): gel po vyřezání fragmentů o přibližné délce 500 bp.

**Tabulka 6:** Složení použité PCR směsi.

Reakce Q5	μl
NFRNext Q5 HotStart Hifi PCR Master Mix	25
Universal PCR Primer	10
Index Primer	10
DNA	15

**Tabulka 7:** Teplotní cyklus použitý při PCR.

Cyklus	Q5
98°C	30 sec
98°C	10 sec
65°C	1 min 15 sec
65°C	30 sec
7°C	∞

} 8x

### 2.2.2.3.2 Obohacení (Enrichment)

V následujících krocích byly ke vzorkům nahybridizovány (byly obohaceny o) předem nadesignované proby nasyntetizované firmou MYcroarray pomocí protokolu MYbaits – Sequence Enrichment for Targeted Sequencing (MYcroarray) (Gnirke *et al.*, 2009). Proby byly dodány v podobě tzv. Baits, což jsou jednořetězcové oligonukleotidy RNA o délce 120 bp s navázaným biotinem. Samotný design prob proběhl automaticky pomocí sondovače (Schmickl *et al.*, 2015), kdy byly nejdříve vybrány sekvence s 99,9% identitou spojením sekvencí získaných z genome skimmingu a transcriptomu Curcumu. Ze získaných sekvencí byly filtrováním vytvořeny klastry, kdy byly odstraněny sekvence s  $\geq 90\%$  podobností, aby bylo zabráněno několikanásobnému vzniku prob pro stejný lokus. Dále byla provedena délková selekce, kdy byly ponechány exony o délce  $\geq 120$  bp a lokusy o délce  $\geq 960$  bp. Nakonec byly znovu odfiltrovány úseky sdílející  $\geq 90\%$  identitu. Výsledkem byl soubor ve formátu FASTA, který obsahoval 1180 lokusů se 4618 exony (Weitemier *et al.*, 2014).

V prvním kroku samotného obohacení byly do 0,2ml Lo-bind zkumavek připraveny tři mixy: LM - Library Master Mix, HM – Hybridization Master Mix, CB – Capture Baits Master Mix, jejich složení je uvedeno v tabulce č. 8. Dále byly jednotlivé zkumavky vloženy do termocykleru Mastercycler Pro (Eppendorf), s cyklem uvedeným v Tab č. 10. Zde proběhla denaturace DNA a specifická hybridizace prob na jednotlivé úseky. Schéma postupu práce je uvedeno na obrázku č. 24. Dále byly pomocí streptavidinem potažených kuliček vyselektovány pouze úseky DNA, kde došlo k hybridizaci s biotinem opatřenými probami. Pak byly vytvořené komplexy pomocí magnetu vytaženy a promyty. Promytí bylo provedeno pro odstranění nespecifických DNA molekul. Výsledkem byl získ obohacené DNA o cílové lokusy. Dále byly tyto úseky namnoženy pomocí PCR. Použitý mix a cyklus jsou uvedeny v tabulce č. 10 a 11, mix byl napipetován do celkového objemu 50  $\mu$ l za použití speciálních primerů uvedených v Tab. č. 9. Nakonec byla namnožená DNA promyta pomocí QIA quick PCR Purification kitu do celkového objemu 30  $\mu$ l a změřena na fluorimetru Qubit 2,0 (Invitrogen). Takto připravené vzorky byly odeslány na Masarykovu univerzitu v Brně do Středoevropského technologického institutu – CEITEC, kde byly vzorky sekvenovány na přístroji Illumina MiSeq 2x 150 PE. Všechny kroky obohacení probíhaly v RNA sterilním prostředí. Těchto podmínek bylo dosaženo vysvícením UV lampou a aplikací dekontaminačního roztoku RNaseZAP<sup>™</sup> (Sigma) do uzavřeného boxu.

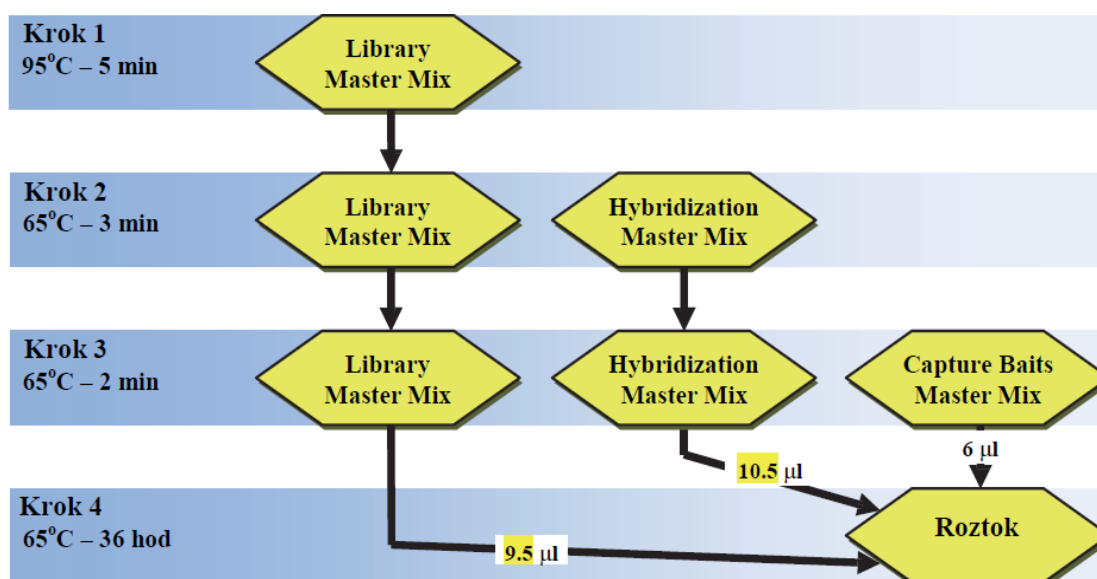
Přípravu knihovny a obohacení provedly Mgr. Monika Pospíšilová a Ing. Lenka Flašková v DNA laboratořích situovaných v Benátské 2, Univerzita Karlova v Praze.

**Tabulka 8:** Složení jednotlivých mixů.

LM – Library Master Mix	μl
Block #1	2,5
Block #2	2,5
Block #3	0,6
DNA	5,9
HM – Hybridization Master Mix	μl
HYB #1	20
HYB #2	0,8
HYB #3	8
HYB #4	0,8
CB – Capture Baits Master Mix	
Capture Probe	
RNase Block	

**Tabulka 9:** Sekvence primerů použitých v Post-Capture Amplification.

Sekvence primerů
P5_Zingiber – 5' AATGATACGGCGACCACCGA
P7_Zingiber – 5' CAAGCAGAAGACGGGCATACGA



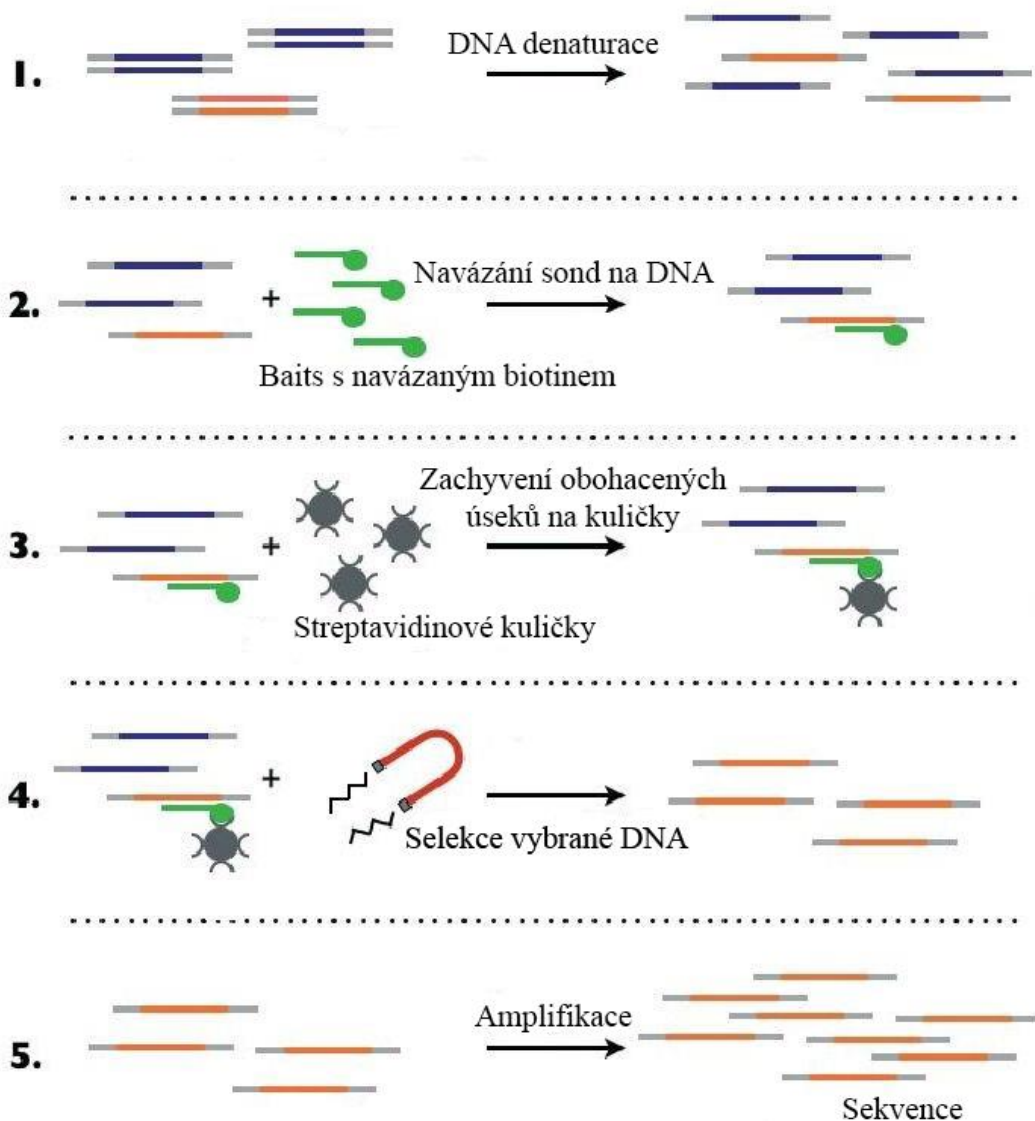
**Obrázek 24:** Náčrtek znázorňující postup přidávání jednotlivých mixů.

**Tabulka 10:** Složení použité PCR směsi.

Post - Capture Amplification	μl
Nuclease-free H2O (Gibco)	8,5 - 18,5
2x KAPA HiFi HotStart ReadyMix	25
P5_Zingiber	1,5
P7_Zingiber	1,5
DNA z knihovny	5 - 15

**Tabulka 11:** Teplotní cyklus použitý při PCR.

Cyklus	
98°C	30 sec
98°C	20 sec
62°C	30 sec
72°C	45 sec
} 12x	
72 °C	5 min
4°C	∞



**Obrázek 25:** Schéma znázorňující proces obohacení. Převzato z: MYbaits Sequence Enrichment for Targeted Sequencing User Manual Version 2.3.1 - 5/22/2014.

### 2.2.2.3.3 Zpracování NGS dat

Ze sekvenátoru byla získána data v podobě raw reads pro 30 vzorků. Raw reads byly nejprve kvalitativně filtrovány, pak byly odstraněny duplikáty pomocí skriptu HybPipe1\_rawprocess.sh (Fér, 2015). Kvalitativní filtrování odstranilo adaptory a ready s kvalitou nižší než Q20 a délkou kratší než 36 bp pomocí programu Trimmomatic 0.32 (Bolger *et al.*, 2014). Duplikáty byly odstraněny pomocí programu FASTX-0.0.13 (Gordon & Hannon, 2010). V dalším kroku byly získané ready namapovány na tzv. pseudoreferenci, což jsou za sebou poskládané exony, které jsou mezi sebou odděleny 400 N. Tento krok byl proveden v programu Geneious 7.0.6 (Kearse *et al.*, 2012). Z programu Geneious byl vyexportován konsenzus s namapovanými ready, z kterého byly vytvořeny pslx soubory, zároveň byla provedena identifikace získaných contigů pomocí programu BLAT (Kent, 2002) a skriptu HybPipe2\_generatepslx.sh (Fér, 2015). V následujícím kroku byl proveden alignment jednotlivých exonů pomocí algoritmu MAFFT-7.029 (Katoh & Standley, 2013). Alignované exony byly zkonkatenovány do jednotlivých lokusů. Vše bylo provedeno prostřednictvím skriptu HybPipe3\_processpslx.sh (Fér, 2015). Potom byly pomocí skriptu HybPipe4\_missingdataremoval.sh (Fér, 2015) odstraněna chybějící data.

Ze sekvenátoru byla získána data v podobě raw reads pro 30 vzorků. Raw reads byly nejprve kvalitativně filtrovány, pak byly odstraněny duplikáty pomocí skriptu HybPipe1\_rawprocess.sh (Fér, in prep.). Kvalitativní filtrování odstranilo adaptory a ready s kvalitou nižší než Q20 a délkou kratší než 36 bp pomocí programu Trimmomatic 0.32 (Bolger *et al.*, 2014). Duplikáty byly odstraněny pomocí programu FASTX-0.0.13 (Gordon & Hannon, 2010). V dalším kroku byly získané ready namapovány na tzv. pseudoreferenci, což jsou za sebou poskládané exony, které jsou mezi sebou odděleny 400 N. Tento krok byl proveden v programu Geneious 7.0.6 (Kearse *et al.*, 2012). Z programu Geneious byl na základě namapovaných readu vyexportován konsenzus, ze kterého byly identifikovány druhově specifické sekvence konkrétních exonů pomocí programu BLAT (Kent, 2002) a skriptu HybPipe2\_generatepslx.sh (Fér, in prep.). V následujícím kroku byly spojeny sekvence všech druhů pro jednotlivé exony a byl proveden jejich alignment pomocí algoritmu MAFFT-7.029 (Katoh & Standley, 2013). Alignované exony byly posléze zkonkatenovány do jednotlivých lokusů podle jejich příslušnosti. Vše bylo provedeno prostřednictvím skriptu HybPipe3\_processpslx.sh (Fér, in prep.) zahrnujícím skript assembled\_exons\_to\_fastas.py (Weitemier *et al.*, 2014). Dále byla pomocí skriptu HybPipe4\_missingdataremoval.sh (Fér, in

prep.) odstraněna chybějící data tak, že byly nejprve odstraněny druhy, které obsahovaly více jak 70 % chybějících dat, následně byly odstraněny alignmenty, kde bylo obsaženo méně jak 75 % druhů. Pomocí skriptu HybPipe5b\_FastTree\_for\_selected.sh (Fér, in prep.) byly pro vybrané geny zkonstruovány jednotlivé gene trees pomocí programu FastTree (Price *et al.*, 2009, 2010), který je schopen analyzovat velké alignmenty za krátký čas. Pro jednotlivé stromy jsou vypočteny lokální podpory pomocí Shimodaira-Hasegawa testu (Shimodaira, 2001), jedná se o obdobu klasického bootstrapu. Lepších výsledků rekonstrukce gene trees dosahuje metoda RaxML, se kterou je ovšem metoda FastTree srovnatelná. V této práci nebyl RaxML použit proto, že doposud neproběhla potřebná optimalizace pro analýzu dat (Fér, in prep.).

Z jednotlivých gene trees byly konstruovány species trees pomocí metod MRL, konkatenace, ASTRAL a ASTRID. Nejprve byly stromy zakořeněny skriptem HybPipe6\_roottrees.sh a dále odhadnuty stromy pomocí skriptů HybPipe7a\_astra\_sh, HybPipe7b\_astrid, HybPipe7c\_mrl.sh a HybPipe7e\_concatenatedFastTree.sh (Fér, in prep.).

Stejný postup analýzy dat proběhl i pro získané chloroplastové geny s tím rozdílem, že jako reference sloužil plastom druhu *Curcum roscoeana*. Dále byly použity dva jiné skripty konkrétně HybPipe2\_generatepslx\_cp.sh a HybPipe3\_processpslx\_cp.sh. Zbytek použitých skriptů byl totožný jako pro analýzu jaderných genů.

#### 2.2.2.3.3.1 Testy monofylie

Pro otestování různých monofyletických skupin byl použit skript, viz Tab. č. 12, kdy bylo vytvořeno 7 potenciálně monofyletických skupin. Konkrétně skupina Para, která testovala, zda rod *Paracostus* je na bázi celé čeledi; skupina Tapein, která testovala, zda jsou monofyletické všechny asijské druhy; skupina Specht\_para, testovala, zda je rod *Paracostus* na bázi asijských druhů; skupina Afr testovala, zda jsou monofyletické africké druhy; skupina Afr\_old\_JA testovala, zda jsou monofyletické africké druhy s rodem *Chamaecostus*, *Dimerocostus* a *Monocostus*; skupina OldJA\_asie testovala monofylii mezi asijskými druhy a rody *Chamaecostus*, *Dimerocostus* a *Monocostus* a Afr\_asie testovala monofylii mezi africkými a asijskými druhy. Pak byly jednotlivé skupiny porovnány s topologiemi každého genového stromu. Výstupem byla matice TRUE/FALSE, pro každou testovanou skupinu.

**Tabulka 12:** Příkazy použité pro testy monofylie (Fér, unpubl.).

```
library (phytools)
library (ape)
#define monophyletic groups of interest

Para=c("Costus-afer_S185","Costus-allenii_S247","Costus-arabicus_S251","Costus-
curvibracteatus_S246","Costus-dinklagei_S364","Costus-glaucus_S69","Costus-
laevis_S265","Costus-ligularis_S267","Costus-lucanusianus_S188","Costus-
phaeotrichus_S306","Costus-pictus_GB","Costus-pulverulentus_S302","Costus-
scaber_S305","Costus-spectabilis_S269","Costus-talbotii_S189","Costus-
tappenbeckianus_S303","Costus-tonkinensis_S268","Costus-vargasii_S304","Costus-
varzearum_S192","Costus-villosissimus_S240","Dimerocostus-argenteus_S190","Chamaecostus-
cuspidatus_S191","Cheilocostus-borneensis_S239","Cheilocostus-candidus_S333","Cheilocostus-
globosus_S184","Cheilocostus-speciosus_S183","Monocostus-uniflorus_S187","Tapeinochilos-
ananassae_S182","Tapeinochilos-dahlii_S365")

Tapein=c("Cheilocostus-borneensis_S239","Cheilocostus-candidus_S333","Cheilocostus-
globosus_S184","Cheilocostus-speciosus_S183","Tapeinochilos-ananassae_S182","Tapeinochilos-
dahlii_S365","Costus-tonkinensis_S268")

Specht_para=c("Paracostus-englerianus_S186","Paracostus-paradoxus_S266","Cheilocostus-
borneensis_S239","Cheilocostus-candidus_S333","Cheilocostus-globosus_S184","Cheilocostus-
speciosus_S183","Tapeinochilos-ananassae_S182","Tapeinochilos-dahlii_S365","Costus-
tonkinensis_S268",)

Afr=c("Costus-afer_S185","Costus-spectabilis_S269","Costus-talbotii_S189","Costus-
dinklagei_S364","Costus-lucanusianus_S188","Costus-phaeotrichus_S306","Costus-
tappenbeckianus_S303","Costus-ligularis_S267")

Afr_old_JA=c("Costus-afer_S185","Costus-spectabilis_S269","Costus-talbotii_S189","Costus-
dinklagei_S364","Costus-lucanusianus_S188","Costus-phaeotrichus_S306","Costus-
tappenbeckianus_S303","Costus-ligularis_S267","Monocostus-uniflorus_S187","Dimerocostus-
argenteus_S190","Chamaecostus-cuspidatus_S191")

OldJA_asie=c("Cheilocostus-borneensis_S239","Cheilocostus-candidus_S333","Cheilocostus-
globosus_S184","Cheilocostus-speciosus_S183","Tapeinochilos-ananassae_S182","Tapeinochilos-
dahlii_S365","Costus-tonkinensis_S268","Monocostus-uniflorus_S187","Dimerocostus-
argenteus_S190","Chamaecostus-cuspidatus_S191")

Afr_asie=c("Costus-afer_S185","Costus-spectabilis_S269","Costus-talbotii_S189","Costus-
dinklagei_S364","Costus-lucanusianus_S188","Costus-phaeotrichus_S306","Costus-
tappenbeckianus_S303","Costus-ligularis_S267","Cheilocostus-borneensis_S239","Cheilocostus-
candidus_S333","Cheilocostus-globosus_S184","Cheilocostus-speciosus_S183","Tapeinochilos-
ananassae_S182","Tapeinochilos-dahlii_S365","Costus-tonkinensis_S268")

#define list of groups (must be defined twice, once as a lists, once as names) - TO BE SOLVED IN
THE FUTURE
groupsToTest=list(Para,Tapein,Specht_para,Afr,Afr_old_JA,OldJA_asie,Afr_asie)
groupsToTestNames=c("Para","Tapein","Specht_para","Afr","Afr_old_JA","OldJA_asie","Afr_asie")
#define trees (all files with *.tre in current directory) - this command will store all tree names in
'trees_files'
trees_files <- dir(pattern="*.tre")
#function reading newick tree and evaluating monophyly of species in a group (tips = "group")
monophyletic <- function(file, monolist) {
```

```

print (file)
#read newick tree
tree = read.newick(file)
#put all species to 'alltips'
alltips <- tree$tip.label
#in 'comparelist' will be only those species from 'monolist' that are present in a tree
comparelist <- alltips[alltips %in% monolist]
#if nr. of species in comparelist is at least two function is.monophyletic is called
if (length(comparelist) > 1) {
  mono <- is.monophyletic(phy = tree, tips = comparelist)
  #otherwise function returns "NA"
} else {
  mono <- NA
}
return(c(mono))
}
#put names of trees to ismonophyl
ismonophyl <- trees_files
#make dataframe (trees names in the first column)
ismonophyl <- data.frame(matrix(unlist(ismonophyl), nrow=(length(ismonophyl)), byrow=T))
#colnames(ismonophyl) <- c("tree")
x <- 0
#loop over all groups
for(i in groupsToTest) {
  x <- x + 1
  #apply monophyly test of a group to all trees
  ismonophyl[[x+1]] <- lapply(trees_files, monophyletic, monolist = i)
}
#add column names to the matrix
colnames(ismonophyl) <- c("tree", groupsToTestNames)
#make matrix
ismonophyl <- as.matrix(ismonophyl)
write.csv(ismonophyl, file="monophylyResults.txt", quote=FALSE, row.names=FALSE)
***SCRIPTEND***

```

## 2.2.3 Vyhodnocení Velikosti genomu a fylogeneze

### 2.2.3.1 Korelace fylogenetické hypotézy s velikostí genomu

Pro výpočet a vizualizaci velikostí genomu na fylogenetickém stromě byl použit program R a balík phytools (Revell, 2012) funkce contMap. Funkce ContMap nakreslí na strom kontinuálně se vyvíjející parametr za použití Maximum likelihood k odhadu stavu znaků ve vnitřních nodech stromu. Tento výpočet byl proveden pro tři stromy. První byl použit strom pocházející z analýzy Datasetu za použití ML. Zde byly zahrnuty vzorky, ke kterým nebyla známá velikost genomu, proto byla použita funkce contMap(method="anc.ML"), kdy se k větvím, kde nejsou známé hodnoty, dopočítá

průměrná hodnota podle okolních známých hodnot. Větve vedoucí ke vzorkům bez velikostí genomu byly obarveny na šedo v programu Adobe Photoshop CS6. Druhý byl species tree sestavený konkatencí a FastTree. Třetím byl strom sestavený metodou ASTRID.

### 2.2.3.2 Rekonstrukce ancestrální velikosti genomu

Pro výpočet ancestrální velikosti genomu pro jednotlivé nody na stromě byl použit program R balík phytools (Revell, 2012) funkce fastAnc. Tento výpočet bylo možné uskutečnit pouze u stromu bez chybějících dat. Byl použit species tree vytvořený metodou ASTRID.

Byly vypočteny škálovací parametry  $\lambda$ =lambda,  $\kappa$ =kappa,  $\delta$ =delta, pro species tree vytvořený metodou ASTRID a species tree vytvořený konkatencí. Škálovací parametry charakterizují evoluci proměnné v tomto případě velikost genomu (Pagel, 1997, 2004). Parametr lambda ( $\lambda$ ) popisuje, zda jsou hodnoty proměnné pro jednotlivé taxony nezávislé na fylogenezi. Tento parametr nabývá hodnot v intervalu od 0 do 1, kdy čím vyšší je tato hodnota, tím více topologie stromu ovlivnila vývoj proměnné. Parametr kappa ( $\kappa$ ) odhaduje, jaký vliv má délka větví na evoluci proměnné.  $\kappa > 1$  znamená, že se delší větve podílejí na evoluci proměnné více než krátké.  $\kappa < 1$  naopak znamená, že na evoluci dané proměnné měly větší vliv krátké větve. V případě, že se  $\kappa = 0$ , evoluce parametru probíhala skokově, naopak pokud se  $\kappa = 1$ , evoluce probíhala pozvolně. Parametr delta ( $\delta$ ) odhaduje, zda evoluci proměnné charakterizuje spíše pomalá akumulace změn v čase, nebo zda došlo k adaptivní radiaci.  $\delta > 1$  odpovídá víceméně konstantním změnám během evoluce,  $\delta < 1$  odpovídá adaptivní radiaci. Odhad škálovacích parametrů byl proveden ML algoritmem v programu R pomocí balíku caper (funkce pgl) (Orme *et al.*, 2012) spolu s vypočtením 95% konfidenčního intervalu.

### 3. VÝSLEDKY

#### 3.1 Měření velikosti genomu

Na průtokovém cytometru bylo celkem analyzováno 146 vzorků. 2C-hodnota se pohybovala v rozmezí 1,07 - 9,41 pg. Nejmenší 2C-hodnota byla naměřena pro druh *Costus talbotii*, naopak nejvyšší pro druh *Costus cf. ligularis*. Nejmenší průměrná 2C-hodnota byla naměřena u rodu *Monocostus* K.Schum. 1,56 pg a rodu *Dimerocostus* Kuntze 1,65 pg, nejvyšší potom pro rod *Paracostus* C.D.Specht 5,07 pg. Souhrn všech naměřených hodnot je uveden v Tab. č. 13. Největší rozpětí hodnot bylo naměřeno pro rody *Costus* L., *Paracostus* C.D.specht a *Cheilocostus* C.D.Specht viz Obr. č. 26.

**Tabulka 13:** Naměřené absolutní velikosti genomu (2C) pro čeleď *Costaceae*. St = standartd, B = *Bellis perennis*, S = *Solanum pseudocapsicum*, 2C-hodnoty jsou uvedeny v pikogramech.

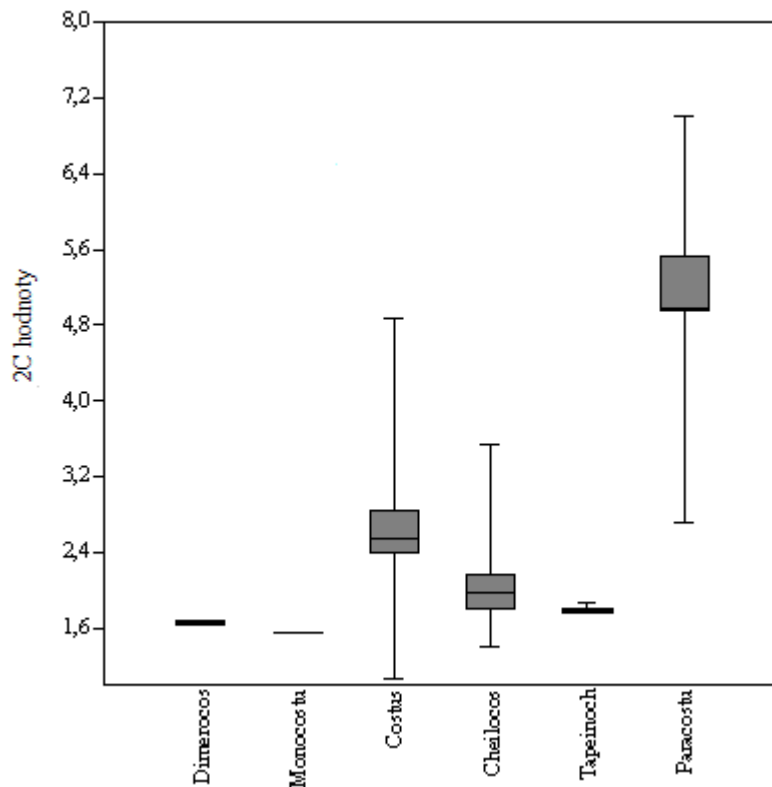
FCM kód	Identifikační číslo	rod	druh	St	2C
<b>Rod Chamaecostus</b>					
FCM 1257	JLS 1935	<i>Chamaecostus</i>	<i>cuspidatus</i>	B	2,7836
<b>Rod Dimerocostus</b>					
FCM 1343	E19750179	<i>Dimerocostus</i>	<i>argenteus</i>	B	1,6543
FCM 1413	US 1994712	<i>Dimerocostus</i>	<i>strobilaceus</i>	B	1,6407
FCM 1523	R3137	<i>Dimerocostus</i>	<i>strobilaceus white</i>	S	1,6490
FCM 599	S19990324	<i>Dimerocostus</i>	<i>strobilaceus 'Yellow'</i>	B	1,6707
Průměr					<b>1,6537</b>
<b>Rod Monocostus</b>					
FCM 1431	US 1994725	<i>Monocostus</i>	<i>uniiflorus</i>	S	1,5609
FCM 588	JLS-431	<i>Monocostus</i>	<i>uniiflorus</i>	B	1,5605
Průměr					<b>1,5607</b>
<b>Rod Costus</b>					
FCM 1307	S20060821	<i>Costus</i>	<i>afer</i>	S	4,8190
FCM 1611	Innsb, ex München	<i>Costus</i>	<i>afer (Munchen)</i>	S	4,8459
FCM 1433	US 1994671	<i>Costus</i>	<i>aff.afer</i>	B	4,7916
FCM 1436	US 1994675	<i>Costus</i>	<i>aff. guanaiensis</i>	B	2,9001
FCM 1491	L88,0426	<i>Costus</i>	<i>aff. curvibracteatus basal</i>	B	2,3723
FCM 1492	R3032	<i>Costus</i>	<i>aff. curvibracteatus terminal</i>	B	2,3383
FCM 1293	JLS 1950	<i>Costus</i>	<i>aff. lasius</i>	B	2,4670

FCM 1297	S20100776	<i>Costus</i>	<i>aff. lucanusianus (yellow fls)</i>	S	3,6631
FCM 1550	R3245	<i>Costus</i>	<i>aff. villosissimus</i>	B	2,6275
FCM 1254	JLS 2084	<i>Costus</i>	<i>allenii 'Red Stem'</i>	S	3,5336
FCM 1316	E19973663	<i>Costus</i>	<i>amazonicus</i>	B	2,6248
FCM 1488	L82,1252	<i>Costus</i>	<i>amazonicus</i>	B	2,8344
FCM 1510	R3124	<i>Costus</i>	<i>arabicus</i>	B	2,6623
FCM 1515	R3213	<i>Costus</i>	<i>barbatus</i>	B	2,3806
FCM 1410	US 1994678	<i>Costus</i>	<i>bracteatus</i>	B	2,6699
FCM 1613	3485 ex Innsbruck	<i>Costus</i>	<i>brasiliensis</i>	S	4,8486
FCM 1241	JLS 1930	<i>Costus</i>	<i>cf. scaber</i>	B	2,4245
FCM 1432	US 1994703	<i>Costus</i>	<i>comosus</i>	B	2,3978
FCM 1549	R3208	<i>Costus</i>	<i>comosus "Cliff Dweller"</i>	B	2,2674
FCM 1517	R3212	<i>Costus</i>	<i>comosus "Cliff Dweller"</i>	B	2,2371
FCM 1312	E19902066	<i>Costus</i>	<i>comosus (barbatus)</i>	B	2,4062
FCM 1315	E19973660	<i>Costus</i>	<i>curvibracteatus</i>	B	2,4777
FCM 1304	S20040282	<i>Costus</i>	<i>deistelii</i>	S	4,8079
FCM 1612	17560 ex Innsbruck	<i>Costus</i>	<i>dinklagei</i>	S	4,8237
FCM 1440	US 1994668	<i>Costus</i>	<i>dubius</i>	B	4,8450
FCM 1305	S20040277	<i>Costus</i>	<i>erythrocoryne</i>	S	2,3504
FCM 1332	E19750176	<i>Costus</i>	<i>glaucus</i>	B	2,4336
FCM 1246	JLS 2085	<i>Costus</i>	<i>guianensis var, tarmicus 'Marmallade'</i>	B	2,6289
FCM 1251	JLS 2098	<i>Costus</i>	<i>chartaceus</i>	S	2,3631
FCM 1428	US 1990016	<i>Costus</i>	<i>chartaceus</i>	B	2,3597
FCM 1334	E19751801	<i>Costus</i>	<i>laevis</i>	B	2,5447
FCM 1424	US 1994681	<i>Costus</i>	<i>laevis</i>	B	2,5519
FCM 1521	R3086	<i>Costus</i>	<i>laevis</i>	B	2,4610
FCM 1548	R3135	<i>Costus</i>	<i>laevis</i>	B	2,6222
FCM 1253	R19990780	<i>Costus</i>	<i>lasius</i>	B	2,5702
FCM 1406	US 1994670	<i>Costus</i>	<i>lasius</i>	B	2,6177
FCM 1256	JLS 2117	<i>Costus</i>	<i>lateriflorus</i>	S	1,3706
FCM 1434	US 1998224	<i>Costus</i>	<i>lateriflorus</i>	S	1,3658
FCM 1538	R3126	<i>Costus</i>	<i>longibracteolatus green</i>	B	2,5365
FCM 1429	US 1994673	<i>Costus</i>	<i>malorteanus</i>	B	2,3224
FCM 1490	R2972	<i>Costus</i>	<i>montanus</i>	B	2,5690
FCM 1507	L2003,0187	<i>Costus</i>	<i>nudicaulis</i>	B	2,0321
FCM 1514	M97P45	<i>Costus</i>	<i>phaeotrichus</i>	B	2,5619
FCM 1427	US 1994685	<i>Costus</i>	<i>pictus</i>	B	2,2466
FCM 1335	E20080228	<i>Costus</i>	<i>pulverulentus</i>	B	2,5020
FCM 1539	R3133	<i>Costus</i>	<i>pulverulentus</i>	B	2,5995

FCM 1618	7264 ex Innsbruck	<i>Costus</i>	<i>pulverulentus</i>	B	4,5834
FCM 1435	US 1994677	<i>Costus</i>	<i>scaber</i>	S	3,5854
FCM 1511	R3076	<i>Costus</i>	<i>scaber</i>	B	2,4958
FCM 1537	R3220	<i>Costus</i>	<i>scaber</i>	B	2,3757
FCM 1614	120070ex Innsbruck	<i>Costus</i>	<i>scaber</i>	B	2,4120
FCM 1266	JLS 1873	<i>Costus</i>	<i>sp. 'aff. plicatus, hairy'</i>	B	2,8240
FCM 1268	JLS 1901	<i>Costus</i>	<i>sp. nov, aff plicatus</i>	B	2,5995
FCM 063	E19881617	<i>Costus</i>	<i>spectabilis</i>	B	4,3934
FCM 1314	E19881617	<i>Costus</i>	<i>spectabilis</i>	S	4,1958
FCM 1430	US 1996284	<i>Costus</i>	<i>spectabilis</i>	B	4,2146
FCM 1437	US 2002127	<i>Costus</i>	<i>spicatus</i>	B	2,5416
FCM 1518	R3129	<i>Costus</i>	<i>spiralis, green bract form</i>	B	2,4600
FCM 1291	JLS 2115	<i>Costus</i>	<i>talbotii</i>	S	1,3723
FCM 1522	M07P213	<i>Costus</i>	<i>talbotii</i>	S	1,0741
FCM 1426	US 1994682	<i>Costus</i>	<i>tappenbeckianus</i>	B	4,8710
FCM 1438	US 1994680	<i>Costus</i>	<i>vargasii</i>	B	2,6383
FCM 1519	R3264	<i>Costus</i>	<i>vargasii</i>	B	2,4843
FCM 1301	S20110245	<i>Costus</i>	<i>varzeorum</i>	B	2,9356
FCM 1313	E19822834	<i>Costus</i>	<i>villosissimus</i>	B	2,5516
FCM 1238	JLS 1931	<i>Costus</i>	<i>villosissimus</i>	B	2,7724
FCM 1242	JLS 1931	<i>Costus</i>	<i>villosissimus</i>	B	2,7465
FCM 1439	US 1994707	<i>Costus</i>	<i>vinosus</i>	B	2,4558
FCM 1524	L2001,0523	<i>Costus</i>	<i>vinosus</i>	B	2,5562
FCM 1520	R3092	<i>Costus</i>	<i>wilsonii</i>	B	2,4772
FCM 1615	4405 ex Innsbruck	<i>Costus</i>	<i>woodsonii</i>	B	2,5032
Průměr					<b>2,8563</b>
<b>Rod Cheilocostu (Hellenia)</b>					
FCM 591	JLS-437	<i>Cheilocostus</i>	<i>Candidus sp. Nov.</i>	B	2,3067
FCM 1509	R3170	<i>Cheilocostus</i>	<i>aff. potierae</i>	B	1,9974
FCM 1235	JLS 2198	<i>Costus</i>	<i>aff. speciosus</i>	S	1,8134
FCM 1338	E20040728	<i>Cheilocostus</i>	<i>borneensis</i>	S	2,8922
FCM 1237	JS 74749	<i>Cheilocostus</i>	<i>cf. speciosus</i>	S	1,9513
FCM 1250	JLS 711	<i>Cheilocostus</i>	<i>cf. speciosus Moluky</i>	S	1,9386
FCM 597	S20060634	<i>Cheilocostus</i>	<i>globosus</i>	B	2,2194
FCM 600	S20040135	<i>Cheilocostus</i>	<i>globosus</i>	B	2,1270
FCM 1300	S20040135	<i>Cheilocostus</i>	<i>globosus =FCM 600</i>	S	2,1554
FCM 1489	R3091	<i>Cheilocostus</i>	<i>globosus yellow ? Mellow yellow?</i>	B	2,2842
FCM 1339	E20090617	<i>Cheilocostus</i>	<i>sopuensis</i>	S	1,3961

FCM 590	JLS-584	<i>Cheilocostus</i>	<i>sopuensis</i>	B	1,4115
FCM 1249	JLS 1723	<i>Cheilocostus</i>	<i>sp.</i>	S	1,6771
FCM 1274	JLS 1605	<i>Cheilocostus</i>	<i>sp.</i>	S	1,8040
FCM 1275	JLS 1785	<i>Cheilocostus</i>	<i>sp.</i>	S	1,8309
FCM 592	JLS-436	<i>Cheilocostus</i>	<i>sp.</i>	B	1,7201
FCM 739	JLS 1693	<i>Cheilocostus</i>	<i>sp.</i>	B	1,9754
FCM 1469	M1731	<i>Cheilocostus</i>	<i>sp. nov.</i>	B	2,0999
FCM 1245	JLS 816	<i>Cheilocostus</i>	<i>speciosus</i>	S	3,5354
FCM 1270	JLS 1176	<i>Cheilocostus</i>	<i>speciosus</i>	S	1,9138
FCM 1272	JS 74766	<i>Cheilocostus</i>	<i>speciosus</i>	S	1,9857
FCM 1296	JS 74757	<i>Cheilocostus</i>	<i>speciosus</i>	S	2,1213
FCM 1342	E19751812	<i>Cheilocostus</i>	<i>speciosus</i>	B	1,9373
Průměr					<b>2,0476</b>
<b>Rod Tapeinochilos</b>					
FCM 1458	W89567	<i>Tapeinochilos</i>	<i>aff. dahlii</i>	B	1,8070
FCM 601	20001570*B	<i>Tapeinochilos</i>	<i>anannasae</i>	B	1,7669
FCM 1340	E20070119	<i>Tapeinochilos</i>	<i>dahlii</i>	S	1,7928
FCM 1294	JLS 2078	<i>Tapeinochilos</i>	<i>densum</i>	S	1,8016
FCM 1456	W14P83	<i>Tapeinochilos</i>	<i>palustris</i>	B	1,8655
FCM 1258	JLS 2118	<i>Tapeinochilos</i>	<i>queenslandiae</i>	S	1,7577
FCM 1457	W04P28	<i>Tapeinochilos</i>	<i>solomonensis</i>	B	1,7684
FCM 1318	E20131999	<i>Tapeinochilos</i>	<i>sp.</i>	S	1,7764
Průměr					<b>1,7920</b>
<b>Rod Paracostus</b>					
FCM 1319	E19622228	<i>Paracostus</i>	<i>paradoxus</i>	B	4,9693
FCM 1320	E19773474	<i>Paracostus</i>	<i>paradoxus</i>	B	4,9613
FCM 1333	E19773484	<i>Paracostus</i>	<i>paradoxus</i>	S	4,9564
FCM 1341	E20040947	<i>Paracostus</i>	<i>sp.</i>	B	2,7224
FCM 1512	L93,0288	<i>Paracostus</i>	<i>englerianus</i>	B	5,3056
FCM 1513	M756	<i>Paracostus</i>	<i>paradoxus</i>	B	7,0156
FCM 589	JLS-888	<i>Paracostus</i>	<i>englerianus</i>	B	5,5279
Průměr					<b>5,0655</b>
<b>Nezařazené</b>					
FCM 1508	R2261	<i>Costus</i>	"El Whiskey"	B	2,8725
FCM 1525	R3160	<i>Costus</i>	<i>aff. nitidus</i>	B	2,4035
FCM 1516	R3139	<i>Costus</i>	<i>aff. ricus</i>	B	2,2796
FCM 1317	E19521080	<i>Costus</i>	<i>cf. ligularis</i>	B	9,4139
FCM 1252	JLS 2101	<i>Costus</i>	<i>hieroglyphicus</i>	S	2,2392

FCM 1628	FRI 48457	<i>Costus</i>	<i>oligophyllus</i>	B	1,9769
FCM 1262	JLS 2100	<i>Costus</i>	<i>productus</i> var. <i>productus</i> 'red stems'	B	2,4784
FCM 1240	JLS 1929	<i>Costus</i>	<i>sanguineus</i>	B	2,3945
FCM 1067	MP 17	<i>Costus</i>	<i>sp.</i>	B	1,8514
FCM 1236	JLS 1227	<i>Costus</i>	<i>sp.</i>	S	1,7025
FCM 1239	JLS 2033	<i>Costus</i>	<i>sp.</i>	S	2,9681
FCM 1244	JLS 1870	<i>Costus</i>	<i>sp.</i>	S	2,8468
FCM 1255	JLS 1605	<i>Cheilocostus</i>	<i>sp.</i>	S	1,7941
FCM 1264	No1785	<i>Costus</i>	<i>sp.</i>	S	1,8439
FCM 1265	JLS 972	<i>Cheilocostus</i>	<i>sp.</i>	B	2,5845
FCM 1295	S20060114	<i>Costus</i>	<i>sp.</i>	B	4,7287
FCM 1306	S20041027	<i>Costus</i>	<i>sp. (green fl)</i>	B	2,7080
FCM 1247	S20010315	<i>Costus</i>	<i>sp. Africa</i>	S	4,8243
FCM 1269	S20100779	<i>Costus</i>	<i>sp. Africa</i>	B	2,5297
FCM 1243	JLS 887	<i>Costus</i>	<i>sp. 'blue flowers'</i>	S	2,0319
FCM 1292	JLS 2037	<i>Costus</i>	<i>sp. nov.? (red pulvinus)</i>	S	1,6522
FCM 1248	JLS 1680	<i>Costus</i>	<i>tonkinensis</i>	B	2,5099
FCM 1271	JLS 1690	<i>Costus</i>	<i>tonkinensis</i>	B	2,5068
FCM 1273	JLS 371	<i>Costus</i>	<i>tonkinensis</i>	B	2,5334
FCM 593	JLS-831	<i>Costus</i>	<i>tonkinensis</i>	B	2,4700
FCM 966	JLS 2153	<i>Costus</i>	<i>tonkinensis</i>	B	2,5774
FCM 1299	JLS 831	<i>Costus</i>	<i>tonkinensis =FCM 593</i>	B	2,4915
FCM 1302	JLS 2153	<i>Costus</i>	<i>tonkinensis glabrous</i>	B	2,5447
FCM 1303	JLS 2153	<i>Costus</i>	<i>tonkinensis hairy</i>	B	2,7260
FCM 1267	S20100780	<i>Costus</i>	<i>xx of Gabon</i>	S	2,0109



**Obrázek 26:** Krabicové diagramy všech naměřených 2C-hodnot pro jednotlivé rody čeledi *Costaceae*.

### 3.2 Doplnění druhů a rekonstrukce fylogeneze

Do souhrnného Datasetu bylo použito celkem 104 jedinců. Bylo zařazeno 23 jedinců shodných z předchozí fylogenetické studie (Specht, 2006a) a 62 jedinců pocházející přímo z této studie, z čehož 2 sloužily jako outgroup (*Siphonochilus decora* a *Siphonochilus kirkii*). Dále bylo přidáno 15 nových druhů, které doposud nebyly analyzovány. Kvůli často přítomnému dvojitému signálu a celkově špatně čitelným sekvencím nebyl do analýz zařazen úsek ITS.

Z HybSeq sekvenování byly získány jak jaderné geny, tak i geny chloroplastu. Do analýz jaderných genů bylo použito 33 jedinců, pouze v případě odhadu species tree metodou ASTRAL bylo zařazeno 31 jedinců. Pro analýzu chloroplastových genů bylo zařazeno 32 jedinců. V obou případech sloužil jako outgroup *Siphonochilus aethiopicus*. Druhy byly vybrány tak, aby co nejlépe odrážely diverzitu čeledi *Costaceae* a současně byly zahrnuty taxony, které se doposud neanalyzovaly.

## 3.2.1 Fylogenetické analýzy z klasického sekvenování

### 3.2.1.1 Maximální parsimonie

Celkově byla metoda Maximální parsimonie uplatněna na dva datové soubory. První byl použit výše zmíněný Dataset, z analýzy celkově vyšlo 1585 stromů, délka stromů byla 2950 kroků. Konzistenční index byl  $CI=0,909$  a retenční index byl  $RI=0,910$ . Jeden z nejparsimoničtějších stromů je uveden v příloze 1, konsenzus není uveden, jelikož neměl v podstatě žádnou rozlišovací schopnost.

Druhý soubor dat byl vytvořen ze vzorků z původní analýzy s tím, že byly zařazeny pouze ty druhy, ke kterým byla měřena i velikost genomu. Celkem vzniklo 14 nejparsimonějších stromů. Délka nejkratších stromů byla 2717 kroků. Konzistenční index byl  $CI = 0,704$  a retenční index byl  $RI = 0,735$ . Základní taxony byly dobře odděleny, avšak často s podporami menšími než 50 %. V rodě *Costus* byla topologie velmi nestabilní.

### 3.2.1.2 Maximum likelihood

Strom vytvořený metodou Maximum likelihood dosáhl věrohodnosti -18445,06. Celkové procento chybějících dat je 68,7 %. Topologie jednotlivých taxonů nejlépe odpovídala publikované práci (Specht, 2006) ze všech použitých 3 metod rekonstrukce fylogeneze. Podpora nižší než 70 % byla považována za slabou. Celkově není strom dobře podpořen bootstrapovými hodnotami, což je s největší pravděpodobností způsobeno velkým množstvím chybějících dat. Výsledný strom je uveden v příloze 2.

### 3.2.1.3 Bayesovská analýza

Strom vytvořený na základě Datasetu měl *Average standard deviation of split frequencies* = 0.016277, což je na hranici použitelnosti. Tato hodnota by se měla pohybovat do 0,01 v tomto případě to znamená, že stromy konvergovaly k velmi podobnému výsledku. Průměrná marginální pravděpodobnost jednotlivých běhů byla -18511,44 a -18515,45. Celkový počet uložených stromů byl 149966. Topologie stromu přibližně odpovídá práci

(Specht, 2006), ale pouze pro jednotlivé rody, jinak byl strom bez rozlišovací schopnosti. Výsledný strom je uveden v příloze 3.

### 3.2.2 Fylogenetické analýzy z NGS-HybSeq

#### 3.2.2.1 Odhady species trees

Metodou FastTree bylo získáno 650 stromů pro jednotlivé jaderné geny a 82 stromů pro chloroplastové geny. Z těchto genových stromů byly odhadnuty species trees jednak konkatenací s konstrukcí stromu pomocí FastTree dále metodami MRL, ASTRAL a ASTRID. Stromy vytvořené na základě jaderných genů jsou uvedeny v přílohách 4 - 7, pro chloroplastové v přílohách 8 - 11.

#### 3.2.2.2 Testy monofylie

Při testech monofylie bylo analyzováno 650 genových stromů, pro jednotlivé definované skupiny. Nejlepšího skóre dosáhla skupina Tapeino, která testovala, zda jsou monofyletické asijské druhy s rodem *Tapeinochilos*, kdy toto pattern bylo nalezeno v 53,38 % genových stromů. Ve 21,69 % genových stromů byl rod *Paracostus* bazální, což bylo testováno zařazením všech zástupců čeledi mimo zmíněný *Paracostus* a ve 12 % genových stromů vyšly monofyletické jihoamerické rody *Chamaecostus*, *Dimerocostu* a *Monocostus* s asijskými druhy. Výsledky testu monofylie jsou shrnuty v tabulce č. 14.

**Tabulka 14:** Výsledky testu monofylie. Je uveden celkový počet monofyletických shod pro jednotlivé skupiny z celkového počtu 650 genových stromů. V druhém řádku jsou výsledky vyjádřeny v procentech.

Para	Tapein	Specht_para	Afr	Afr_old_JA	OldJA_asie	Afr_asie
141	347	2	7	1	78	0
21.69 %	53.38 %	0.31 %	1.08 %	0.15 %	12.00 %	0.00 %

### 3.3 Korelace velikosti genomu s fylogenezí

Pro porovnání velikosti genomu s fylogenetickými vztahy čeledi *Costaceae* byly jednotlivé 2C-hodnoty namapovány na stromy vytvořené metodou Maximum likelihood. Poté byly pomocí ML odhadnuty parametry  $\lambda$ ,  $\kappa$ ,  $\delta$ .

#### 3.3.1 Evoluce velikosti genomu

Velikosti genomu byly namapovány na tři stromy. První byl použit strom pocházející z analýzy Datasetu za použití ML viz příloha 12. Druhý byl species tree sestavený konkatenací a FastTree viz příloha 13. Třetím byl strom sestavený metodou ASTRID, viz příloha 14, současně byly pro tento strom vypočteny ancestrální velikosti genomu pro jednotlivé nody pomocí ML, viz příloha 15.

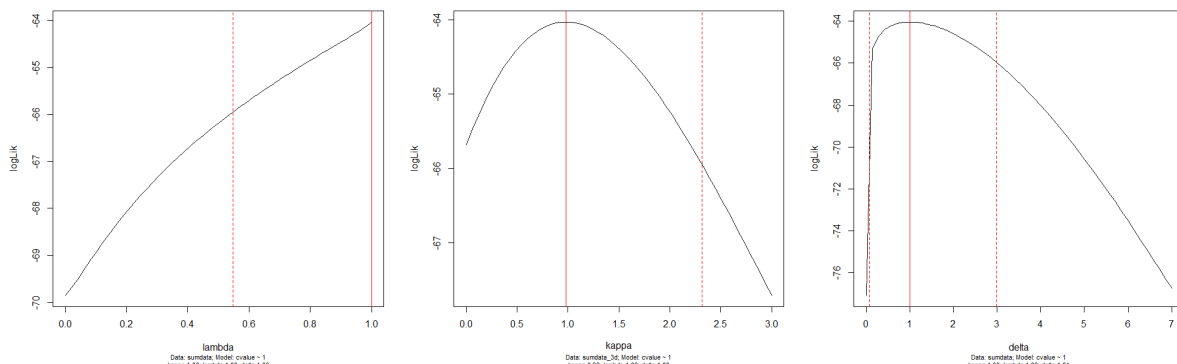
#### 3.3.2 Odhady škálovacích parametrů

Odhady škálovacích parametrů byly vypočteny pro konkatenovaný species tree a species tree sestavený metodou ASTRID. Jednotlivé odhady škálovacích parametrů spolu s 95% konfidenčními intervaly jsou uvedeny v tabulce 15 a 16 spolu s grafickým znázorněním na obrázku 27 a 28. Parametr  $\lambda$  byl defaultně nastaven na 1 s tím, že byl dopočten 95% konfidenční interval (CI). Parametr  $\kappa$  vyšel pouze v případě konkatenovaného species tree, kdy se rovnal 0,980, což naznačuje, že evoluce proměnné měla spíše pozvolnou evoluci. Parametr  $\delta$  vyšel u každého species tree odlišně a jsou spolu v rozporu, čili nelze pomocí tohoto parametru charakterizovat evoluci proměnné.

**Tabulka 15:** Odhady škálovacích parametrů  $\lambda$ ,  $\kappa$  a  $\delta$  pro konkatenovaný species tree spolu s 95% konfidenčními intervaly.

Parametr	Odhad parametru (ML)	95% konfidenční interval
$\lambda$	1.000	0.547 - NA
$\kappa$	0.980	NA, 2.317
$\delta$	1.007	0.072, 2.981

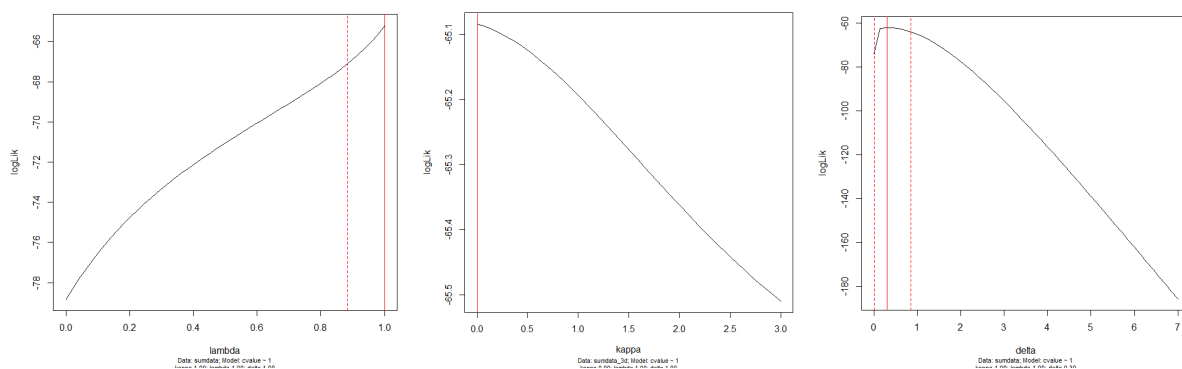
Obr 27



**Obrázek 27:** Grafy znázorňující 95% konfidenční interval parametrů  $\lambda$ ,  $\kappa$  a  $\delta$  pro 2C-hodnoty na konkatenovaném species tree.

**Tabulka 16:** Odhady škálovacích parametrů  $\lambda$ ,  $\kappa$  a  $\delta$  pro species tree vytvořený na základě ASTRID spolu s 95% konfidenčními intervaly.

Parametr	Odhad parametru (ML)	95% konfidenční interval
$\lambda$	1,000	0,884, NA
$\kappa$	0,000	NA, NA
$\delta$	0,304	0,023, 0,847



**Obrázek 28:** Grafy znázorňující 95% konfidenční interval parametrů  $\lambda$ ,  $\kappa$  a  $\delta$  pro 2C-hodnoty na species tree vytvořeném metodou ASTRID.

## 4. DISKUZE

### 4.1 Nomenklatura a taxonomie

V této diplomové práci byla použita některá synonyma k dnes akceptovaným názvům. Jedná se o rod *Hellenia* ( $\equiv$ *Cheilocostus*). Bylo tak učiněno z důvodu, že tento název je platný pouze krátkou dobu a je reflektován pouze v databázi World checklist of selected plant families ([online]), databáze IPNI ([online]) a The Plant List ([online]) tuto změnu zatím nereflektovaly. Dalším důvodem je fakt, že je usilováno o konzervaci jména *Cheilocostus*, viz kapitola 1.4.2.4. Rod *Hellenia* v současnosti zahrnuje čtyři druhy, což zřejmě dostatečně neodpovídá morfologické diverzitě tohoto rodu, jak už naznačuje (Specht, 2006a), kdy druhy *Hellenia speciosa* a *Hellenia globosa* zřejmě představují široké komplexy. Dále byl do této práce zahrnut asijský druh *Costus tonkinensis*, který neodpovídá rodové charakteristice pro *Costus*, který je zastoupen pouze Jižní Americe a Africe. I ve výsledcích fylogenetických analýz tento druh spadá k asijskému rodu *Hellenia*.

### 4.2 Sekvence DNA

Při sekvenování jednotlivých úseků DNA byla optimalizována PCR reakce, viz Obr. č. 22 (obrázky gelů). Navzdory tomu byly v sekvencích často přítomny dvojsignály. V případě ITS se jednalo skoro o všechny sekvence, v tomto případě lze tento jev nejspíše vysvětlit nedokonalou „concerted evolution“, nicméně tento úsek byl součástí práce (Specht *et al.*, 2001; Specht, 2006a), proto bylo předpokládáno, že výsledky budou dobré. Je zřejmé, že i v případě těchto prací ITS nefungovalo optimálně, jelikož pro některé taxony tento úsek chybí. Je otázkou, jak přesně vypadaly použité Specht sekvence. Celkově se při průběžném testování ukázalo, že úsek ITS má malou výpovědní hodnotu o fylogenezi této čeledi. Je to jeden z důvodů, proč nakonec nebyl úsek ITS do analýz zahrnut. Nízká variabilita ITS je do jisté míry překvapením s ohledem na sesterskou čeleď *Zingiberaceae*, kde je tento úsek značně variabilnější rovnající se několikanásobně delšímu *matK* (Pospíšilová, 2012; Hlavatá, 2014). V případě úseku *trnL-F* byl detekován stejný problém, nicméně v menší míře, zato vysvětlení původu dvojsignálu v úseku chloroplastové DNA není triviální. Může se jednat o nedokonalost polymeráz, nicméně v některých případech byl i po několikanásobném

opakování výsledek stále stejný, oproti tomu jiné sekvence byly velmi kvalitní už od začátku. Celkově byla do analýz použita přibližně třetina získaných sekvencí.

V případě metody HybSeq byly nekvalitní sekvence odfiltrovány. Z těchto sekvencí byly pomocí výše zmíněné pipeline získány gene trees. Je třeba říci, že tyto gene trees nebyly dále hodnoceny ani filtrovány. Na základě získaných statistik o stromech je možné provést manuální odstranění některých genů, kdy výsledkem mohou být lepší odhady species tree s většími statistickými podporami. I když v této práci nebyla provedena korekce gene trees, jsou statistické podpory jednotlivých species trees vysoké a mají dobrou rozlišovací schopnost.

### 4.3 Rekonstrukce fylogeneze

V této práci nebyly provedeny ILD (Incongruence Length Difference) testy, které by v případě signifikantního výsledku naznačovaly, že použité sekvence z úseku ITS a *trnL-F* jsou ve vzájemném konfliktu a nebylo by pak vhodné je zařadit do analýz společně. ILD testy nebyly zařazeny proto, že byly pro jednotlivé úseky provedeny v práci (Specht *et al.*, 2002; Specht, 2006a), kdy nevyšly signifikantní a také díky tomu, že do analýz nebyl zahrnut úsek ITS, který by spolu s chloroplastovými úseky mohl potenciálně jevit inkongruence.

Výsledné stromy z analýz MP, ML a Bayesovské analýzy, viz přílohy 1 – 3, vykazovaly stejnou topologii pro oddělení rodu *Monocostus*, *Dimerocostua* a *Chamaecostus* od zbytku zástupců *Costaceae*. Rod *Chamaecostus* měl ve všech případech stejnou topologii, což lze vysvětlit přítomností pouze jednoho vzorku s velkým množstvím chybějících dat, dále totožnou topologii pro všechny stromy měl rod *Monocostus*. Rod *Dimerocostus* tvořil monofyletickou skupinu s dobrými bootstrapovými podporami pouze v případě ML, ve stromech MP a BI vykazovala tato skupina polytomie spolu s rodem *Monocostus*. Stejná topologie všech stromů byla detekována i pro oddělení rodu *Cheilocostus* a *Tapeinochilos* od zbývajících skupin *Paracostus* a *Costus*. Rod *Tapeinochilos* byl v případě MP a BI s nízkými podporami, ale beze zbytku oddělen, v případě MP byl *Tapeinochilos anannasae* přiřazen k druhům komplexu *Cheilocostus globosus*. Rod *Cheilocostus* vyšel ve všech analýzách jako parafyletický, kdy je na bázi oddělen komplex *Cheilocostus globosus* a od něj jsou odděleny rod *Tapeinochilos* a komplex *Cheilocostus speciosus* společně s druhy *Cheilocostus aff. Potierae* a *Cheilocostus sopusensis*. Topologie rodu *Paracostus* byla nestabilní i po vyřazení

vzorku *Paracostus paradoxus* původem z datasetu (Specht, 2006a), který obsahoval jen krátkou sekvenci, a kde zbytek byla chybějící data. Tato skupina se ve všech analýzách zařadila na bázi ke skupině *Costus*, což je v rozporu s publikovanou prací (Specht, 2006a), kdy skupina *Paracostus* tvořila bázi ke skupinám *Cheilocostus* a *Tapeinochilos*. Největší skupina *Costus* vyšla ve všech analýzách jako nejodvozenější. Všechny taxony odpovídají již dříve publikované práci (Specht 2006a), ale pouze zařazením do jihoamerického rodu *Costus*. Vnitřní vztahy byly velmi málo podpořeny, nebo úplně bez podpory. Důvodem může být jednak velký podíl chybějících dat, tak i fakt, že tato skupina rychle radiovala a vybrané úseky nejsou dostatečně variabilní, aby uspokojivě vysvětlovaly vztahy v rámci této skupiny. Celkově největším problémem všech v této práci provedených analýz z úseku *trnL-F* je nízká statistická podpora větví, což je s velkou pravděpodobností způsobeno velkým podílem chybějících dat, která přesahovala 60 %.

Výsledky fylogenetických analýz založených na NGS datech jsou značně uspokojivější. Získané species trees mají na základě různých přístupů ve všech případech pro jednotlivé rody stejnou topologii, rozdílů jsou patrné až v rámci vztahu na vnitrorodové úrovni, viz přílohy 4 - 11. Tyto výsledky nenaznačují nějakou fylogenetickou souvislost s různými adaptacemi na opylování a je tak v souladu s prací (Specht *et al.*, 2012), která též postuluje homoplasie v přechodu z opylování hmyzen na opylování ptáky.

Zásadním a novým zjištěním je, že rod *Paracostus* je s největší pravděpodobností na bázi celé čeledi. Tento výsledek je v rozporu jak s publikovanou prací Specht (2006a), tak s výsledky získanými na základě úseku *trnL-F*. Co se vnitrorodového vztahu týče se svým disjunktním rozšířením v Africe a Asii se buď může jednat o reliktní druhy, kdy mohl být společný předek pro tento rod široce rozšířen a v současnosti můžeme pozorovat zbytky širšího areálu, nebo se může jednat o jednorázovou disperzi a to buď Americko-Asijskou nebo Asijsko-Americkou. Dále všechny fylogenetické analýzy v této práci naznačují, že rod *Hellenia* není monofyletický, nýbrž se jedná o parafyletickou skupinu, do které je vmezeřen rod *Tapeinochilos*, který tvoří sesterskou skupinu s druhem *Hellenia speciosa* ( $\equiv$  *Cheilocostus speciosus*). Tuto hypotézu potvrzují i testy monofylie, kdy tento stav byl přítomen pro 53 % jaderných genových stromů. Tento výsledek je v rozporu s navrženou fylogenezí Specht (2006), kde je rod *Hellenia* monofyletický. Zajímavým zjištěním též bylo, že stejnou topologii Specht publikovala v práci zabývající se datováním čeledi *Costaceae* (Specht, 2006b), kdy tento fakt nikterak nekommentuje, současně při předběžných testech variability jednotlivých úseků, kdy byla použita data z práce (Specht, 2006a) získaná

z treeBASE, bylo pozorováno stejné pattern. Je tedy otázkou, jak přesně Specht data analyzovala, že výsledky ukazovaly na monofyletičnost rodu *Hellenia*. Co se morfologie rodů *Hellenia* a *Tapeinochilos* týká, je nakloněna spíše hypotéze monofyletického rodu *Hellenia*. Lze najít morfologické podobnosti pro *Hellenia speciosa* a *Hellenia globosa* v jejich rozšíření na asijském kontinentě, kde jsou opylovány včelami. V kontrastu s rodem *Tapeinochilos* s centrem diverzity na Nové Guinei a jejím blízkém okolí, kdy tyto rostliny opylují ptáci. U rodů *Cheilocostus*, *Dimerocostus* a *Monocostus* je pozorována stejná topologie jako v publikaci (Specht, 2006a).

#### 4.4 Velikost genomu

V této diplomové práci naměřené velikosti genomu (2C 1,07 – 9,41 pg) patří podle Soltise *et. al.* (2003) do kategorie velmi malých až malých genomů. Řadí se tak k mnohým zástupcům s podobnou velikostí genomu z čeledi *Zingiberaceae* podčeledi *Zingiberoideae*, jež je charakteristická svou distribucí v tropickém sezonním lese, kde tvoří dormantní stádia. V čeledi *Costaceae* se dormantní stádia nevyskytují, nicméně jsou zde zástupci, kteří se též vyskytují v tropických sezonních lesech a mají sezónně opadavé listy. Naopak druhá podčeleď *Alpinoideae* má signifikantně větší velikosti genomu, přičemž se vyskytuje pouze v tropickém deštném lese (Specht & Stevenson, 2006; Fér *et.al*, unpubl.). Překvapivým zjištěním bylo naměření největších genomových velikostí (2C ~ 5 pg) z celé čeledi u rodu *Paracostus*, který je typický svým vůbec nejmenším vzrůstem ze všech zástupců čeledi *Costaceae* a zahrnuje dva druhy *Paracostus englerianus* vyskytující se v Africe a *Paracostus paradoxus* s typickým výskytem na Borneu. I tímto lze poukázat na fakt, že jsou tyto dva druhy nejspíš blízce příbuzné, i když mají odlišný areál rozšíření. Naopak pro rod *Dimerocostus*, jež zahrnuje zástupce dosahující největšího vzrůstu, byly naměřeny jedny z nejmenších 2C-hodnot (2C ~ 1,7 pg). Toto zjištění naznačuje, že velikost genomu v této čeledi s největší pravděpodobností nesouvisí z velikostí rostlin, ale spíše odpovídá procesům probíhajícím na chromozomální úrovni.

## 4.5 Velikost genomu a fylogeneze

Pro porovnání, zda velikosti genomu koreluje s fylogenezí, byl použit vlastní Dataset, kde byly přiřazeny velikosti genomu pouze k odpovídajícím vzorkům. Zbylé taxony, původem z práce (Specht, 2006a), které byly k tvorbě stromu též použity, aby bylo dosaženo rozčlenění jednotlivých skupin, byly ponechány bez přiřazené 2C-hodnoty, viz příloha 12, protože samotný úsek *trnL-F* neměl dostatečnou výpovědní hodnotu o fylogenetických vztazích v rámci čeledi.

Dále byly velikosti genomu namapovány na species trees na základě konkatenace a ASTRID, navíc byly pro proměnnou vypočteny škálovací parametry. Na strom zkonstruovaný metodou ASTRID byly ještě pomocí ML vypočteny ancestrální genomové velikosti, viz přílohy 13 - 15.

Výsledky ze všech analýz naznačují, že velikost genomu hrála významnou roli v evoluci čeledi *Costaceae*, kdy jednotlivé rody s výjimkou rodu *Costus* vykazovaly vzájemně podobné genomové velikosti a lišily se od rodů blízké příbuzných. Výsledky velikostí genomu pro rod *Costus* naznačují, že původní africké druhy mají větší genomové velikosti a druhy jihoamerické radiace prodělaly redukci ve velikosti genomu. Z výsledků popisujících trendy v evoluci genomu je zřejmé, že se fylogeneze podílela na evoluci tohoto parametru, nicméně další charakteristiky parametrů  $\kappa$  a  $\delta$  nelze uspokojivě interpretovat.

## 4.6 Velikost genomu a počet chromozómů

Z literatury bylo zjištěno poměrně dost údajů o počtech chromozómů pro jednotlivé druhy čeledi *Costaceae* ( $2n=18, 27, 28, 36, 72$ ), (Mahanty, 1970; Mass, 1972; Mass, 1977; Gill, 1978; Subrahmanyam and Khoshoo, 1986; Guerra, 1988; Prathepa, 1995; Vovides & Lascurain, 1995; Edeoga and Okoli, 2000). Při srovnání těchto údajů s fylogenetickými vztahy a zároveň velikostí genomu pro jednotlivé rody, je možno nalézt určité konsekvence. Pro rody *Chamaecostus* a *Tapeinochilos* s jedněmi z nejmenších genomů bylo napočítáno  $2n=18$  čili nejmenší spočtený počet. Pro vzájemně příbuzné rody *Monocostus* a *Dimerocostus* s též malými velikostmi genomu, bylo spočítáno  $2n=27$ . V rodě *Hellenia* byla zjištěna celá škála počtů chromozómů, ačkoli velikosti genomu jsou přibližně konstantní. V tomto případě může jít buď o nedostatečné podchycení celkové variability velikostí genomu, nebo a to je

pravděpodobnější o skutečnost, že extrémně velké počty chromozomů jsou zapříčiněny šlechtěním, jelikož druh *Hellenia speciosa* je velmi hojně pěstována a splňuje zpět do volné přírody. I díky tomuto faktu má dnes tento druh o hodně větší areál výskytu, než byly původní lokality. Například přítomnost tohoto druhu v Jižní Americe může být důsledkem zavlečení nikoli důsledkem disperze. Pro rod *Paracostus* s největší naměřenou velikostí genomu bylo zjištěno  $2n=36$ , což odpovídá tetraploidní úrovni. Takto velký genom může být výhodný na místech, kde je velké zastínění, jelikož je takto efektivnější fotosyntéza, což přesně tento rod splňuje. Pro rod *Costus* bylo zjištěno  $2n=18, 27, 36$  a velká škála velikostí genomu, která by mohla v některých případech odpovídat jednotlivým počtům chromozomů. Např. pro *Costus afer* byla  $2C$ -hodnota  $\sim 4,8$  pg a zaznamenáno  $2n=18, 36$ . Teoreticky je to možné, protože se jedná o nejmenší druhy čeledi *Costaceae*.

## ZÁVĚR

Výsledky fylogenetických analýz ukazují na zásadní rozdíly v chápání příbuzenských vztahů čeledi *Costaceae*, kdy vychází jako bazální skupina rod *Paracostus* a asijský rod *Cheilocostus* se jeví být parafyletický, kdy se do něj vmezeřil rod *Tapeinochilos*.

Z výsledků měření absolutní velikosti genomu bylo zjištěno, že zástupci čeledi *Costaceae* mají spíše malý genom a jednotlivé 2C-hodnoty se pohybují v rozmezí 1,07 - 9,41 pg. Naměřené hodnoty byly namapovány na fylogenetické stromy, kde se ukázalo, že velikost genomu celkem dobře odráží fylogenetické vztahy v rámci této čeledi. U jednotlivých druhů v rámci rodu je velikost genomu srovnatelná, oproti velikostem genomu mezi jednotlivými rody se liší. Největší velikost genomu byla překvapivě naměřená v rodě *Paracostus*, který je charakteristický svým velmi malým vzrůstem, oproti zbylým rodům v rámci této čeledi. Jedna z nejmenších velikostí genomu byla naměřena v rodě *Dimerocostus*, který dosahuje nejvyššího vzrůstu.

## SEZNAM SYMBOLŮ A ZKRATEK

BI	Bayesian inference, Bayesovská analýza
bp	base pair, komplementární páry bází
GBIF	Global Biodiversity Information Facility
GTR	Generalised time-reversible
Hyb-Seq	Sequencing using Hybridization-based enrichment, Sekvenování, kdy pro tvorbu obohacených knihoven bylo použito hybridizace
HGT	Horizontal gene transfer
ILS	Incomplete lineage sorting
ITS	Internal transcribed spacer, vnitřní přepisovaný spacer
LSC	Long single copy (region)
ML	Maximum likelihood, maximální věrohodnost
m.n.m	metrů nad mořem
MP	Maximum parsimony, maximální parsimonie
NGS	Next-Generation Sequencing, sekvenování nové generace (sekvenování další generace)
treeBASE	databáze (A Database of Phylogenetic Knowledge)
<i>trnL-trnF</i>	<i>trnL</i> intron, 3' <i>trnL</i> exon a <i>trnL-F</i> intergenový spacer

## POUŽITÁ LITERATURA

- Bayzid MS, Warnow T. 2012.** Estimating Optimal Species Trees from Incomplete Gene Trees Under Deep Coalescence. *Journal of Computational Biology* **19**: 591–605.
- Beaulieu JM, Leitch IJ, Patel S, Pendharkar A, Knight C a. 2008a.** Genome size is a strong predictor of cell size and stomatal density in angiosperms. *New Phytologist* **179**: 975–986.
- Beaulieu JM, Leitch IJ, Patel S, Pendharkar A, Knight CA. 2008b.** Genome size is a strong predictor of cell size and stomatal density in angiosperms. *New Phytologist* **179**: 975–986.
- Beaulieu JM, Moles AT, Leitch IJ, Bennett MD, Dickie JB, Knight CA. 2007.** Correlated evolution of genome size and seed mass. *New Phytologist* **173**: 422–437.
- Bennett MD. 1987.** Variation in genomic form in plants and its ecological implications. *New phytologist* **106**: 177–200.
- Bennett MD, Bhandol P, Leitch IJ. 2000.** Nuclear {DNA} amounts in angiosperms and their modern uses—807 new estimates. *Annals of Botany* **86**: 859–909.
- Bennetzen JL. 2000.** Transposable element contributions to plant gene and genome evolution. *Plant Molecular Biology* **42**: 251–269.
- Bennetzen JL, Kellogg EA. 1997.** Do plants have a one-way ticket to genomic obesity? *The Plant Cell* **9**: 1509.
- Bolger AM, Lohse M, Usadel B. 2014.** Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114–2120.
- Cronn R, Knaus BJ, Liston A, Maughan PJ, Parks M, Syring J V, Udall J. 2012.** Targeted enrichment strategies for next-generation plant biology. *American Journal of Botany* **99**: 291–311.
- Degnan JH, Rosenberg NA. 2009.** Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology and Evolution* **24**: 332–340.
- Devos KM, Brown JKM, Bennetzen JL. 2002.** Genome size reduction through illegitimate recombination counteracts genome expansion in Arabidopsis. *Genome Research* **12**: 1075–1079.
- Dolezel J, Bartos J, Voglmayr H, Greilhuber J. 2003.** Nuclear {DNA} content and genome size of trout and human. *Cytometry. Part A: The Journal of the International Society for Analytical Cytology* **51**: 127–128; author reply 129.

- Doležel J, Greilhuber J, Suda J. 2007.** Estimation of nuclear {DNA} content in plants using flow cytometry. *Nature protocols* **2**: 2233–2244.
- Dutta S.** Comment on the identity of *Hellenia speciosa* (J. Koenig ex Smith) S. Dutta & *Hellenia speciosa* (J. Koenig ex Smith) S. Dutta var *dilnavazii* (MR Almeida & SM Almeida) S. Dutta.
- Edeoga HO, Okoli BE. 2000.** Chromosome numbers of *Costus lucanusianus* (Costaceae) in Nigeria. *Folia Geobotanica* **35**: 315–318.
- Gill LS. 1978.** Chromosome numbers of angiosperms in Tanzania II. *Adansonia* **18**: 19–24.
- Gnrke A, Melnikov A, Maguire J, Rogov P, LeProust EM, Brockman W, Fennell T, Giannoukos G, Fisher S, Russ C, et al. 2009.** Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nature biotechnology* **27**: 182–9.
- Gordon A, Hannon GJ. 2010.** Fastx-toolkit. FASTQ/A short-reads pre-processing tools. Unpublished [http://hannonlab.cshl.edu/fastx\\_toolkit](http://hannonlab.cshl.edu/fastx_toolkit).
- Govaerts R. 2013.** *Hellenia* Retz., the correct name for *Cheilocostus* CD Specht (Costaceae). *Phytotaxa* **151**: 63–64.
- Greilhuber J. 2005.** Intraspecific variation in genome size in angiosperms: Identifying its existence. *Annals of Botany* **95**: 91–98.
- Grotkopp E, Rejmánek M, Sanderson MJ, Rost TL. 2004.** EVOLUTION OF GENOME SIZE IN PINES (PINUS) AND ITS LIFE-HISTORY CORRELATES: SUPERTREE ANALYSES. *Evolution* **58**: 1705–1729.
- Grover CE, Wendel JF. 2010.** Recent insights into mechanisms of genome size change in plants. *Journal of Botany* **2010**.
- Hall T. 1999.** BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41**: 95–98.
- Heled J, Drummond AJ. 2010.** Bayesian Inference of Species Trees from Multilocus Data. *Molecular Biology and Evolution* **27**: 570–580.
- Herben T, Suda J, Klimešová J, Mihulka S, \vR'iha P, Š'imová I. 2012.** Ecological effects of cell-level processes: genome size, functional traits and regional abundance of herbaceous plant species. *Annals of botany*: mcs099.
- Hodgson JG, Sharafi M, Jalili A, D'iaz S, Montserrat-Mart'i G, Palmer C, Cerabolini B, Pierce S, Hamzehee B, Asri Y, et al. 2010.** Stomatal vs. genome size in angiosperms: the somatic tail wagging the genomic dog? *Annals of Botany* **105**: 573–584.
- Huang H, He Q, Kubatko LS, Knowles LL. 2010.** Sources of error inherent in species-tree

estimation: Impact of mutational and coalescent effects on accuracy and implications for choosing among different methods. *Systematic Biology* **59**: 573–583.

**Huelsenbeck JP, Crandall KA. 1997.** Phylogeny Estimation and Hypothesis Testing Using Maximum Likelihood. *Annual Review of Ecology and Systematics* **28**: 437–466.

**Huelsenbeck JP, Ronquist F. 2001.** MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics (Oxford, England)* **17**: 754–755.

**K. H. 2014.** The evolution of {Elettariopsis} ({Zingiberaceae}).

**Katoh K, Standley DM. 2013.** MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution* **30**: 772–780.

**Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, et al. 2012.** Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**: 1647–1649.

**Kent WJ. 2002.** BLAT - The BLAST-like alignment tool. *Genome Research* **12**: 656–664.

**Knight CA, Ackerly DD. 2002.** Variation in nuclear DNA content across environmental gradients: A quantile regression analysis. *Ecology Letters* **5**: 66–76.

**Knight CA, Beaulieu JM. 2008.** Genome size scaling through phenotype space. *Annals of Botany* **101**: 759–766.

**Knight CA, Molinari NA, Petrov DA. 2005.** The large genome constraint hypothesis: evolution, ecology and phenotype. *Annals of Botany* **95**: 177–190.

**Kress WJ. 1990.** The Phylogeny and Classification of the Zingiberales. *Annals of the Missouri Botanical Garden* **77**: 698–721.

**Küeck P, Koenig A. 2014.** FASconCAT-G\_v1.02.pl.

**Larsen K. 1998.** Flowering Plants {·} Monocotyledons: Alismatanae and Commelinanae (except Gramineae). In: Kubitzki K, ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 128–132.

**Leitch IJ, Bennett MD. 2007.** Genome size and its uses: the impact of flow cytometry. *Flow cytometry with plant cells: analysis of genes, chromosomes and genomes*: 153–176.

**Leitch IJ, Chase MW, Bennett MD. 1998.** Phylogenetic Analysis of DNA C-values Provides Evidence for a Small Ancestral Genome Size in Flowering Plants. *Annals of Botany* **82**: 85–94.

**Leitch AR, Leitch IJ. 2012.** Ecological and genetic factors linked to contrasting genome dynamics in seed plants. *New Phytologist* **194**: 629–646.

- Leitch AR, Leitch IJ, Trimmer M, Guignard MS, Woodward G. 2014.** Impact of genomic diversity in river ecosystems. *Trends in Plant Science* **19**: 361–366.
- Levin DA, Funderburg SW. 1979.** Genome size in angiosperms: temperate versus tropical species. *American Naturalist*: 784–795.
- Liu L. 2008.** BEST: Bayesian estimation of species trees under the coalescent model. *Bioinformatics* **24**: 2542–2543.
- Liu L, Yu L. 2011.** Estimating species trees from unrooted gene trees. *Systematic Biology* **60**: 661–667.
- Liu L, Yu L, Edwards S V. 2010.** A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. *BMC evolutionary biology* **10**: 302.
- Liu L, Yu L, Pearl DK, Edwards S V. 2009.** Estimating species phylogenies using coalescence times among sequences. *Systematic Biology* **58**: 468–477.
- M P. 2012.** Evoluční velikosti genomu v rodě {Globba} ({Zingiberaceae}).
- Maddison W. 1997.** Gene trees in species trees. *Systematic Biology* **46**: 523–536.
- Mahanty HK. 1970.** A Cytological Special Study Reference of the to Zingiberales with Their Introduction Taxonomic position : According to Bentham and Hooker ' s General Plantarum ( 1883 ) the Zingiberales as a whole were considered to be one family among the Monocotyledon.
- Marhold K, Suda J. 2001.** Analýza multivariačních dat v taxonomii (Fenetické metody). **2**: 1 – 134.
- Mass PJM. 1972.** Costoideae (Zingiberaceae) Flora Neotropica Monographs 8. : 1–140.
- Mass PJM. 1977.** *Flora Neotropica*. New York: The New York Botanical Garden.
- Masterson J. 1994.** Stomatal size in fossil plants: evidence for polyploidy in majority of angiosperms. *Science (New York, N.Y.)* **264**: 421–424.
- Mayol M, Rosselló JA. 2001.** Why nuclear ribosomal DNA spacers (ITS) tell different stories in Quercus. *Molecular Phylogenetics and Evolution* **19**: 167–176.
- Metzker ML. 2010.** Sequencing technologies - the next generation. *Nature reviews. Genetics* **11**: 31–46.
- Mirarab S, Reaz R, Bayzid MS, Zimmermann T, S. Swenson M, Warnow T. 2014.** ASTRAL: Genome-scale coalescent-based species tree estimation. *Bioinformatics*.
- Mirarab S, Warnow T. 2015.** ASTRAL-II: Coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. *Bioinformatics*.i44–i52.
- Morgan HD, Westoby M. 2005.** The relationship between nuclear DNA content and leaf

strategy in seed plants. *Annals of Botany* **96**: 1321–1330.

**Nguyen N, Mirarab S, Warnow T. 2012.** MRL and SuperFine+MRL: new supertree methods. *Algorithms for Molecular Biology* **7**: 3.

**Nieto Feliner G, Rosselló JA. 2007.** Better the devil you know? Guidelines for insightful utilization of nrDNA ITS in species-level evolutionary studies in plants. *Molecular Phylogenetics and Evolution* **44**: 911–919.

**Ohri D. 2005.** Climate and growth form: the consequences for genome size in plants. *Plant Biology* **7**: 449–458.

**Orme CDL, Freckleton RP, Thomas GH, Petzoldt T, Fritz SA, Isaac NJB. 2012.** CAPER: Comparative Analyses of Phylogenetics and Evolution in R. *Methods in Ecology and Evolution* **3**: 145–151.

**Philippe H, Snell EA, Baptiste E, Lopez P, Holland PWH, Casane D. 2004.** Phylogenomics of eukaryotes: Impact of missing data on large alignments. *Molecular Biology and Evolution* **21**: 1740–1752.

**Price MN, Dehal PS, Arkin AP. 2009.** Fasttree: Computing large minimum evolution trees with profiles instead of a distance matrix. *Molecular Biology and Evolution* **26**: 1641–1650.

**Price MN, Dehal PS, Arkin AP. 2010.** FastTree 2--approximately maximum-likelihood trees for large alignments. *PloS one* **5**: e9490.

**Puttick MN, Clark J, Donoghue PCJ. 2015.** Size is not everything: rates of genome size evolution, not C-value, correlate with speciation in angiosperms. *Proceedings. Biological sciences / The Royal Society* **282**: 20152289–.

**Rejmánek M. 1996.** A theory of seed plant invasiveness: The first sketch. *Biological Conservation* **78**: 171–181.

**Rejmánek M. 2000.** Invasive plants: approaches and predictions. *Austral Ecology* **25**: 497–506.

**Revell LJ. 2012.** phytools: an {R} package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* **3**: 217–223.

**dos Santos Guerra M. 1988.** Characterization of different types of condensed chromatin in *Costus* (Zingiberaceae). *Plant systematics and evolution* **158**: 107–115.

**Shendure J, Ji H. 2008.** Next-generation DNA sequencing. *Nat Biotechnol.* 2008;26(10):1135–45. quencing. *Nature biotechnology* **26**: 1135–1145.

**Shimodaira H. 2001.** Multiple comparisons of log-likelihoods and combining nonnested

models with applications to phylogenetic tree selection. *COMMUNICATIONS IN STATISTICS-THEORY AND METHODS* **30**: 1751–1772.

**Schemske DW. 1980.** The evolutionary significance of extrafloral nectar production by *Costus woodsonii* (Zingiberaceae): an experimental analysis of ant protection. *The Journal of Ecology*: 959–967.

**Schmickl R, Liston A, Zeisek V, Oberlander K, Weitemier K, Straub SCK, Cronn RC, Dreyer LL, Suda J. 2015.** Phylogenetic marker development for target enrichment from transcriptome and genome skim data: The pipeline and its application in southern African *Oxalis* (Oxalidaceae). *Molecular Ecology Resources*.

**Smith S a, Donoghue MJ. 2008.** Rates of molecular evolution are linked to life history in flowering plants. *Science (New York, N.Y.)* **322**: 86–89.

**Soltis DE, Soltis PS, Bennett MD, Leitch IJ. 2003.** Evolution of genome size in the angiosperms. *American Journal of Botany* **90**: 1596–1603.

**Specht CD. 2006a.** Systematics and Evolution of the Tropical Monocot Family Costaceae (Zingiberales): A Multiple Dataset Approach. *Systematic Botany* **31**: 89–106.

**Specht CD. 2006b.** Gondwanan vicariance or dispersal in the tropics? The biogeographic history of the tropical monocot family Costaceae (Zingiberales). *Aliso* **22**: 631–642.

**Specht CD, Kress WJ, Stevenson DW, DeSalle R. 2001.** A molecular phylogeny of Costaceae (Zingiberales). *Molecular Phylogenetics and Evolution* **21**: 333–345.

**Specht CD, Stevenson DW. 2006.** A new phylogeny-based generic classification of Costaceae (Zingiberales). *Systematic Botany* **31**: 153–163.

**Specht, Yockteng R, Almeida A, Kirchoff B, Kress J, Kress J. 2012.** Homoplasy, Pollination, and Emerging Complexity During the Evolution of Floral Development in the Tropical Gingers (Zingiberales). *Botanical Review* **78**: 440–462.

**Stamatakis A. 2014.** RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics (Oxford, England)* **30**: 1312–1313.

**Straub SCK, Parks M, Weitemier K, Fishbein M, Cronn RC, Liston A. 2012.** Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. *American Journal of Botany* **99**: 349–364.

**Subrahmanyam G V, Khoshoo TN. 1986.** Cytological Studies in the Genus *Costus* L. India *Costus*, a genus of perennial rhizomatous herbs, is pan-tropical in distribution, although the majority of its species are concentrated in Africa and especially America (Holttum 1950). There are about 100 species: 737–748.

- Suda J, Meyerson LA, Leitch IJ, Pyšek P. 2015.** The hidden side of plant invasions: the role of genome size. *New Phytologist* **205**: 994–1007.
- Swift H. 1950.** The {Constancy} of {Desoxyribose} {Nucleic} {Acid} in {Plant} {Nuclei}. *Proceedings of the National Academy of Sciences* **36**: 643–654.
- Swofford DL. 2003.** PAUP\*: phylogenetic analysis using parsimony, version 4.0b10. *21 Libro*: 11pp.
- Szöllosi GJ, Daubin V. 2012.** Modeling gene family evolution and reconciling phylogenetic discord. *Methods in Molecular Biology* **856**: 29–51.
- Šimová I, Herben T. 2012.** Geometrical constraints in the scaling relationships between genome size, cell size and cell cycle length in herbaceous plants. *Proceedings of the Royal Society B: Biological Sciences* **279**: 867–875.
- Šmarda P, Hejzman M, Březinová A, Horová L, Steigerová H, Zedek F, Bureš P, Hejzmanová P, Schellberg J. 2013.** Effect of phosphorus availability on the selection of species with different ploidy levels and genome sizes in a long-term grassland fertilization experiment. *New Phytologist* **200**: 911–921.
- Thomas CA. 1971.** The genetic organization of chromosomes. *Annual Review of Genetics* **5**: 237–256.
- Thompson K. 1990.** Genome Size Seed Size and Germination Temperature in Herbaceous Angiosperms. *Evolutionary Trends in Plants* **4**: 113–116.
- Thomson RC, Shaffer HB. 2010.** Sparse supermatrices for phylogenetic inference: Taxonomy, alignment, rogue taxa, and the phylogeny of living turtles. *Systematic Biology* **59**: 42–58.
- Tomáš Fér. 2015.** HybPipe - Pipeline for Hyb-Seq data processing and tree building.
- Tonini J, Moore A, Stern D. 2015.** Concatenation and species tree methods exhibit statistically indistinguishable accuracy under a range of simulated conditions. *PLOS Currents Tree of Life* **7**: doi: 10.1371/currents.tol.34260cc27551a527b124ec5f.
- Vachaspati P, Warnow T. 2015.** ASTRID: Accurate Species TRees from Internode Distances. *BMC Genomics* **16**: 1–13.
- Vovides AP, Lascurain M. 1995.** Números cromosómicos de cuatro especies de {Costus} ({Costaceae}), una de {Calathea}, una de {Maranta} y una de {Stromanthe} ({Marantaceae}). *Acta Botánica Mexicana*.
- Weitemier K, Straub SCK, Cronn RC, Fishbein M, Schmickl R, McDonnell A, Liston A. 2014.** Hyb-Seq: Combining Target Enrichment and Genome Skimming for Plant

Phylogenomics. *Applications in Plant Sciences* **2**: 1400042.

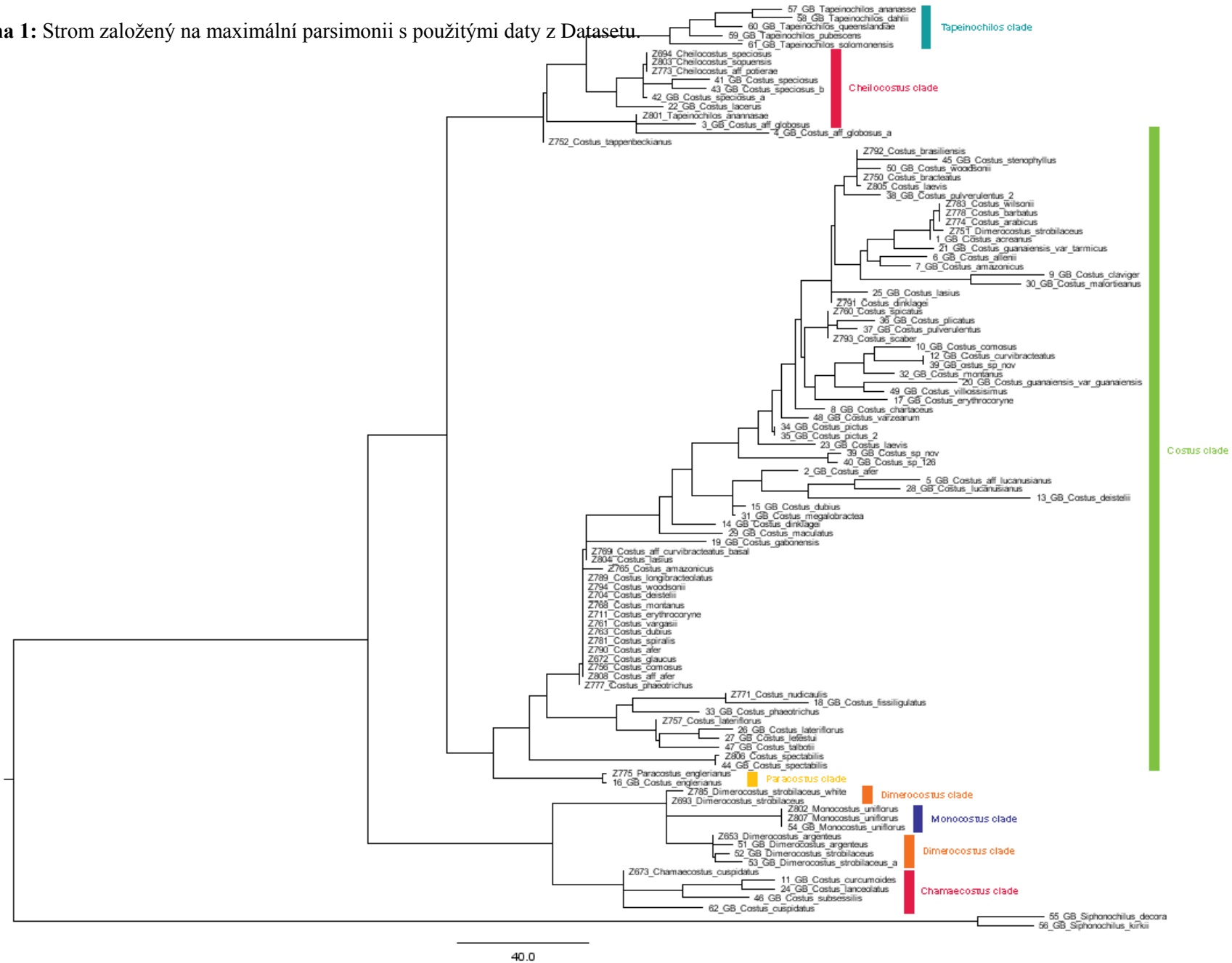
**Wendel JF, Jackson SA, Meyers BC, Wing RA. 2016.** Evolution of plant genome architecture. *Genome biology* **17**: 1.

**White TJ, Bruns S, Lee S, Taylor J. 1990.** Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *PCR Protocols: A Guide to Methods and Applications*: 315–322.

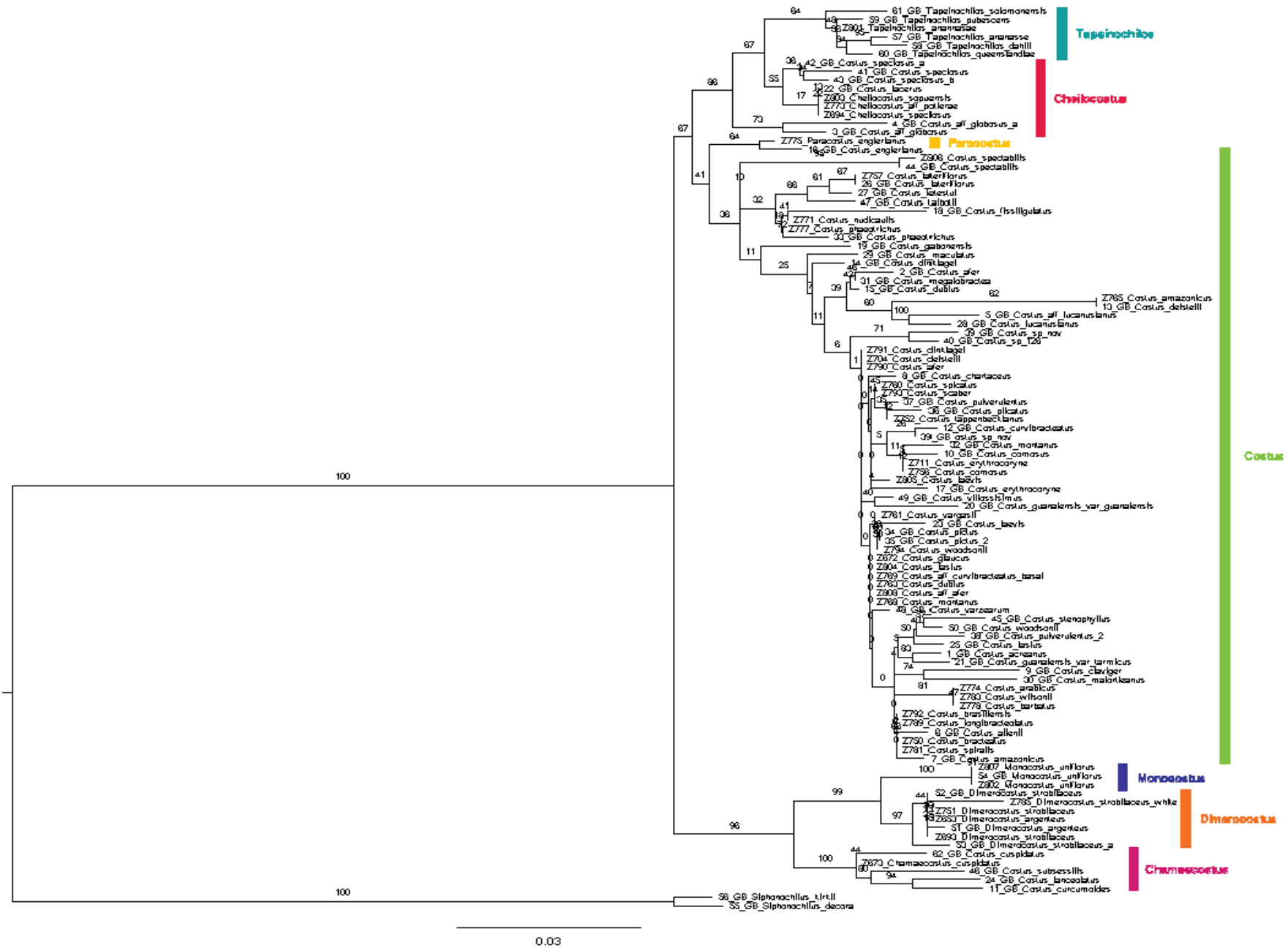
**Xi Z, Liu L, Davis CC. 2015.** The impact of missing data on species tree estimation. *Molecular biology and evolution*: msv266–.

# PŘÍLOHY

Příloha 1: Strom založený na maximální parsimonii s použitými daty z Datasetu.

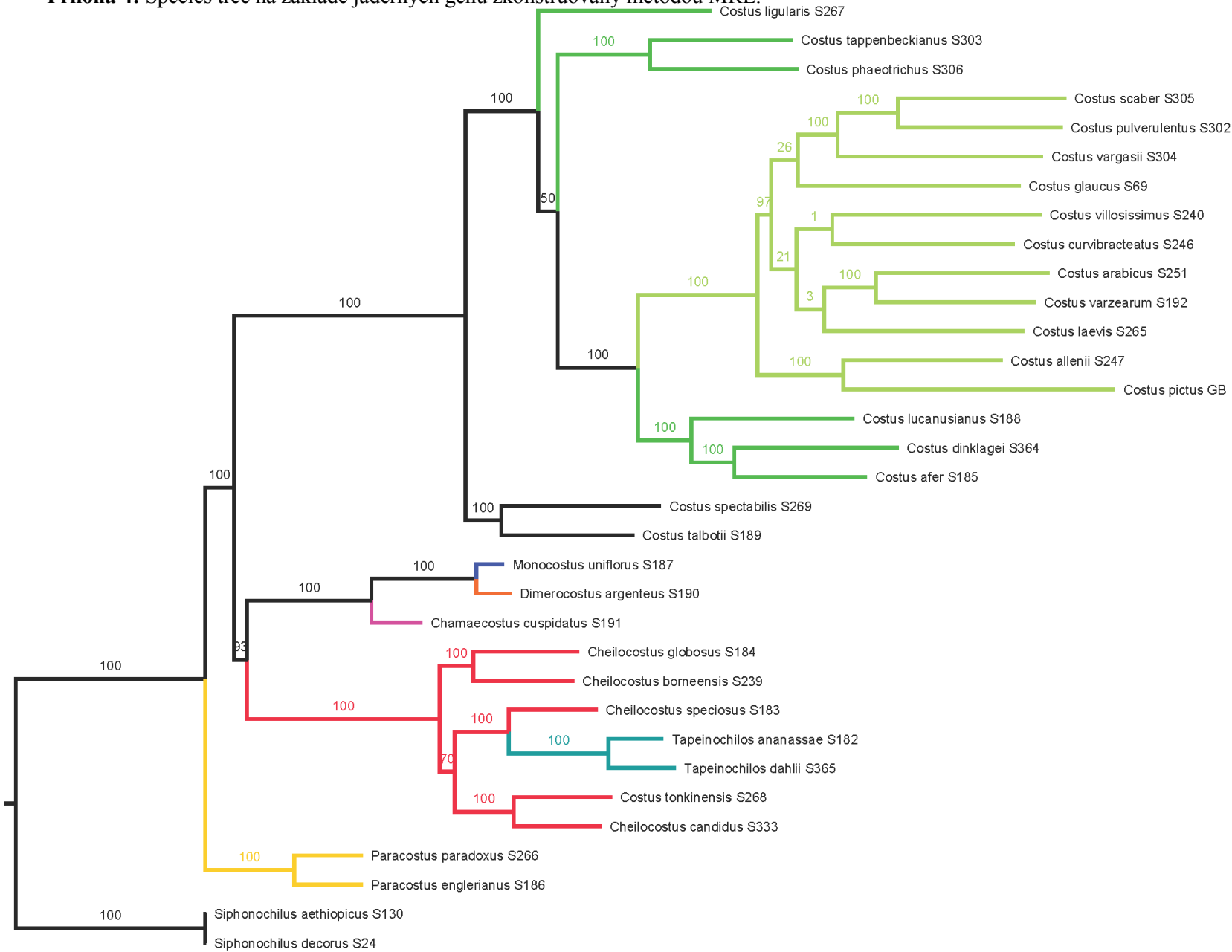


**Příloha 2:** Strom založený na maximum likelihood s použitými daty z Datasetu.



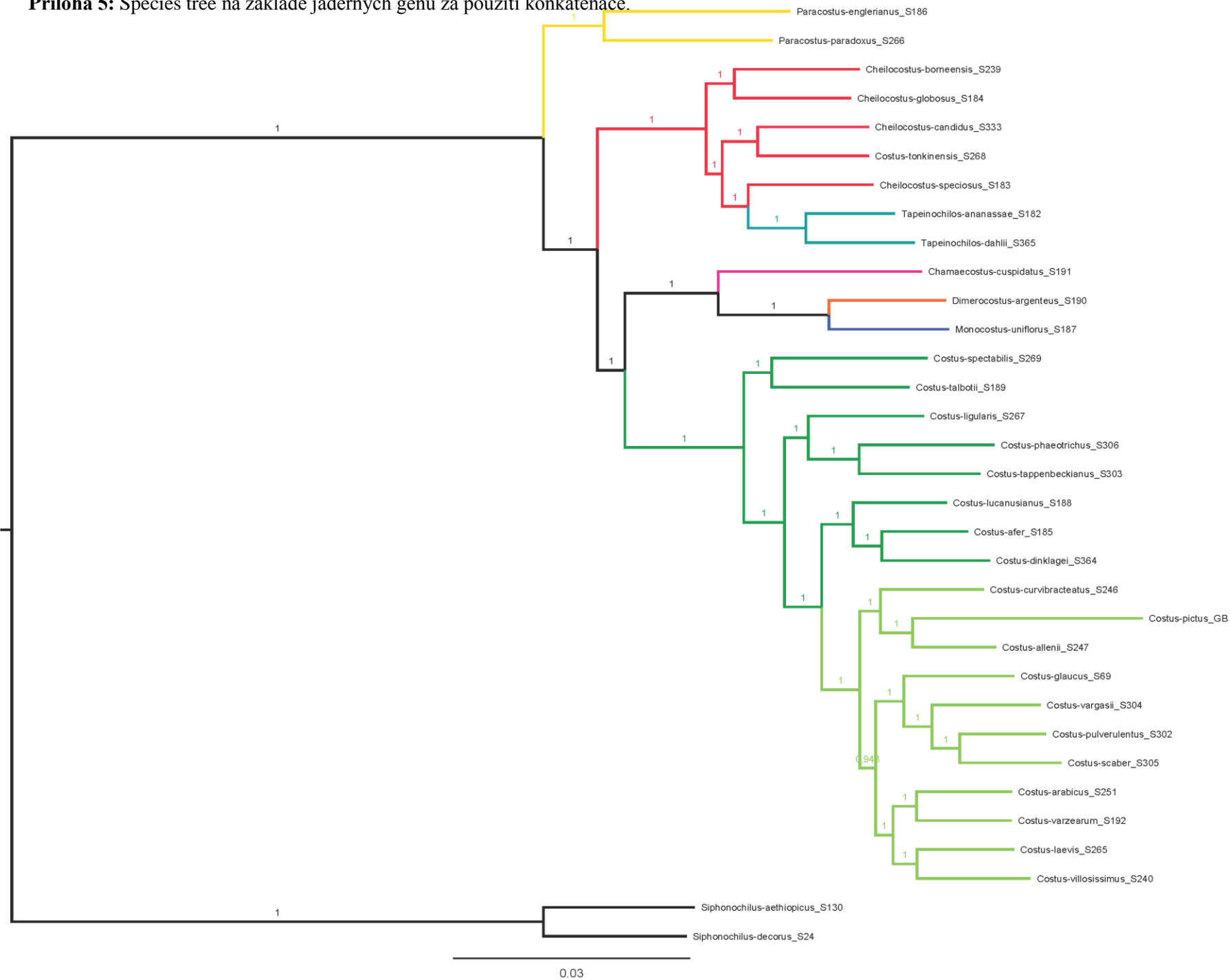


**Příloha 4:** Species tree na základě jaderných genů zkonstruovaný metodou MRL.

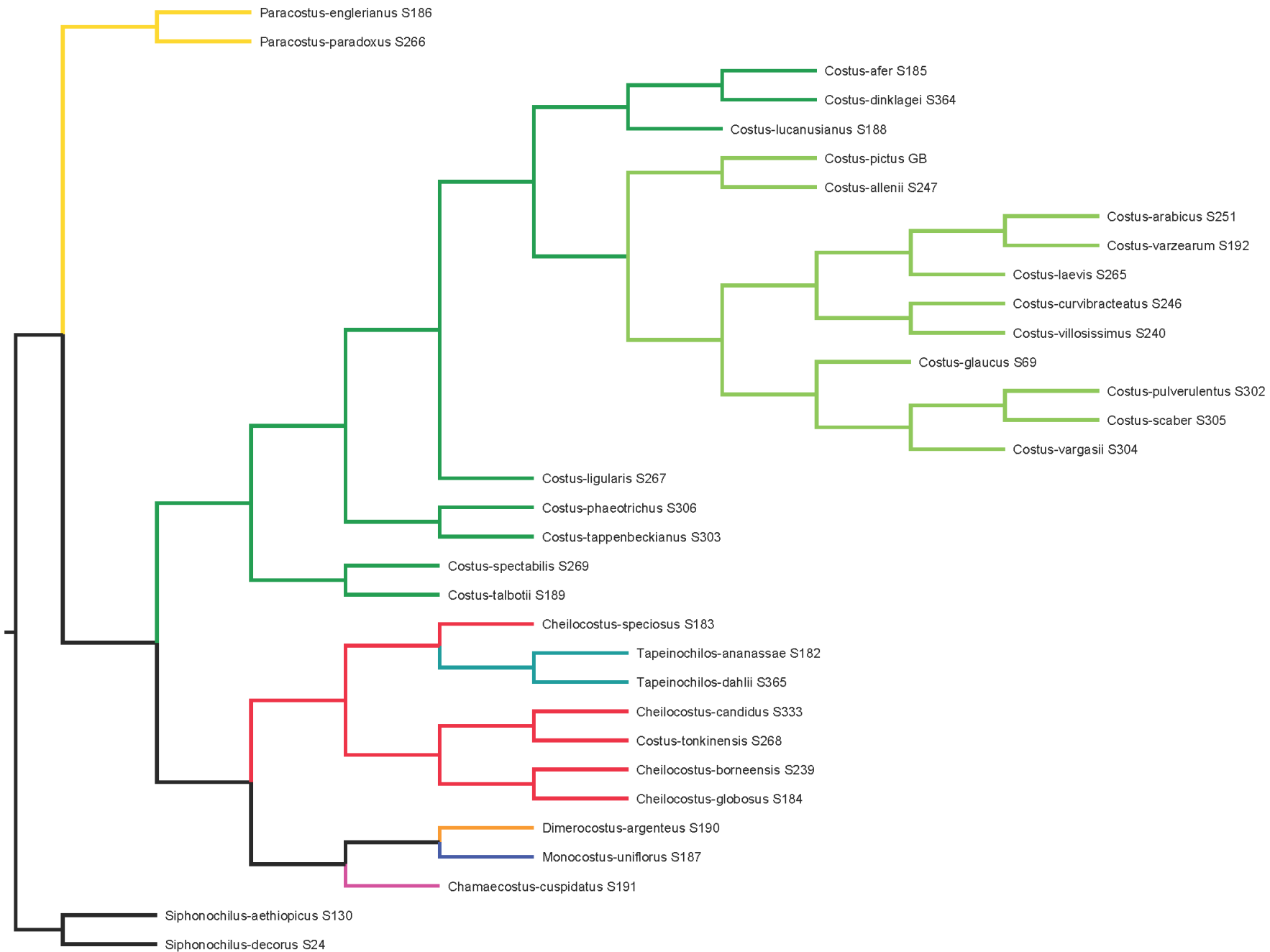


0.06

# Příloha 5: Species tree na základě jaderných genů za použití konkaténace.



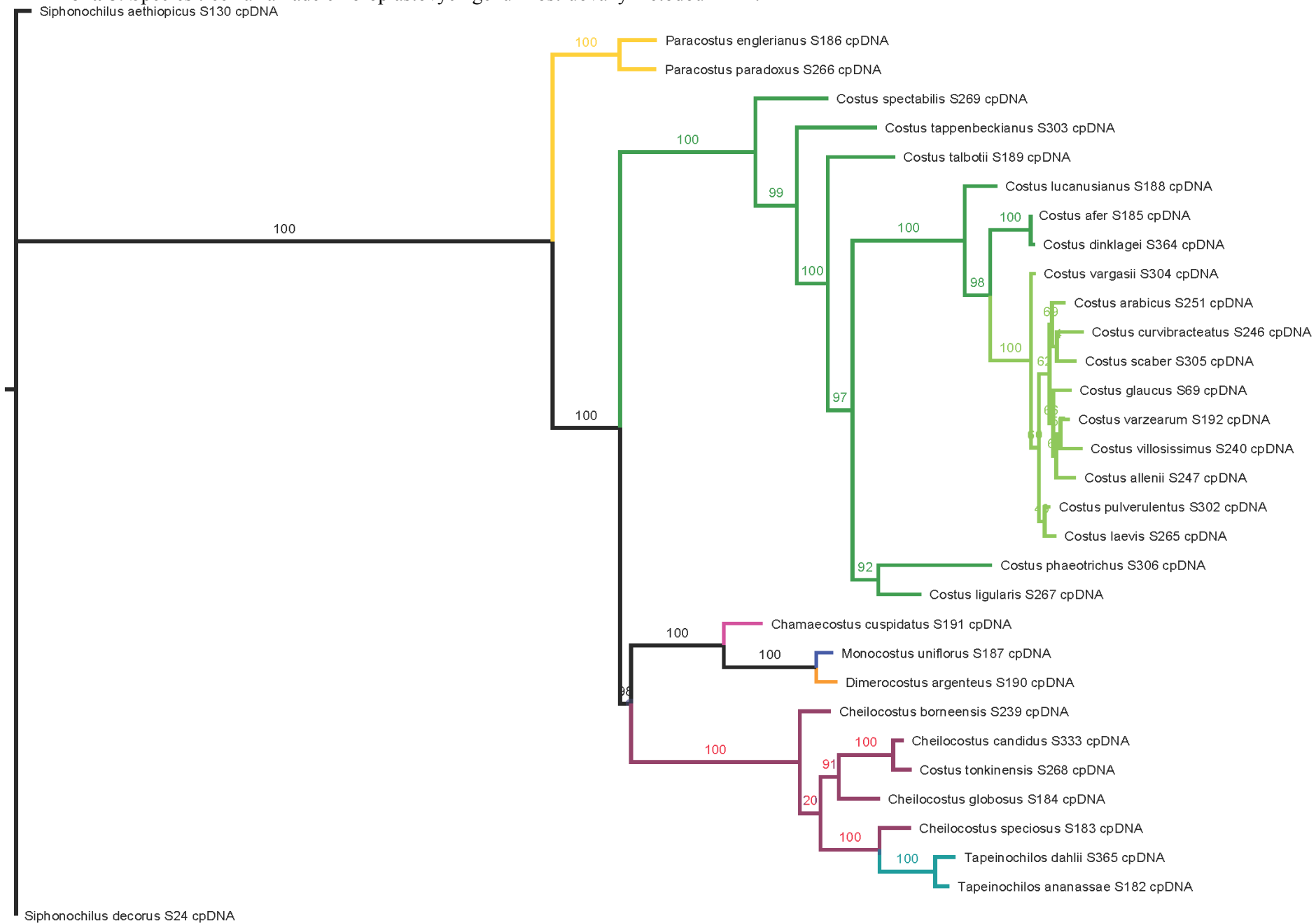
**Příloha 6:** Species tree na základě jaderných genů zkonstruován metodou ASTRID.



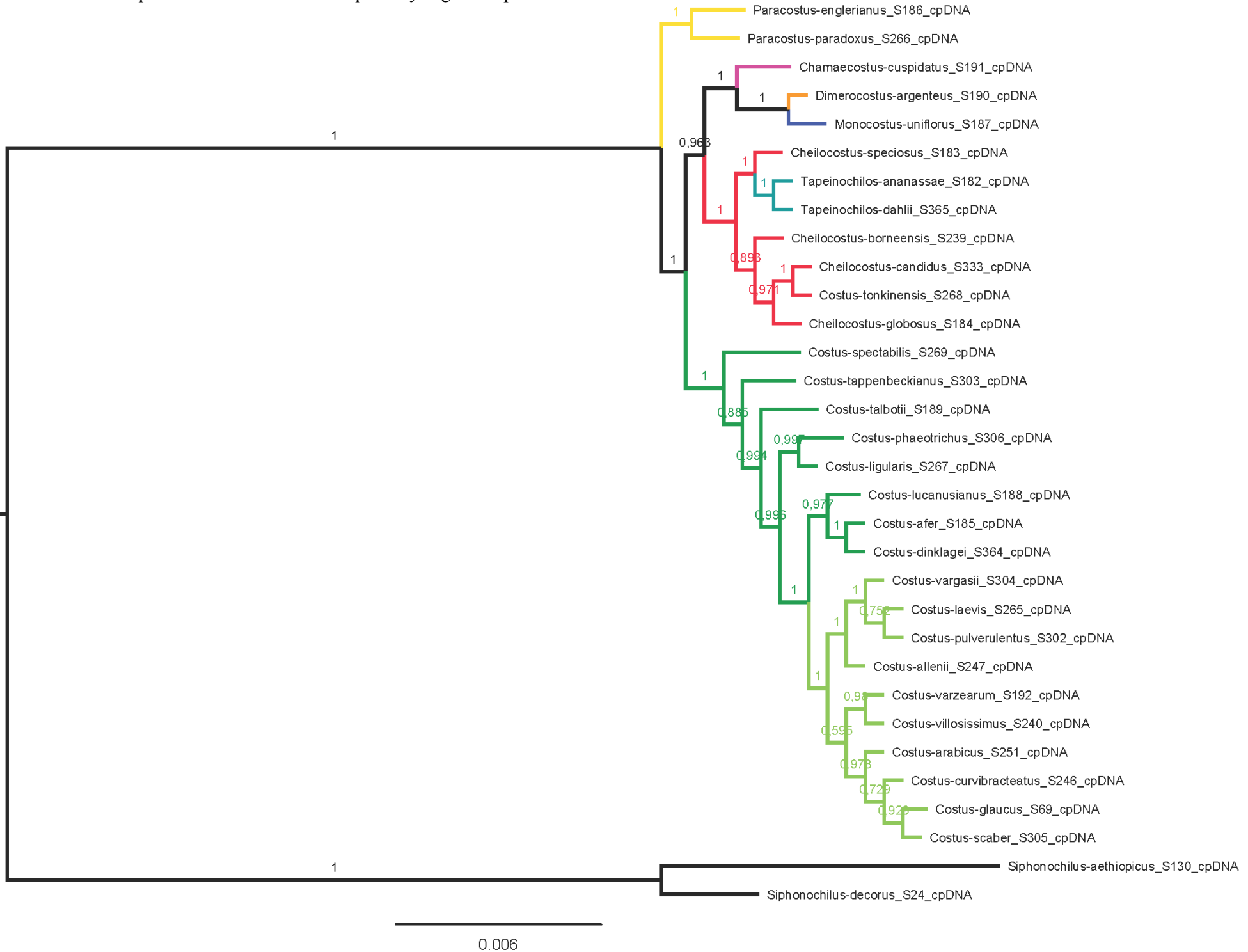
2.0



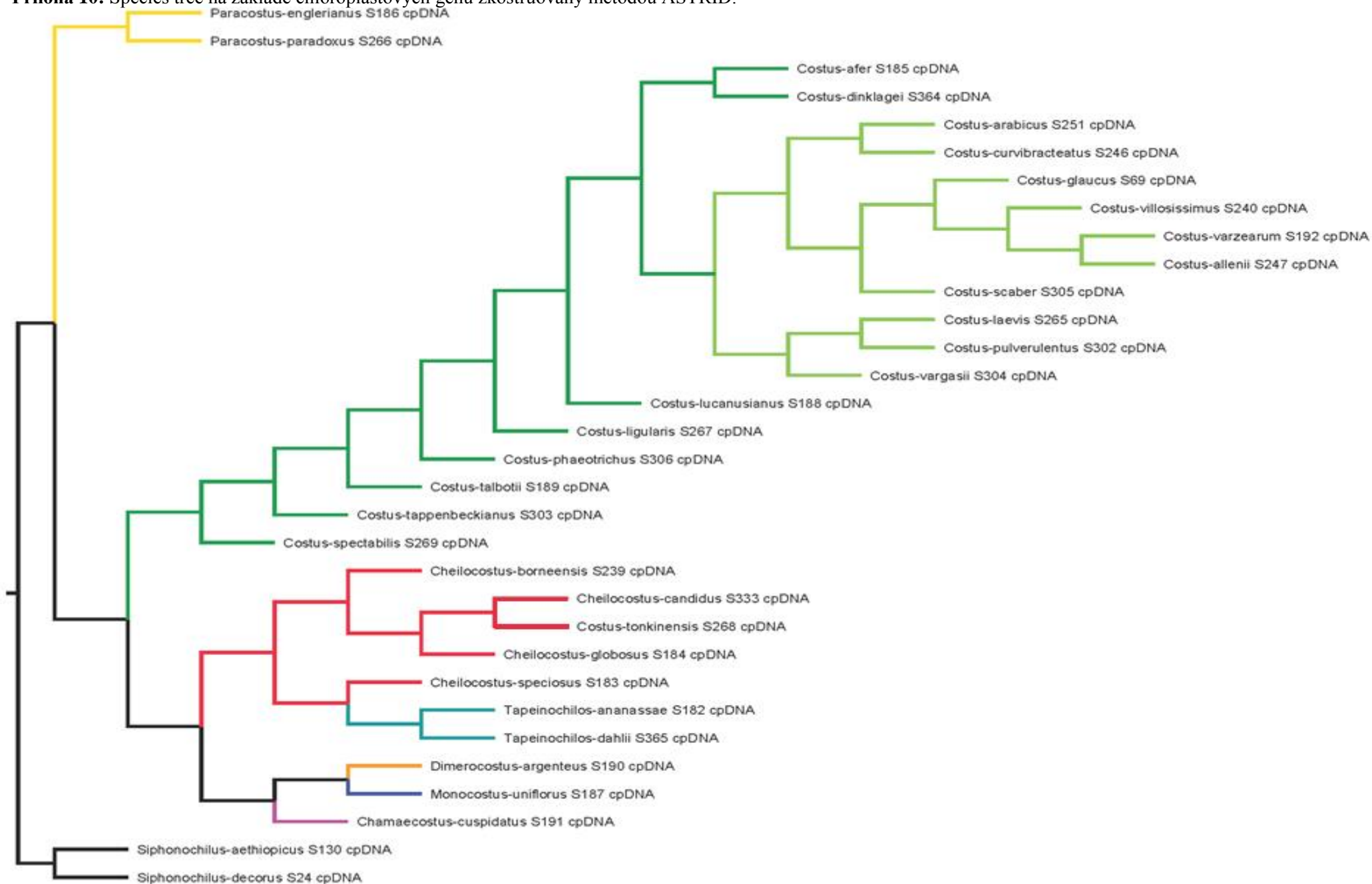
**Příloha 8:** Species tree na základě chloroplastových genů zkonstruovaný metodou MRL.



**Příloha 9:** Species tree na základě chloroplastových genů za použití konkaténace.



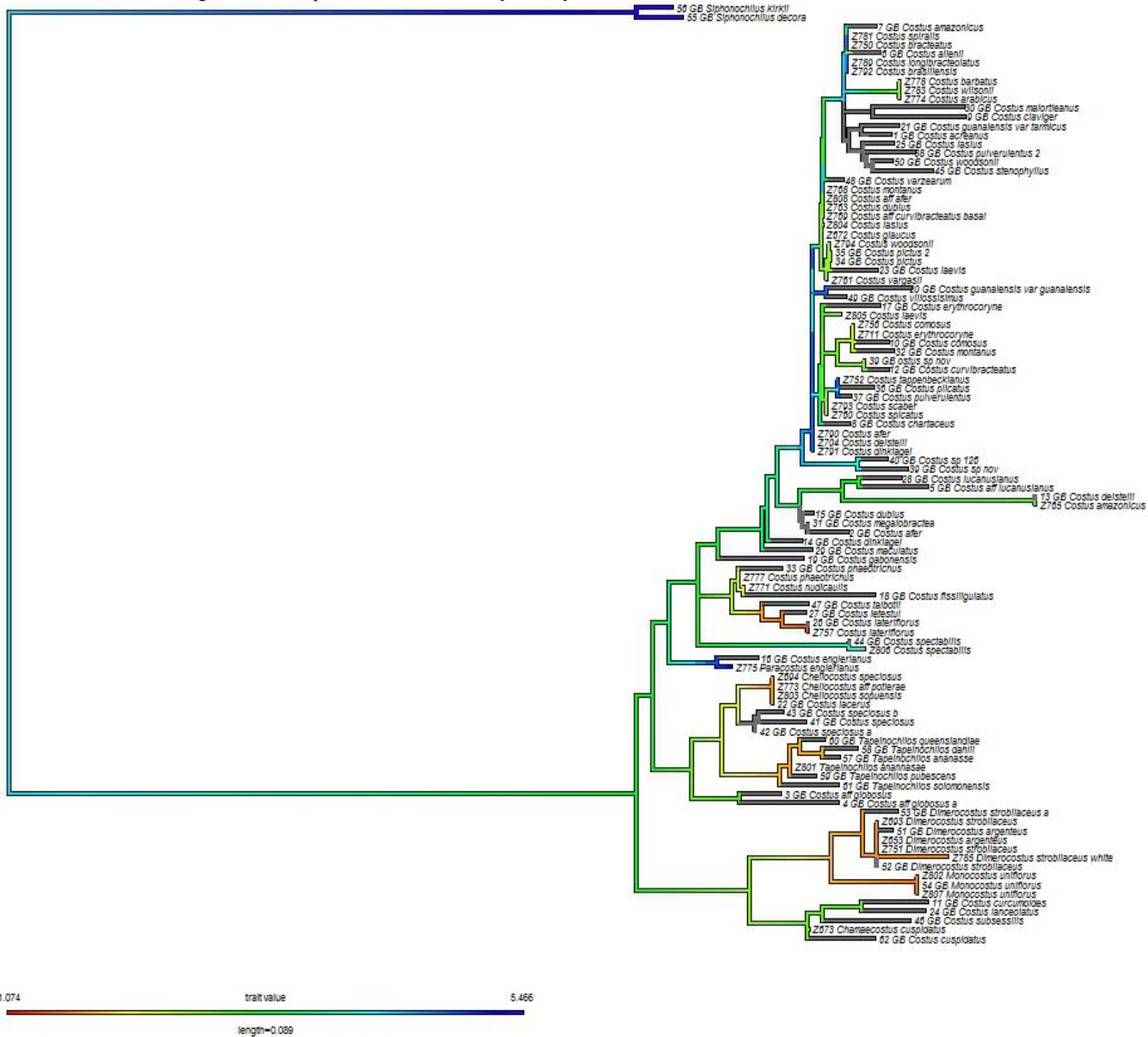
**Příloha 10:** Species tree na základě chloroplastových genů zkonstruovaný metodou ASTRID.



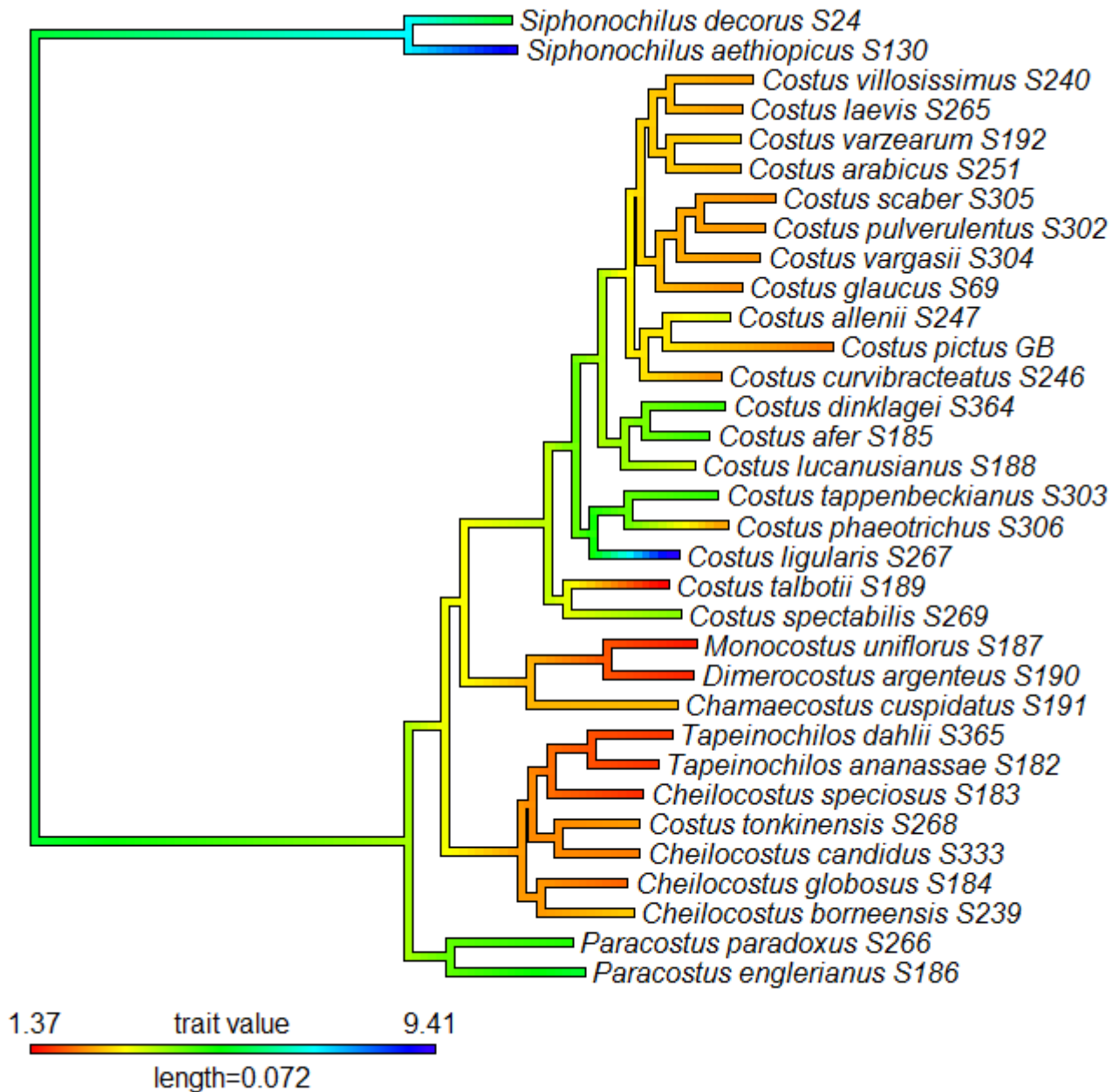
2.0



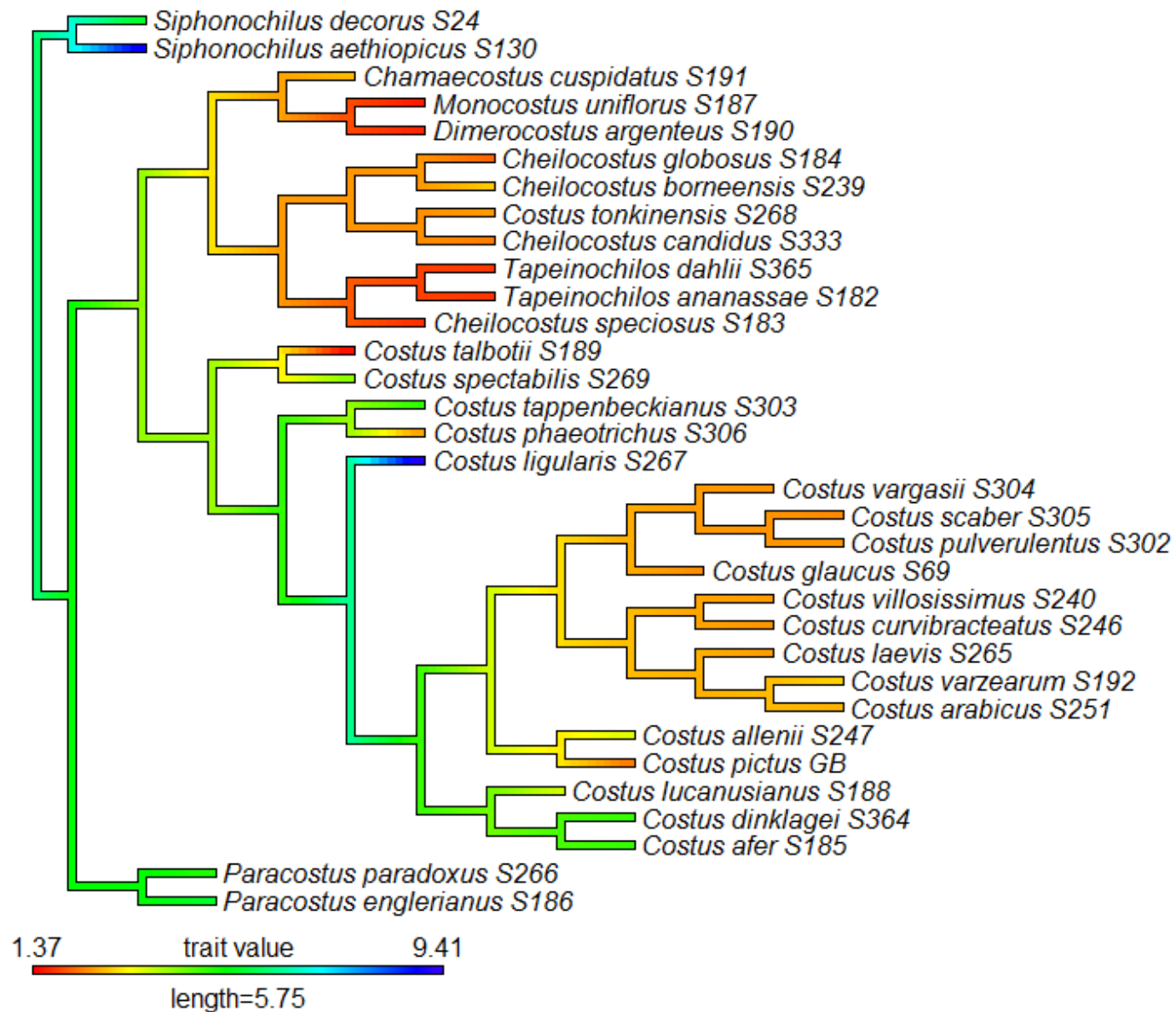
Příloha 12: Velikosti genomu namapované na ML strom vytvořený z Datasetu.



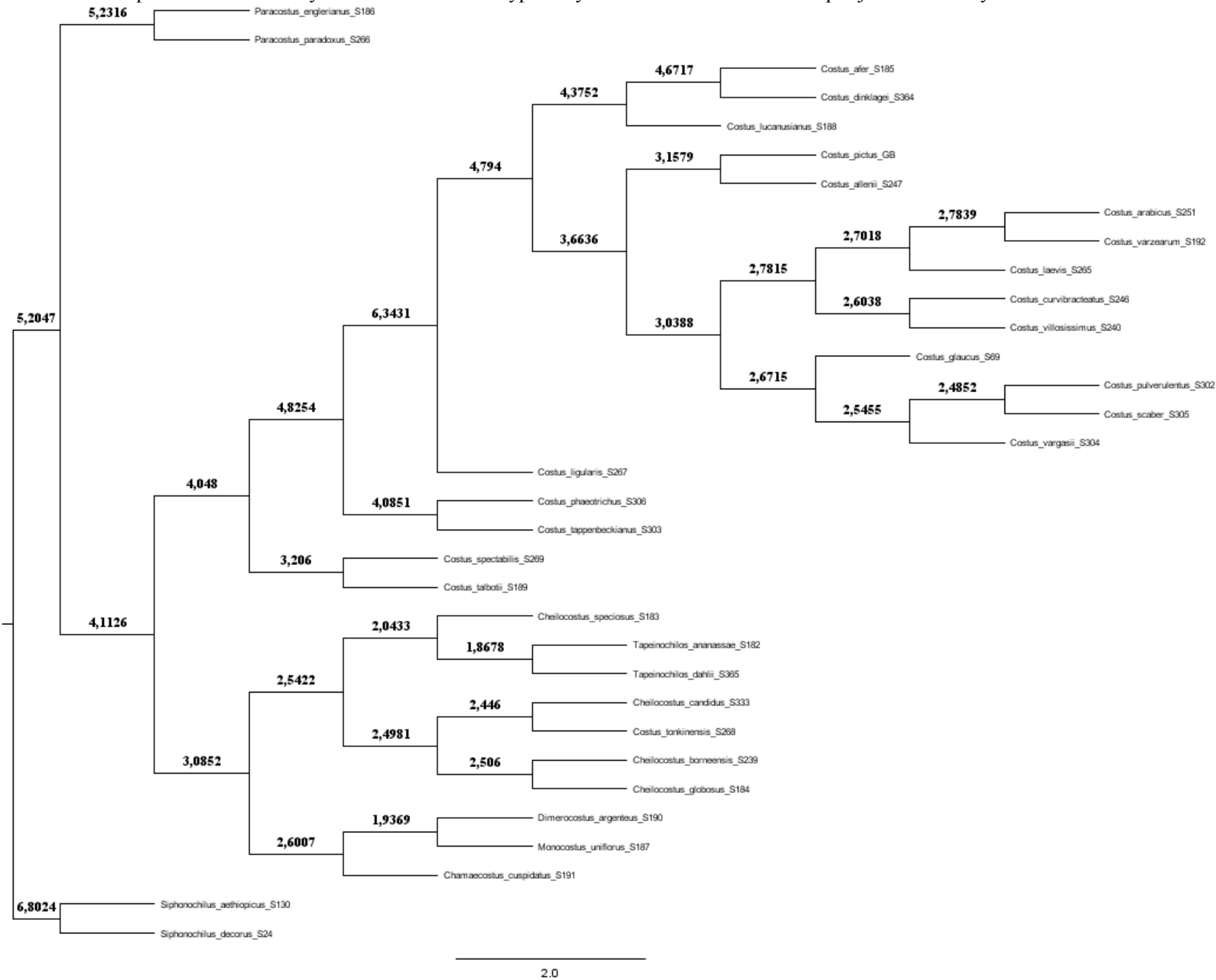
**Příloha 13:** Species tree konkatovaný (FastTree) s namapovanými velikostmi genomu.



**Příloha 14:** Species tree odhadnutý metodou ASTRID s namapovanými velikostmi genomu.



**Příloha 15:** Species tree odhadnutý metodou ASTRID s vypočtenými ancestrálními velikostmi pro jednotlivé nody.



2.0