

Univerzita Karlova
Přírodovědecká fakulta

Studijní program: Biologie

Studijní obor: Botanika



Bc. Jana Skopalíková

Využití metody Hyb-Seq pro rekonstrukci retikulární vnitrorodové fylogeneze:
příklad z polyploidního rodu *Curcuma* L. (*Zingiberaceae*)

Application of Hyb-Seq method for reconstruction of reticulate infrageneric phylogeny:
example from polyploid genus *Curcuma* L. (*Zingiberaceae*)

Diplomová práce

Školitel: Mgr. Tomáš Fér, Ph.D.

Praha, 2017

PROHLÁŠENÍ

Prohlašuji, že jsem diplomovou 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, 15. 8. 2017

Podpis:

PODĚKOVÁNÍ

Na tomto místě bych chtěla poděkovat především svému školiteli Mgr. Tomáši Férovi, Ph.D. za odborné vedení a cenné rady jak při analýze molekulárních dat, tak při následném psaní diplomové práce. Dále bych ráda poděkovala Mgr. Monice Pospíšilové za odvedenou laboratorní práci při přípravě knihoven, Mgr. Elišce Záveské, Ph.D. a Mgr. Janě Leong-Škorničkové, Ph.D. za cenné rady a připomínky k práci. Dále také Mgr. Evě Skopalíkové a Bc. Václavu Kubaljakovi za pomoc s odhalením překlepů a přehlédnutých chyb a Národní Gridové Infrastruktury MetaCentrum (projekt LM2015042 E-infrastruktura CESNET, program Projekty velkých infrastruktur pro VaVaI) za poskytnutí přístupu k výpočetním a úložným zdrojům. Finanční prostředky pro vytvoření této práce byly poskytnuty Grantovou agenturou České republiky (grant č. 14-13541S, Evoluce, diversifikace a biogeografie polyploidní čeledi *Zingiberaceae*: synergie fylogenomiky, cytogenomiky, geologické a klimatické historie).

ABSTRAKT

Tato diplomová práce se zabývá výzkumem fylogeneze hybridně polyploidního rodu *Curcuma* z čeledi *Zingiberaceae* s využitím next-generation sekvenování, které umožňuje získat stovky až tisíce jaderných lokusů pro tvorbu fylogenetických stromů, což se jeví jako vhodnější přístup oproti dosud hojně využívané cpDNA a ITS lokusu, zvláště u hybridních a polyploidních skupin.

K získání dat pro fylogenetické analýzy byla využita metoda Hybridization-based sequencing (Hyb-Seq), která kombinuje cílené obohacení (target enrichment) a mělké sekvenování (genome skimming). K analýze dat byla využita především pipeline HybPhyloMaker specificky vyvinutá pro práci s Hyb-Seq daty.

Bylo osekvenováno celkem 27 druhů rodu *Curcuma* a tři druhy tvořící outgroup. Fylogenetické stromy získané ze všech 1 154 a 811 vybraných jaderných low-copy genů vykazují vysoké podpory jednotlivých uzlů, oproti tomu fylogeneze získané z celého chloroplastového genomu a rDNA cistronu jsou v některých větvích podpořené méně a vykazují inkongruence v topologii oproti stromům z jaderných low-copy genů. Fylogenetické sítě vytvořené z jaderných genů, lineage movement analýza a testy monofylie jsou ve shodě s dříve publikovanou teorií o hybridním meziliniiovém původu 3 druhů – *C. vamana*, *C. myanmarensis* a *C. roscoeana*. Tyto metody také ukázaly pravděpodobně hybridní vznik druhu *C. candida* z rodičovských druhů spadajících do skupiny *Curcuma I* a na bázi podrodu *Curcuma*.

KLÍČOVÁ SLOVA

Next-generation sekvenování, Hyb-Seq, kurkuma, fylogeneze, polyploidie, hybridizace, jaderné low-copy geny, plastom, ribozomální DNA

ABSTRACT

This master thesis focuses on the phylogeny of hybridogenous and polyploid genus *Curcuma* from family *Zingiberaceae* using Next-Generation Sequencing data from hundreds to thousands nuclear loci. This approach seems to be better than widely used cpDNA and ITS sequencing especially in the case of hybridogenous and polyploid groups.

Data for phylogeny reconstruction were generated using Hybridization-based sequencing (Hyb-Seq) method which combines target enrichment and genome skimming strategies. Data analysis was performed primarily using pipeline HybPhyloMaker especially created for Hyb-Seq data analysis.

Twenty-seven species from the genus *Curcuma* and three outgroup species were sequenced in this work. Phylogenetic trees based on all 1 154 and 811 selected nuclear low-copy genes show high support values of all nodes which is in contrast to plastome and rDNA phylogeny with lower support values in some nodes and incongruences in topology compared to low-copy genes phylogeny. Phylogenetic networks inferred from low-copy genes, lineage movement analysis and monophyly tests agree with published hypotheses of interlineage hybrid origin of three species – *C. vamana*, *C. myanmarensis* and *C. roscoeana*. These analyzes show likely hybrid origin of *C. candida* too with parents from the group *Curcuma I* and basal species of the subgenus *Curcuma*.

KEY WORDS

Next-generation sequencing, Hyb-Seq, turmeric, phylogeny, polyploidy, hybridisation, nuclear low-copy genes, plastome, ribosomal DNA

OBSAH

1	ÚVOD	8
1.1	Cíl práce.....	10
1.2	Problematika druhových stromů a sítí	11
1.3	Sekvenování nové generace.....	13
1.3.1	Příprava DNA knihovny.....	13
1.3.2	Hybridization-based sequencing (Hyb-Seq)	14
1.3.3	Sekvenační platformy.....	16
1.4	Analýza dat	16
1.4.1	Assembly	17
1.4.2	Fylogenetické stromy	17
1.4.2.1	Koalescenční metody	18
1.4.2.2	Supertree metody	19
1.4.2.3	Konkatenace.....	19
1.4.3	Pipeline HybPhyloMaker	19
1.5	Rod <i>Curcuma</i> L.	20
1.5.1	Morfologická charakteristika a ekologie.....	21
1.5.2	Polyploidie a hybridizace	21
1.5.3	Chromozomové počty	22
1.5.4	Fylogeneze rodu <i>Curcuma</i>	23
2	MATERIÁL A METODY	25
2.1	Rostlinný materiál.....	25
2.2	Metody.....	25
2.2.1	Molekulární analýzy.....	25
2.2.1.1	Izolace DNA	25
2.2.1.2	Příprava Hyb-Seq knihovny.....	27
2.2.1.3	Obohacení Hyb-Seq knihovny a sekvenování	28
2.2.2	Zpracování dat.....	30
2.2.2.1	Analýza jaderných low-copy genů.....	30
2.2.2.2	Analýza chloroplastového genomu	34
2.2.2.3	Fylogeneze ribozomální DNA	36
2.2.2.4	Srovnání species trees z různých molekulárních markerů	37

2.2.2.5	Testování hybridního původu vybraných druhů	37
3	VÝSLEDKY	42
3.1	Fylogenetické analýzy	45
3.1.1	Fylogeneze odvozená z jaderných low-copy genů.....	45
3.1.2	Fylogeneze založená na chloroplastové DNA.....	45
3.1.3	Fylogeneze ribozomální DNA	46
3.1.4	Srovnání chloroplastové, ribozomální a jaderné fylogeneze.....	50
3.1.5	Rekonstrukce fylogeneze na základě podsouboru genů.....	53
3.2	Testování hybridního původu vybraných druhů.....	55
3.2.1	Fylogenetické sítě.....	55
3.2.2	Lineage movement analýza.....	57
3.2.3	Testy monofylie.....	58
4	DISKUZE	60
4.1	Úspěšnost sekvenování.....	60
4.2	Rekonstrukce fylogeneze.....	61
4.2.1	Jaderné low-copy geny	61
4.2.2	Chloroplastová DNA a ribozomální cistron.....	62
4.2.3	In/kongruence mezi jednotlivými fylogenezemi.....	64
4.2.4	Redukované datasety	66
4.3	Testování hybridního původu vybraných druhů.....	68
4.3.1	Fylogenetické stromy a sítě.....	68
4.3.2	Metody shrnující odlišné topologie nukleárních genů	70
5	ZÁVĚR.....	72
6	SEZNAM POUŽITÉ LITERATURY	74
7	PŘÍLOHY	83

SEZNAM ZKRATEK

ASTRAL	Accurate Species Tree Algorithm
ASTRID	Accurate Species Trees from Internode Distances
BI	Bayesovská analýza
BLAT	BLAST-like alignment tool
bp	pár bází
cpDNA	chloroplastová DNA
DNA	deoxyribonukleová kyselina
DOGMA	Dual Organellar GenoMe Annotator
ETS	external transcribed spacer
ExaML	Exascale Maximum Likelihood
Hyb-Seq	Hybridization-based sequencing
ILS	incomplete lineage sorting
IR	obrácená repetice (z angl. „inverted repeat“)
ITS	internal transcribed spacer
LSC	large single copy region
MAFFT	Multiple Alignment using Fast Fourier Transform
ML	Maximum Likelihood
MP	Maximum Parsimony
MPL	Maximum Pseudo-Likelihood
MRL	Matrix Representation with Likelihood
MRP	Matrix Representation with Parsimony
mtDNA	mitochondriální DNA
nt	nukleotid/y
NTS	nontranscribed spacer
PCR	polymerázová řetězová reakce (z angl. „polymerase chain reaction“)
RAxML	Randomized Axelerated Maximum Likelihood
rDNA	ribozomální DNA
RF distance	Robinson-Fouldova vzdálenost
SNP	jednodukleotidový polymorfismus (z angl. „single nucleotide polymorphism“)
SSC	small single copy region
UPGMA	Unweighted Pair Group Method
UTR	nepřekládaná oblast (z angl. „untranslated region“)

1 ÚVOD

Molekulární fylogenetika založená na porovnávání nukleotidových sekvencí se stala důležitým nástrojem taxonomie a evoluční biologie. Do nedávné doby byly hojně využívány především chloroplastové (cpDNA) a jaderné ribozomální (rDNA) lokusy, nejčastěji ITS (internal transcribed spacer). Ty se však nejeví jako nejvhodnější pro odvozování fylogeneze hybridních a polyploidních druhů kvůli uniparentální dědičnosti cytoplazmatických genomů (Small et al., 2004) a tíhnutí biparentálně děděné ribozomální DNA (deoxyribonukleové kyseliny) ke vzájemné evoluci (tzv. „concerted evolution“, Alvarez and Wendel, 2003; Wendel et al., 1995), čímž vzniká rozpor mezi fylogenezí odvozenou z těchto markerů a skutečnými vztahy mezi druhy (např. Huang et al., 2010; Schmickl et al., 2016). Naopak vhodnější se zdají být biparentálně děděné jaderné geny, tzv. low-copy markery (v buňce přítomná pouze 1 nebo několik kopií daného lokusu oproti repetitivní DNA s mnoha kopiemi), které nepodléhají tak snadno vzájemné evoluci a mohou vykazovat vyšší variabilitu než chloroplastové úseky (Small et al., 2004). Bohužel ani studium evolučních vztahů na základě nezávislých jaderných genů není bez problémů, jednotlivé genové stromy se mohou od skutečných vztahů mezi druhy (tzv. druhových stromů, species trees) lišit. Příčinou těchto odlišností může být například incomplete lineage sorting, genové duplikace a ztráty, horizontální přenos nebo hybridizace mezi druhy (Degnan and Rosenberg, 2009; Maddison, 1997).

Využívání low-copy markerů spolu s klasickým Sangerovým sekvenováním má i několik metodických úskalí, jako je problematický vývoj primerů pro nové lokusy nebo nutnost klonování pro odlišení jednotlivých alel (Sang, 2002; Small et al., 2004). Kvůli tomu a také kvůli značné finanční náročnosti Sangerova sekvenování, se ke konstrukci fylogenetických stromů tradičně využívalo pouze několik jednotek, maximálně desítek genů, čímž někdy vznikaly fylogenetické stromy obsahující špatně vyřešené nebo málo podpořené větve (např. Syring et al., 2007).

Velká část překážek byla s příchodem next-generation sekvenování (NGS) překonána – např. vývoj specifických primerů ani klonování díky odlišnému principu metod není nutné. Vysoký výkon NGS umožňuje osekvenovat stovky až tisíce lokusů několika desítek jedinců v jednom sekvenačním běhu a fylogenetické stromy konstruované z těchto dat vykazují obecně vysoké podpory jednotlivých větví (např. Bombarely et al., 2014; Cannon et al., 2015). Některé problémy však stále přetrvávají – rekonstrukce komplexních retikulátních

vztahů je i s použitím mnoha stovek jaderných markerů velkou výzvou, a to kvůli evolučním fenoménům zapříčiňujícím nesourodost mezi fylogenezemi z jednotlivých jaderných genů. Zatímco některé příčiny inkongruence mezi gene trees a species tree lze modelovat pomocí moderních analytických přístupů (např. incomplete lineage sorting a modely založené na koalescenční teorii; např. Mirarab et al., 2014; Vachaspati and Warnow, 2015), jiné příčiny (hybridizace) a kombinace více evolučních fenoménů se modelují a analyzují velmi problematicky (např. Yu et al., 2013). S využíváním NGS metod vyvstala také některá nová úskalí – např. analýza dat je náročná jak na výpočetní zdroje, tak na vývoj nových programů a pipeline, protože tradičně využívané přístupy jsou často nevhodné pro velké datasety získané next-generation sekvenováním (např. Mirarab et al., 2014; Pickrell and Pritchard, 2012; Vachaspati and Warnow, 2015).

Jako modelová skupina pro tuto diplomovou práci byl vybrán rod *Curcuma* L. z čeledi *Zingiberaceae*, který zahrnuje jak hybridní, tak polyploidní druhy. Rod sestává ze čtyř linií – *Curcuma*, *Hitcheniopsis*, *Ecomata* a *Pierreana* (Záveská et al., 2012), ploidie jednotlivých druhů roste od 2x do 15x, povětšinou se však jedná o diploidizované paleohexaploidy (Leong-Škorničková et al., 2007) a několik druhů je považováno za meziliniové hybridy (Záveská et al., 2016). Tyto poznatky mohou být využity při testování nových přístupů, jako je analýza dat z next-generation sekvenování a metody Hyb-Seq. V rámci diplomové práce budou na rodu *Curcuma* prezentovány výhody a úskalí využití metody Hyb-Seq pro účely rekonstrukce retikulární fylogeneze, jakož i detekce případných hybridogenních druhů v datasetu.

K získání dat byla využita metoda Hyb-Seq (Hybridization-based sequencing), která kombinuje target enrichment (na hybridizaci založené obohacení stovek až tisíců cílových sekvencí ve vzorku) a genome skimming (sekvenování oblastí s přirozeně vysokým zastoupením v buňce jako je chloroplastová nebo ribozomální DNA; Weitemier et al., 2014). Celkem 27 druhů rodu *Curcuma* a tři druhy tvořící outgroup byly obohaceny o 1 180 low-copy genů s 4 618 exony a vzorky byly následně osekvenovány na platformě Illumina. K analýze jaderných, chloroplastových i ribozomálních dat bylo využito několik přístupů – pipeline HybPhyloMaker (Fér, 2016), *de-novo* assembly i mapování na referenci pro tvorbu fylogenetických stromů, dále testy monofylie, lineage movement analýza a konstrukce fylogenetických sítí pro odhalení potenciálně hybridních druhů a analýza podsouboru genů pro zhodnocení množství genů potřebných pro tvorbu adekvátního species tree.

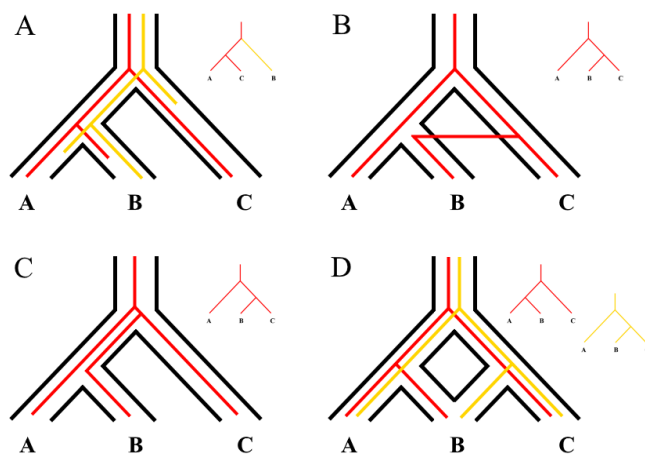
1.1 CÍL PRÁCE

Cílem této diplomové práce je zodpovědět následující otázky:

1. Je vhodnější a efektivnější rekonstruovat species tree/species network na základě mnoha stovek jaderných genů nežli za použití několika málo tradičně používaných molekulárních markerů (ITS lokus, cpDNA atd.)? Tj. jsou tyto stromy lépe rozlišené a disponují vyšší statistickou podporou?
2. Jsou fylogenetické stromy založené na různých typech molekulárních markerů (jaderná, chloroplastová a ribozomální DNA) ve vzájemné shodě?
3. Kolik jaderných low-copy genů je potřeba pro kvalitní rekonstrukci species tree a které současné metody rekonstrukce species tree jsou k tomu nejvhodnější?
4. Lze na základě několika stovek jaderných genů lépe detekovat potenciální hybridogenní druhy?
5. Jak přispěla tato data k pochopení retikulární evoluce rodu *Curcuma*?

1.2 PROBLEMATIKA DRUHOVÝCH STROMŮ A SÍTÍ

Fylogenetické stromy jsou grafickým znázorněním příbuzenských vztahů mezi různými taxonomickými jednotkami. Při rekonstrukci fylogeneze založené na nukleotidových sekvencích se nejčastěji využívají genové stromy (gene trees) a stromy druhů (species trees). Jak se již v minulosti ukázalo, jednotlivé části genomu mohou podstupovat odlišný vývoj díky různému typu dědičnosti (uniparentální versus biparentální), rekombinaci (výměně DNA mezi alelami/lokusy) nebo odlišným substitučním rychlostem jednotlivých úseků DNA (Alvarez and Wendel, 2003; Small et al., 2004), proto není vhodné odvozovat strom druhů pouze na základě jednoho genu nebo DNA lokusu. Rekonstrukci fylogeneze založené na větším množství lokusů však často doprovází problém inkongruencí mezi jednotlivými genealogiemi. Odlišné topologie mohou způsobovat evoluční fenomény jako je incomplete lineage sorting (ILS), hybridizace, horizontální přenos nebo genové duplikace a ztráty (Obrázek 1; shrnuto např. v Degnan and Rosenberg, 2009; Maddison, 1997).



Obrázek 1: Procesy způsobující nesoulad mezi topologiemi genových stromů. Genové stromy jsou vykresleny červenou a žlutou barvou, stromy druhů jsou černé. **A:** genová duplikace (zmožnění genu a následný samostatný vývoj jeho dvou kopií), **B:** horizontální genový přenos (příjem sekvence DNA od nerodičovského jedince), **C:** incomplete lineage sorting (zachování ancestrálního polymorfismu – existence více alel daného lokusu, které přetrvaly během druhové divergence) a **D:** hybridizace (křížení jedinců různých druhů; převzato a upraveno z Degnan and Rosenberg, 2009; Kubatko, 2009).

Horizontální genový přenos je typický pro jednobuněčné organismy (bakterie a kvasinky), které jsou schopné výměny genetické informace i s jedinci, kteří nejsou jejich přímými potomky. U rostlin je tento jev obecně vzácný a jeho nejznámější formou je tzv. chloroplast capture, stav kdy jedinec disponuje jaderným genomem jednoho druhu a chloroplastovým genomem druhu jiného, což způsobuje cytonukleární diskordance při odvozování fylogeneze z různých molekulárních markerů (např. Parks et al., 2009). Tradičně je tento fenomén vysvětlován introgrésí (např. Tsitrone et al., 2003) mezi rodičovskými taxony, avšak v některých případech známky introgrese odhaleny nejsou. Možné vysvětlení poskytuje hypotéza „přirozeného roubování“ (natural grafting), kde se výměna organelárních

genomů uskutečňuje po přímém kontaktu kořenů či stonků dvou blízce příbuzných druhů (Stegemann et al., 2012).

Incomplete lineage sorting je naopak velmi častým jevem, ke kterému dochází v případě zachování ancestrálního polymorfismu i po diverzifikaci linií, tzn. přítomnosti většího počtu alel na lokus, které přetrvaly během druhové divergence. Genealogie postižené ILS vykazují mylné podobnosti mezi méně příbuznými druhy, což může vést k chybné rekonstrukci fylogeneze založené na těchto genech (Degnan and Rosenberg, 2009; Maddison, 1997). Vzhledem k častému výskytu ILS byly vytvořeny moderní analytické přístupy schopné zahrnout vliv ILS při rekonstrukci species tree pomocí modelů založených na koalescenční teorii (např. Mirarab et al., 2014; Vachaspati and Warnow, 2015).

Genové duplikace a ztráty – zmnožení genu a následný samostatný vývoj jeho dvou kopií, jsou dalším fenoménem způsobujícím nesoulad mezi odvozenými genealogiemi v případě, že nedojde k zachycení ortologních kopií např. z důvodu extinkce, přeměny na pseudogen nebo získání nové funkce některých kopií u různých druhů (Maddison, 1997). Genové duplikace vytváří podobný konflikt jako ILS, a proto by mohly koalescenční metody překonat i tento problém (Záveská et al., 2016). Kromě jednotlivých genových duplikací, může docházet i k polyploidizaci – celogenomové duplikaci, po níž typicky následuje rediploidizace (ztráty, neofunkcionalizace nebo subfunkcionalizace jednotlivých genů), což zvyšuje obtížnost rozlišování ortologních sekvencí od paralogních (Dufresne et al., 2014). Možným řešením je sekvenování striktně ortologních sekvencí (Small et al., 2004) nebo zahrnutí pouze jednokopiových genů do fylogenetických analýz – např. vyfiltrovaných při designu hybridizačních proub pro Hyb-Seq (Weitemier et al., 2014), viz dále.

Hybridizace je jev, kdy dojde ke křížení jedinců různých druhů za vzniku životaschopného potomka, který disponuje geny s evoluční historií obou rodičů. Jedna z hlavních metod identifikace hybridních druhů je porovnávání topologie jaderných a chloroplastových lokusů, případně několika nezávislých jaderných genů (např. Fehrer et al., 2007; Pelsler et al., 2010; Renoult et al., 2009). Vzhledem ke své uniparentální dědičnosti může chloroplastový genom v kombinaci s biparentálně děděnými jadernými markery poskytnout důkaz o hybridizaci nebo introgresi a může být využit k identifikaci mateřských taxonů (např. Pardo et al., 2004). Odlišení hybridizace od ostatních evolučních fenoménů by mělo být možné také na základě nalezení celé řady jaderných genů vykazující jednotnou topologii, která se však liší od topologie majoritní (Linder and Rieseberg, 2004). K metodám zobrazujícím odlišnou topologii jednotlivých taxonů na základě vstupního souboru genových

stromů patří např. lineage movement analýza (Smith and Dunn, 2008), která byla použita k testování alternativních fylogenetických hypotéz u cichlid (Schwarzer et al., 2009) nebo liliovců (Bleidorn et al., 2009).

K zobrazení retikulátních vztahů mezi hybridními druhy a jejich rodiči, které jsou v rámci klasického fylogenetického stromu často nepostihnutelné, byly vyvinuty metody konstruující fylogenetické sítě. K nejvyužívanějším z nich patří Neighbor-Net (Bryant and Moulton, 2004) – velmi rychlá distanční metoda založená na Neighbour-Joining algoritmu, která byla využita např. v pracích Bombarely et al. (2014) nebo Závěská et al. (2016). K dalším z nich patří STEM-Hy (Kubatko et al., 2009) nebo PhyloNet, softwarový balíček k rekonstrukci a analýze retikulátních evolučních vztahů (Than et al., 2008), schopný tvorby fylogeneze jak pomocí algoritmu Maximum Parsimony (MP; Yu et al., 2013), tak Maximum Likelihood (ML; Yu et al., 2014), Maximum Pseudo-Likelihood (MPL; Yu and Nakhleh, 2015) i Bayesovské analýzy (BI; Wen et al., 2016a). PhyloNet se kromě hybridizace snaží zahrnout také vliv dalšího evolučního fenoménu, konkrétně incomplete lineage sorting a byl využit např. v práci Wen et al. (2016b). Podle srovnání některých PhyloNet metod a Stem-Hy dosahuje nejlepších výsledků metoda ML-PhyloNet, která je bohužel velmi výpočetně náročná, naopak nejhorší výsledky (špatné rekonstrukce vztahů téměř ve všech případech) byly zaznamenány u metody MP-PhyloNet (Kamneva and Rosenberg, 2017). BI-PhyloNet a MPL-PhyloNet metody hodnoceny nebyly, avšak BI-PhyloNet trpí také vysokou výpočetní náročností (Wen et al., 2016a), která je překonána u metody MPL-PhyloNet (Yu and Nakhleh, 2015).

1.3 SEKVENOVÁNÍ NOVÉ GENERACE

V posledním desetiletí se začíná v biologii uplatňovat tzv. sekvenování nové generace. Jedná se o metody, pomocí kterých je možné osekvenovat paralelně miliony molekul za zlomek času a ceny oproti klasickému Sangerovu sekvenování (Glenn, 2011; Sanger et al., 1977), které umožňuje v jednom běhu nejvýkonnějšího sekvenátoru (Applied Biosystems 3730) získat sekvenci pouze 96 fragmentů DNA (Hutchison, 2007).

1.3.1 Příprava DNA knihovny

Vlastnímu procesu sekvenování předchází příprava vzorků – tzv. tvorba DNA knihovny. Metod k tomu určených je celá řada od jednoduchých – pouhé nafragmentování DNA a naligování adaptorů, až po složité metody zahrnující mnoho kroků. Se snižujícími se

náklady na sekvenování nové generace je dnes již možno osekvenovat celé genomy v reálném čase, avšak pro fylogenetické studie zahrnující větší počet taxonů je tento přístup stále velmi finančně náročný a množství vygenerovaných dat by převyšovalo možnosti jejich úspěšného zpracování. Z tohoto důvodu se ve fylogenetice využívají metody snižující genomovou komplexitu. K nejpoužívanějším z nich patří amplikonové sekvenování (Bybee et al., 2011; Meyer et al., 2008), mělké sekvenování (Genome skimming; Straub et al., 2012), cílené obohacení (Target enrichment; Gnirke et al., 2009; Okou et al., 2007), transkriptomové sekvenování (RNA-Seq; Marioni et al., 2008) a metody založené na restrikci DNA (RAD-Seq, GBS a další; Baird et al., 2008; Elshire et al., 2011). Každá výše zmíněná metoda má své výhody i nevýhody, šíří a možnosti použití (shrnuto např. v Lemmon and Lemmon, 2013).

1.3.2 Hybridization-based sequencing (Hyb-Seq)

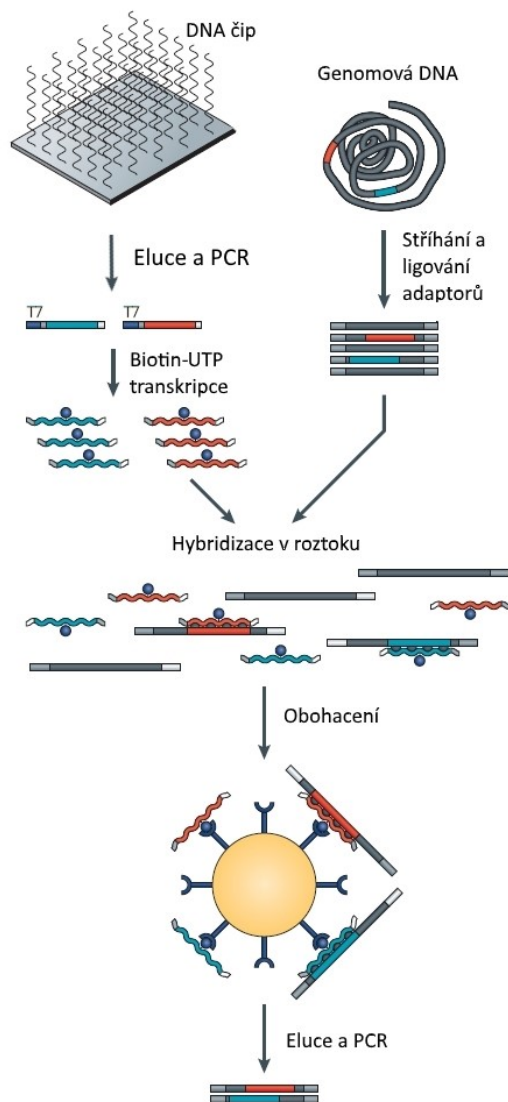
Pro účely této diplomové práce byla vybrána metoda Hyb-Seq, která kombinuje dva přístupy – mělké sekvenování (Straub et al., 2012) a cílené obohacení (Gnirke et al., 2009), díky kterému dojde ke snížení komplexity vzorku a přednostnímu osekvenování předem vybrané oblasti genomu – několika set low-copy jaderných genů (Weitemier et al., 2014). Hloubka pokrytí u mělkého sekvenování závisí na relativním zastoupení molekul DNA ve vzorku, čímž dojde k hlubokému osekvenování vysokokopiové frakce – ribozomové DNA, mitochondriálního (mtDNA) a chloroplastového genomu (Straub et al., 2012). Kombinací obou přístupů jsou získána sekvenační data jak jaderných low-copy markerů, tak z high-copy oblastí genomu (Weitemier et al., 2014). Metoda Hyb-Seq byla již použita například v práci Heyduk et al. (2016) nebo Schmickl et al. (2016) a přinesla vysoce podpořené fylogeneze. Několik set jaderných low-copy genů by proto mohlo přinést zlepšení i v chápání retikulátních fylogenetických vztahů, které jsou z rodu *Curcuma* známy (Záveská et al., 2012).

Hyb-Seq metodu můžeme rozdělit na dvě části – design hybridizačních sond a vlastní přípravu knihovny pro next-generation sekvenování. Design sond se provádí na základě porovnávání genomu získaného mělkým sekvenováním s transkriptomem blízkého příbuzného taxonu s cílem vybrat několik set/tisíc exonů z low-copy genů. Proto se odfiltrují úseky, kde se transkriptomová data namapovala na více než jedno místo v genomu a také všechny lokusy s více než 90% sekvenční identitou, aby se zabránilo obohacování příliš podobných úseků. Posledním požadavkem na hybridizační sondy je jejich délka – např. 120 bp (párů bází) pro jednotlivé exony a 960 bp v součtu všech exonů pro daný lokus. Minimální velikost

jednotlivých exonů byla stanovena tak, aby byla dostatečná pro konstrukci hybridizačních sond, a délka lokusů tak, aby byly dostatečně informativní v následné rekonstrukci fylogeneze. Při designu hybridizačních sond je také nutné dbát na to, aby byly odfiltrovány úseky z organelárních genomů, které jsou v buňce zastoupeny ve větším množství kopií, a proto je není třeba dále obohacovat (Weitemier et al., 2014).

Samotná příprava knihovny zahrnuje izolaci DNA, fragmentaci, ligování adaptorů s barcody (4 – 6 bp dlouhá sekvence umožňující následné rozlišení sekvencí z jednotlivých vzorků), selekci fragmentů požadované délky (Straub et al., 2012) a poolování (smíchání všech vzorků do jedné zkumavky v ekvimolárním poměru). Následuje hybridizace DNA v roztoku (tzv. „solution hybridization“) s biotinylovanými RNA (ribonukleová kyselina) sondami. Po inkubaci jsou sondy zachyceny pomocí magnetických kuliček obalených streptavidinem, které váží biotin z RNA sond. Posledními kroky jsou eluce DNA fragmentů, amplifikace pomocí polymerázové řetězové reakce (PCR) a sekvenování (Obrázek 2; Gnirke et al., 2009).

Výhodou cíleného obohacení je malé množství chybějících dat (Lemmon and Lemmon, 2013) a možnost využití metody i na degradované vzorky, např. herbářové položky (Sousa et al., 2014). Kromě cílených low-copy lokusů (exonů) jsou získávána data i z přilehlých rychleji se vyvíjejících nekódujících oblastí, jako jsou introny nebo 5' a 3' nepřekládaná oblast (UTR), které mohou být využity při rekonstrukci fylogeneze na nižších taxonomických úrovních. Díky kombinaci s mělkým sekvenováním jsou získány sekvence náležící i rDNA cistronu a organelárním genomům, jejichž evoluce se může lišit



Obrázek 2: Princip cíleného obohacení v roztoku – DNA sondy nasynthetizované na čipu se odštěpí a namnoží pomocí PCR. In vitro transkripce se sondy přepíše do RNA a začlení se biotin-UTP. Po hybridizaci jsou RNA sondy s cílovou DNA vycíhávány magnetickými kuličkami se streptavidinem vážící se na biotin. Následuje eluce DNA fragmentů a PCR amplifikace (převzato a upraveno z Metzker, 2010).

od evoluce jaderného genomu (Weitemier et al., 2014). Jednou z mála nevýhod Hyb-Seq metody a cíleného sekvenování obecně může být vysoká cena přípravy hybridizačních sond (Cronn et al., 2012).

1.3.3 Sekvenační platformy

Platformem pro sekvenování nové generace se v posledním desetiletí objevilo hned několik. Do tzv. metod druhé generace, u kterých je nutné před sekvenováním fragmenty namnožit pomocí PCR, patří 454 pyrosekvenování, Illumina, SOLiD a Ion Torrent, ke třetí generaci sekvenující jednotlivé molekuly v reálném čase pak řadíme PacBio a nanopórové sekvenování. Popis fungování jednotlivých metod stejně jako jejich výhody a nevýhody jsou shrnuty v práci Glenn (2011).

Pro účely této diplomové práce byla vybrána platforma Illumina (www.illumina.com; Bentley et al., 2008), která je v současné době nejvyužívanější sekvenační platformou nové generace. Vyniká především svou nízkou cenou za gigabázi dat, vysokým počtem readů v jednom sekvenačním běhu a malou mírou chybovosti (Glenn, 2014).

1.4 ANALÝZA DAT

Fylogenetika využívající data next-generation sekvenování bývá založená buď na identifikaci jednonukleotidových polymorfismů (SNP), nebo vytváří fylogenetické stromy z celých sekvencí jednotlivých lokusů. Metoda využívající celé sekvence se hodí spíše pro otázky na vyšších fylogenetických úrovních (např. Schmickl et al., 2016; Weitemier et al., 2014), oproti tomu SNP jsou využívány především na vnitrodruhové úrovni (např. Li et al., 2012; Hand et al., 2012).

Pro rekonstrukci fylogeneze rodu *Curcuma* byla vybrána metoda využívající celé sekvence jednotlivých lokusů, jež zahrnuje především kroky jako je filtrování kvality primárních readů, assembly readů do kontigů (delší sekvence využitelné pro fylogenetické analýzy – celé geny, exony...), alignment pozic v rámci sekvencí jednotlivých druhů/jedinců a tvorba fylogenetického stromu na základě jednoho nebo více lokusů. Jednotlivé kroky je možné provádět zvlášť s využitím množství programů nebo pomocí tzv. pipeline – většinou několika na sebe navazujících skriptů, které značně usnadňují práci s jednotlivými programy a velkým množstvím dat, např. PHYLUCE (metoda pro analýzu konzervovaných genomových lokusů; Faircloth, 2016), HybPiper (nástroj k assembly genů a identifikaci paralogních kopií; Johnson et al., 2016) nebo aTRAM (pipeline pro assembly genů ze

vzdálených taxonů; Allen et al., 2015). V této práci byla využita především pipeline HybPhyloMaker (Fér, 2017; www.github.com/tomas-fer/HybPhyloMaker), která byla specificky vyvinuta pro práci s Hyb-Seq daty a pro tvorbu fylogenetických stromů z velkého množství (několika set až tisíc) jaderných low-copy genů.

1.4.1 Assembly

Pro assembly jednotlivých kontigů z krátkých readů se využívá dvou základních přístupů – mapování na referenci a *de-novo* assembly. Při přístupu mapování na referenci dochází k rekonstrukci výsledného lokusu z krátkých readů mapováním na základě podobnosti s již existující sekvencí (referencí), při *de-novo* assembly se výsledná sekvence sestavuje na základě podobnosti mezi jednotlivými ready bez apriorní znalosti referenční sekvence (McCormack et al., 2013).

V dnešní době se při odvozování fylogeneze využívá především mapování na referenci, avšak bylo zjištěno, že v některých případech je vhodnější využít přístupu *de-novo* assembly. K těmto případům patří: 1) dostupná referenční sekvence příliš vzdálená od cílových druhů (Martin and Wang, 2011), 2) možnost detekování strukturních změn (např. Zhou et al., 2017) přítomných např. v cpDNA a 3) separace paralogních sekvencí (Johnson et al., 2016), jejichž větší zastoupení je předpokládáno u rDNA lokusu. Proto byly v této práci využity obě metody assembly u cpDNA a rDNA cistronu a hodnocena jejich prospěšnost při odvozování vztahů v rodu *Curcuma*.

1.4.2 Fylogenetické stromy

Rozvoj next-generation sekvenování a stále se zvětšující objemy získaných dat vedly také k rozvoji metod a programů schopných práce s velkými daty v reálném čase. Ke dvěma hlavním skupinám patří koalescenční metody, které se snaží zohlednit vliv incomplete lineage sorting, jednoho z hlavních fenoménů způsobujícího nesoulad mezi genovými stromy, a metoda konkatenace, kde ILS zohledněn není. Mnoho studií hodnotilo výkon jednotlivých metod a došlo k odlišným závěrům – v některých případech konkatenace překonala koalescenční metody (např. Bayzid and Warnow, 2012), jinde naopak dosáhla špatných výsledků (např. Kubatko et al., 2007), případně výsledky značně závisely na testovaných podmínkách – míra ILS aj. (např. Kupczok et al., 2010; Mirarab et al., 2014). V případě nízkého fylogenetického signálu v datech vykazuje většina metod sníženou přesnost při rekonstrukci species tree, více jsou však tímto jevem ovlivněny metody

rekonstruuující species trees z jednotlivých genových stromů oproti metodě konkatenace (Bayzid and Warnow, 2013). Výkon všech metod se zlepšuje s využitím většího množství lokusů a v přítomnosti menšího množství ILS (Bayzid and Warnow, 2013), avšak s rostoucí mírou chybějících dat naopak přesnost klesá (Kupczok et al., 2010). Proto je doporučováno využít několik různých metod a srovnat jimi odvozené fylogeneze (Kupczok et al., 2010). Některé z nejčastěji využívaných metod k rekonstrukci fylogeneze na základě NGS dat jsou stručně popsány níže.

1.4.2.1 Koalescenční metody

Koalescenční metody se snaží zohlednit vliv incomplete lineage sorting při odvozování species trees pomocí modelování multi-species coalescence. Můžeme je rozdělit na tzv. summary methods (např. ASTRAL a ASTRID; Mirarab et al., 2014; Vachaspati and Warnow, 2015), které vytváří species tree na základě vstupních genových stromů a na koestimační metody (např. *BEAST; Heled and Drummond, 2010), které jako vstupní data využívají sekvenční alignmenty a konstruuují současně genové stromy i strom druhů. Vzhledem k extrémní výpočetní náročnosti koestimačních metod pro datasety o >100 genech (Bayzid and Warnow, 2012), jsou pro analýzy dat na genomové úrovni využívány především summary metody.

ASTRAL

Accurate Species Tree Algorithm (ASTRAL) je velmi rychlá, statisticky konzistentní a přesná metoda odhadu stromu druhů, která jako vstupní data využívá nezakořeněné genové stromy. Jako species tree je pomocí heuristického vyhledávání zvolen strom, který souhlasí s největším počtem kvartetových stromů vytvořených na základě vstupních gene trees. Délka analýzy genomových dat se pohybuje v řádu minut a species trees vytvořené pomocí metody ASTRAL jsou přesnější než u jiných summary metod. Co se týká porovnání s konkatenací, v případě střední a vyšší míry ILS dosahuje lepších výsledků ASTRAL, v případě nízké ILS pak konkatenace (Mirarab et al., 2014).

ASTRID

Metoda Accurate Species Trees from Internode Distances (ASTRID) odhadující species tree pomocí distančních metod je založená na NJst algoritmu (Liu and Yu, 2011) – k výpočtu matice vzdáleností využívá průměrný počet internodií mezi dvojicemi taxonů ze všech genových stromů. Jedná se o další z velmi přesných koalescenčních metod

využitelných i pro velké datasety – délka analýzy je dokonce ještě několikanásobně kratší než u metody ASTRAL (Vachaspati and Warnow, 2015).

1.4.2.2 Supertree metody

Supertree metody kombinují topologie jednotlivých genových stromů do species tree. Na základě vstupních stromů je vytvořena binární matice, která slouží pro následnou fylogenetickou analýzu. Matice je konstruována tak, že se u každé větve jednotlivých genových stromů určí, zda se daný taxon nachází na jedné straně (1) nebo na straně druhé (0). Mezi nejznámější patří Matrix Representation with Parsimony (MRP) a Matrix Representation with Likelihood (MRL). Obě metody jsou velmi rychlé, avšak MRL dosahuje lepších výsledků (Nguyen et al., 2012).

1.4.2.3 Konkatenace

Konkatenace, také tzv. superalignment metoda, vytváří species tree ze „superalignmentu“, který vznikl konkatenací jednotlivých genových alignmentů a je analyzován jako jeden lokus ignorující možné odlišnosti v evoluční historii původních alignmentů. Příkladem metod založených na konkatenaci, které jsou využitelné i pro velké datasety, může být RAxML (Randomized Axelerated Maximum Likelihood; Stamatakis, 2014) a ExaML (Exascale Maximum Likelihood; Kozlov et al., 2015) využívající algoritmus Maximum Likelihood nebo FastTree s implementací Neighbour-Joining algoritmu (Price et al., 2009). Program FastTree je rychlejší, avšak méně přesný při odvozování fylogeneze v hůře podpořených uzlech než RAxML (Price et al., 2010). Problém výpočetní náročnosti u velkých datasetů řeší metoda ExaML, která je založená na „odlehčeném“ RAxML algoritmu a disponuje vysokou výpočetní rychlostí (Kozlov et al., 2015).

1.4.3 Pipeline HybPhyloMaker

Velká část analýzy dat v této práci byla prováděna pomocí pipeline HybPhyloMaker (Fér, 2016), která byla vyvinuta pro tvorbu species trees na základě velkého počtu jaderných, případně i chloroplastových genů. Skládá se z osmi základních kroků reprezentovaných jedním, případně několika alternativními skripty a rozšířením z několika dalších skriptů.

V prvním kroku dochází k odstranění adaptorů z primárních readů, filtrování kvality (oříznutí konců, kde je průměrné Phred quality score menší než 20 a úplnému odstranění sekvencí kratších než 36 bp pomocí programu Trimmomatic; Bolger et al., 2014)

a vyfiltrování duplicitních readů (Weitemier et al., 2014). Pro mapování readů na referenční sekvenci, tzv. pseudoreferenci – sekvenci práb oddělených 400 N (any) pozicemi (Straub et al., 2012), je využíván program Geneious (Kearse et al., 2012). Po alignmentu (vzájemném přiřazení odpovídajících si pozic DNA) pomocí programu MAFFT (Multiple Alignment using Fast Fourier Transform; Katoh et al., 2002) je provedena konkatenace exonů jednotlivých genů a úseky obklopující exony jsou vyřazeny z analýzy. Následuje vyřazení genů s příliš vysokým množstvím chybějících dat a tvorba genových stomů (Fér, 2016). Fylogenetické stromy druhů jsou konstruovány pomocí několika přístupů – koalescenčních metod ASTRAL (Mirarab et al., 2014) a ASTRID (Vachaspati and Warnow, 2015), supertree metody MRL (Nguyen et al., 2012) a konkatenace vypočtené pomocí metody Maximum Likelihood v programu ExaML (Kozlov et al., 2015) případně FastTree (Price et al., 2009, 2010).

1.5 ROD *CURCUMA* L.

Řád *Zingiberales*, kam spadá i čeleď *Zingiberaceae* a rod *Curcuma*, čítá 8 čeledí (*Musaceae*, *Heliconiaceae*, *Strelitziaceae*, *Lowiaceae*, *Marantaceae*, *Cannaceae*, *Zingiberaceae* a *Costaceae*), okolo 90 rodů a více než 2 200 druhů jednoděložných rostlin (Kress, 1990; Kress et al., 2001). Čeleď *Zingiberaceae* zahrnující okolo 50 rodů a 1 300 druhů má centrum diverzity v Asii. Jedná se o vytrvalé byliny rostoucí od nížin po horské oblasti, především jako podrost v tropických lesích. K nejdůležitějším rodům patří zázvor (*Zingiber officinale* Roscoe), kurkuma (*Curcuma longa* L.) a kardamovník (*Elettaria cardamomum* (L.) Maton; Škorničková, 2007).

Rod *Curcuma* v dnešní době čítá okolo 120 druhů (Škorničková et al., 2004) a dělí se na tři podrody – *Curcuma* K. Schum., *Hitcheniopsis* (Baker) K. Schum. a *Ecomatae* Škorničk. & Šída f. (Leong-Škorničková et al., 2015; Záveská et al., 2012), která zahrnuje dvě molekulárně, avšak ne morfologicky, odlišitelné linie – *Ecomata* a *Pierreana* (Záveská et al., 2012). Oblastí přirozeného výskytu je jižní a jihovýchodní Asie s centrem diverzity v Thajsku a Indii a několika druhy rozšířenými také v Číně, Austrálii a jižním Pacifiku (Škorničková, 2007).

Nejvýznamnějším druhem z hlediska hospodářského využití je *Curcuma longa* L., jejíž oddenky obsahují fenolické látky (kurkuminoidy) s protizánětlivou a antioxidační aktivitou (např. Ammon and Wahl, 1991; Ruby et al., 1995), dále je z nich získáván žlutý prášek využívaný v kuchyni (součást koření kari) a jako textilní barvivo. K dalším významným druhům patří také *C. aromatica* Salisb., *C. amada* Roxb., *C. caesia* Roxb.,

C. aeruginosa Roxb., *C. zanthorrhiza* Roxb. a *C. zedoaria* Roscoe – zdroj koření nebo léčiv (např. Škorníčková, 2007; Yasuda et al., 1988) a *C. alismatifolia* Gagnep., *C. rhabdota* Sirirugsa & Newman. nebo *C. roscoeana* Wall. pěstované jako okrasné rostliny (Paisooksantivatana et al., 2001; Škorníčková, 2007).

1.5.1 Morfologická charakteristika a ekologie

Kurkumy jsou vytrvalé středně vysoké byliny s oddenky. Většina druhů dorůstá výšky 0,5 až 1,5 m, můžeme zde však najít také malé druhy, jako je *Curcuma bhatii* s 10 – 20 cm, nebo naopak druhy 2,5 – 3,5 m vysoké, jako je *C. latifolia* (Škorníčková, 2007). Oddenky jsou buď jednoduché nebo větvené (Valeton, 1918), zvenku světle hnědé, uvnitř nabývající mnoha barev – od bílé, přes oranžovou až po modrou a fialovou (Škorníčková, 2007). Listy mají kopinatý, vejčitý nebo podlouhlý tvar a na bázi tvoří pochvu. Květenství je koncové tyrsoidní (hrozen z vrcholíků) s listeny ve vrcholové části nabývajícími výrazných barev (Valeton, 1918) pravděpodobně sloužící k lákání opylovačů, kterými je převážně hmyz. Květy jsou zygomorfni s jednou fertlní tyčinkou – dvěma bočními staminodii a pyskem. Doba kvetení je velmi krátká – do 10 hodin, plodem jsou tobolky se semeny opatřenými bílým arillem (Škorníčková, 2007; Valeton, 1918).

Jednotlivé druhy rostou převážně v nížinách a tvoří důležitou složku polostinných biotopů v monzunové Asii, jako jsou lesní lemy, křoviny, kokosové a arekové lesy nebo břehy řek. Některé druhy jsou také pionýrské – obsazují narušená stanoviště, jako jsou okraje cest a příkopy. Rostliny rozmnožující se sexuálně (náležící k nižším ploidním úrovním) obsazují především primární stanoviště, sterilní druhy (vyšší ploidie) rostou na sekundárních narušených stanovištích (Škorníčková, 2007).

1.5.2 Polyploidie a hybridizace

Hybridizace a polyploidizace patří u rodu *Curcuma* k nejdůležitějším speciálním procesům. K hybridizaci dochází v přírodě často, přednostně mezi blízkce příbuznými druhy (Škorníčková, 2007; Závěská, 2014), čehož se využívá i v zahradnictví k produkci nových kultivarů (např. Anuntalabhochai et al., 2007; Ketmaro et al., 2012). Práce Závěská et al. (2016) však naznačuje, že by mohlo v některých případech docházet i k hybridizaci vzdáleně příbuzných druhů z různých podrodů.

Polyploidie u rodu *Curcuma* je tradičně zmiňována zejména v souvislosti se zástupci podrodu *Curcuma* (Islam, 2004; Joseph et al., 1999; Ramachandran, 1969). V tomto podrodu

můžeme najít tzv. nižší polyploidy ($2n = 6x$, diploidní druhy v tomto podrodu známy nejsou), tak i tzv. vyšší polyploidy ($2n = 9x, 11x, 12x$ a $15x$; Leong-Škorníčková et al., 2007) pravděpodobně allopolyploidního původu (Záveská, 2014). Předpokládá se, že nonaploidi vznikli splynutím redukované a neredukované gamety hexaploidních druhů a spolu s dalšími hexaploidy dali vznik druhům s ploidii $12x$ a $15x$ (Leong-Škorníčková et al., 2007). Současná data však ukazují i určitou míru polyploidizace v rámci podrodů *Hitcheniopsis* a *Ecomatae*. Zatímco u podrodu *Ecomatae* byli pozorováni zejména jedinci s $2n = 6x$ (stupeň ploidie odpovídající tzv. nižším polyploidům z podrodu *Curcuma*), zástupci podrodu *Hitcheniopsis* byli charakterizováni zejména jako diploidní, zato s četným výskytem aneuploidů (Leong-Škorníčková et al., 2007; Šída et al., unpubl.).

Zjištění stáří a počtu polyploidizačních událostí, které proběhly před oddělením rodu nebo po jeho diverzifikaci, může být přínosné při snaze pochopit strukturu genomu a reprodukční strategii u jednotlivých druhů (Záveská, 2014). Např. v podrodu *Curcuma* můžeme najít druhy se stejným počtem chromozomů ($x = 42$), avšak lišící se rozmnožovací strategií (sexuální nebo vegetativní) a diploidními versus polyploidními profily pozorovanými u jaderných genů (Záveská et al., 2012, 2016).

1.5.3 Chromozomové počty

Rozpětí pozorovaných počtů chromozomů je díky rozdílné ploidii jednotlivých druhů značné – od $2n = 20$ až po $2n = 105$ s mnoha mezistupni (např. Islam, 2004; Joseph et al., 1999; Leong-Škorníčková et al., 2007). V době, kdy bylo nejvíce chromozomových počtů známo u zástupců podrodu *Curcuma*, bylo jako základní chromozomové číslo rodu nejčastěji uváděno $x = 21$ (Islam, 2004; Joseph et al., 1999; Ramachandran, 1969), po dalším výzkumu (Leong-Škorníčková et al., 2007) však bylo navrženo nové chromozomové číslo, a to $x = 7$. Zatímco při $x = 21$ je $2n = 42$ považováno za diploidní cytotyp, $2n = 63$ za triploidní apod., při $x = 7$ je na všechny cytotypy nahlíženo jako na paleopolyploidní, tudíž $2n = 42$ odpovídá hexaploidnímu cytotypu, $2n = 63$ nonaploidnímu atd. Oba koncepty lze nicméně aplikovat pouze na zástupce podrodu *Curcuma* a *Ecomatae*, neboť zástupci podrodu *Hitcheniopsis* svými chromozomovými počty neodpovídají násobkům $x = 7$ ani $x = 21$. Mezi nejčastější ploidie a počty chromozomů patří v rámci podrodu *Curcuma* $2n = 6x = 42$ a $2n = 9x = 63$, dále pak $2n = 12x = 84$, $2n = 11x = 77$ a $2n = 15x = 105$, u podrodu *Ecomatae* $2n = 42$. U podrodu *Hitcheniopsis* byly nalezeny převážně aneuploidní počty chromozomů jako $2n = 20, 24, 28, 32, 34, 36, 40$ nebo 56 (Leong-Škorníčková et al., 2007; Šída et al., unpubl.).

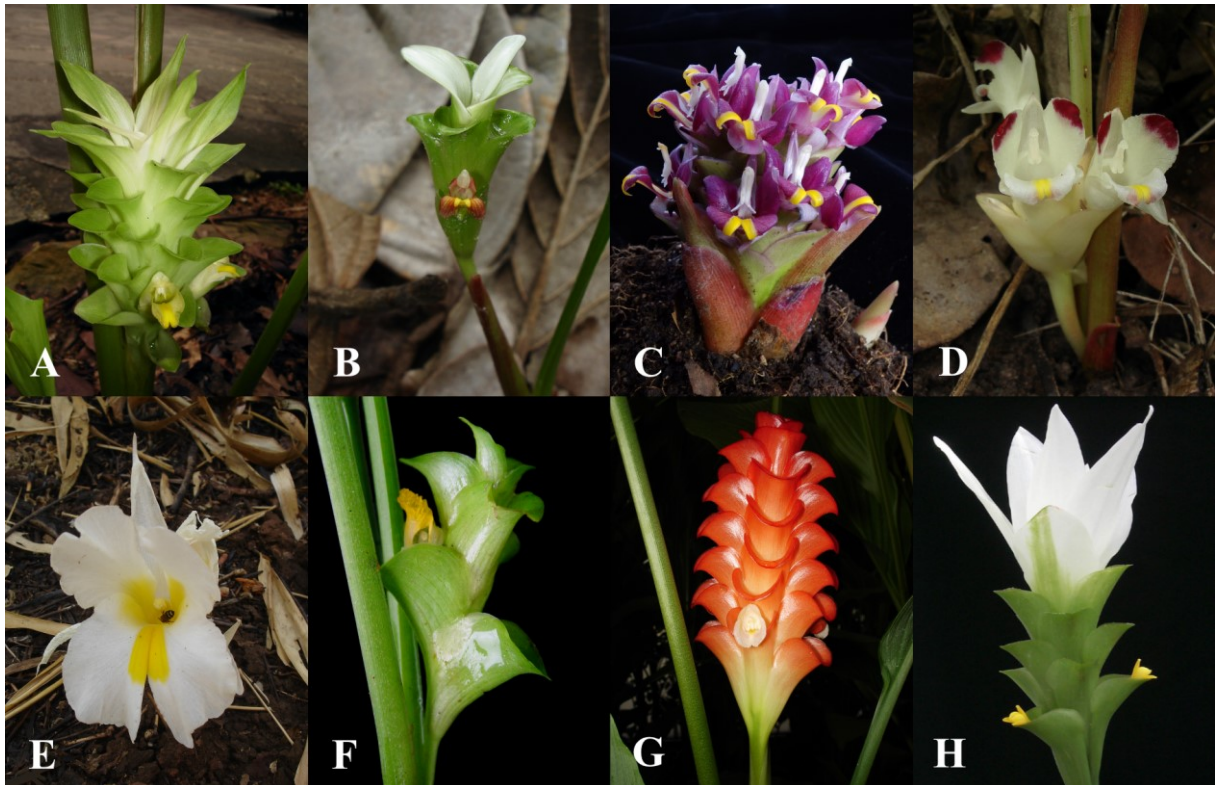
1.5.4 Fylogeneze rodu *Curcuma*

První fylogenetická studie zabývající se vztahy v rámci čeledi *Zingiberaceae* popsala rod *Curcuma* jako parafyletický a navrhla jeho rozčlenění na několik menších rodů (Kress et al., 2002). Druhá práce, která zkoumala fylogenezi tribu *Zingiberae*, také potvrdila parafyletičnost rodu *Curcuma*, avšak navrhla zahrnutí dříve vyčleňovaných rodů *Hitchenia* Wall., *Smithatris* W.J. Kress & K. Larsen a *Stahlianthus* Kuntze do rodu *Curcuma* (Ngamriabsakul et al., 2003). Obě tyto práce byly založeny na jednom rDNA lokusu a jednom cpDNA lokusu. Na to navazující studie Záveská et al. (2012), která si klade za cíl vyřešit tento molekulárně taxonomický problém. Fylogenezi rodu *Curcuma* a příbuzných rodů byla vyšetřena pomocí ITS a tří cpDNA lokusů a jejím výsledkem bylo konečné zařazení rodů *Hitchenia*, *Smithatris*, *Laosanthus*, *Stahlianthus* a dvou druhů z rodu *Kaempferia* do rodu *Curcuma*.

Pomocí metody AFLP byly dále hodnoceny vztahy v rámci podčeledi *Curcuma* obsahující vyšší i nižší polyploidní druhy, které byly rozděleny podle velikosti genomu na tři skupiny – I, II a III (Leong-Škorníčková et al., 2007). Výsledky této analýzy odhalily dvě základní linie – první obsahovala vyšší a nižší polyploidy spadající do skupiny velikosti genomu I, do druhé linie byly zařazeny hexaploidi ze skupiny II a III (Záveská et al., 2011). Pomocí sekvenování ITS a tří chloroplastových lokusů byla vyšetřována vnitrorodová fylogeneze, která odhalila čtyři hlavní skupiny – *Ecomata*, *Pierreana*, *Hitcheniopsis* a *Curcuma* (příkladové druhy spadající do jednotlivých linií zobrazeny na Obrázku 3), která se dále dělila na dvě další skupiny, a to *Curcuma I* a *Curcuma II*, avšak vztahy mezi liniemi nebyly vyřešeny spolehlivě (Záveská et al., 2012). ITS region vykazoval vysokou vnitroindividuální variabilitu a chloroplastové lokusy naopak nedisponovaly dostatečnou variabilitou (Záveská, 2014). Vyřešit vztahy mezi skupinami pomohly čtyři jaderné low-copy markery a jako sesterské skupiny byly určeny *Curcuma* s *Hitcheniopsis* a *Ecomata* + *Pierreana* (Záveská et al., 2016).

Na základě topologie stromu získaného z ITS sekvencí a inkongruencí mezi celkem šesti genovými stromy (Záveská et al., 2012, 2016) byly v rodu *Curcuma* odhaleny dva typy hybridních druhů – hybridy vzniklé z blízké příbuzných druhů (spadajících do jedné vývojové linie) a hybridy vzniklé z druhů spadajících do různých linií. Hybridy z blízké příbuzných druhů byli detekováni ve všech molekulárních liniích, avšak ne vždy s vysokou podporou. Z podrodu *Curcuma* můžeme jmenovat např. *C. montana*, *C. aromatica*, *C. angustifolia* a *C. aeruginosa*, z podrodu *Hitcheniopsis* pak např. *C. parviflora*, *C. harmandii* a *C. thorelii*

(Záveská et al., 2012). Jako hybridní druhy s rodiči spadajícími do různých podrodů byly označeny *C. vamana*, *C. roscoeana* a *C. myanmarensis* s rodiči ze skupin *Curcuma* a *Hitcheniopsis* (Záveská et al., 2016). a jako další potenciálně hybridní druh byla navržena *C. candida* (pravděpodobně hybridní druhy viz Obrázek 3), která se v ITS fylogenezi zařadila do skupiny *Pierreana*, ale v cpDNA stromu do skupiny *Curcuma* (Záveská et al., 2012). Pomocí čtyř jaderných low-copy genů její hybridní původ však potvrzen nebyl – *C. candida* zde vykazovala postavení v rámci podrodu *Curcuma* (Záveská et al., 2016).



Obrázek 3: A: *Curcuma longa* L. (podrod *Curcuma*), B: *C. gracillima* (podrod *Hitcheniopsis*), C: *C. ecomata* (podrod *Ecomata*) D: *C. pierreana* (podrod *Ecomata*), E: *C. candida*, F: *C. vamana*, G: *C. roscoeana*, H: *C. myanmarensis*. A–G foto Jana Leong-Škorníčková, H – převzato z: <https://si.edu>.

2 MATERIÁL A METODY

2.1 ROSTLINNÝ MATERIÁL

Pro účely této diplomové práce bylo vybráno 27 druhů rodu *Curcuma*, 2 druhy příbuzného rodu *Camptandra* a *Siphonchilus aethiopicus* jako outgroup. Druhy byly vybírány tak, aby pokryly rodovou diverzitu – všechny čtyři molekulární linie včetně tří potenciálně hybridních druhů. Z podrodu *Curcuuma* bylo osekvenováno celkem 12 druhů, z podrodu *Hitcheniopsis* 4 druhy, z podrodu *Ecomatae* 13 druhů – konkrétně z linie *Ecomata* 6 a z linie *Pierreana* 7 druhů a potenciálně hybridní druhy – *C. vamana*, *C. myanmarensis* a *C. roscoeana*. Rostlinný materiál byl získán především z terénních sběrů Mgr. Jany Leong-Škorničkové, Ph.D., nebo z rostlin pěstovaných v Královské botanické zahradě v Edinburghu a Botanické zahradě v Singapuru. Do práce byla také zahrnuta transkriptomová data druhu *C. longa*, která byla získána v rámci projektu 1000 Plants (<http://www.onekp.com>). Detailní informace jako identifikační číslo, lokalita sběru materiálu, ploidie nebo příslušnost k molekulární linii viz Tabulka 1.

2.2 METODY

2.2.1 Molekulární analýzy

2.2.1.1 Izolace DNA

Izolace DNA byla provedena ze sušeného rostlinného materiálu pomocí Invisorb Spin Plant Mini Kit (Invitek). Přibližně 20 mg usušeného listu bylo drceno wolfram-karbidovými kuličkami v mlýnku Mixer Mill 400 MM (Retsch) po dobu 3 minut při 30 otáčkách/s. Další kroky izolace DNA probíhaly podle pokynů výrobce kromě několika následujících změn:

1. Místo 400 µl bylo přidáno 450 µl pufru Lysis Buffer P
2. Krok přidání Wash Bufferu II, centrifugace a vylití supernatantu byl zopakován dvakrát
3. Eluce byla provedena také dvakrát:
 - I. eluce: v 70 µl Elution Buffer D s délkou inkubace 30 minut
 - II. eluce: v 50 µl Elution Buffer D s inkubací 45 minut

Pro izolaci DNA byla použita centrifuga 5415 R (Eppendorf), Thermomixer comfort (Eppendorf). Výsledná koncentrace DNA byla změřena na přístroji Nanodrop™ 1000 Spectrofotometr (Thermo Fisher Scientific) a posléze na fluorimetru Qubit® 2.0 (Invitrogen). Izolace DNA byla provedena v DNA laboratoři Katedry botaniky Přírodovědecké fakulty Univerzity Karlovy v Praze (PřF UK).

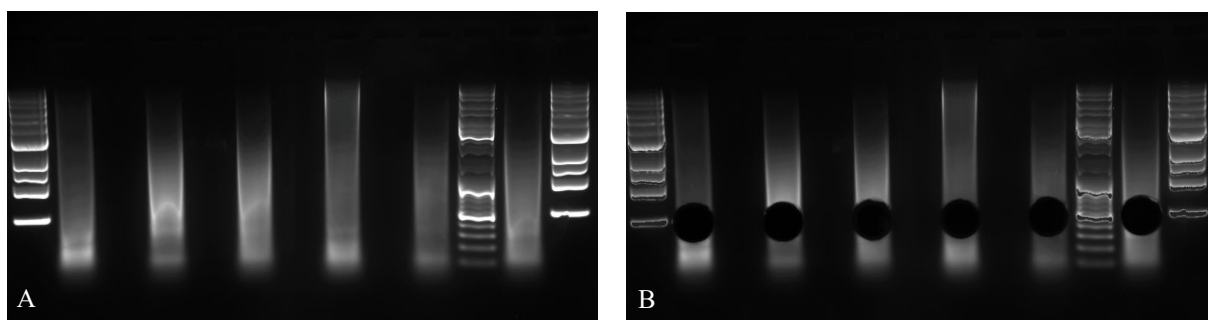
Tabulka 1. Rostlinný materiál zahrnutý v této práci. Linie – značí jednu ze čtyř hlavních molekulárních linií v rodu *Curcuma* (*Curcuma*, *Hitcheniopsis*, *Ecomata* a *Pierreana*), u podrodu *Curcuma* v závorce uvedena skupina velikosti genomu CI nebo CII, možné hybridní druhy označeny hvězdičkou. Identif. číslo – sběrné číslo rostliny, Ploidie – závorka označuje odhadovanou ploidiu odvozenou u jiného jedince téhož druhu. Data získána z: Leong-Škorničková et al. (2007), Záveská et. al. (2012, 2016) a Šída (unpubl.).

Druh	Linie	Identif. číslo	Lokalita	Ploidie
<i>Curcuma sp. nov aff. aeruginosa</i>	<i>Curcuma (CI)</i>	JLS 793	neznámá	9x
<i>Curcuma alismatifolia</i> Gagnep.	<i>Hitcheniopsis</i>	JLS 427	Thajsko	2x
<i>Curcuma bhatii</i> (R. M. Sm.) Škorničk. & M. Sabu	<i>Curcuma (CII)</i>	JLS 73446	Indie	6x
<i>Curcuma bicolor</i> Mood & K. Larsen	<i>Ecomata</i>	JLS 2215	Thajsko	(6x)
<i>Curcuma candida</i> (Wall.) Techapr. & Škorničk.	<i>Curcuma*</i>	JLS 606	Thajsko	6x
<i>Curcuma kannanorensis</i> R. Ansari et al.	<i>Curcuma (CII)</i>	JLS 84144	Indie	(6x)
<i>Curcuma corniculata</i> Škorničk.	<i>Ecomata</i>	JLS 1752	Laos	-
<i>Curcuma ecomata</i> Craib	<i>Ecomata</i>	T-11-09	Thajsko	6x
<i>Curcuma flammea</i> Škorničk.	<i>Ecomata</i>	JLS 1777	Laos	-
<i>Curcuma flaviflora</i> S.Q. Tong	<i>Ecomata</i>	T-11-113	Thajsko	6x
<i>Curcuma glans</i> K. Larsen & Mood	<i>Ecomata</i>	T-11-42	Thajsko	(6x)
<i>Curcuma gracillima</i> Gagnep.	<i>Hitcheniopsis</i>	Tran 48	Vietnam	2x
<i>Curcuma involucrata</i> (King ex Baker) Škorničk.	<i>Hitcheniopsis</i>	JLS 73246	Čína	(6x)
<i>Curcuma longa</i> L.	<i>Curcuma (CI)</i>	Soltis 2933	neznámá	(9x)
<i>Curcuma myanmarensis</i> (W.J. Kress) Škorničk.	<i>Curcuma*</i>	Kress 99-6507	Myanmar	6x
<i>Curcuma newmanii</i> Škorničk.	<i>Pierreana</i>	JLS 473	Vietnam	6x
<i>Curcuma oligantha</i> Trimen	<i>Curcuma (CII)</i>	JLS 73223	Srí Lanka	11x
<i>Curcuma pambrosima</i> Škorničk. & N.S. Lý	<i>Pierreana</i>	Ly 316	Vietnam	6x
<i>Curcuma parviflora</i> Wall.	<i>Hitcheniopsis</i>	MFN 2606	Thajsko	(2x)
<i>Curcuma petiolata</i> Roxb.	<i>Curcuma (CI)</i>	JLS 22	Thajsko	6x
<i>Curcuma pierreana</i> Gagnep.	<i>Pierreana</i>	JLS 73352	Thajsko	(6x)
<i>Curcuma plicata</i> Wall. ex Baker	<i>Curcuma (CI)</i>	Tran 80	Vietnam	(6x)
<i>Curcuma rhomba</i> Mood & K. Larsen	<i>Pierreana</i>	LNS 368	Vietnam	(6x)
<i>Curcuma roscoeana</i> Wall.	<i>Curcuma*</i>	MFN 2590	Thajsko	(6x)
<i>Curcuma singularis</i> Gagnep.	<i>Pierreana</i>	MFN 2413	Kambodža	6x
<i>Curcuma vamana</i> M. Sabu & Mangaly	<i>Hitcheniopsis*</i>	JLS 84156	Indie	2x
<i>Curcuma vitellina</i> Škorničk. & H.Đ. Trần	<i>Pierreana</i>	Tran 70	Vietnam	6x
<i>Curcuma xanthella</i> Škorničk.	<i>Pierreana</i>	LNS 348	Vietnam	(6x)
<i>Camptandra latifolia</i> Ridl.	-	M-11-8	Malajsie	-
<i>Camptandra ovata</i> Ridl.	-	M-11-3	Malajsie	-
<i>Siphonochilus aethiopicus</i> (Schweinf.) B.L.Burt	-	MFN & JLS 1995	Afrika	-

2.2.1.2 Příprava Hyb-Seq knihovny

Fragmentace DNA pro přípravu knihovny byla provedena sonikací na přístroji M220 Focused-ultrasonicator™ (Covaris). Pro získání fragmentů o průměrné délce 500-600 bp byly použity parametry: 50 W, 2% duty factor, 200 cyklů při teplotě 6°C po dobu 95 sekund. Ověření délky fragmentů bylo provedeno elektroforézou na 0,8% agarózovém gelu s ethidium bromidem v 1xTAE pufru při napětí 8 V/cm s velikostním standardem O'GeneRuler™ 100bp DNA Ladder Plus (Thermo Fisher Scientific) a Quick-Load® 1 kb DNA Ladder (New England BioLabs).

Příprava DNA knihovny byla provedena pomocí NEBNext Ultra DNA Library Prep Kit for Illumina (New England BioLabs). Prvním krokem byla oprava kohezních konců po štěpení DNA, A tailing a ligace adaptorů provedená podle pokynů výrobce. Následovalo přečištění produktů pomocí QIAquick PCR Purification Kit (Qiagen) podle pokynů výrobce kromě rozpuštění DNA ve 30 µl ddH₂O místo 50 µl pro lepší nanášení celého DNA vzorku na 1% agarózový gel z 1x low EDTA/TAE pufru s GelRed. Elektroforéza proběhla při napětí 8 V/cm, jako velikostní žebříček byl použit O'GeneRuler™ DNA Ladder Mix (Thermo Fisher Scientific). Z gelu byly vyřezány produkty o velikosti 500-600 bp (Obrázek 4). Následovalo další přečištění pomocí QIAquick Gel Extraction Kit (Qiagen) podle pokynů výrobce, pouze místo inkubace při 50°C po dobu 10 minut byly vzorky inkubovány při pokojové teplotě 15-30 minut (do rozpuštění gelu) a eluce DNA probíhala 5 minut do 20 µl ddH₂O.



Obrázek 4. Příklad selekce fragmentů o délce okolo 500-600 bp vyřezáváním z 1% agarózového gelu. **A:** jednotlivé vzorky před vyřezáním, **B:** gel po vyřezání terčíků příslušné velikosti.

Dále byly produkty namnoženy a oindexovány pomocí PCR s Q5 Hot Start HiFi PCR Master Mixem (součást NEBNext Ultra DNA Library Prep Kit for Illumina) a index primerů NEBNext Multiplex Oligos for Illumina (Set 1, NEB #E7335; složení reakční směsi a použitý cyklus viz Tabulka 2 a 3). Namnožené produkty byly dvakrát přečištěny pomocí Agencourt AMPure XP kitu (Beckman Coulter) v poměru 0,65:1 a 0,75:1 kit:DNA. Eluce byla provedena dvakrát – 1. do 40 µl ddH₂O, 2. do 25 µl ddH₂O. Výsledná kvalita byla ověřena

na 1% agarózovém gelu z 1x TAE pufru s GelRed při napětí 8 V/cm s velikostním standardem O'GeneRuler™ 100bp DNA Ladder Plus (Thermo Fisher Scientific) a Quick-Load® 1 kb DNA Ladder (New England BioLabs).

Tabulka 2. Složení PCR směsi s NEBNext Q5 HotStart Hifi polymerázou

Chemikálie	μl
NEBNext Q5 HotStart Hifi PCR Master Mix	25
Universal PCR primer (10 μM)	5
Index primer (10 μM)	5
DNA	15

Tabulka 3. Teplotní cyklus použitý při PCR s NEBNext Q5 HotStart Hifi polymerázou

Teplota	Doba	} 8x
98°C	30 s	
98°C	10 s	
65°C	1 min 15 s	
65°C	30 s	
7°C	∞	

Koncentrace DNA byla změřena na fluorimetru Qubit® 2.0 (Invitrogen) a jednotlivé vzorky byly zpoolovány do jedné zkumavky tak, aby výsledná knihovna obsahovala stejné množství DNA (v ng) od každého vzorku. Následovala koncentrace knihovny na objem 6 μl pomocí přístroje IR Concentrator NB 503CIR (N-Biotek).

Pro přípravu knihovny byl využit Mastercycler Pro (Eppendorf), centrifuga 5415 R (Eppendorf) a pro vizualizaci vzorků na agarózových gelech Standard UVT-28 M Transilluminator (Herolab), kamera Gel Logic 100 (Kodak) a softwar 1D Image Analysis (Kodak).

2.2.1.3 Obohacení Hyb-Seq knihovny a sekvenování

Knihovna DNA byla obohacena pomocí hybridizačních sond o 1 180 lokusů s 4 618 exony a celkovou délkou 1 571 800 bp, které byly navrženy pomocí skriptu Sondovač (Schmickl et al., 2016) z transkripromových a genome skimming dat druhů *C. ecomata* (genome skimming) a *C. longa* (transkriptom) podle doporučení ve Weitemier et al. (2014). Navržení hybridizačních sond provedl Mgr. Tomáš Fér, Ph.D. a Roswitha Elisabeth Schmickl, Ph.D. a nasynthetizovala je firma MYcroarray (Ann Arbor, Michigan, USA).

Obohacení DNA knihovny bylo provedeno pomocí kitu MYbaits® In-Solution Sequence Capture for Targeted High-Throughput Sequencing (MYcroarray). Do zkumavek byly připraveny tři mixy: Library Master Mix, Hybridization Master Mix, Capture Baits Master Mix (složení viz Tabulka 4). Podle schématu v Tabulce 5 byly jednotlivé Master mixy inkubovány a posléze smíchány do jedné zkumavky, kde probíhala hybridizace po dobu 26 hodin.

Tabulka 4. Složení Library Master Mix, Hybridisation Master Mix a Capture Baits Master Mix pro obohacení Hyb-Seq knihovny

Library Master Mix		Hybridisation Master Mix		Capture Baits Master Mix	
Block #1	2,5 µl	HYB #1	20 µl	Capture Baits	5 µl
Block #2	2,5 µl	HYB #2	0,8 µl		
Block #3	0,6 µl	HYB #3	8 µl	RNase Block	1 µl
DNA knihovna	5,9 µl	HYB #4	0,8 µl		

Po skončení hybridizace byly cílové (nahybridizované) fragmenty DNA odděleny z roztoku pomocí magnetických kuliček Dyneabeads® My One™ Streptavidin C1 magnetic beads (Thermo Fisher Scientific) potažených streptavidinem, který tvoří vazbu s biotinem inkorporovaným v hybridizačních sondách (Capture Baits). V prvním kroku bylo 50 µl magnetických kuliček třikrát promyto pomocí 200 µl Binding Buffer, poté rozpuštěno ve 20 µl Binding Buffer a inkubováno 45 minut při 65 °C. Následovalo přidání 500 µl Wash Buffer II, inkubace 5 minut při 65 °C a odpipetování supernatantu, které proběhlo celkem třikrát, a rozpuštění peletky v 30 µl GIBCO™ Distilled Sterile Water (Thermo Fisher Scientific).

Pomocí PCR s 2x KAPA HiFi HotStart ReadyMix byly cílové fragmenty DNA namnoženy (složení PCR směsi a teplotní cyklus viz Tabulka 6

Tabulka 5. Cyklus a schéma hybridizační reakce. LM – Library Master Mix, HM – Hybridization Master Mix, CB – Capture Baits Master Mix.

Teplota	95 °C	65 °C	65 °C	65 °C
Čas	5 min	3 min	2 min	26 hod
	LM	LM	LM	9,5 µl LM
		HM	HM	+ 10,5 ml HM
			CB	+ 6 ml CB

a 7) s primery P5_Zingiber: 5'-AATGATACGGCGACCACCGA a P7_Zingiber: 5'-CAAGCAGAAGACGGCATAACGA, posléze přečištěny s využitím QIAquick PCR Purification Kit (Qiagen) podle pokynů výrobce kromě eluce DNA do 30 µl GIBCO™ Distilled Sterile Water (Thermo Fisher Scientific) místo 50 µl a změřeno množství DNA ve vzorku pomocí fluorimetru Qubit® 2.0 (Invitrogen).

Při obohacení Hyb-Seq knihovny byl využit Mastercycler Pro (Eppendorf), Thermomixer comfort (Eppendorf) a centrifuga 5415 R (Eppendorf). Příprava Hyb-Seq knihovny a její obohacení bylo provedeno v DNA laboratoři Katedry botaniky PřF UK Mgr. Monikou Pospíšilovou a sekvenování vzorků proběhlo ve Středoevropském technologickém institutu (CEITEC) v Brně na přístroji Illumina MiSeq 2x 150 bp paired-end reads.

Tabulka 6. Složení PCR směsi s 2x KAPA HiFi Hot Start ReadyMix

Chemikálie	μl
2x KAPA HiFi HotStart ReadyMix	25
PCR primer mix (10 μM)	1,5
Hyb-Seq knihovna	5 - 15
GIBCO™ Distilled Sterile Water	8,5 - 18,5

Tabulka 7. Teplotní cyklus použitý při PCR s 2x KAPA HiFi Hot Start ReadyMix

Teplota	Doba	}12x
98°C	30 s	
98°C	20 s	
62°C	30 s	
72°C	45 s	
72°C	5 min	
4°C	∞	

2.2.2 Zpracování dat

2.2.2.1 Analýza jaderných low-copy genů

Tvorba fylogenetických stromů z kompletního datasetu

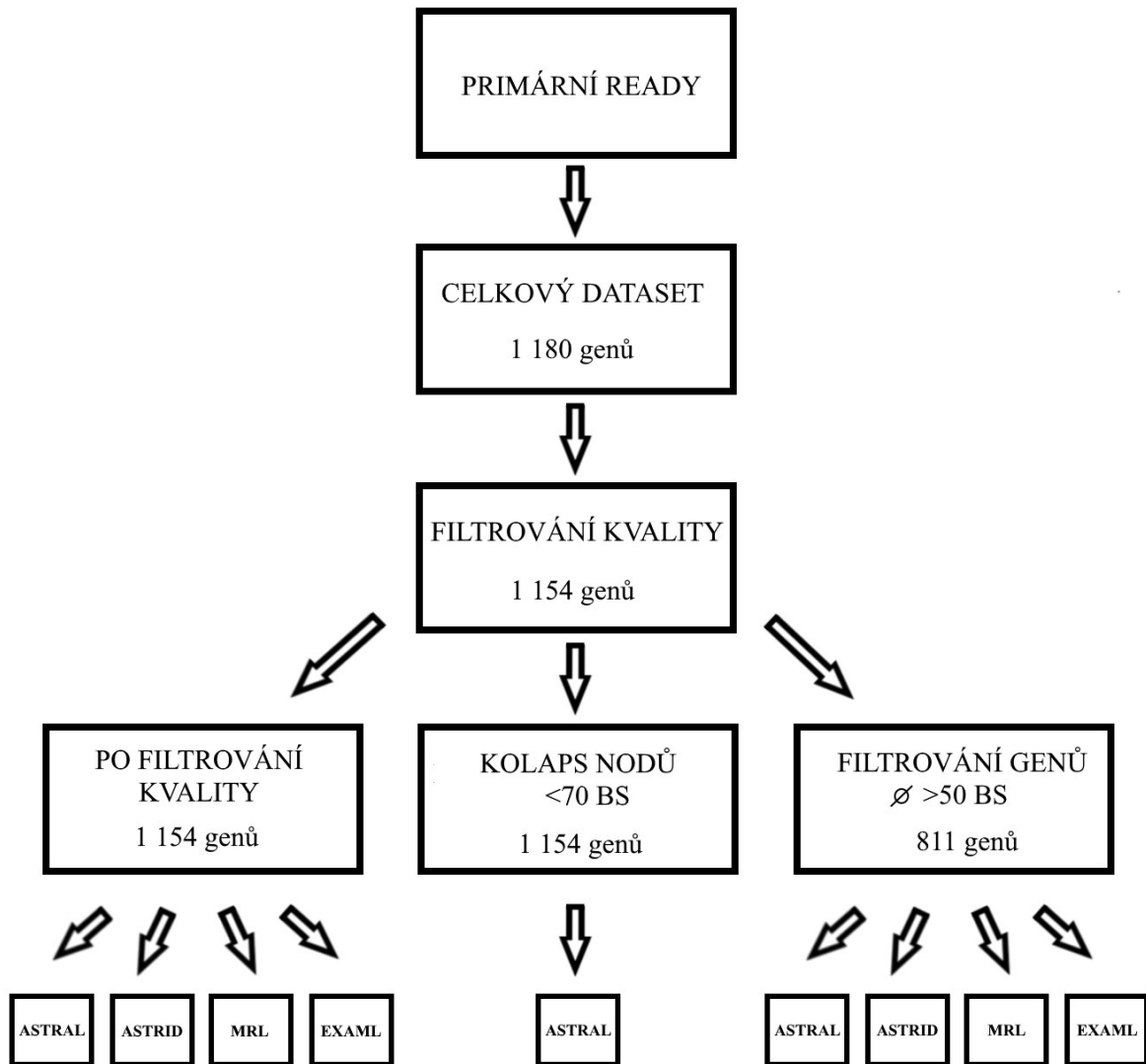
Fylogenetické stromy z přibližně tisíce jaderných low-copy genů byly vytvořeny pomocí pipeline HybPhyloMaker 1.2.2. DNA sekvence (primární ready) všech 30 druhů získané metodou Hyb-Seq byly nejprve pomocí skriptu HybPhyloMaker1_rawprocess.sh podrobeny filtrování a ořezání adaptorů (Fér, 2016). PhiX ready byly odstraněny za pomoci programů Bowtie 2 2.2.4 (Langmead and Salzberg, 2012) a SAMtools 0.1.19 (Li et al., 2009) a duplicitní ready pomocí programu FASTX 0.0.13 (Gordon & Hannon, 2010). Pro filtrování kvality a ořez adaptorů byl využit program Trimmomatic 0.32 (Bolger et al., 2014), přičemž byly odstraněny sekvence s kvalitou nižší než Q20 anebo kratší než 36 bp. Jako další produkt tohoto skriptu byla vygenerována tabulka se souhrnnými statistikami, jako je původní počet nebo množství odfiltrovaných readů.

Druhým krokem bylo namapování readů s využitím programu Geneious 7.0.6 (Kearse et al., 2012) na tzv. pseudoreferenci – sekvenci exonů využitých při tvorbě hybridizačních prób navzájem oddělených 400 N. Nastavení pro mapování bylo ponecháno podle doporučení v manuálu z Fér (2016; Allow gaps: Maximum Per Read 15 %, Word Length 14, Ignore words repeated more than 20 times, Maximum Mismatches Per Read 30%, Maximum Gap Size 10, Index Word Length 12 a Maximum Ambiguity 4). Vyexportovaná konsenzuální sekvence (Threshold: 0% – Majority a bez zaškrtnutí možnosti „Trim to reference“ a „Ignore Gaps“) byla dále zpracována skriptem HybPhyloMaker2_generatepslx.sh, který rozdělí konsenzuální sekvenci vzorků na jednotlivé exony, ty jsou pomocí programu BLAT 34 (BLAST-like alignment tool; Kent, 2002) porovnány se sekvencemi původních exonů

(minimum similarity threshold = 90) a výsledky jsou uloženy jako soubory „PSLX“, které jsou následně využívány jako vstupní soubory dalších analýz (Fér, 2016). V tomto kroku byla do analýzy také zahrnuta také transkriptomová data z *Curcuma-longa_JQCX* (1000 Plants; <http://www.onekp.com>).

Pomocí skriptu `HybPhyloMaker3_processpslx.sh` byly sekvence jednotlivých exonů s využitím MAFFT 7.029 s automatickým nastavením (Katoh and Standley, 2013) zalignovány a jednotlivé exony byly zkonkatenovány do genů. V dalším kroku byla odstraněna chybějící data za použití skriptu `HybPhyloMaker4_missingdataremoval.sh`. Byly ponechány sekvence genů obsahujících méně než 20 % chybějících dat a geny obsahující alespoň 75 % analyzovaných druhů. Následovalo vytvoření genových stromů pomocí skriptu `HybPhyloMaker5a_RAxML_for_selected.sh` (Fér, 2016) využívající metodu Maximum Likelihood implementovanou v programu RAxML 8.2.4 (Stamatakis, 2014) s 1 000 bootstrapovými replikacemi a souhrnné tabulky a histogramy byly vygenerovány pomocí skriptu `HybPhyloMaker5a2_RAxML_trees_summary.sh` (Fér, 2016) s využitím programu R 3.2.3 (R Core Team, 2017). Následovalo zakořenění stromů druhem *Siphonochilus aethiopicus* pomocí Newick Utilities 1.6 (Junier and Zdobnov, 2010) a sloučení všech vygenerovaných genových stromů do jednoho souboru skriptem `HybPhyloMaker6_roottrees.sh`. Skripty `HybPhyloMaker7a_astral.sh`, `HybPhyloMaker7b_astrid.sh`, `HybPhyloMaker7c_mrl.sh` a `HybPhyloMaker7f_concatenated-ExaML.sh` byly využity pro odhad species tree z 1 154 genů získaných po filtrování kvality pomocí programu ASTRID 1.0 (Vachaspati and Warnow, 2015) s 1 000 bootstrapovými replikacemi, ASTRAL 4.10.10 (Mirarab et al., 2014), metody MRL (Nguyen et al., 2012) a konkatenace v programu ExaML 3.0.15 (Kozlov et al., 2015) se 100 bootstrapovými replikacemi. Pro tvorbu stromu v programu ExaML (Kozlov et al., 2015) byl využit dataset rozdělený na partitions po jednotlivých genech připravený pomocí programu PartitionFinder2 (Lanfear et al., 2014) s nastavením `branchlengths = linked`; `models = GTR+G` a `model_selection = AICc` a pro tvorbu MRL stromu pak MRP matice (Mirarab, 2015) a RAxML 8.2.4 s 1 000 bootstrapovými replikacemi (Stamatakis, 2014). Dále byly vytvořeny tzv. „update“ fylogenetické stromy z 811 genů – došlo k vyřazení celkem 343 genů s nízkou průměrnou bootstrapovou podporou (<50). Byl použit skript `HybPhyloMaker8_update_trees.sh` a dále pak skripty `HybPhyloMaker6_roottrees.sh`, `HybPhyloMaker7a_astral.sh`, `HybPhyloMaker7b_astrid.sh`, `HybPhyloMaker7c_mrl.sh` a `HybPhyloMaker7f_concatenatedExaML.sh`, vždy s 1 000 bootstrapovými replikacemi kromě konkatenace, kde bylo použito pouze 100

replikací (Fér, 2016). Jako poslední byl vytvořen strom metodou ASTRAL (Mirarab et al., 2014) z 1 154 genových stromů se zkolabovanými nody s bootstrapovou podporou <70, kolaps proveden v programu TreeCollapserCL 4 (Hodcroft, 2016), pomocí skriptu HybPhyloMaker10a_Ast-ral_on_collapsed.sh (Fér, unpubl.), schéma analýzy dat jaderných low-copy genů na Obrázku 5.



Obrázek 5. Schéma analýzy jaderných low-copy genů.

Tvorba fylogenetických stromů z podsouboru genů

Pro zhodnocení počtu genů potřebných k věrohodné rekonstrukci species tree byly vytvořeny podsoubory genů – 5, 20, 100, 200 a 500, na jejichž základě byly odvozovány fylogenetické stromy s využitím čtyř metod rekonstrukce species trees – ASTRAL, ASTRID, MRL a konkatence ExaML. Ke zhodnocení vlivu hybridních druhů s rodiči ze vzdálených

skupin na přesnost odvozované fylogeneze byly vytvořeny dva datasety – kompletní s hybridními druhy a kontrolní bez hybridů.

Dvakrát 1 154 genových stromů se 100 bootstrapovými replikacemi vytvořených pomocí programu RAxML 8.2.4 (Stamatakis, 2014) v rámci pipeline HybPhyloMaker (Fér, 2016) bylo využito jako vstupní datasety pro fylogenetické analýzy založené na podsouboru genů. První dataset zahrnoval všech 31 druhů, z druhého kontrolního datasetu byly odstraněny pravděpodobné hybridní druhy – *C. myanmarensis*, *C. roscoeana* a *C. vamana*. Pomocí webového serveru www.random.org byly jednotlivé genové stromy náhodně seřazeny a pro tvorbu fylogenetických stromů bylo vybráno vždy 5, 20, 100 a 500 genů s deseti opakováními. Vybrány byly geny 1-100, 101-200 atd. pro analýzu 100 genů, 1-5, 101-105 atd. pro analýzu založenou na pěti genech a 51-70, 151-170 atd. pro analýzu 20 genů – tzn. každý genový strom mohl být zahrnut pouze v jednom opakování dané analýzy. Protože celkový počet genů byl pouze 1 154, byly pro analýzu 500 genů kombinovány jednotlivé stovky genů tak, aby každý gen byl zahrnut v pěti analýzách. Genové stromy byly založeny druhem *Siphonochilus aethiopicus* pomocí Newick Utilities 1.6 (Junier and Zdobnov, 2010) a pro všechny datasety vytvořeny fylogenetické stromy pomocí programu ASTRID 1.0 (Vachaspati and Warnow, 2015), ASTRAL 4.10.10 (Mirarab et al., 2014), metody MRL (Nguyen et al., 2012) a konkatence v programu ExaML 3.0.15 (Kozlov et al., 2015) s využitím skriptů pipeline HybPhyloMaker 1.2.2 (Fér, 2016). Bylo získáno 10 species trees se 100 bootstrapovými replikacemi pro každou kategorii počtu vstupních genů a metodu tvorby fylogenetického stromu, mezi kterými byla spočítána Robinson-Fouldova vzdálenost (RF distance) definovaná jako nejmenší počet transformací, které přemění jednu topologii stromu na druhou (Robinson and Foulds, 1981) pomocí programu PhyBin 0.3 (Newton and Newton, 2013). Jako Clustering option byla zvolena metoda UPGMA (Unweighted Pair Group Method) a data byla analyzována v modu „tolerant“ kvůli některým chybějícím taxonům ve fylogenetických stromech. RF distance (Robinson and Foulds, 1981) byla také spočítána mezi stromy vzniklými z podsouboru genů a fylogenetickými stromy založenými na 1 154 jaderných low-copy genech se 100 bootstrapovými replikacemi. Jednotlivé vzdálenosti byly vyděleny maximální RF vzdáleností: $RF_{\max} = 2(n - 3)$, kde n = počet taxonů zahrnutých v analýze – pro dataset obsahující hybridy $RF_{\max} = 56$, pro kontrolní dataset $RF_{\max} = 50$ a vynásobeny 100 pro získání relativní RF vzdálenosti v procentech, které byly vyneseny do krabicových grafů (boxplotů) pomocí programu PAST 2.17 (Hammer et al., 2001).

Všechny úlohy pipeline Fér (2016) kromě lokálního mapování readů v programu Geneious (Kearse et al., 2012) byly prováděny na strojích Národní Gridové Infrastruktury MetaCentrum (<https://metavo.metacentrum.cz/>).

2.2.2.2 Analýza chloroplastového genomu

Tvorba fylogenetického stromu z protein kódujících oblastí

Pipeline HybPhyloMaker 1.2.2 byla využita také k tvorbě chloroplastového fylogenetického stromu založeného na kódujících chloroplastových lokusech. Vyfiltrované ready pomocí skriptu HybPhyloMaker1_rawprocess.sh (Fér, 2016) byly namapovány v programu Geneious 7.0.6 (Kearse et al., 2012) na pseudoreferenci obsahující pouze kódující úseky genomu *C. roscoeana* (Barrett et al., 2014) oddělené 400N s nastavením shodným jako u mapování jaderných genů.

Dále bylo využito skriptů HybPhyloMaker2_generatepslx.sh, kde byla do analýzy zahrnuta *Curcuma-longa_JQCX* (1000 Plants; <http://www.onekp.com>) a naopak vyřazena *Camptandra-latifolia_S4* (z důvodu velkého množství chybějících dat), HybPhyloMaker3_processpslx.sh, HybPhyloMaker4_missingdataremoval.sh, HybPhyloMaker5a_RAxML_for_selected.sh, HybPhyloMaker5a2_RAxML_trees_summary.sh, HybPhyloMaker6_roottrees.sh a Hyb-PhyloMaker7f_concatenatedExaML.sh (Fér, 2016). Při filtrování dat byly ponechány sekvence genů obsahujících méně než 75 % chybějících dat a geny obsahující alespoň 75 % analyzovaných druhů a pro tvorbu fylogenetického stromu byl využit konkatenovaný dataset obsahující pouze kódující úseky 86 genů taktéž rozdělené na partitions po jednotlivých genech. Fylogenetický strom byl vytvořen v programu ExaML 3.0.15 (Kozlov et al., 2015) s 1 000 bootstrapovými replikacemi. Dataset byl rozdělen na partitions po jednotlivých genech připravený pomocí programu PartitionFinder2 (Lanfear et al., 2014) s nastavením branchlengths = linked; models = GTR+G a model_selection = AICc.

Tvorba fylogenetického stromu z celé cpDNA – mapování na referenci

Jako referenční genom pro tvorbu fylogenetického stromu mapováním na referenci byl využit chloroplastový genom *C. roscoeana* (Barrett et al., 2014), na který byly postupně namapovány ready ze 30 druhů vyfiltrované pomocí skriptu HybPhyloMaker1_rawprocess.sh (Fér, 2016). Pro krok mapování byl využit program Geneious 7.0.6 (Kearse et al., 2012) s nastavením shodným jako u jaderných low-copy lokusů. Vyexportované konsenzuální sekvence (Threshold: 0 %) byly zalignovány pomocí webového serveru MAFFT 7 (Katoh and

Standley, 2013) s defaultním nastavením kromě matice 1PAM a upraveny v programu BioEdit 7.2.5 (Hall, 1999). Anotace byla provedena pomocí programu DOGMA (Dual Organellar GenoMe Annotator; Wyman et al., 2004) a z datasetu obsahujícího large single copy (LSC) region, small single copy region (SSC) i invertovanou repetici (IR) rozděleného na partitions (geny, r-RNA a nekódující úseky – t-RNA a partitions kratší než 75 bp byly sloučeny se sousední partition) byl pomocí programu RAxML 8.2.4 (Stamatakis, 2014) s 1 000 bootstrapovými replikacemi a modelem GTR+G vytvořen species tree.

Tvorba fylogenetického stromu z celé cpDNA – de-novo assembly

Pro zkonstruování fylogenetického stromu pomocí *de-novo* assembly byly využity ready 30 druhů namapované na referenci *C. roscoeana* (Barrett et al., 2014) v předešlé analýze (Mapování na referenci). Ty byly vyexportovány v podobě BAM souborů a pomocí skriptu HybPipe10_spadesaabacas.sh (Fér, unpubl.) podrobeny *de-novo* assembly v programu SPAdes 3.6.0 (Bankevich et al., 2012) s hodnotami k-mer = 55, 77, 89 a threads = 6 a následně byly vybrány kontigy delší než 1 000 bp a 800–1 000 bp dlouhé kontigy s průměrným pokrytím 3 a více, které byly opět namapovány na plastom *C. roscoeana* (Barrett et al., 2014) v programu Geneious 7.0.6 (Kearse et al., 2012) s následujícími parametry: High sensitivity, Allow gaps: Maximum Per Read 20 %, Word Length 12, Maximum Mismatches Per Read 40 %, Maximum Gap Size 100, Index Word Length 11 a Maximum Ambiguity 16. Po tomto kroku mělo následovat rozdělení kontigů přecházejících přes konec reference na dva pro lepší mapování obou částí, avšak většina vzorků přecházela pouze 89 bp (z ~160 000 bp), proto byla ořezána reference a 89 bp dlouhý úsek přemístěn na opačný konec referenční sekvence. Mapování na referenci v programu Geneious 7.0.6 (Kearse et al., 2012) bylo zopakováno, tentokrát s Highest sensitivity (Allow gaps: Maximum Per Read 20 %, Word Length 10, Maximum Mismatches Per Read 50 %, Maximum Gap Size 500, Index Word Length 10 a Maximum Ambiguity 16). Z analýzy byly vyřazeny čtyři vzorky (Camptandra-latifolia_S4, Curcuma-gracillima_S1, Curcuma-singularis_S2 a Curcuma-vitellina_S3), jejichž délka kontigů nepřesahovala 25 % délky reference. Konce jednotlivých kontigů, které se špatně shodovaly s referencí byly ořezány a sekvence zalignovány pomocí webového serveru MAFFT 7 (Kato and Standley, 2013) s defaultním nastavením kromě matice 1PAM a ořezány v programu BioEdit 7.2.5 (Hall, 1999). Anotace chloroplastového genomu byla provedena pomocí programu DOGMA (Wyman et al., 2004) a z datasetu obsahující LSC, SSC i IR region rozděleného na partitions (viz Mapování

na referenci) byl pomocí programu RAxML 8.2.4 (Stamatakis, 2014) s 1 000 bootstrapovými replikacemi a modelem GTR+G vytvořen species tree.

2.2.2.3 *Fylogeneze ribozomální DNA*

Tvorba fylogenetického stromu z rDNA – de-novo assembly

Nejprve byly ready 30 druhů vyfiltrované pomocí skriptu HybPhylo-Maker1_rawprocess.sh (Fér, 2016) namapovány na hybridní referenci složenou z ETS (external transcribed spacer) *C. alismatifolia* (Hayakawa et al., 2011) a zbylé části rDNA cistronu *Elettaria cardamomum* (Fér, unpubl.) s využitím programu Geneious 7.0.6 (Kearse et al., 2012) a shodného nastavení s mapováním jaderných low-copy lokusů. Tyto ready byly vyexportovány v podobě BAM souborů a pomocí skriptu HybPipe10_spadesabacas.sh (Fér, unpubl.) podrobeny *de-novo* assembly v programu SPAdes 3.6.0 (Bankevich et al., 2012) s hodnotami k-mer = 55, 77, 89 a threads = 6 a následně byly vybrány kontigy delší než 1 000 bp a 800–1 000 bp dlouhé kontigy s průměrným pokrytím 3 a více, které byly namapovány zpět na hybridní referenci v programu Geneious 7.0.6 (Kearse et al., 2012) s nastavením High sensitivity. Špatně namapované konce byly odstraněny, kontigy zalignovány pomocí webového serveru MAFFT 7 (Katoh and Standley, 2013) s defaultním nastavením kromě matice 1PAM a ořezány koncové sekvence v programu BioEdit 7.2.5 (Hall, 1999). Dataset byl rozdělen na sedm partitions (ETS; 18S; ITS1; 5,8S; ITS2; 26S a NTS) a pomocí programu RAxML 8.2.4 (Stamatakis, 2014) s 1 000 bootstrapovými replikacemi a modelem GTR+G vytvořen species tree.

Tvorba fylogenetického stromu z rDNA – mapování na referenci

Pro mapování na referenci byly využity ready namapované v předešlé analýze (*De-novo* assembly) a jako reference sloužila sekvence *C. rhomba* vzniklá při *de-novo* assembly ořezaná na původní délku hybridní referenční sekvence (byly odstraněny oba přesahující konce, které s rDNA pravděpodobně nesouvisí). Mapování bylo provedeno v programu Geneious 7.0.6 (Kearse et al., 2012) se shodným nastavením jako při mapování jaderných low-copy genů. Vyexportované konsenzuální sekvence (Threshold: 0 %) byly zalignovány pomocí webového serveru MAFFT 7 (Katoh and Standley, 2013) s defaultním nastavením kromě matice 1PAM a ořezány koncové sekvence v programu BioEdit 7.2.5 (Hall, 1999). Dataset byl rozdělen na sedm partitions (ETS; 18S; ITS1; 5,8S; ITS2; 26S a NTS) a pomocí

programu RAxML 8.2.4 (Stamatakis, 2014) s 1 000 bootstrapovými replikacemi a modelem GTR+G vytvořen species tree.

Všechny úlohy pipeline Fér (2016) kromě lokálního mapování readů v programu Geneious (Kearse et al., 2012) a výpočet fylogenetických stromů pomocí programu RAxML 8.2.4 (Stamatakis, 2014) byly prováděny na strojích Národní Gridové Infrastruktury MetaCentrum (<https://metavo.metacentrum.cz/>). Alignmenty jednotlivých jaderných genů využitých k rekonstrukci fylogeneze, stejně jako cpDNA a rDNA alignmenty jsou uvedeny v samostatné Příloze.

2.2.2.4 Srovnání species trees z různých molekulárních markerů

K porovnání topologií species trees vytvořených z různých molekulárních markerů byl vybrán strom z 811 konkatenovaných jaderných low-copy genů, strom z celého chloroplastového genomu z mapování na referenci a strom z ribozomálního cistronu z *de-novo* assembly. Pomocí grafického editoru paint.net (www.getpaint.net) byly vytvořeny dva obrázky srovnávající topologii jaderného species tree s plastomovým stromem a jaderného species tree s ribozomálním stromem.

2.2.2.5 Testování hybridního původu vybraných druhů

Pro testování přítomnosti hybridních druhů v datasetu byly provedeny tři různé analýzy. První z nich byla tvorba fylogenetických sítí zobrazujících retikulární vztahy mezi jednotlivými druhy pomocí Neighbor-Net algoritmu (Bryant and Moulton, 2004) a programu PhyloNet (Than et al., 2008), druhá testování přítomnosti předem definovaných monofyletických skupin druhů v jednotlivých genových stromech a třetí lineage movement analýza (Smith and Dunn, 2008) zobrazující procento výskytu testovaného druhu v alternativních pozicích konsenzuálního stromu na základě vstupního souboru genových stromů.

Fylogenetické sítě

Pro vizualizaci retikulárních vztahů v jednotlivých datasetech byly vytvořeny fylogenetické sítě s využitím algoritmu Neighbor-Net (Bryant and Moulton, 2004) implementovaného v programu SplitsTree 4.14.4 (Huson and Bryant, 2005) s uncorrected pairwise distance maticí. Neighbor-Net analýza byla provedena pro konkatenovaný dataset z 1 154 i 811 jaderných genů a kódujících sekvencí chloroplastové DNA vytvořených pomocí pipeline HybPhyloMaker (Fér, 2016), dále z rDNA a cpDNA sekvencí vzniklých jak

mapováním na referenci, tak pomocí *de-novo* assembly, celkem bylo tedy vytvořeno sedm sítí. Pro lepší znázornění vztahů mezi jednotlivými druhy kurkum byly z analýzy vyřazeny taxony tvořící outgroup.

Pomocí Maximum Pseudo-Likelihood přístupu pro odvozování fylogenetických sítí (Yu and Nakhleh, 2015) v programu PhyloNet 3.6.1 (Than et al., 2008) bylo vytvořeno pět fylogenetických sítí z 1 154 jaderných low-copy genů 28 druhů rodu *Curcuma* a dvou druhů rodu *Camptandra* (vzdálenější outgroup byl vyřazen pro možné ovlivnění analýzy) konstruovaných metodou RAxML (Stamatakis, 2014) s maximálně 20 % chybějících dat na lokus a přítomností alespoň 75 % taxonů. Sítě byly odvozovány pouze z topologií jednotlivých genů a jednoho runu pseudo-likelihood analýzy s nastavením maximálně čtyř hybridizačních událostí.

Lineage movement analýza

Jako vstupní genové stromy pro tuto analýzu byly použity stromy z jaderných low-copy genů vytvořené pomocí pipeline HybPhyloMaker 1.2.2 (Fér, 2016). Bylo získáno celkem 970 se 100 bootstrapovými replikacemi, které obsahovaly všech 29 taxonů (druhy *Siphonochilus aethiopicus* a *Camptandra latifolia* byly z analýzy vyřazeny kvůli velkému procentu chybějících dat) a maximálně 20 % chybějících pozic u jednotlivých genů. Následovala lineage movement analýza v programu PhyUtility 2.2 (Smith and Dunn, 2008). Jako konsenzuální strom byl využit konkatenovaný Maximum Likelihood species tree vytvořený v programu ExaML (Kozlov et al., 2015) založený na 1 154 genech. Procenta přítomnosti druhů na odlišných pozicích v genových stromech byly sečteny pro sedm skupin – *Curcuma I* a *Curcuma II*, podrod *Curcuma* celkově, *Hitcheniopsis*, *Ecomata* a *Pierreana* a „Hybridi“. Zařazení jednotlivých druhů podle Tabulky 1 a sběrnou nemonofyletickou skupinu „Hybridi“ tvořily tři potenciálně meziliniovní hybridy *C. myanmarensis*, *C. roscoeana* a *C. vamana*, *C. candida* byla pro celkový souhrn zařazena do skupiny *Curcuma II* kvůli její bazální pozici v konsenzuálním stromu, pro detailní tabulku zabývající se pouze hybridními druhy vyčleněna zvlášť.

Testování monofylie

Vstupní genové stromy z jaderných low-copy genů pro testování monofylie jednotlivých skupin byly vytvořeny pomocí pipeline HybPhyloMaker 1.2.2 metodou RAxML se 100 bootstrapovými replikacemi (Fér, 2016). Při filtrování dat byly ponechány sekvence genů obsahujících méně než 20 % chybějících dat a geny obsahující alespoň 75 %

analyzovaných druhů. Byly vytvořeny celkem čtyři datasety, na kterých byla monofylie později testována:

1. Všechny druhy kromě 3 meziliniových hybridů – *C. myanmarensis*, *C. roscoeana* a *C. vamana* pro testování hybridního vzniku *C. candida* s rodiči spadajícími do skupiny

Curcuma I a *Curcuma II* a pro kontrolní testy nehybridních druhů

2. Všechny druhy kromě *C. roscoeana* a *C. vamana* pro testování hybridního vzniku *C. myanmarensis* s rodiči spadajícími do podrodů *Curcuma* a *Hitcheniopsis*
3. Všechny druhy kromě *C. myanmarensis* a *C. vamana* pro testování hybridního vzniku *C. roscoeana* s rodiči spadajícími do podrodů *Curcuma* a *Hitcheniopsis*
4. Všechny druhy kromě *C. roscoeana* a *C. myanmarensis* pro testování hybridního vzniku *C. vamana* s rodiči spadajícími do podrodů *Curcuma* a *Hitcheniopsis*

Ze souboru genových stromů byly nejprve vybrány a odstraněny ty, které neobsahovaly testovaný druh pomocí skriptu `grepoutgroup.sh` (Fér, unpubl.; Tabulka 8). Jednalo se pouze o malé množství genů (do 40 genů na jeden dataset, obvykle však kolem 10 genů). Dále byly zkolabovány větve s bootstrapovou podporou nižší než 70 pomocí programu `TreeCollapserCL 4` (Hodcroft, 2016) a testy monofylie probíhaly v programu `R 3.3.3` (R Core Team, 2017) pomocí funkce „`is.monophyletic`“ z knihovny „`ape`“ implementované ve skriptu `monophyly.R` (Fér, unpubl.; viz Tabulka 9). Pro testování monofylie bylo zvoleno celkem 12 skupin:

1. C – druhy podrodu *Curcuma* bez testovaného druhu
2. Chyb – podrod *Curcuma* včetně testovaného druhu
3. H – druhy podrodu *Hitcheniopsis* bez testovaného druhu
4. Hhyb – druhy podrodu *Hitcheniopsis* a testovaný druh
5. CHhyb – druhy podrodů *Curcuma* a *Hitcheniopsis* včetně testovaného druhu
6. E – druhy linie *Ecomata* bez testovaného druhu
7. Ehyb – druhy linie *Ecomata* a testovaný druh
8. P – druhy linie *Pierreana* bez testovaného druhu
9. Phyb – druhy linie *Pierreana* a testovaný druh

Tabulka 8. Příkazy použité při filtrování stromů neobsahujících příslušného hybrida - `grepoutgroup.sh` (Fér, unpubl.), hybrid = jméno hybridního druhu.

```
for i in $(ls *Assembly*); do
echo $i
if grep -q "hybrid" $i; then
echo -e "$i\t1" >> geneswithoutgroup.txt
else
echo -e "$i\t0" >> geneswithoutgroup.txt
fi
done
```


10. EPhyb – druhy linie *Ecomata* a *Pierreana* včetně testovaného druhu

11. C1 – skupina *Curcuma I* bez testovaného druhu

12. C1hyb – *Curcuma I* včetně testovaného druhu

Výstupem byla TRUE/FALSE matice s hodnotami pro každý gen a testovanou skupinu a geny obsahující danou monofyletickou skupinu byly sečteny. Jako geny, kde se vyskytuje hybrid na bázi podrodu *Curcuma*, byly považovány geny monofyletické jak pro skupinu C, tak pro Chyb a jako geny, kde je hybrid ve skupině *Curcuma II* ty geny, které splnily podmínku: $C1 = \text{true} \wedge C1\text{hyb} = \text{false} \wedge C = \text{false} \wedge \text{Chyb} = \text{true}$. Jako geny, kde hybrid na bazální pozici obou podrodů *Curcuma* a *Hitcheniopsis*, byly považovány geny, které splnily podmínku: $C = \text{true} \wedge \text{Chyb} = \text{false} \wedge H = \text{true} \wedge H\text{hyb} = \text{false} \wedge CH\text{hyb} = \text{true}$.

Tabulka 9. Příkazy použité při testech monofylie – skript monophyly.R (Fér, unpubl.). Za "hybrid_Sxxx" byly dosazovány jednotlivé druhy (které byly dříve odstraněny z původních skupin).

```
library (ape)
library (maps)
library (phytools)

E=c ("Curcuma-bicolor_S155","Curcuma-corniculata_S317","Curcuma-glans_S316","Curcuma-
flaviflora_S315","Curcuma-flammea_S318","Curcuma-ecomata_S162")

P=c ("Curcuma-pambrosima_S319","Curcuma-newmanii_S314","Curcuma-xanthella_S320","Curcuma-
vitellina_S3","Curcuma-singularis_S2","Curcuma-rhomba_S160","Curcuma-pierreana_S313")

Ehyb=c ("Curcuma-hybrid_Sxxx","Curcuma-bicolor_S155","Curcuma-corniculata_S317","Curcuma-
glans_S316","Curcuma-flaviflora_S315","Curcuma-flammea_S318","Curcuma-ecomata_S162")

Phyb=c ("Curcuma-hybrid_Sxxx","Curcuma-pambrosima_S319","Curcuma-newmanii_S314","Curcuma-
xanthella_S320","Curcuma-vitellina_S3","Curcuma-singularis_S2","Curcuma-
rhomba_S160","Curcuma-pierreana_S313")

EPhyb=c ("Curcuma-hybrid_Sxxx","Curcuma-bicolor_S155","Curcuma-corniculata_S317","Curcuma-
glans_S316","Curcuma-flaviflora_S315","Curcuma-flammea_S318","Curcuma-
ecomata_S162","Curcuma-pambrosima_S319","Curcuma-newmanii_S314","Curcuma-
xanthella_S320","Curcuma-vitellina_S3","Curcuma-singularis_S2","Curcuma-
rhomba_S160","Curcuma-pierreana_S313")

C=c ("Curcuma-plicata_S204","Curcuma-petiolata_S175","Curcuma-oligantha_S157","Curcuma-
aeruginosa_S169","Curcuma-bhatii_S165","Curcuma-candida_S156","Curcuma-
cannanorensis_S164","Curcuma-longa_JQCX")

Chyb=c ("Curcuma-hybrid_Sxxx","Curcuma-plicata_S204","Curcuma-petiolata_S175","Curcuma-
oligantha_S157","Curcuma-aeruginosa_S169","Curcuma-bhatii_S165","Curcuma-
candida_S156","Curcuma-cannanorensis_S164","Curcuma-longa_JQCX")

H=c ("Curcuma-parviflora_S159","Curcuma-alismatifolia_S158","Curcuma-
involucrata_S205","Curcuma-gracillima_S1")

Hhyb=c ("Curcuma-hybrid_Sxxx","Curcuma-parviflora_S159","Curcuma-
alismatifolia_S158","Curcuma-involucrata_S205","Curcuma-gracillima_S1")

CHhyb=c ("Curcuma-hybrid_Sxxx","Curcuma-plicata_S204","Curcuma-petiolata_S175","Curcuma-
oligantha_S157","Curcuma-aeruginosa_S169","Curcuma-bhatii_S165","Curcuma-
candida_S156","Curcuma-cannanorensis_S164","Curcuma-longa_JQCX","Curcuma-
parviflora_S159","Curcuma-alismatifolia_S158","Curcuma-involucrata_S205","Curcuma-
gracillima_S1")

C1=c ("Curcuma-aeruginosa_S169","Curcuma-longa_JQCX","Curcuma-plicata_S204","Curcuma-
petiolata_S175")
```

```

Clhyb=c ("Curcuma-aeruginosa_S169","Curcuma- hybrid_Sxxx","Curcuma-longa_JQCX","Curcuma-
plicata_S204","Curcuma-petiolata_S175")

groupsToTest=list(E, Ehyb, P, Phyb, EPhyb, C, Chyb, H, Hhyb, CHhyb,C1,C1hyb)
groupsToTestNames=c("E","Ehyb","P","Phyb","EPhyb","C","Chyb","H","Hhyb","CHhyb","C1","C1hyb
")

trees_files <- dir(pattern="*.result")
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)
  } else {
    mono <- NA
  }
  return(c(mono))
}
ismonophyl <- trees_files
ismonophyl <- data.frame(tree=trees_files)
x <- 0
for(i in groupsToTest) {
  x <- x + 1
  ismonophyl[[x+1]] <- lapply(trees_files, monophyletic, monolist = i)
}
colnames(ismonophyl) <- c("tree", groupsToTestNames)
ismonophyl <- as.matrix(ismonophyl)
write.csv(ismonophyl, file="monophylyResults.txt", quote=FALSE, row.names=FALSE)

```

3 VÝSLEDKY

Celkově bylo osekvenováno 27 druhů rodu *Curcuma*, 2 druhy rodu *Camptandra* (*C. latifolia* a *C. ovata*) a *Siphonochilus aethiopicus* jako outgroup. Bylo získáno celkem 260 375–3 119 433 readů na vzorek s průměrem 1 803 532 readů. Při filtrování kvality bylo vyřazeno od 2,3 % do 8,1 % readů s průměrnou hodnotou 4,6 % readů, při filtrování duplikát pak 3,7–41,0 % s průměrem 14,0 % readů. Při mapování na pseudoreferenci tvořenou sekvencemi jaderných low-copy genů došlo k namapování od 83 037 do 1 399 470 readů na vzorek, což tvoří od 18,0 do 63,9 % readů s průměrem 52,1 %. Na chloroplastový genom se namapovalo od 613 do 110 819 readů, což tvoří 0,3–4,7 % readů s průměrnou hodnotou 2,1 % a na rDNA cistron pak 671–28 481, což tvoří 0,2–2,3 % s průměrem 0,9 % (podrobnosti viz Tabulka 10).

Celkové množství namapovaných jaderných genů s méně než 20 % chybějících údajů se pohybovalo mezi 274 geny (23,2 %) u outgroup a 1 173 geny (99,4 %) s průměrem 1 115 genů (94,5 %). Množství chybějících dat v těchto genech dosahovalo hodnot od 0,5 % do 6,8 % s průměrem 1,5 % (Tabulka 11). Délka assembly jednotlivých jaderných low-copy genů činila od 939 do 4 023 bp s průměrem 1 335 bp. Rozpětí průměrné bootstrapové hodnoty jednotlivých genových stromů získaných metodou RAxML bylo široké – od 16 do 88, a průměrná hodnota bootstrapu všech genových stromů činila 56, pro dataset „update“ pak 62. Množství variabilních pozic rostlo od 2,6 % do 26,7 % s průměrem 15,4 % a množství parsimonicky informativních pozic od 1,6 % do 14,0 % s průměrem 6,4 % (Příloha 1).

Chybějící data u chloroplastové DNA tvořila pro alignment z protein kódujících oblastí od 0 % do 88,4 % s průměrem 9,9 %, u *de-novo* assembly 0–96,9 % s průměrem 13,4 % a mapování na referenci 0–73,6 %, průměrně 7,0 %. U rDNA *de-novo* assembly tvořila pak chybějící data od 0 % do 4,9 % s průměrem 0,6 % a u mapování na referenci 1–19,1 %, průměrně 1,0 %. Průměrná sekvenční hloubka u jaderných genů dosahovala od 7 do 100 readů s celkovým průměrem 57 readů, u cpDNA od 1 do 93 readů s celkovým průměrem 30 readů a u rDNA 14–387 readů, průměrně pak 168 readů (Tabulka 11).

Tabulka 10. Výsledky sekvenování. Celkový počet readů (po odstranění PhiX), po filtrování kvality a duplikátů a po mapování na exonovou (sloupec exony), chloroplastovou (sloupec cpDNA) a ribozomální (sloupec rDNA) referenci. Procento vyfiltrovaných a duplicitních readů je vypočítáno z původního množství readů, procento namapovaných readů (% exonů, % cpDNA a % rDNA) je vypočteno z počtu readů bez duplikátů.

Druh	Počet readů	Po filtr. kvality (% readů)	Bez duplikát (% readů)	Exony (% exonů)	cpDNA (% cpDNA)	rDNA (% rDNA)
<i>Cam. latifolia</i>	260 375	240 248 (92,3 %)	151 514 (63,1 %)	83 037 (54,8 %)	613 (0,4 %)	671 (0,4 %)
<i>Cam. ovata</i>	1 385 720	1 322 322 (95,4 %)	1 097 167 (83,0 %)	598 277 (54,5 %)	8 215 (0,7 %)	2 448 (0,2 %)
<i>C. aeruginosa</i>	2 465 310	2 376 497 (96,4 %)	2 218 659 (93,4 %)	1 163 150 (52,4 %)	51 467 (2,3 %)	23 485 (1,1 %)
<i>C. alismatifolia</i>	1 311 693	1 268 631 (96,7 %)	1 113 216 (87,7 %)	586 944 (52,7 %)	22 307 (2,0 %)	7 713 (0,7 %)
<i>C. bhatii</i>	1 972 144	1 920 002 (97,4 %)	1 682 260 (87,6 %)	889 182 (52,9 %)	41 305 (2,5 %)	8 340 (0,5 %)
<i>C. bicolor</i>	1 145 699	1 096 507 (95,7 %)	963 640 (87,9 %)	541 518 (56,2 %)	24 727 (2,6 %)	10 956 (1,1 %)
<i>C. candida</i>	1 794 107	1 745 894 (97,3 %)	1 518 480 (87,0 %)	927 113 (61,1 %)	32 166 (2,1 %)	9 850 (0,6 %)
<i>C. cannanorensis</i>	1 692 178	1 644 365 (97,2 %)	1 463 490 (89,0 %)	771 061 (52,7 %)	47 583 (3,3 %)	11 534 (0,8 %)
<i>C. corniculata</i>	2 918 677	2 780 558 (95,3 %)	2 508 162 (90,2 %)	1 298 605 (51,8 %)	46 130 (1,8 %)	21 951 (0,9 %)
<i>C. ecomata</i>	1 755 851	1 707 004 (97,2 %)	1 496 658 (87,7 %)	786 637 (52,6 %)	59 455 (4,0 %)	16 793 (1,1 %)
<i>C. flammea</i>	2 299 465	2 142 017 (93,2 %)	1 990 607 (92,9 %)	1 022 252 (51,4 %)	36 836 (1,9 %)	26 565 (1,3 %)
<i>C. flaviflora</i>	2 799 988	2 622 940 (93,7 %)	2 361 244 (90,0 %)	1 277 710 (54,1 %)	110 819 (4,7 %)	17 075 (0,7 %)
<i>C. glans</i>	2 266 087	2 134 605 (94,2 %)	1 987 865 (93,1 %)	1 016 403 (51,1 %)	52 981 (2,7 %)	24 487 (1,2 %)
<i>C. gracillima</i>	787 278	755 690 (96,0 %)	490 639 (64,9 %)	276 827 (56,4 %)	4 162 (0,8 %)	1 806 (0,4 %)
<i>C. involucrata</i>	2 171 646	2 055 225 (94,6 %)	1 913 459 (93,1 %)	752 511 (39,3 %)	41 544 (2,2 %)	18 089 (0,9 %)
<i>C. myanmarensis</i>	1 560 209	1 515 266 (97,1 %)	1 337 314 (88,3 %)	706 019 (52,8 %)	47 213 (3,5 %)	14 814 (1,1 %)
<i>C. newmanii</i>	2 424 989	2 270 239 (93,6 %)	2 089 222 (92,0 %)	1 183 921 (56,7 %)	27 449 (1,3 %)	23 040 (1,1 %)
<i>C. oligantha</i>	1 414 928	1 369 496 (96,8 %)	1 231 368 (89,9 %)	667 142 (54,2 %)	31 619 (2,6 %)	12 350 (1,0 %)
<i>C. pambrosima</i>	2 513 479	2 388 655 (95,0 %)	2 190 830 (91,7 %)	1 182 157 (54,0 %)	45 343 (2,1 %)	26 542 (1,2 %)
<i>C. parviflora</i>	1 157 939	1 098 092 (94,8 %)	949 035 (86,4 %)	483 323 (50,9 %)	29 100 (3,1 %)	7 980 (0,8 %)
<i>C. petiolata</i>	645 192	623 285 (96,6 %)	597 317 (95,8 %)	294 849 (49,4 %)	25 373 (4,2 %)	1 765 (0,3 %)
<i>C. pierreana</i>	3 119 433	2 948 843 (94,5 %)	2 662 622 (90,3 %)	1 399 470 (52,6 %)	40 512 (1,5 %)	28 481 (1,1 %)
<i>C. plicata</i>	2 604 155	2 494 570 (95,8 %)	2 313 497 (92,7 %)	1 020 931 (44,1 %)	66 038 (2,9 %)	11 767 (0,5 %)
<i>C. rhomba</i>	1 730 534	1 635 534 (94,5 %)	1 394 537 (85,3 %)	691 528 (49,6 %)	25 627 (1,8 %)	17 664 (1,3 %)
<i>C. roscoeana</i>	1 932 644	1 883 789 (97,5 %)	1 644 168 (87,3 %)	806 907 (49,1 %)	30 307 (1,8 %)	15 982 (1,0 %)
<i>C. singularis</i>	979 892	936 313 (95,6 %)	552 931 (59,1 %)	353 536 (63,9 %)	1 620 (0,3 %)	12 732 (2,3 %)
<i>C. vama</i>	1 997 538	1 951 100 (97,7 %)	1 608 787 (82,5 %)	950 361 (59,1 %)	32 628 (2,0 %)	7 876 (0,5 %)
<i>C. vitellina</i>	709 195	652 052 (91,9 %)	406 078 (62,3 %)	252 628 (62,2 %)	1 877 (0,5 %)	8 336 (2,1 %)
<i>C. xanthella</i>	2 558 735	2 404 037 (94,0 %)	2 185 154 (90,9 %)	1 138 237 (52,1 %)	18 956 (0,9 %)	25 614 (1,2 %)
<i>S. aethiopicus</i>	1 730 873	1 652 933 (95,5 %)	1 591 660 (96,3 %)	286 335 (18,0 %)	29 085 (1,8 %)	7 575 (0,5 %)
průměr	1 803 532	1 721 224 (95,4 %)	1 523 719 (86,0 %)	780286 (52,1 %)	34435 (2,1 %)	14 143 (0,9 %)

Tabulka 11. Výsledky mapování readů. Map. nDNA = počet nukleárních genů s chybějícími daty do 20 %; Chyb. nDNA = průměr chybějících dat u jednotlivých Map. nDNA genů. Chyb. cpDNA/rDNA = průměrný počet chybějících údajů u cpDNA/rDNA (% N v nejdelším příslušném alignmentu).

Druh	Map. nDNA (% z 1180)	Chyb. nDNA	Chyb. cpDNA geny	Chyb. cpDNA reference	Chyb. cpDNA de-novo	Chyb. rDNA reference	Chyb. rDNA de-novo	Ø pokrytí nDNA	Ø pokrytí cpDNA	Ø pokrytí rDNA
<i>Cam. latifolia</i>	731 (61,9 %)	6,8 %	88,4 %	73,6 %	96,9 %	3,6 %	0 %	7	1	14
<i>Cam. ovata</i>	1 129 (95,7 %)	1,4 %	2,4 %	0,8 %	30,6 %	0,5 %	0 %	44	7	49
<i>C. aeruginosa</i>	1 173 (99,4 %)	0,6 %	0 %	0 %	0 %	0 %	0,4 %	84	47	307
<i>C. alismatifolia</i>	1 153 (97,7 %)	1,1 %	0 %	0 %	0 %	0 %	0 %	43	20	93
<i>C. bhatii</i>	1 163 (98,6 %)	0,7 %	0 %	0 %	0 %	0 %	0,1 %	66	38	156
<i>C. bicolor</i>	1 151 (97,5 %)	1,2 %	0 %	0,1 %	0 %	0 %	0,7 %	40	20	103
<i>C. candida</i>	1 171 (99,2 %)	0,8 %	0 %	0 %	0 %	0 %	0 %	70	29	133
<i>C. kannanorensis</i>	1 172 (99,3 %)	0,8 %	0 %	0 %	0 %	0 %	1,3 %	57	43	177
<i>C. corniculata</i>	1 172 (99,3 %)	0,6 %	0 %	0 %	0 %	0 %	0 %	94	42	227
<i>C. ecomata</i>	1 165 (98,7 %)	1,0 %	0 %	0 %	0 %	0 %	0,4 %	58	54	182
<i>C. flammea</i>	1 167 (98,9 %)	0,7 %	0 %	0 %	0 %	0 %	0 %	74	33	242
<i>C. flaviflora</i>	1 170 (99,2 %)	0,5 %	0 %	0,1 %	0 %	0 %	0 %	92	93	226
<i>C. glans</i>	1 169 (99,1 %)	0,8 %	0 %	0,2 %	0 %	0 %	0 %	74	47	261
<i>C. gracillima</i>	1 125 (95,3 %)	2,5 %	44,5 %	27,6 %	78,5 %	0 %	0 %	22	4	25
<i>C. involucrata</i>	1 157 (98,1 %)	1,6 %	0 %	0,1 %	0 %	0 %	0 %	54	34	211
<i>C. myanmarensis</i>	1 169 (99,1 %)	1,0 %	0 %	0 %	0 %	0 %	0,8 %	52	43	157
<i>C. newmanii</i>	1 165 (98,7 %)	0,6 %	0 %	0 %	0 %	0 %	1,0 %	84	24	272
<i>C. oligantha</i>	1 167 (98,9 %)	0,8 %	0 %	0 %	0 %	0,2 %	1,1 %	49	26	164
<i>C. pambrosima</i>	1 166 (98,8 %)	0,7 %	0 %	0,1 %	0 %	0 %	0,2 %	86	39	316
<i>C. parviflora</i>	1 141 (96,7 %)	1,3 %	0 %	0,1 %	0 %	0,8 %	0 %	36	26	94
<i>C. petiolata</i>	1 135 (96,2 %)	2,0 %	0 %	0 %	0 %	0 %	4,9 %	22	23	29
<i>C. pierreana</i>	1 166 (98,8 %)	0,7 %	0 %	0,1 %	0 %	0 %	0 %	100	36	387
<i>C. plicata</i>	1 165 (98,7 %)	1,0 %	0 %	0 %	0 %	0 %	0,7 %	74	58	187
<i>C. rhomba</i>	1 162 (98,5 %)	1,1 %	0 %	0,4 %	0 %	0 %	0 %	51	23	186
<i>C. roscoeana</i>	1 163 (98,6 %)	0,9 %	0 %	0 %	0 %	0 %	0,1 %	59	28	186
<i>C. singularis</i>	1 151 (97,5 %)	2,2 %	79,9 %	52,9 %	94,3 %	19,1 %	1,1 %	28	1	65
<i>C. vamana</i>	1 166 (98,8 %)	0,8 %	0 %	0 %	0 %	0,3 %	0 %	70	24	94
<i>C. vitellina</i>	1 120 (94,9 %)	3,2 %	80,1 %	53,0 %	93,8 %	0 %	4,9 %	21	2	48
<i>C. xanthella</i>	1 165 (98,7 %)	0,8 %	0 %	0,4 %	0,9 %	0 %	0,1 %	81	16	314
<i>S. aethiopicus</i>	274 (23,2 %)	6,3 %	1,4 %	1,3 %	5,9 %	5,7 %	1,3 %	22	24	150
Průměr	1 115 (94,5 %)	1,5 %	9,9 %	7,0 %	13,4 %	1,0 %	0,6 %	57	30	168

3.1 FYLOGENETICKÉ ANALÝZY

3.1.1 Fylogeneze odvozená z jaderných low-copy genů

Celkem byly vytvořeny čtyři fylogenetické stromy z 1 154 („původní“) jaderných low-copy genů, čtyři „update“ stromy z 881 genů – metodami ASTRID (Příloha 2), ASTRAL (Příloha 3), konkatenace metodou Maximum Likelihood (Obrázek 6A a Příloha 4B) a MRL (Příloha 5) a jeden strom metodou ASTRAL (Příloha 4A) z 1 154 genů se zkolabovanými nody s bootstrapovou podporou <70 vždy z 28 druhů rodu *Curcuma* a 3 druhů outgroup.

Ve všech devíti species trees se vytvořily čtyři hlavní molekulární linie – *Curcuma* se sesterskou *Hitcheniopsis* a sesterské *Ecomata* a *Pierreana* a v rámci podrodu *Curcuma* došlo také k oddělení skupin *Curcuma I* a *Curcuma II*. Hodnoty bootstrapové podpory a posterior probability většiny větví byly vysoké, v intervalu od 95 do 100, resp. 0,95–1, asi 15 % všech větví mělo pak podporu nižší a tyto nízké podpory byly rovnoměrně rozděleny mezi jednotlivé stromy kromě obou MRL stromů, kde hodnoty bootstrapové podpory dosáhly 100 % u všech nodů.

Pravděpodobní meziliniovní hybridy (vznikly z rodičů spadajících do podrodů *Curcuma* a *Hitcheniopsis*) ve všech stromech zaujali pozici na bázi podrodu *Hitcheniopsis*, avšak v některých případech doprovázenou nižší bootstrapovou podporou. *Curcuma candida* – druh, možná taktéž hybridního původu s rodiči spadající do různých linií podrodu *Curcuma*, stojí v šesti případech na samotné bázi celého podrodu *Curcuma*, ve zbylých dvou pak na bázi skupiny *Curcuma I* a jeho topologie v rámci jednotlivých druhových stromů je také provázena nižším bootstrapem. Co se týká vztahů uvnitř skupin, všechny metody odhadu species tree vytvořily mírně odlišné fylogenetické stromy liší se topologiemi některých vnitřních větví, které jsou doprovázené (kromě MRL stromů) sníženou bootstrapovou podporou, nebo hodnotu posterior probability, příslušného nodu.

3.1.2 Fylogeneze založená na chloroplastové DNA

Fylogeneze cpDNA byla odvozena na základě tří přístupů k rekonstrukci sekvencí – z celého chloroplastového genomu mapováním na referenci (Obrázek 7A) a pomocí *de-novo* assembly (Příloha 6A) a také pouze z protein kódujících oblastí (Příloha 6B) cpDNA. Pro mapování na referenci bylo využito 30 druhů (bez *C. longa*) s celkovou délkou alignmentu 140 241 nukleotidů (nt), pro *de-novo* assembly 26 druhů (bez

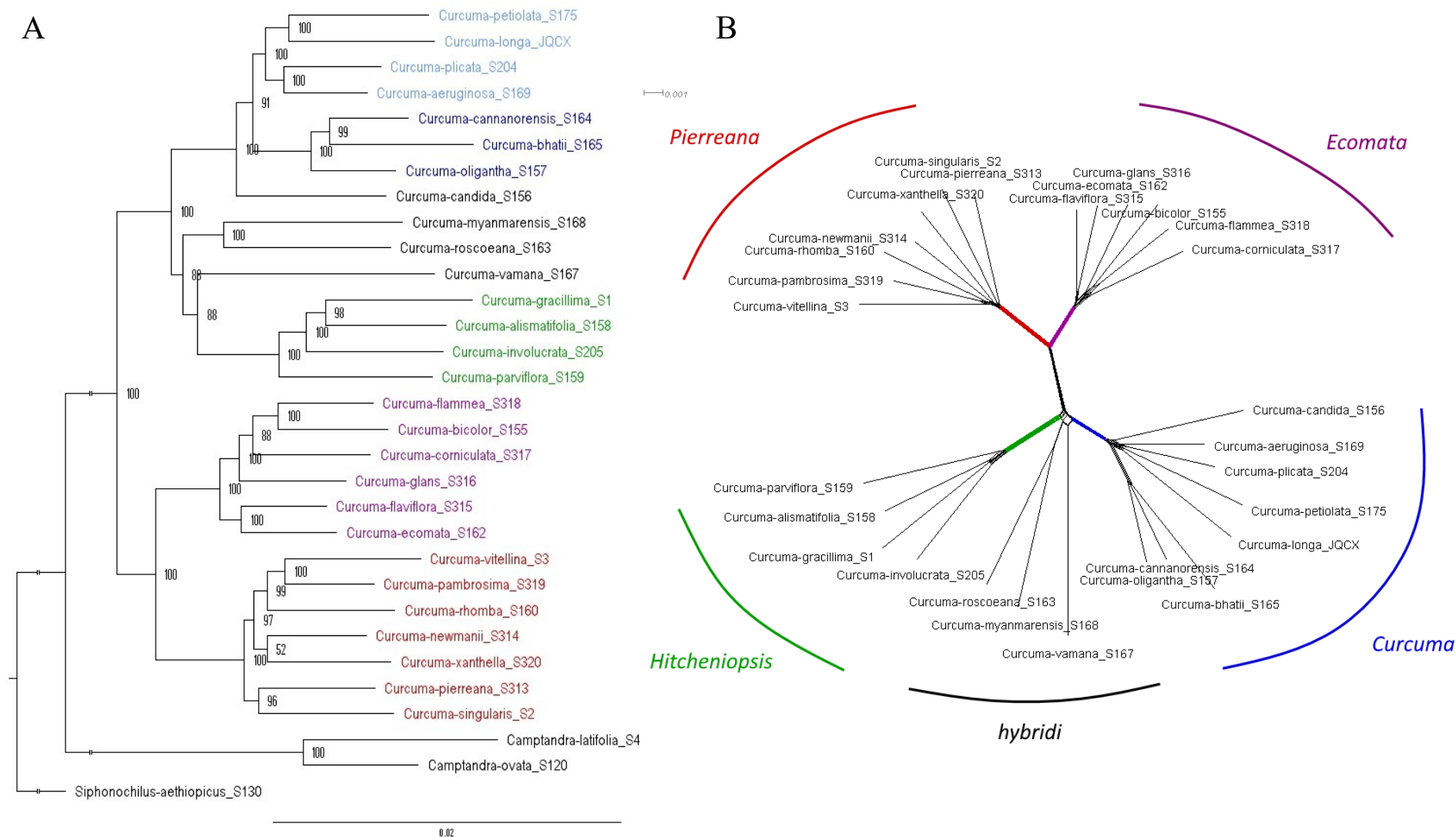
C. longa, *C. gracillima*, *C. singularis*, *C. vitellina* a *Camptandra latifolia*) s délkou alignmentu 141 714 nt a pro protein kódující oblasti 30 druhů (vyřazena *Camptandra latifolia*) z 86 genů s délkou alignmentu 78 919 nt. Druhy vyřazené z jednotlivých analýz vykazovaly příliš velké množství chybějících dat a fylogenetické stromy byly odvozovány s využitím metody Maximum Likelihood.

Ve všech třech fylogenetických stromech se vytvořily čtyři hlavní skupiny. *Curcuma* se oddělila od ostatních tří jako první, následovala *Ecomata* a *Pierreana* tvořila sesterskou skupinu k *Hitcheniopsis*. *Curcuma candida* zaujímala pozici v rámci skupiny *Curcuma I* a tři meziliniioví hybridy stáli na bázi podrodu *Curcuma*. Bootstrapové podpory jednotlivých větví byly obecně vysoké, avšak topologie některých vnitřních nodů zůstala nepodpořená a lišila se mezi jednotlivými stromy, především u species tree vytvořeného z kódujících genů od ostatních dvou.

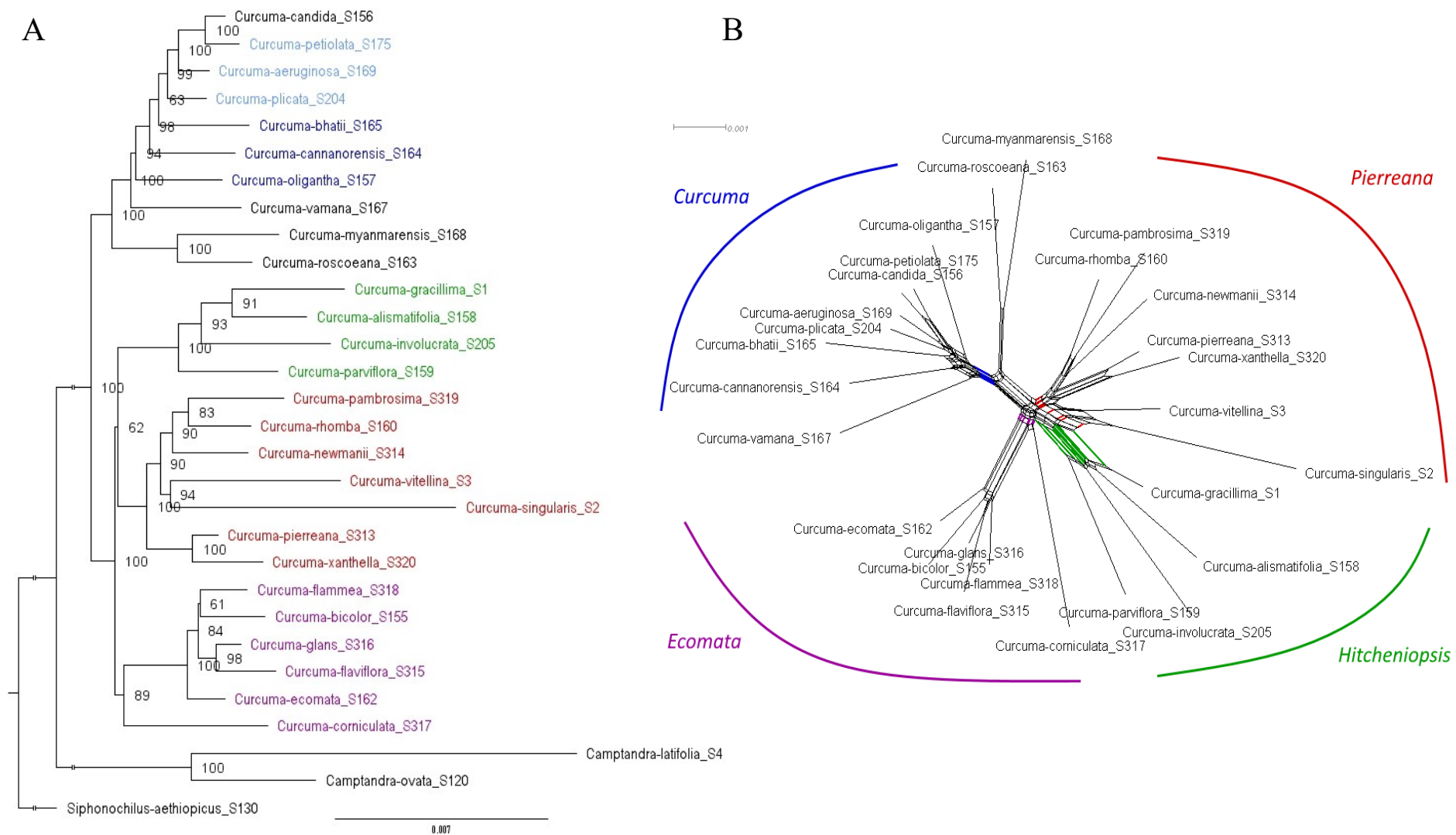
3.1.3 Fylogeneze ribozomální DNA

Odvození fylogenetického stromu z ribozomální DNA bylo provedeno pomocí dvou přístupů – *de-novo* assembly (Obrázek 8A) a mapování na referenci (Příloha 7A) s využitím metody Maximum Likelihood. V obou případech bylo zahrnuto 30 druhů (bez *C. longa*) a délka alignmentu pro *de-novo* assembly čítala 6 943 nt a pro mapování na referenci 6 885 nt.

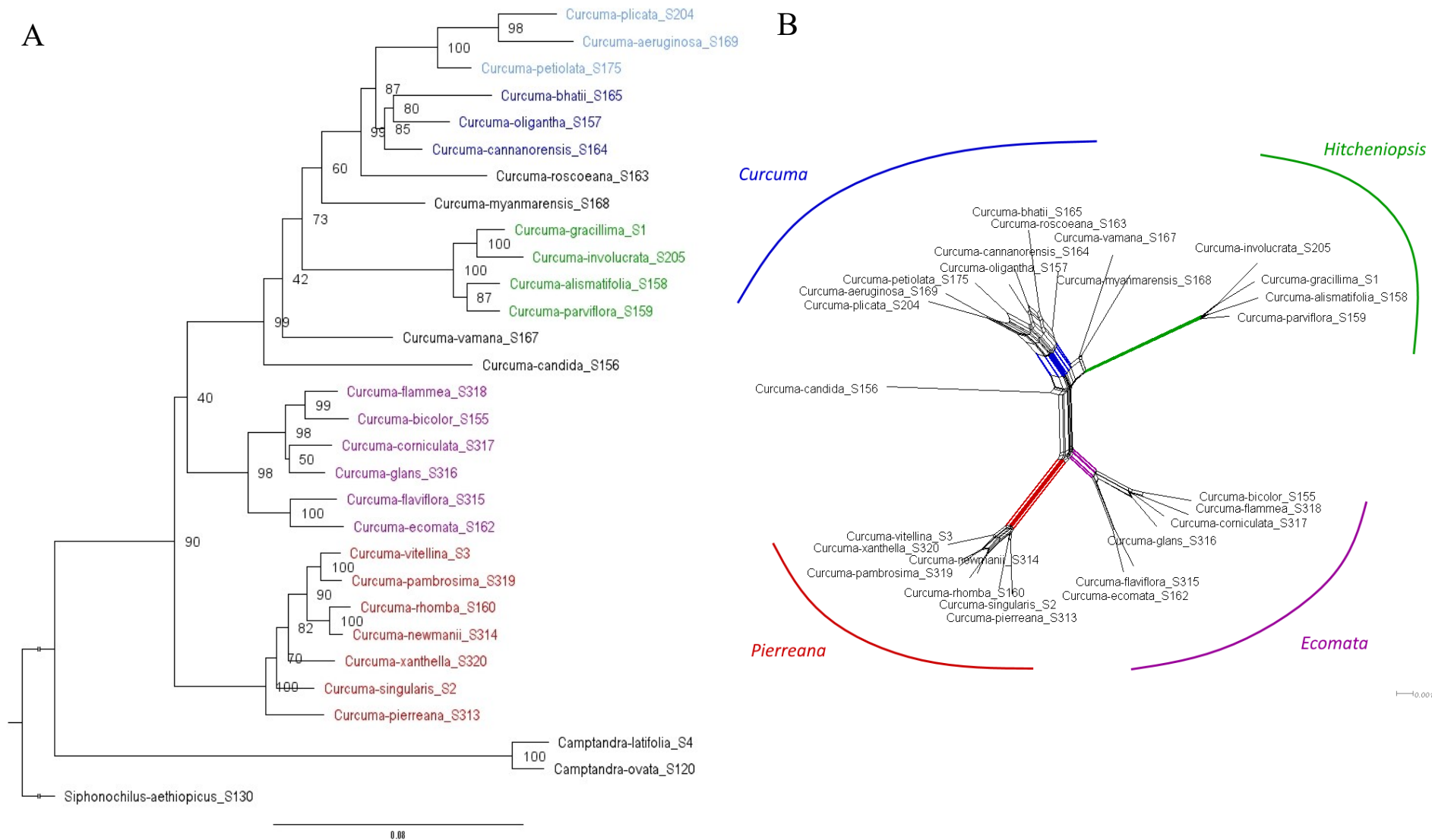
V obou species trees se vytvořily čtyři hlavní skupiny – *Curcuma* se sesterskou *Hitcheniopsis*, ke které (avšak s velmi nízkou podporou) přistupovala *Ecomata* a následně pak i *Pierreana* v případě *de-novo* assembly, v případě mapování na referenci tvořily skupiny *Ecomata* a *Pierreana* clade, který disponoval mírnou podporou. V obou případech byla *C. roscoeana* a *C. myanmarensis* zahrnuta ve skupině *Curcuma* a *C. vamana* zaujímala bazální pozici ke spojeným liniím *Curcuma* a *Hitcheniopsis*. Odlišného postavení v obou stromech nabývala pak *C. candida*, která stála v případě *de-novo* assembly také na pozici po spojení linií *Curcuma* a *Hitcheniopsis*, v případě mapování na referenci na bazální pozici skupiny *Hitcheniopsis*, avšak s velmi nízkou bootstrapovou podporou. V rámci čtyř hlavních skupin vykazovalo několik druhů také odlišné postavení, avšak některé vztahy byly velmi špatně podpořeny (BS <50) a celková podpora rDNA stromů byla nižší než u chloroplastových nebo jaderných fylogenezí.



Obrázek 6. A: Update species tree z 811 konkatenovaných jaderných genů vytvořený metodou Maximum Likelihood se 100 bootstrapovými replikacemi, B: Neighbor-Net analýza z 1 154 jaderných low-copy genů. Hlavní linie označeny barevně – *Curcuma I* světle modrou, *Curcuma II* tmavě modrou, *Hitcheniopsis* zeleně, *Ecomata* fialově, *Pierreana* červeně a hybridní druhy černou barvou.



Obrázek 7. A: Species tree z celého chloroplastového genomu získaný mapováním na referenci *C. roscoeana* a pomocí metody Maximum Likelihood s 1 000 bootstrapovými replikacemi. Hlavní linie označeny barevně – *Curcuma I* světle modrou, *Curcuma II* tmavě modrou, *Hitcheniopsis* zeleně, *Ecomata* fialově, *Pierreana* červeně a hybridní druhy černou barvou. B: Neighbor-Net analýza stejného datasetu.



Obrázek 8. A: Species tree z rDNA cistronu získaný pomocí *de-novo* assembly a metody Maximum Likelihood s 1 000 bootstrapovými replikacemi. Hlavní linie označeny barevně – *Curcuma I* světle modrou, *Curcuma II* tmavě modrou, *Hitcheniopsis* zeleně, *Ecomata* fialově, *Pierreana* červeně a hybridní druhy černou barvou. **B:** Neighbor-Net analýza stejného datasetu.

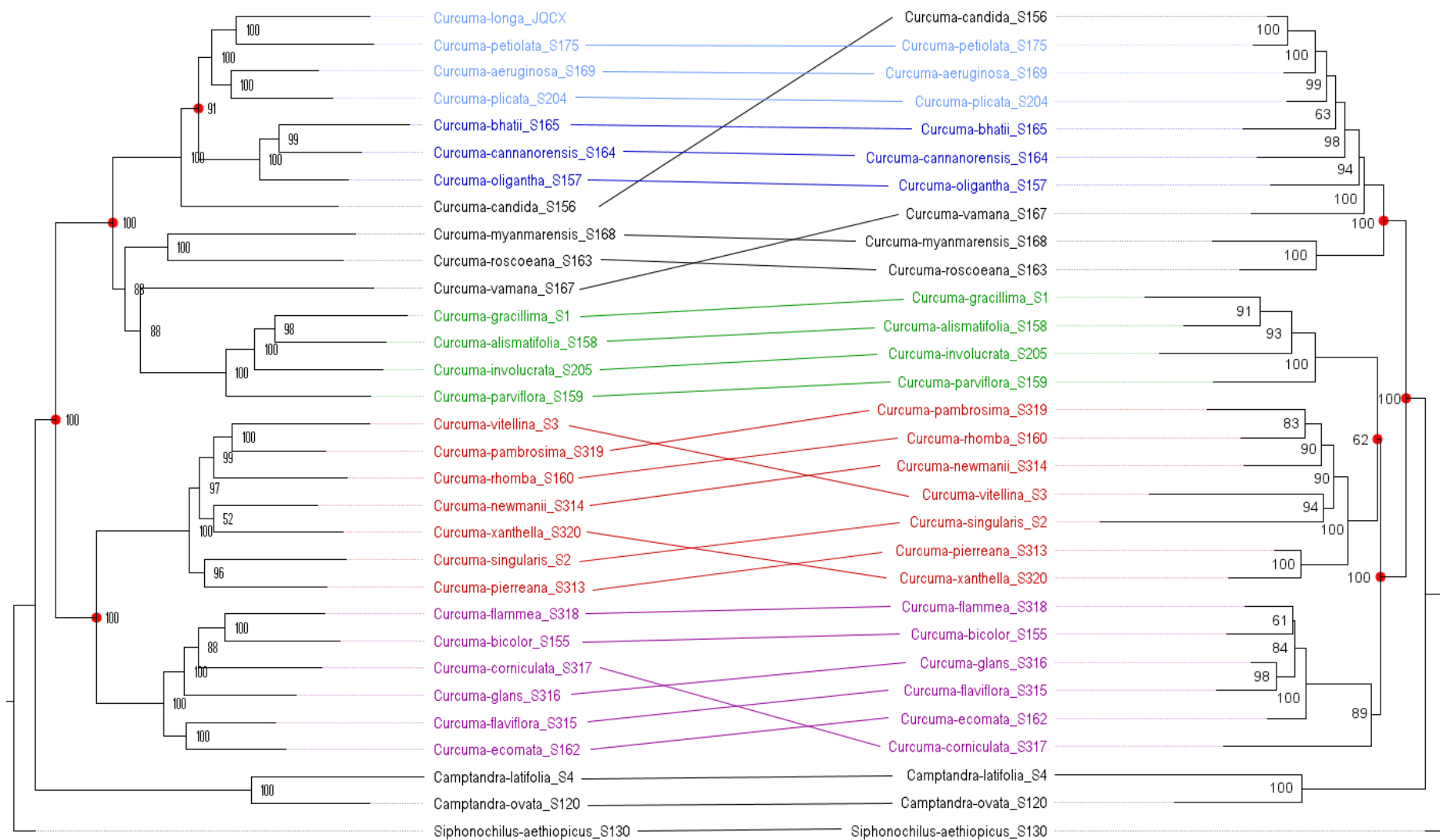
3.1.4 Srovnání chloroplastové, ribozomální a jaderné fylogeneze

Obrázek 9 a 10 znázorňují odlišnosti v topologii mezi vybranými fylogenetickými stromy – srovnání species tree z 811 konkatenovaných jaderných low-copy genů se stromem z celého chloroplastového genomu z mapování na referenci (Obrázek 9) a srovnání species tree z jaderných low-copy genů se stromem z ribozomálního cistronu z *de-novo* assembly (Obrázek 10).

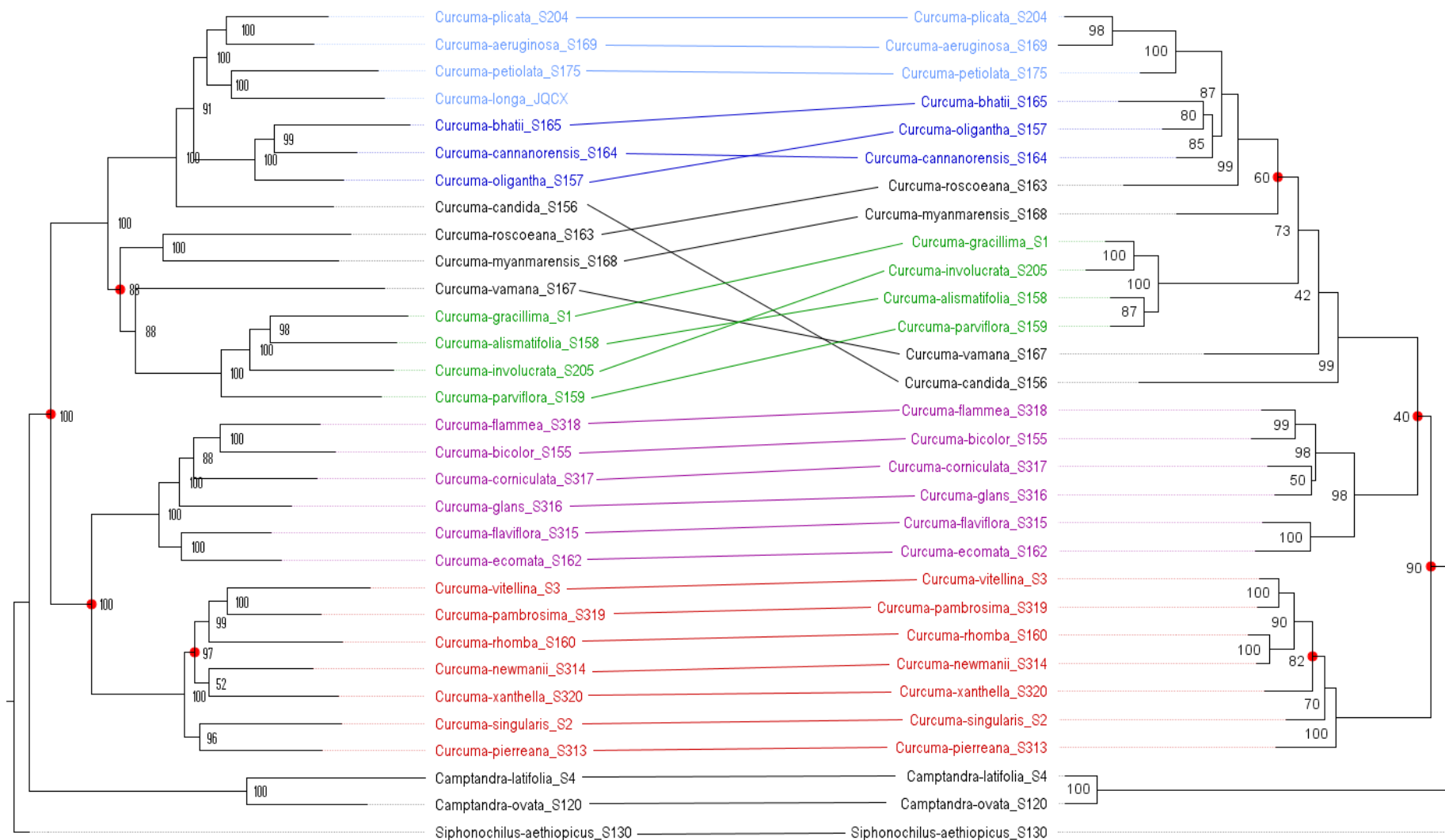
Ve všech fylogenetických stromech se vytvořily čtyři základní skupiny – *Curcuma*, *Hitcheniopsis*, *Ecomata* a *Pierreana*. Co se týká postavení jednotlivých linií, tak v případě stromu z jaderných genů dochází k oddělení dvou sesterských skupin *Ecomata* a *Pierreana* od zbylých dvou *Curcuma* s *Hitcheniopsis*, v případě chloroplastového genomu dojde nejprve k oddělení podrodu *Curcuma* od ostatních tří skupin a posléze pak odstupuje *Ecomata* od sesterských *Hitcheniopsis* a *Pierreana*. U fylogenetického stromu z ribozomální DNA dochází nejprve k oddělení linie *Pierreana* od ostatních tří a posléze *Ecomata* od sesterských *Curcuma* a *Hitcheniopsis*, avšak tato topologie není podpořená.

Pravděpodobně hybridní druhy – *C. myanmarensis*, *C. roscoeana*, *C. vamana* i *C. candida* zaujímají taktéž různé pozice. *Curcuma candida* stojí ve fylogenezi založené na jaderných genech na bázi celého podrodu *Curcuma*, v chloroplastovém stromu naopak v její vrcholové skupině (*Curcuma I*) a v ribozomálním stromu na bázi po spojení podrodů *Curcuma* a *Hitcheniopsis*. Ostatní tři hybridy zaujímají ve stromu z jaderných genů pozici na bázi podrodu *Hitcheniopsis*, v chloroplastovém stromu na bázi podrodu *Curcuma* a v ribozomální fylogenezi stojí *C. myanmarensis* a *C. roscoeana* na bazální pozici podrodu *Curcuma*, kdežto *C. vamana* na bazální pozici po spojení podrodů *Curcuma* a *Hitcheniopsis*.

Postavení některých druhů v rámci skupin se mezi jednotlivými stromy také mírně liší, např. *C. corniculata*, která ve stromu z jaderných genů zaujímá pozici uprostřed linie *Ecomata* se nachází v chloroplastovém stromu na její bázi nebo druhy *C. singularis* a *C. vitellina* tvořící clade v chloroplastové fylogenezi, stojí ve species tree z jaderných genů jedna ve vrcholové pozici a druhá na bázi skupiny *Pierreana*.



Obrázek 9. Srovnání topologie ML stromu z 811 konkatenovaných jaderných low-copy genů se 100 bootstrapovými replikacemi (vlevo) a ML stromu z celého chloroplastového genomu z mapování na referenci s 1 000 bootstrapovými replikacemi (vpravo). Hlavní linie označeny barevně – *Curcuma I* světle modrou, *Curcuma II* tmavě modrou, *Hitcheniopsis* zeleně, *Ecomata* fialově, *Pierreana* červeně a hybridní druhy černou barvou. Hlavní topologické změny označeny červeným puntíkem.

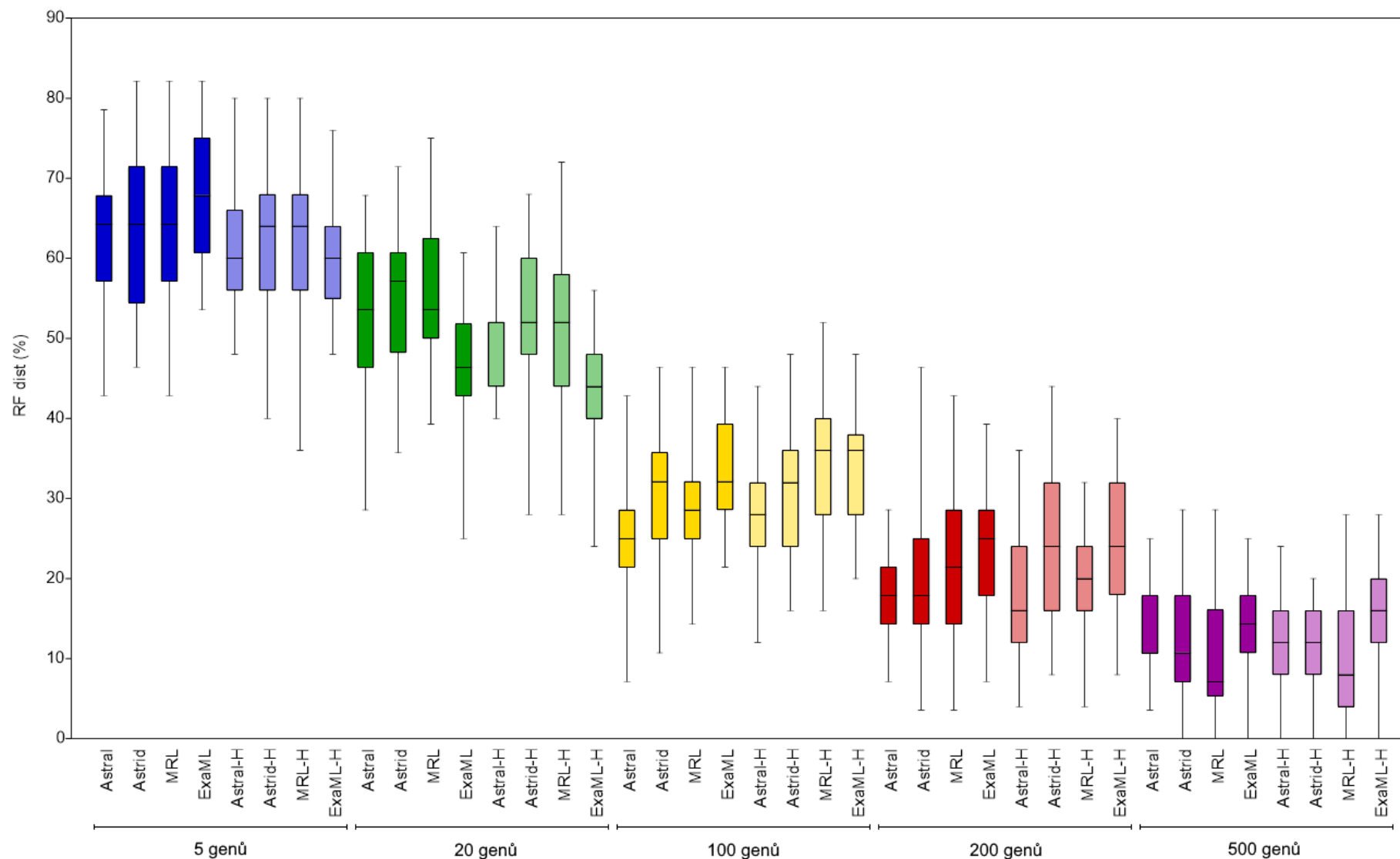


Obrázek 10. Srovnání topologie ML stromu z 811 konkatenovaných jaderných low-copy genů se 100 bootstrapovými replikacemi (vlevo) a ML stromu z ribozomálního cistronu z *de-novo* assembly s 1 000 bootstrapovými replikacemi (vpravo). Hlavní linie označeny barevně – *Curcuma I* světle modrou, *Curcuma II* tmavě modrou, *Hitcheniopsis* zeleně, *Ecomata* fialově, *Pierreana* červeně a hybridní druhy černou barvou. Hlavní topologické změny označeny červeným puntíkem.

3.1.5 Rekonstrukce fylogeneze na základě podsouboru genů

K odhadu množství genů potřebných pro tvorbu adekvátních fylogenetických stromů byly vytvořeny a hodnoceny fylogenetické stromy založené na 5, 20, 100, 200 a 500 genech. Celkem 10 species trees pro každou z kategorií – 5, 20, 100, 200 a 500 náhodně vybraných low-copy genových stromů, bylo vytvořeno čtyřmi metodami rekonstrukce species trees – ASTRAL, ASTRID, MRL a Maximum Likelihood (konkrétně ExaML) pro dva různé datasety – plný dataset obsahující 31 druhů a dataset s vyřazenými pravděpodobnými meziliniiovými hybridy, *C. myanmarensis*, *C. roscoeana* a *C. vamana*. Robinson-Fouldova vzdálenost sloužící jako ukazatel podobnosti jednotlivých species trees (viz Obrázek 11 a Příloha 8 a 9), kde Obrázek 11 znázorňuje srovnání RF distancí v závislosti na počtu genů použitých při tvorbě stromu druhů, Příloha 8 srovnání mezi jednotlivými metodami tvorby species trees a Příloha 9 zobrazuje průměrné hodnoty RF distancí jak mezi jednotlivými stromy, tak mezi stromy z podsouboru genů a celkovým fylogenetickým stromem z 1 154 jaderných low-copy genů.

Z Obrázku 11 je patrné, že RF distance mezi jednotlivými stromy klesá s rostoucím počtem genových stromů zahrnutých v analýze – jak pro dataset obsahující hybridní druhy, tak pro dataset bez hybridů, zhruba od 60–70 % pro 5 genů po 10–20 % pro 500 genů, přičemž hodnoty RF distancí obou datasetů dosahují podobných hodnot. U pěti a 20 genů dataset bez hybridů vykazuje o něco nižší hodnoty RF distancí, u 100 genů je tomu však naopak a u 200 a 500 genů záleží na použité metodě tvorby fylogenetických stromů. Co se týká srovnání jednotlivých metod, dosahují RF distance také podobných hodnot a žádná nevykazuje výrazně odlišné výsledky, pouze mírně vyšších hodnot kromě 20 genů dosahuje metoda Maximum Likelihood. Vzdálenost mezi fylogenetickým stromem z 1 154 genů a z jednotlivých genových podsouborů vykazuje také snižující se tendenci s rostoucím počtem genů bez závislosti na použité metodě, avšak dosahuje zhruba o 5–10 % nižších průměrných hodnot než srovnatelné kategorie vzdálenosti mezi jednotlivými stromy z podsouboru genů.



Obrázek 11. Srovnání RF distancí v závislosti na počtu genů. RF distance vyjádřené jako procento z maximální RF distance mezi 10 druhovými stromy získanými vždy z 5, 20, 100, 200 a 500 jaderných low-copy genů metodami ASTRAL, ASTRID, MRL a ExaML. „-H“ značí dataset bez přítomnosti meziliniových hybridů *C. myanmarensis*, *C. roscoeana* a *C. vamana*. Jednotlivé počty genů vyvedeny stejnou barvou, dataset neobsahující hybridní druhy je ve světlejších odstínech.

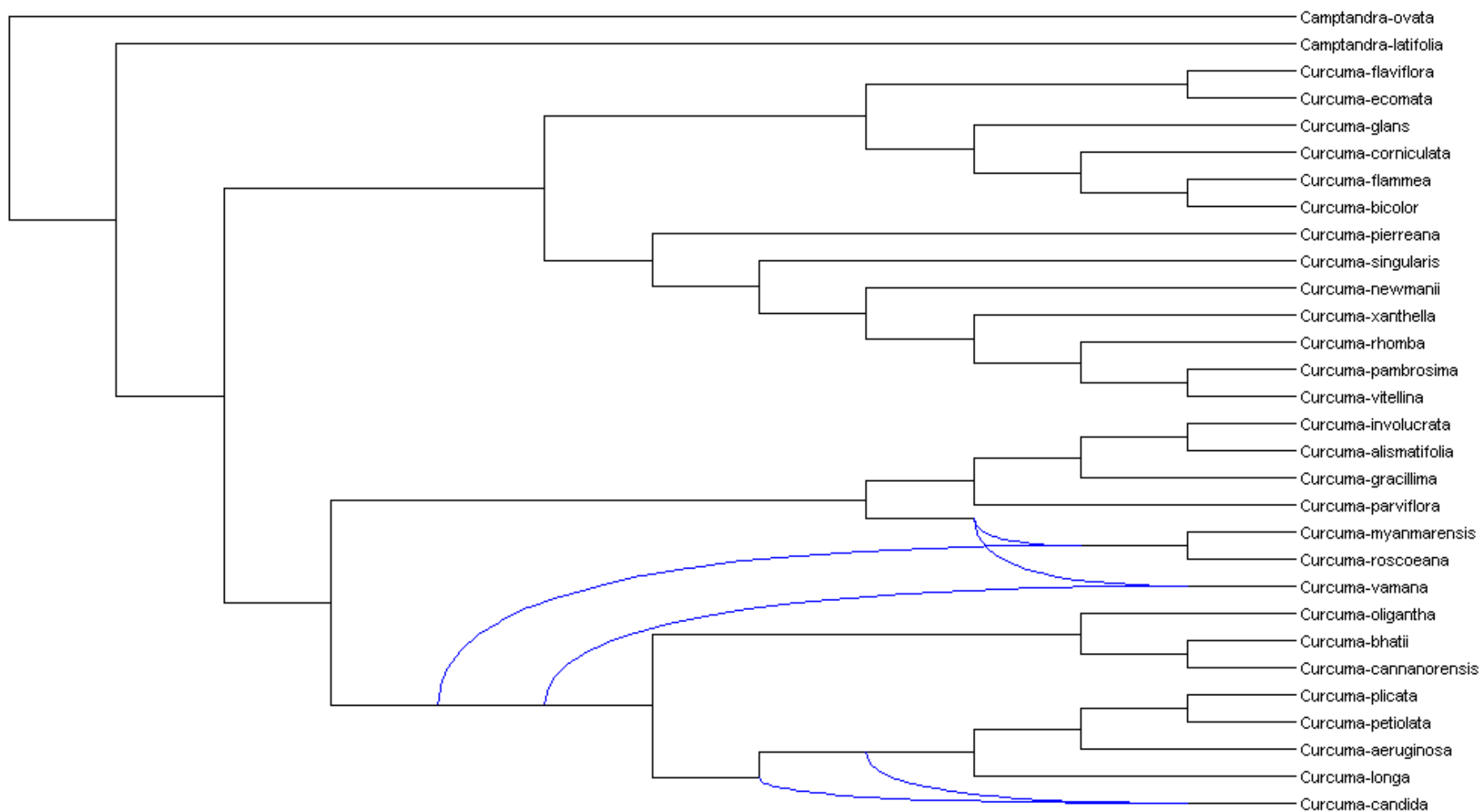
3.2 TESTOVÁNÍ HYBRIDNÍHO PŮVODU VYBRANÝCH DRUHŮ

Testování přítomnosti hybridních druhů v datasetu bylo provedeno třemi metodami – pomocí Neighbor-Net analýzy a MPL-PhyloNet metody, které zobrazují zesíťované vztahy mezi jednotlivými druhy. Druhá lineage movemet analýza ukazuje množství genových stromů, ve které druh zaujímá odlišnou topologii od konsenzuálního stromu a třetí testy monofylie počítají množství genových stromů, kde jsou předem definované skupiny (zahrnující i možný hybridní druh) monofyletické.

3.2.1 Fylogenetické sítě

Neighbor-Net analýza byla provedena s využitím všech tří částí genomu – konkatenovaných jaderných low-copy genů (Obrázek 6B), celého chloroplastového lokusu (Obrázek 7B) a ribozomální DNA (Obrázek 8B a Příloha 7B). Fylogenetické sítě získané metodou Neighbor-Net z 1 154 (Obrázek 6B) i 811 low-copy genů vykazovaly čtyři hlavní skupiny shodné s fylogenetickými stromy, *C. roscoeana*, *C. vamana* a *C. myanmarensis* stály mezi skupinami *Hitcheniopsis* a *Curcuma* a *C. candida* zaujímala bazální pozici ve skupině *Curcuma* se zesíťováním ke skupině *Curcuma I*. Skupiny vytvořené v rámci fylogenetických sítí z chloroplastového genomu se shodovaly se skupinami a postavením druhů v rámci cpDNA stromů. Ve fylogenetických sítích z rDNA se utvořily čtyři skupiny shodné s fylogenetickými stromy, *C. roscoeana* zaujala postavení na bázi skupiny *Curcuma*, *C. vamana* a *C. myanmarensis* stály mezi skupinami *Hitcheniopsis* a *Curcuma* a *C. candida* vykazovala bazální pozici ke spojeným liniím *Curcuma* a *Hitcheniopsis*.

Fylogenetické sítě získané metodou Maximum Pseudo-Likelihood v programu PhyloNet dosáhly hodnoty total log probability od -2 272 299 do -2 272 253, na Obrázku 12 je znázorněna síť s nejvyšší hodnotou. Ve všech pěti odvozených sítích byly jako hybridní druhy rozpoznány *C. myanmarensis*, *C. roscoeana* a *C. vamana* se vznikem z podrodů *Curcuma* a *Hitcheniopsis*. Společný vznik druhů *C. myanmarensis* a *C. roscoeana* byl zjištěn u všech pěti sítí, vznik *C. vamana* na ostatních dvou druhích byl nezávislý. V jedné síti byly jako hybridní rozpoznány také druhy *C. longa* a *C. petiolata* vzniklé z linie *Curcuma I* a předka skupiny *Curcuma I* a v síti s nejvyšší hodnotou total log probability byl jako hybridní identifikován také druh *C. candida* s rodiči z linie *Curcuma I* a předka skupiny *Curcuma I*. Co se týká postavení ostatních druhů v rámci fylogenetické sítě, vykazovaly některé z nich odlišné postavení u některých sítí.



Obrázek 12. Fylogenetická síť z 1 154 low-copy jaderných genů získaná pomocí Maximum Pseudo-Likelihood v programu PhyloNet. Hybridní původ čtyř druhů znázorněn modrými liniemi.

3.2.2 Lineage movement analýza

Základ pro lineage movement analýzu zobrazující procento výskytu daného druhu v alternaticních pozicích konsenzuálního species tree tvořilo 970 genových stromů získaných metodou Maximum Likelihood z jaderných low-copy genů s 29 taxony. Dva druhy tvořící outgroup, *Siphonochilus aethiopicus* a *Camptandra latifolia*, nebyly do analýzy zařazeny kvůli velkému množství chybějících dat, kvůli kterému by se značně snížil počet vstupních genových stromů, protože do analýzy mohly být zahrnuty pouze stromy obsahující všechny taxony.

Výsledky lineage movement analýzy jsou shrnuty ve dvou tabulkách – první zobrazuje pouze hybridní druhy (Tabulka 12), v druhé je celkový souhrn (Příloha 10). Procenta přítomnosti druhů na jednotlivých pozicích byla sečtena pro sedm skupin – *Curcuma I* a *Curcuma II*, podrod *Curcuma* celkově, *Hitcheniopsis*, *Ecomata*, *Pierreana* a „Hybridi“. Zařazení jednotlivých druhů je podle Tabulky 1, skupina „Hybridi“ je sběrná nemonofyletická skupina tří potenciálně meziliniových hybridů *C. myanmarensis*, *C. roscoeana* a *C. vamana* stojících na bázi linie *Hitcheniopsis*. *Curcuma candida* byla pro celkový souhrn zařazena do skupiny *Curcuma II* kvůli její bazální pozici v konsenzuálním stromu.

Všechny pravděpodobně nehybridní druhy vykazují pohyb především v rámci skupiny, do které samy spadají (viz Příloha 10), v nízké míře pak i do ostatních skupin – zastoupení genů v rámci sesterské skupiny (*Curcuma* + *Hitcheniopsis* + hybridy a *Ecomata* + *Pierreana*) čítá 0,2–1,5 % kromě zastoupení druhů podrodu *Curcuma* ve skupině hybridů, kde se jedná o 1,4–2,9 % a zastoupení v nesesterských liniích tvoří 0–0,6 %.

Tři potenciální meziliniovní hybridy – *C. myanmarensis*, *C. roscoeana* a *C. vamana* vykazují největší míru pohybu také v rámci své skupiny, která stojí na bázi linie *Hitcheniopsis*. Do podrodu *Curcuma* spadají v 4,4–6,8 % genových stromů, zastoupení ve skupině *Curcuma I* a *Curcuma II* je podobné (viz Tabulka 11) a na ostatních pozicích linie *Hitcheniopsis* se objevují v 1,2–3,3 % případů. *Curcuma candida* zaujímající v konsenzuálním stromu úplně bazální pozici – ((*Curcuma I* + *Curcuma II*) *C. candida*), vykazovala přítomnost v rámci skupiny *Curcuma I* v 35,2 % genů, v rámci skupiny *Curcuma II* v 4,3 % genů, v rámci skupiny hybridů v 4,4 % případů a v rámci *Hitcheniopsis* v 0,6 % stromů. Zastoupení v rámci nesesterských skupin *Ecomata* a *Pierreana* dosahuje u všech hybridů 0–0,5 % stejně jako u nehybridních druhů.

Tabulka 12. Výsledky lineage movement analýzy založené na 970 stromech z jaderných low-copy genů čtyř potenciálně hybridních druhů rodu *Curcuma*. Skupiny odpovídají molekulárním liniím (viz Tabulka 1), sběrnou skupinu „Hybridi“ tvoří potenciálně meziliniovní hybridy – *C. myanmarensis*, *C. roscoeana* a *C. vamana*. Hodnoty přítomnosti druhů v alternativních pozicích od konsenzuálního stromu jsou uvedeny v procentech a zaokrouhleny.

Skupina	<i>C. candida</i>	<i>C. myanmarensis</i>	<i>C. roscoeana</i>	<i>C. vamana</i>
<i>Curcuma I</i>	35,2	2,3	3,1	3,1
<i>Curcuma II</i>	4,3	2,2	3,2	3,7
<i>Curcuma</i>	39,5	4,4	6,3	6,8
<i>Hitcheniopsis</i>	0,6	1,2	1,9	3,3
<i>Hybridi</i>	4,4	50,7	51,5	13,7
<i>Ecomata</i>	0,0	0,2	0,2	0,4
<i>Pierreana</i>	0,2	0,2	0,5	0,2

3.2.3 Testy monofylie

U 1 154 stromů získaných z jaderných low-copy genů metodou Maximum Likelihood se 100 bootstrapovými replikacemi a se zkolabovanými větvemi s bootstrapovou podporou <70 byla testována přítomnost jednotlivých druhů rodu *Curcuma* v rámci 12 předdefinovaných skupin. Výsledky přítomnosti jednotlivých druhů ve všech testovaných skupinách jsou shrnuty v Příloze 12, v Tabulce 13 jsou pak shrnuty výskyty potenciálně hybridních druhů v rámci nejdůležitějších skupin.

V analýze nehybridních druhů bylo pozorováno majoritní množství genových stromů, ve kterých analyzovaný druh spadal do stejné skupiny, do které patřil podle apriorní definice skupin. Monofyletičnost těchto druhů s jinými skupinami je ojedinělá (detaily viz Příloha 11). *Curcuma candida* ve velkém počtu případů spadá do podrodu *Curcuma*, monofyletickou skupinu s ostatními liniemi tvoří také zřídka. Co se týká postavení v rámci tohoto podrodu, 29 genů vykazuje monofylii druhů skupiny *Curcuma I* a *C. candida*, v rámci skupiny *Curcuma II* se *C. candida* objevuje v 10 případech a bazální pozici vykazuje v 41 genových stromech. Druhy skupiny *Curcuma I* vykazují monofylii s touto skupinou ve 28–29 genech, se skupinou *Curcuma II* ve 2–6 genech a na bazální pozici stojí v 7–17 genových stromech. Druhy skupiny *Curcuma II* vykazují monofylii s touto skupinou v 16–19 genech, se skupinou *Curcuma I* ve 2 genech a na bazální pozici stojí v 3–7 genových stromech.

Ostatní hybridní druhy – *C. myanmarensis*, *C. roscoeana* a *C. vamana* tvoří monofyletickou skupinu s podrodem *Curcuma* ve 108–135 případech, s podrodem *Hitcheniopsis* v 138–199 případech a na bazální pozici od obou podrodů stojí v 131–156

genových stromech (Tabulka 13), oproti tomu ostatní druhy podrodů *Curcuma* a *Hitcheniopsis* stojí na bazální pozici obou podrodů v 0–14 případech.

Tabulka 13. Počty genů podporující jednotlivé skupiny získané při testech monofylie čtyř hybridních druhů – *C. candida*, *C. myanmarensis*, *C. roscoeana* a *C. vamana* založených na 1 154 jaderných low-copy genech se zkolabovanými větvemi s bootstrapovou podporou <70. *Curcuma* = geny s výskytem hybridu vrámci podrodu *Curcuma*, *Hitcheniopsis* = geny s výskytem hybridu vrámci podrodu *Hitcheniopsis*, báze C + H = geny s hybridem stojícím na bázi spojených podrodů *Curcuma* a *Hitcheniopsis*. *Curcuma I* = geny s výskytem hybridu vrámci skupiny velikosti genomu *Curcuma I*, *Curcuma II* = geny s výskytem hybridu vrámci skupiny velikosti genomu *Curcuma II*, báze C = bazální postavení hybridu vrámci podrodu *Curcuma*.

skupina	<i>C. myanmarensis</i>	<i>C. roscoeana</i>	<i>C. vamana</i>	skupina	<i>C. candida</i>
<i>Curcuma</i>	108	112	135	<i>Curcuma I</i>	29
<i>Hitcheniopsis</i>	138	138	199	<i>Curcuma II</i>	10
báze C + H	148	156	131	báze C	41

4 DISKUZE

4.1 ÚSPĚŠNOST SEKVENOVÁNÍ

Rozpětí počtu získaných readů u jednotlivých vzorků je značné – více než desetinásobný rozdíl mezi minimální a maximální hodnotou. Celkem pět vzorků dosáhlo hodnoty nižší než 1 000 000 readů (při průměru ~1 800 000), což u čtyř z nich činilo problémy v některých fylogenetických analýzách. Jedná se o druhy *Camptandra ovata*, *Curcuma gracilima*, *C. singularis* a *C. vamana*, které vykazovaly taktéž velké množství vyfiltrovaných duplicitních readů, kolem 40 % (při průměru ~15 %) a nízké množství readů náležící chloroplastovému genomu a rDNA cistronu. U cpDNA dosáhlo pokrytí těchto vzorků hodnoty 1–4x, při průměru 30x, u rDNA 14–65x, při průměrném pokrytí 168x, což zapříčinilo velké množství chybějících dat v analýzách chloroplastového genomu, a dokonce vyřazení těchto vzorků z analýzy u *de-novo* assembly, kde množství chybějících dat přesáhlo 75 %. U rDNA nižší pokrytí sice způsobilo větší množství chybějících dat, avšak k vytvoření téměř kompletního cistronu s využitím obou metod došlo. Co se týká jaderných genů, bylo procento chybějících dat těchto vzorků zvýšeno jen mírně – proces obohacení zřejmě proběhl dostatečně. Naopak zde hrála roli vzdálenost od reference – hybridizačních prób, které byly designovány ze sekvencí rodu *Curcuma* (Fér and Schmickl, unpubl.), a proto měly oba druhy rodu *Camptandra* a *Siphonochilus aethiopicus* nejvyšší množství chybějících dat.

Velké rozpětí v počtu readů, množství duplikátů a zastoupení vysokokopiové frakce by mohlo být způsobeno zařazením různých vzorků do různých knihoven s ostatními druhy čeledi *Zingiberaceae*. Čtyři problematické vzorky byly připravovány a sekvenovány společně, a proto jsou kvalita a množství knihovny spolu s nepřesností při kvantifikaci slabých knihoven nejpravděpodobnějšími příčinami výše zmíněných rozdílů.

Co se týká porovnání s ostatními pracemi využívajícími metodu target enrichment, bylo dosaženo dobré účinnosti zachycení jaderných low-copy genů. V této práci se jednalo v průměru o 52 % z celkového počtu readů, u ostatních studií tvořily cílové sekvence od 33 do 63 % readů (Heyduk et al., 2016; Johnson et al., 2016; Schmickl et al., 2016; Weitemier et al., 2014). Množství cpDNA a rDNA readů bylo sice nižší než u ostatních studií, avšak množství chybějících dat je srovnatelné (Heyduk et al., 2016; Schmickl et al., 2016; Weitemier et al., 2014), což souhlasí s průměrnou hodnotou pokrytí dostatečnou k dobré assembly stanovenou na 30x u chloroplastové DNA a 40x u rDNA (Straub et al., 2012).

4.2 REKONSTRUKCE FYLOGENEZE

4.2.1 Jaderné low-copy geny

Fylogenetické stromy založené na jaderných low-copy genech vykazují obecně vysoké bootstrapové podpory a konzistenci napříč všemi použitými metodami – koalescenčním přístupem (ASTRAL a ASTRID), supertree metodě (MRL) i konkatenci (Maximum Likelihood), což je ve shodě s dalšími pracemi využívajícími next-generation sekvenování k rekonstrukci fylogeneze (např. Ai and Kang, 2015; Heyduk et al., 2016; Léveillé-Bourret et al., 2017; Schmickl et al., 2016).

Topologie jednotlivých linií v rámci rodu *Curcuma* (sesterské podrody *Curcuma* s *Hitcheniopsis* a linie *Ecomata* + *Pierreana*) se shoduje ve všech vytvořených stromech a koresponduje s již publikovanou fylogenezí (Záveská et al., 2016). Větší morfologická podobnost podrodu *Curcuma* s liniemi *Ecomata* a *Pierreana* oproti druhům fylogeneticky sesterské skupiny *Hitcheniopsis* byla vysvětlována jako vznik linií *Curcuma* a *Ecomata* + *Pierreana* ze společného předka a až následné odštěpení vrcholové skupiny *Hitcheniopsis* od linie *Curcuma* (Záveská et al., 2016).

I přes vysoké množství genů využitých pro odvození jednotlivých species trees se na nižších úrovních objevují drobné topologické odlišnosti téměř u každé z hlavních linií, které jsou obvykle doprovázeny snížením bootstrapové podpory nebo posterior probability. Obdobných výsledků bylo dosaženo v práci na rodu *Oxalis* (Schmickl et al., 2016), kde se fylogenetické stromy vytvořené jednotlivými metodami také liší v topologii 2 – 3 nodů, což může být způsobeno určitou mírou nepřesnosti při odvozování species tree, jak bylo prokázáno ve studiích Kupczok et al. (2010) a Vachaspati and Warnow (2015) nebo hybridizačními událostmi mezi druhy – vzhledem k vysokému množství genů zahrnutých v jednotlivých analýzách se jeví další příčiny způsobující nízké podpory u species tree, např. nedostatek signálu v datech nebo vliv ILS, což je náhodný a nesystematický proces, jako nepravděpodobné. Oproti tomu hybridizační události systematické jsou – produkují velké množství genů se shodnou topologií lišící se od topologie majoritní (Linder and Rieseberg, 2004). Jako nejlepší (obsahující nejmenší počet špatně podpořených a od topologie odvozené většinou ostatních metod se lišících cladů) byl vybrán strom vytvořený konkatencí 811 genů, který obsahoval pouze jednu špatně podpořenou topologii – clade tvořící *C. newmanii* a *C. xanthella*, kteřto v ostatních případech přistupuje s vysokou podporou nejdříve *C. newmanii* a posléze až *C. xanthella* k jádru skupiny *Pierreana*.

Tři pravděpodobně hybridní druhy (*C. myanmarensis*, *C. roscoeana* a *C. vamana*) se ukázaly jako bližší skupině *Hitcheniopsis* u fylogeneze z NGS dat, což se liší od práce Závěská et al. (2016), kde byly tyto druhy na základě 4 jaderných genů, ITS a cpDNA lokusu přiřazeny bližší skupině *Curcuma*, avšak s velmi nízkou podporou. Vztahy hybridních druhů byly ve species trees konstruovaných z jaderných low-copy genů ve většině případů vysoce podpořeny, avšak v některých topologiích, zvláště z „update datasetu“ (genů s průměrným bootstrapem >50), byla statistická podpora těchto nodů mírně snížena, což může zpochybňovat výhradní vztah druhů ke skupině *Hitcheniopsis* a naznačovat právě jejich hybridní původ. Postavení druhu *C. candida* se také liší mezi oběma pracemi, v Závěská et al. (2016) se tento druh zařadil do skupiny *Curcuma I*, zde stojí na bázi celé skupiny *Curcuma*.

I přes mírné inkongruence mezi stromy vytvořenými pomocí jednotlivých metod se jedná o velký pokrok v pochopení vztahů v rámci rodu *Curcuma*, které jsou vysoce podpořeny oproti dříve publikované fylogenezi (Závěská et al., 2012) založené na tradičně používaných markerech, jako je ITS sekvence a několik málo chloroplastových lokusů. Díky tomu se metoda Hyb-Seq (Weitemier et al., 2014) zahrnující okolo tisíce low-copy nukleárních genů jeví jako vhodná i pro odvozování fylogenetických vztahů u komplexních skupin zahrnující polyploidní, případně hybridní druhy, jejichž odhalování v rámci datasetu bude diskutováno níže.

4.2.2 Chloroplastová DNA a ribozomální cistron

Vzhledem k možným strukturním změnám v chloroplastovém genomu a paralogním sekvencím u ribozomálního lokusu byly cpDNA a rDNA sekvence analyzovány jak pomocí *de-novo* assembly, tak mapováním na referenci se zahrnutím kódujících i nekódujících oblastí. V případě cpDNA byla fylogeneze odvozena také pouze z protein kódujících oblastí a hodnocena schopnost jednotlivých metod vyřešit fylogenetické vztahy v rodu *Curcuma*.

Čtyři skupiny vytvořené v rámci chloroplastové fylogeneze byly shodné, jak z hlediska přítomnosti jednotlivých druhů, tak vzájemné topologie, napříč všemi použitými metodami. Linie *Hitcheniopsis*, *Ecomata* a *Pierreana* zde tvoří podpořenou monofyletickou skupinu, což je v souladu s dříve publikovanými výsledky fylogeneze cpDNA využívající pouze tři chloroplastové lokusy, kde byla tato topologie podpořena pouze slabě (Závěská et al., 2012). Vztahy v rámci těchto tří linií však vykazují i u NGS dat sníženou bootstrapovou podporu a krátké větve.

Všechny tři cpDNA stromy z NGS dat měly většinově vysoké bootstrapové podpory, podobně jako v dalších studiích využívající celý/téměř celý chloroplastový genom k odvození vnitrorodové fylogeneze (např. Parks et al., 2009; Schmickl et al., 2016; Weitemier et al., 2014). Některé nody však vykazovaly odlišnou topologii a/nebo sníženou bootstrapovou podporu v závislosti na použité metodě. Jako nejvhodnější přístup u rodu *Curcuma* se ukázalo mapování na referenci, které vykazovalo nejvyšší bootstrapové podpory (minimální hodnota 60) a nejmenší množství chybějících dat, které bylo u většiny druhů a všech metod zanedbatelné, avšak u čtyř vzorků s celkově nízkým počtem readů se hodnoty chybějících dat lišily i o několik desítek procent. U *de-novo* assembly tyto druhy obsahovaly vysoké množství chybějících dat (i více než 90 %), a proto byly z analýzy vyřazeny. Stejných výsledků – získání většího množství dat pro následné fylogenetické analýzy pomocí mapování na referenci, bylo dosaženo také u rodu *Asclepias* v práci Weitemier et al. (2014).

Odlišnou topologii mezi jednotlivými stromy vykazoval nejbližší příbuzný druhu *C. gracillima* – ve fylogenezi z mapování na referenci jím byla *C. alismatifolia*, z kódujících oblastí pak *C. involucrata* (s *de-novo* assembly bohužel srovnání provést nelze, protože byl druh *C. gracillima* z důvodu chybějících dat vyřazen). Podpora cladu u kódujících oblastí byla sice 100 %, oproti 91 % u mapování na referenci, avšak tento strom byl odvozen z téměř dvojnásobně dlouhého alignmentu a *C. gracillima* obsahovala o 15 % méně chybějících dat než u kódujících oblastí. Další topologické odlišnosti vykazovalo pět druhů ze skupiny *Ecomata*, jejichž vztahy byly shodné u metody *de-novo* assembly a stromu z kódujících oblastí plastomu oproti mapování na referenci. U všech metod vykazovala tato skupina velmi krátké větve a sníženou bootstrapovou podporu, což značí nedostatek signálu v datech (Wiens et al., 2008), avšak nejvyšší podpory dosahuje metoda mapování na referenci. Obecně nízkými podporami a krátkými větvemi trpěla také skupina *Pierreana* ve stromu z kódujících oblastí, avšak data z nekódujících oblastí u mapování na referenci tento problém překonaly.

Výsledky této práce založené na celém chloroplastovém genomu nicméně ukazují velké zlepšení oproti poslední publikované fylogenezi založené pouze na třech chloroplastových lokusech (Záveská et al., 2012), podobně jako v práci Parks et al. (2009), kde celogenomová fylogeneze přinesla vysoké rozlišení oproti dřívějším studiím na základě několika málo lokusů a ověřili tím pozitivní vztah mezi délkou matice a její rozlišovací schopností. Bohužel ani celý chloroplastový genom nemusí poskytovat dostatek variability pro vyřešení všech větví s vysokou podporou, jako bylo ukázáno zde a např. u rodu *Sabal*

(Heyduk et al., 2016) nebo *Fragaria* (Njuguna et al., 2013), což může ukazovat na nedávný vznik vyšetřované skupiny, podobně jako např. u rodu *Hieracium* (Fehrer et al., 2007).

Fylogeneze ribozomálního cistronu založené na *de-novo* assembly a mapování na referenci poskytly také částečně se lišící výsledky doprovázené obecně nižšími bootstrapovými podporami. Odlišnosti jsou patrné již na úrovni topologie jednotlivých skupin, které ovšem ani v jednom případě nebyly dobře podpořeny. Další odlišné topologie a nízké bootstrapové podpory se objevují okolo postavení potenciálně hybridních druhů a v rámci skupin *Ecomata* a *Pierreana*. Mírně odlišných výsledků bylo také dosaženo oproti fylogenezi získané z ITS lokusu (Záveská et al., 2012), především u druhů *C. candida* a *C. vamana*, jejichž přesné postavení je doprovázeno nižšími statistickými podporami u ITS lokusu i NGS dat. Na základě NGS dat sekvence rDNA cistronu *C. candida* spadá s vysokou podporou do příbuznosti skupin *Curcuma* a *Hitcheniopsis*, oproti postavení v rámci linie *Pierreana* odvozené na základě lokusu ITS.

Podobných výsledků u ribozomální fylogeneze bylo dosaženo v práci Schmickl et al. (2016), kde byly polytomie a špatně podpořené větve ribozomálního cistronu interpretovány jako výsledek nedostatku parsimonicky informativních míst a s tím kontrastující vysoká variabilita ITS lokusu. Oproti chloroplastovému genomu a celkové délce jaderných low-copy genů je ribozomální cistron také mnohonásobně kratší a jak bylo ukázáno v práci Parks et al. (2009) s délkou alignmentu roste i rozlišovací schopnost získané fylogeneze.

4.2.3 In/kongruence mezi jednotlivými fylogenezemi

Topologie stromů založených na různých částech genomu – jaderné low-copy geny, chloroplastová DNA a ribozomální cistron, vykazují značné odlišnosti, především co se týká postavení jednotlivých linií a pravděpodobně hybridních druhů, v menší míře pak i v postavení druhů v rámci jednotlivých skupin.

Topologie skupin získaná na základě jaderných genů vykazuje vysokou podporu a je ve shodě s prací Záveská et al. (2016). Topologie vzniklá na základě chloroplastového genomu je sice podobná topologii založené na třech lokusech v práci Záveská et al. (2012), avšak není příliš podpořená. Topologie skupin z rDNA cistronu je v případě mapování na referenci shodná s jadernou fylogenezí, v případě *de-novo* assembly se postavení skupin liší, avšak obě tyto varianty vykazují nízkou statistickou podporu podobně jako vztahy v práci Záveská et al. (2012) založené na ITS lokusu.

Čtyři druhy navržené jako hybridní na základě odlišného postavení v jednotlivých fylogenetických stromech u předchozích studií (Záveská et al., 2012, 2016) vykazovaly odlišnou topologii i ve vztazích získaných na základě NGS dat. V celkovém species tree založeném na jaderných low-copy genech stojí *C. candida* na bázi celého podrodu *Curcuma* a ostatní tři druhy (*C. myanmarensis*, *C. roscoeana* a *C. vamana*) zaujímají s vysokou podporou bazální postavení ke skupině *Hitcheniopsis*. Oproti tomu v rDNA stromu se *C. vamana* a *C. candida* objevují na bázi po spojení skupin *Curcuma* a *Hitcheniopsis* s nízkou podporou a *C. myanmarensis* s *C. roscoeana* zaujímají pozici na bázi linie *Curcuma*. V chloroplastovém lokusu je naopak *C. candida* součástí skupiny *Curcuma I* a ostatní tři druhy stojí na bázi linie *Curcuma*.

Co se týká postavení druhů v rámci jednotlivých skupin, *C. corniculata* zaujímá v chloroplastové fylogenezi pozici na bázi podrodu *Ecomata*, naopak ve species tree z jaderných genů stojí ve vrcholové části skupiny, což by mohlo být způsobeno buď procesem chloroplast capture, nebo incomplete lineage sorting chloroplastového genomu. Další odlišnou topologii zaujímá *C. singularis* a *C. vitellina*, které tvoří skupinu v chloroplastovém stromu, avšak v jaderném species tree stojí *C. singularis* na bazální pozici skupiny *Pierreana*, naopak *C. vitellina* v postavení vrcholovém. Oba tyto druhy vykazují vysoké množství chybějících dat a dlouhé větve v chloroplastové fylogenezi, proto by se mohlo jednat o artefakt dlouhých větví podobně jako u rodu *Oxalis* (Schmickl et al., 2016). Možným vysvětlením je také incomplete lineage sorting, případně chloroplast capture, protože v práci Záveská et al. (2012) tyto dva druhy vykazují podobné postavení, bohužel nepříliš podpořené.

Cytonukleární discordance byly v poslední době zaznamenávány hojně (např. v pracích Huang et al., 2010; Schmickl et al., 2016; Weitemier et al., 2014), což souhlasí s tvrzením, že organelární fylogeneze samostatně nejsou schopné reprezentovat skutečný species tree (Heyduk et al., 2016; Schmickl et al., 2016). Odlišnosti mezi fylogenezemi získanými z jaderné a chloroplastové DNA bývají interpretovány jako chloroplast capture (např. Parks et al., 2009), incomplete lineage sorting nebo hybridizační události (např. Fehrer et al., 2007; Pelsner et al., 2010; Renoult et al., 2009). Skutečná příčina inkongruencí u rodu *Curcuma* nebyla testována, avšak na základě velmi krátkých větví jednotlivých skupin u chloroplastové fylogeneze způsobené zřejmě rychlou radiací jednotlivých skupin (Záveská, 2014) by mohlo být stejně jako v případě rodu *Oxalis* (Schmickl et al., 2016) usuzováno na incomplete lineage sorting. Hybridizační události jsou také v tomto případě vysoce

pravděpodobné jak z pozorování v terénu, tak z molekulárních studií, na jejichž základě byl navržen a testován hybridní vznik několika druhů (Záveská et al., 2012, 2016).

Odlišnosti stromů vzniklých z rDNA cistronu od vztahů odvozených z jaderných low-copy genů mohou být způsobeny procesy probíhajícími u rDNA jako je rekombinace, nebo nedokončená homogenizace homeologních lokusů, což souhlasí s prací Záveská et al. (2012), kde u některých druhů rodu *Curcuma* bylo získáno klonováním více alel ITS lokusu. Z NGS dat byla získána pouze jedna majoritní alela na druh i v případě *de-novo* assembly jednotlivých readů, kde byly využity pro rekonstrukci fylogeneze pouze dlouhé sekvence (> 1 000 bp) a podrobnější analýza kratších úseků by mohla odhalit také vnitroindividuální variabilitu ITS, ETS nebo NTS lokusu. Další možnou příčinou způsobující nízké podpory jednotlivých větví a špatně vyřešenou fylogenezi ribozomálního lokusu je nedostatek signálu v datech (Schmickl et al., 2016), a proto se nejeví jako vhodné využití ribozomálního cistronu, případně jeho částí, k odvozování celkové fylogeneze skupin obsahující polyploidní a hybridní druhy, jak bylo popsáno např. již v práci Alvarez and Wendel (2003).

4.2.4 Redukované datasey

S příchodem metod next-generation sekvenování byl překonán hlavní problém fylogenetiky – nedostatek dat. V dnešní době je možné osekvenovat celé organelární genomy, rDNA cistron a stovky až tisíce nezávislých jaderných genů v jednom sekvenačním běhu a na jejich základě pak vytvářet fylogenetické stromy (např. Heyduk et al., 2016; Schmickl et al., 2016; Weitemier et al., 2014). Z toho vyvstává otázka: kolik nezávislých lokusů (genů) je potřeba pro kvalitní rekonstrukci fylogeneze? K jejímu zodpovězení byly vytvořeny species trees z 5, 20, 100, 200 a 500 genů a spočítána Robinson-Fouldova vzdálenost mezi stromy jako ukazatel, jak blízké nebo vzdálené jsou si topologie jednotlivých species trees vytvořené ze stejného počtu náhodně vybraných genů.

Trend snižující se RF vzdálenosti s rostoucím množstvím využitých genů napovídá, že čím vyšší počet genů je zahrnut do analýzy, tím je z nich odvozená fylogeneze přesnější, což souhlasí s výsledky v práci Léveillé-Bourret et al. (2017). RF distance ze stromů odvozených pomocí jednotlivých metod (koalescenčních – ASTRAL a ASTRID, supertree – MRL a supermatrix metody – Maximum Likelihood z konkatenovaného datasetu) sice vykazují drobné rozdíly, avšak není zde patrný žádný trend nejlepší nebo nejhorší metody. Stejně tak nebyly tyto distance ovlivněny přítomností tří meziliniových hybridů v datasetu – dataset s i bez hybridních druhů vykazoval obdobné výsledky.

Rozdíl v přesnosti odvození fylogeneze z 5 až 500 genů je značný, u 5 genů se jedná o zhruba 60% odlišnost mezi jednotlivými stromy, respektive zhruba o 50% odlišnost od stromu získaného z kompletního datasetu (1 154 genů), u stromů z 500 genů se jedná pouze asi o 10%, respektive 8% odlišnost. Přesnost odvozených fylogenezí se s rostoucím počtem genů zvyšuje zhruba lineárně, stejně jako v případě Léveillé-Bourret et al. (2017), kde klesá RF distance z 35 % u 5 genů po 7 % u 200 genů (porovnáváno se stromem z 460 genů). Rozdílná míra Robinson-Fouldovy vzdálenosti mezi stromy odvozenými ze stejného počtu genů (35 % oproti 60 % u pěti genů atd.) by mohla být způsobena přítomností hybridních a polyploidních druhů v kurkumovém datasetu nebo obecněji vyšší mírou inkongruence mezi jednotlivými genovými stromy, případně vyšší mírou ILS. I celkový počet genů zahrnutých v analýze by mohl mít potenciálně vliv na konečný výsledek – při použití polovičního počtu genů z celkového datasetu dosahuje RF distance okolo 7 % v obou datasetech a při zhruba čtvrtině celkového počtu genů okolo 10 % (Léveillé-Bourret et al., 2017), respektive asi 13 % u kurkum. Značná část odlišností mezi odvozovanými fylogenezemi by mohla být způsobena také chybovostí jednotlivých metod při rekonstrukci species tree – v práci Kupczok et al. (2010) bylo prokázáno, že při odvozování fylogeneze z genů s vysokou mírou ILS dosahují koalescenční i konkatenáčnické metody chybovosti 20 % a více a při testování spolehlivosti metod ASTRAL a ASTRID byla zjištěna zvyšující se přesnost odvozené fylogeneze s rostoucím počtem genů zahrnutých v analýze – v případě střední hodnoty ILS klesala RF distance z ~40 % u 10 genů po ~5 % u 800 genů (Vachaspati and Warnow, 2015).

Dalším vysvětlením by mohla být menší informativnost jednotlivých genů – v této práci jednalo o vnitrorodovou fylogenezi, u Léveillé-Bourret et al. (2017) o fylogenezi cladu zahrnující tři odlišné rody. Vliv hloubky fylogeneze na počet genů potřebných pro její správnou rekonstrukci je patrný také z dalších prací, kde byl zjištěn minimální počet 50 genů potřebných pro odvození fylogeneze krytosemenných (Zeng et al., 2014) a vztahů v rámci čeledi *Vitaceae* (Wen et al., 2013), ale 120 genů pro odvození vnitrorodové fylogeneze rodu *Oryza* (Zou et al., 2008) a 175 u rodu *Primulina* (Ai and Kang, 2015). Tento trend je pozorovatelný i u kurkumového datasetu, kde při využití pouze pěti genů pro tvorbu fylogenetického stromu se druhy zařadí do správných skupin a při využití 20 genů tyto skupiny vykazují i správnou topologii. Pro zařazení druhů do správné skupiny v rámci podrodu *Curcuma* (skupiny *Curcuma I* a *Curcuma II*) u všech opakování bylo potřeba i 200 genů, což souhlasí s poznatky o komplikovanosti vztahů v rámci tohoto podrodu díky časté hybridizaci a polyploidizaci (Škorníčková, 2007; Záveská et al., 2012) a ani v případě použití

500 genů nebyly vždy vytvořené fylogenetické stromy shodné s výsledným stromem z 1 154 jaderných low-copy genů.

Počet potřebných genů může ovlivňovat i počet taxonů zahrnutý do jednotlivých analýz a komplexnost odvozované fylogeneze – vnitrorodové fylogeneze byly odvozovány pouze pomocí 11 druhů (Ai and Kang, 2015; Zou et al., 2008) a fylogeneze *Vitaceae* s pomocí 15 druhů (Wen et al., 2013). V kurkumovém datasetu a práci Léveillé-Bourret et al. (2017) bylo zahrnuto okolo 30 taxonů, čímž vzrostl i počet vyšetřovaných nodů a množství možných inkongruencí mezi topologiemi jednotlivých genových stromů, což podporují i výsledky v práci Ai and Kang (2015), kde se po odsranění tří problematických druhů snížil počet potřebných genů více než dvojnásobně (ze 175 na 75). Obecně lze tedy doporučit využití většího množství genů na nižších úrovních fylogeneze a při vyšší komplexnosti vyšetřované skupiny, u vnitrorodové fylogeneze polyploidních a hybridogenních skupin jako je rod *Curcuma* pak alespoň 500 genů.

4.3 TESTOVÁNÍ HYBRIDNÍHO PŮVODU VYBRANÝCH DRUHŮ

4.3.1 Fylogenetické stromy a sítě

I když není vhodné využívat fylogenezi chloroplastového genomu nebo rDNA lokusu jako celkový species tree a odlišnosti mezi těmito stromy mohly vzniknout nejen hybridizací, ale i pomocí dalších evolučních fenoménů (Degnan and Rosenberg, 2009; Maddison, 1997), mohou být tyto inkongruence prvním signálem ukazujícím na možný hybridní původ příslušných jedinců/druhů (např. Pardo et al., 2004; Závěská et al., 2012). Na základě porovnání fylogenetických stromů ze tří různých částí genomu (cpDNA, rDNA a jaderné geny) byly odhaleny čtyři druhy vykazující odlišné postavení v rámci jednotlivých skupin. *Curcuma myanmarensis*, *C. roscoeana* a *C. vamana*, které vykazovaly postavení buď na bázi skupiny *Curcuma*, nebo na bázi skupiny *Hitcheniopsis* a *C. candida*, která vykazovala dokonce tři alternativní pozice – na bázi podrodu *Curcuma*, vrcholovém postavení ve skupině *Curcuma I* a na bázi po spojení skupin *Curcuma* a *Hitcheniopsis*. Postavení *C. candida* se také lišilo ve species trees vytvořených pomocí různých metod z jaderných lokusů – ve dvou případech stál tento druh na bázi skupiny *Curcuma I*, v ostatních pěti případech na bázi celého podrodu *Curcuma* a obě tyto topologie byly často doprovázeny sníženou statistickou podporou. Tyto čtyři druhy byly vyšetřovány jako potenciálně hybridní již v předešlé práci

(Záveská et al., 2016) na základě čtyř jaderných genů, ITS a cpDNA a kromě druhu *C. candida* byly všechny identifikovány jako pravděpodobní meziliniovní hybridy.

Fylogenetické sítě získané metodou Neighbor-Net se většinou shodovaly se vztahy odvozenými v rámci fylogenetických stromů, avšak síť z jaderných low-copy genů ukázala kromě čtyř oddělených fylogenetických linií také zesíťované vztahy tří druhů – *C. myanmarensis*, *C. roscoeana* a *C. vamana*, a jejich postavení uprostřed mezi liniemi *Curcuma* a *Hitcheniopsis*. Jako hybridní s původem z báze podrodů *Curcuma* a *Hitcheniopsis*, byly tyto druhy odvozeny i při analýze MPL v programu PhyloNet, což značně podporuje hypotézu jejich vzniku meziliniovým křížením. Co se týká vztahů mezi druhy, tvoří *C. myanmarensis* a *C. roscoeana* jak ve fylogenetických sítích (Neighbor-Net i MPL-PhyloNet), tak ve všech fylogenetických stromech z jaderných low-copy genů i chloroplastového lokusu jasnou skupinu, kdežto *C. vamana* přistupuje k ostatním druhům samostatně, což podporuje i geografie – stenoendemické rozšíření druhu *C. vamana* asi 2 000 km vzdálené od překrývajících se areálů ostatních dvou druhů (Záveská et al., 2012, 2016).

Curcuma candida zaujímala ve fylogenetické síti získané metodou Neighbor-Net z jaderných low-copy genů pozici na bázi podrodu *Curcuma* se značným zesíťováním směrem ke skupině *Curcuma I*, avšak ne se skupinou *Curcuma II*, což ukazuje na hybridizační událost mezi skupinou *Curcuma I* a druhem/linií stojící na bázi celého podrodu *Curcuma*, avšak v této studii buď nezahrnuté, případně vyhynulé. V síti vytvořené pomocí rDNA lokusu je patrné také zesíťování směrem k linii *Pierreana*, což mohlo zapříčinit zařazení *C. candida* do této skupiny na základě ITS lokusu (Záveská et al., 2012). Specifická morfologie *C. candida* nejvíce podobná druhům z linie *Pierreana* by mohla podporovat hypotézu zesíťování díky introgresi, avšak na základě testů monofylie a lineage movement analýzy se to zdá jako nepravděpodobné. Zesíťování směrem k linii *Pierreana* by spíše mohl způsobit incomplete lineage sorting vzhledem k pravděpodobnému vzniku linií *Curcuma* a *Ecomata* + *Pierreana* ze společného předka (Záveská et al., 2016). Hybridní vznik druhu *C. candida* podporuje také jedna síť (s nejvyšší hodnotou total log probability) z MPL analýzy v programu PhyloNet s hybridizační událostí mezi skupinou *Curcuma I* a předkem skupiny *Curcuma I*. Vzhledem k recentnějšímu vzniku tohoto druhu oproti ostatním hybridům, komplexním vztahům a menšímu množství signálu v datech (viz testy monofylie) by bylo vhodné se zaměřit důkladněji na podrod *Curcuma* – zvětšit sampling jak do počtu druhů, tak do počtu zahrnutých jedinců jednoho druhu a ověřit možné rodičovské druhy/linie na větším datasetu.

4.3.2 *Metody shrnující odlišné topologie nukleárních genů*

K podpoření hypotézy o hybridním vzniku čtyř druhů, případně k odhalení dalších hybridů byly provedeny dva testy sumarizující množství genových stromů vykazující alternativní topologie. Lineage movement analýza byla velmi rychlá a pro její provedení nebyly potřeba apriori žádné další informace. Pro testy monofylie bylo nutné vytvořit několik genových sad k testování každého hybridního druhu zvlášť a také bylo nezbytné předem definovat monofyletické skupiny, což vyžadovalo jistou znalost testované skupiny. Oproti lineage movement analýze bylo možné provést testy monofylie např. pouze s geny vykazující silný fylogenetický signál (zde využity stromy se zkolabovanými větvemi s BS <70), což zvyšuje věrohodnost výsledků, avšak může vést k vyřazení velkého množství dat, zvláště na nižších úrovních vyšetřované fylogeneze a u skupin, kde proběhla rychlá radiace, což je zřejmě i případ rodu *Curcuma* (Záveská, 2014).

Lineage movement analýza i testy monofylie podporují výsledky získané z fylogenetických stromů a sítí, tj. přítomnost čtyř hybridních druhů – *C. candida*, *C. myanmarensis*, *C. roscoeana* a *C. vamana*, další potenciálně hybridní druhy v datasetu nalezeny nebyly.

Lineage movement analýza ukázala přítomnost druhu *C. candida* v 35 % genových stromů v rámci skupiny *Curcuma I*, do skupiny *Curcuma II* spadala v 5 % případů a podobné množství genů připadlo i na skupinu „Hybridí“. V konsenzuálním stromu zaujímala *C. candida* bazální pozici k podrodu *Curcuma*, tzn. na tuto pozici připadalo asi 55 % genů. To mírně kontrastuje s výsledky testů monofylie, které podporují topologii *C. candida* jako součást skupiny *Curcuma I* pouze u ~30 genů, součást skupiny *Curcuma II* u 10 genů a bazální postavení k podrodu *Curcuma* v ~40 případech z ~1 100 genů. Nutno podotknout, že skupina *Curcuma I* byla monofyletická pouze u ~30 genů, *Curcuma II* u ~20 genů a celý podrod *Curcuma* pouze u ~550 genů (ostatní podrody vykazovaly monofylii v ~500 – ~800 případech). Nízké množství genů podporující monofylii jednotlivých skupin (především *Curcuma I* a *Curcuma II*) je způsobeno zřejmě nedostatkem signálu v datech (nízká podpora jednotlivých nodů) v důsledku rychlé radiace rodu (Záveská, 2014), což souhlasí s výsledky práce Heyduk et al. (2016), kde bylo nalezeno velké množství genů se špatně podpořenými topologiemi u rodu *Sabal*, který také prodělal rychlou radiaci, případně i častými hybridizačními událostmi v rodu *Curcuma* (Škorníčková, 2007; Záveská et al., 2012).

Co se týká druhů *C. myanmarensis*, *C. roscoeana* a *C. vamana*, lineage movement analýza ukázala přítomnost těchto druhů v rámci podrodu *Curcuma* u ~5–7 % a v rámci *Hitcheniopsis* u ~1,5–3 % genů. Zařazení hybridů k linii *Hitcheniopsis* podporuje také jejich postavení v konsenzuálním stromu na bázi této skupiny a velké množství genů s topologiemi v rámci vlastní skupiny „Hybridi“, kam spadá *C. vamana* v ~14 %, *C. myanmarensis* a *C. roscoeana* v ~50 % případů. U těchto tří druhů se naopak zdají průkaznější testy monofylie, které odhalily >100 genů spadajících do podrodu *Curcuma*, >140 genů monofyletických s podrodem *Hitcheniopsis* a >130 genů vykazující pozici na bázi po spojení linií *Curcuma* a *Hitcheniopsis*, které se u ostatních druhů téměř nevyskytují. Topologie těchto genů může být zapříčiněna rekombinací mezi genovými kopiemi z jednotlivých rodičovských linií, nebo reflektovat hlubokou koalescenci jako pozůstatek rychlé radiace jednotlivých druhů z hlavních linií (Záveská et al., 2016), případně vznik těchto druhů z předků obou rodičovských linií, což podporují i výsledky MPL analýzy v programu PhyloNet ukazující na jejich vznik z báze obou linií.

Detekování potenciálně hybridních druhů na základě několika set low-copy genů není bez problémů – některé klasicky využívané přístupy a programy nejsou téměř schopné pracovat s velkými datasety v reálném čase (např. tvorba fylogenetických sítí pomocí MPL-PhyloNet; Yu and Nakhleh, 2015), jiné nedávají konzistentní výsledky (např. MP-PhyloNet; Kamneva and Rosenberg, 2017; Yu et al., 2013) a některé metody použitelné u NGS dat nejsou vhodné pro kladené otázky a vyšetřovanou skupinu z této práce, např. čtyřtaxonová D-statistika (Green et al., 2010), testování jednoho hybridního druhu vzhledem k jeho dvěma možným rodičům a outgroup, která byla využita u rodu *Carex* (Escudero et al., 2014) nebo mapování readů hybridu na referenci získanou z potenciálních nehybridních rodičovských druhů (Bombarely et al., 2014). Kombinací tří různých přístupů – tvorby fylogenetických sítí metodou Neighbor-Net (Bryant and Moulton, 2004) a MPL-PhyloNet (Yu and Nakhleh, 2015), lineage movement analýzy (Smith and Dunn, 2008) a testů monofylie, bylo dosaženo shodných výsledků – detekce čtyř hybridních druhů s rodiči spadajícími do různých skupin, avšak vývoj nových analýz na odhalování hybridních druhů z NGS dat se jeví jako potřebný. Možností ověření hybridního původu druhů pomocí nebioinformatických analýz by mohla být FISH (fluorescence *in situ* hybridisation) analýza (Chester et al., 2010) se sondami navrženými na základě liniově specifických repetitivních sekvencí (identifikace např. pomocí programu RepeatExplorer; Novak et al., 2013) z readů získaných v rámci necílové frakce Hyb-Seq dat.

5 ZÁVĚR

S využitím metody Hyb-Seq byly v této práci ověřeny vztahy mezi hlavními molekulárními liniemi u rodu *Curcuma* a odvozeny vztahy mezi druhy v rámci jednotlivých linií. Výsledky fylogenetických analýz na základě next-generation sekvenování ukazují velké zlepšení v rozlišení a statistické podpoře species trees z několika stovek jaderných low-copy genů v porovnání s tradičně používanými markery, jako je ITS sekvence nebo několik málo chloroplastových lokusů.

Inkongruence mezi fylogenezemi odvozenými z různých částí genomu (jaderné low-copy geny vs. chloroplastový genom a rDNA cistron) jsou u rodu *Curcuma* časté – liší se především vztahy mezi hlavními molekulárními liniemi a postavení čtyř hybridních druhů. Fylogeneze založená na jaderných low-copy genech vykazuje vysoké statistické podpory, oproti tomu vztahy odvozené z cpDNA a rDNA obsahují krátké větve a nízké podpory některých uzlů, zvláště s inkongruentní topologií, což dokládá nevhodnost využití organelární a ribozomální fylogeneze k reprezentaci celkového species tree, zvláště u hybridních a polyploidních skupin.

Počet genů nezbytný pro věrohodnou rekonstrukci species tree závisí především na hloubce odvozované fylogeneze – správné vztahy mezi čtyřmi hlavními liniemi rodu *Curcuma* jsou zrekonstruovány již při využití nízkého počtu genů (okolo 20), avšak pro věrohodné odvození vztahů v rámci jednotlivých skupin je nutné zahrnout mnohem vyšší počet genů (500 nebo i více). Významnou roli hraje zřejmě i komplexita studovaných fylogenetických vztahů, kdy zejména přítomnost hybridních druhů může vnést do výsledného species tree určitou míru nejasnosti i za použití několika set genů. Čtyři metody reprezentující tři hlavní přístupy k rekonstrukci fylogeneze (koalescenční metody, supertree a konkatenace) porovnávané v rámci této práce dosáhly ve všech případech obdobné míry nepřesnosti snižující se s rostoucím počtem genů zahrnutých do dané analýzy, proto z jich nelze vybrat nejvhodnější metodu.

Stovky jaderných low-copy genů poskytují velké množství dat pro detekci případných hybridních druhů v datasetu, nicméně samotné provedení může být stále náročné vzhledem k tomu, že některé klasicky využívané analýzy nejsou pro velké datasety vhodné a jiné nemusí dát vždy přesvědčivé výsledky. S pomocí několika typů analýz byla ověřena hypotéza tří hybridních druhů (*C. myanmarensis*, *C. roscoeana* a *C. vamana*) vzniklých meziliniovým křížením rodičovských druhů spadajících do podrodů *Curcuma* a *Hitcheniopsis*. Byl také

odhalen čtvrtý potenciálně hybridogenní druh – *C. candida*. V tomto případě se nejedná o hybrida vzniklého křížením druhů z odlišných podrodů, ale z různých skupin v rámci podrodu *Curcuma*. Jak se v této práci ukázalo, *C. candida* pravděpodobně vznikla z rodičů spadajících do skupiny *Curcuma I* a na bázi celého podrodu *Curcuma*.

6 SEZNAM POUŽITÉ LITERATURY

- Ai, B., and Kang, M. (2015). How Many Genes are Needed to Resolve Phylogenetic Incongruence? *Evolutionary Bioinformatics* 185.
- Allen, J.M., Huang, D.I., Cronk, Q.C., and Johnson, K.P. (2015). aTRAM - automated target restricted assembly method: a fast method for assembling loci across divergent taxa from next-generation sequencing data. *BMC Bioinformatics* 16, 98.
- Alvarez, I., and Wendel, J.F. (2003). Ribosomal ITS sequences and plant phylogenetic inference. *Mol. Phylogenet. Evol.* 29, 417–434.
- Ammon, H.P., and Wahl, M.A. (1991). Pharmacology of *Curcuma longa*. *Planta Med.* 57, 1–7.
- Anuntalabhochai, S., Sitthiphrom, S., Thongtaksin, W., Sanguansermisri, M., and Cutler, R.W. (2007). Hybrid detection and characterization of *Curcuma* spp. using sequence characterized DNA markers. *Scientia Horticulturae* 111, 389–393.
- Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A., and Johnson, E.A. (2008). Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. *PLOS ONE* 3, e3376.
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., et al. (2012). SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J Comput Biol* 19, 455–477.
- Barrett, C.F., Specht, C.D., Leebens-Mack, J., Stevenson, D.W., Zomlefer, W.B., and Davis, J.I. (2014). Resolving ancient radiations: can complete plastid gene sets elucidate deep relationships among the tropical gingers (Zingiberales)? *Ann. Bot.* 113, 119–133.
- Bayzid, M.S., and Warnow, T. (2012). Estimating optimal species trees from incomplete gene trees under deep coalescence. *J. Comput. Biol.* 19, 591–605.
- Bayzid, M.S., and Warnow, T. (2013). Naive binning improves phylogenomic analyses. *Bioinformatics* 29, 2277–2284.
- Bentley, D.R., Balasubramanian, S., Swerdlow, H.P., Smith, G.P., Milton, J., Brown, C.G., Hall, K.P., Evers, D.J., Barnes, C.L., Bignell, H.R., et al. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 456, 53–59.
- Bleidorn, C., Podsiadlowski, L., Zhong, M., Eeckhaut, I., Hartmann, S., Halanych, K.M., and Tiedemann, R. (2009). On the phylogenetic position of Myzostomida: can 77 genes get it wrong? *BMC Evolutionary Biology* 9, 150.
- Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120.
- Bombarely, A., Coate, J.E., and Doyle, J.J. (2014). Mining transcriptomic data to study the origins and evolution of a plant allopolyploid complex. *PeerJ* 2, e391.

- Bryant, D., and Moulton, V. (2004). Neighbor-Net: An Agglomerative Method for the Construction of Phylogenetic Networks. *Mol Biol Evol* 21, 255–265.
- Bybee, S.M., Bracken-Grissom, H., Haynes, B.D., Hermansen, R.A., Byers, R.L., Clement, M.J., Udall, J.A., Wilcox, E.R., and Crandall, K.A. (2011). Targeted amplicon sequencing (TAS): a scalable next-gen approach to multilocus, multitaxa phylogenetics. *Genome Biol Evol* 3, 1312–1323.
- Cannon, S.B., McKain, M.R., Harkess, A., Nelson, M.N., Dash, S., Deyholos, M.K., Peng, Y., Joyce, B., Stewart, C.N., Rolf, M., et al. (2015). Multiple Polyploidy Events in the Early Radiation of Nodulating and Nonnodulating Legumes. *Mol Biol Evol* 32, 193–210.
- Chester, M., Leitch, A.R., Soltis, P.S., and Soltis, D.E. (2010). Review of the Application of Modern Cytogenetic Methods (FISH/GISH) to the Study of Reticulation (Polyploidy/Hybridisation). *Genes* 1, 166–192.
- Cronn, R., Knaus, B.J., Liston, A., Maughan, P.J., Parks, M., Syring, J.V., and Udall, J. (2012). Targeted enrichment strategies for next-generation plant biology. *Am. J. Bot.* 99, 291–311.
- Degnan, J.H., and Rosenberg, N.A. (2009). Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology & Evolution* 24, 332–340.
- Dufresne, F., Stift, M., Vergilino, R., and Mable, B.K. (2014). Recent progress and challenges in population genetics of polyploid organisms: an overview of current state-of-the-art molecular and statistical tools. *Mol. Ecol.* 23, 40–69.
- Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S., and Mitchell, S.E. (2011). A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLOS ONE* 6, e19379.
- Escudero, M., Eaton, D.A.R., Hahn, M., and Hipp, A.L. (2014). Genotyping-by-sequencing as a tool to infer phylogeny and ancestral hybridization: A case study in *Carex* (Cyperaceae). *Molecular Phylogenetics and Evolution* 79, 359–367.
- Faircloth, B.C. (2016). PHYLUCE is a software package for the analysis of conserved genomic loci. *Bioinformatics* 32, 786–788.
- Fehrer, J., Gemeinholzer, B., Chrték, J., and Bräutigam, S. (2007). Incongruent plastid and nuclear DNA phylogenies reveal ancient intergeneric hybridization in *Pilosella* hawkweeds (Hieracium, Cichorieae, Asteraceae). *Mol. Phylogenet. Evol.* 42, 347–361.
- Glenn, T.C. (2011). Field guide to next-generation DNA sequencers: FIELD GUIDE TO NEXT-GEN SEQUENCERS. *Molecular Ecology Resources* 11, 759–769.
- Gnirke, A., Melnikov, A., Maguire, J., Rogov, P., LeProust, E.M., 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. *Nat Biotech* 27, 182–189.
- Green, R.E., Krause, J., Briggs, A.W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M.H.Y., et al. (2010). A Draft Sequence of the Neandertal Genome. *Science* 328, 710–722.

- Hall, T. (1999). {BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT}. pp. 95–98.
- Hammer, O., Harper, D.A.T., and Ryan, P.D. (2001). PAST: Paleontological Statistics Software Package for Education and Data Analysis. *Palaeontologia Electronica*.
- Hand, M.L., Cogan, N.O.I., and Forster, J.W. (2012). Genome-wide SNP identification in multiple morphotypes of allohexaploid tall fescue (*Festuca arundinacea* Schreb). *BMC Genomics* *13*, 219.
- Hayakawa, H., Minaniya, Y., Ito, K., Yamamoto, Y., and Fukuda, T. (2011). Difference of Curcumin Content in *Curcuma longa* L. (Zingiberaceae) Caused by Hybridization with Other *Curcuma* Species. *American Journal of Plant Sciences* *2*, 111.
- Heled, J., and Drummond, A.J. (2010). Bayesian Inference of Species Trees from Multilocus Data. *Mol Biol Evol* *27*, 570–580.
- Heyduk, K., Trapnell, D.W., Barrett, C.F., and Leebens-Mack, J. (2016). Phylogenomic analyses of species relationships in the genus *Sabal* (Arecaceae) using targeted sequence capture. *Biological Journal of the Linnean Society* *117*, 106–120.
- Huang, H., He, Q., Kubatko, L.S., and Knowles, L.L. (2010). Sources of error inherent in species-tree estimation: impact of mutational and coalescent effects on accuracy and implications for choosing among different methods. *Syst. Biol.* *59*, 573–583.
- Huson, D.H., and Bryant, D. (2005). Application of Phylogenetic Networks in Evolutionary Studies. *Molecular Biology and Evolution* *23*, 254–267.
- Hutchison, C.A. (2007). DNA sequencing: bench to bedside and beyond. *Nucl. Acids Res.* *35*, 6227–6237.
- Islam, M.A. (2004). Genetic diversity of the genus *Curcuma* in Bangladesh and further biotechnological approaches for in vitro regeneration and long-term conservation of *C. longa* germplasm. PhD Thesis, University of Hannover.
- Johnson, M.G., Gardner, E.M., Liu, Y., Medina, R., Goffinet, B., Shaw, A.J., Zerega, N.J.C., and Wickett, N.J. (2016). HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment1. *Appl Plant Sci* *4*.
- Joseph, R., Joseph, T., and Jose, J. (1999). Karyomorphological studies in the genus *Curcuma* Linn. *Cytologia* *64*, 313–317.
- Junier, T., and Zdobnov, E.M. (2010). The Newick utilities: high-throughput phylogenetic tree processing in the UNIX shell. *Bioinformatics* *26*, 1669–1670.
- Kamneva, O., and Rosenberg, N. (2017). Simulation-Based Evaluation of Hybridization Network Reconstruction Methods in the Presence of Incomplete Lineage Sorting. *Evolutionary Bioinformatics* *13*.
- Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* *30*, 772–780.

- Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 30, 3059–3066.
- 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, W.J. (2002). BLAT--the BLAST-like alignment tool. *Genome Res.* 12, 656–664.
- Ketmaro, S., Taychasinpitak, T., Mongkolchaiyaphruek, A., and Wongchaochant, S. (2012). Effect of Colchicine on Increasing Pollen Viability in a *Curcuma* Hybrid (*Curcuma sparganifolia* × *C. parviflora*). *Kasetsart J. (Nat. Sci.)* 46, 363–370.
- Kozlov, A.M., Aberer, A.J., and Stamatakis, A. (2015). ExaML version 3: a tool for phylogenomic analyses on supercomputers. *Bioinformatics* 31, 2577–2579.
- Kress, W.J. (1990). The phylogeny and classification of the Zingiberales. *Ann. Missouri Bot. Gard* 698–721.
- Kress, W.J., Prince, L.M., Hahn, W.J., and Zimmer, E.A. (2001). Unraveling the evolutionary radiation of the families of the Zingiberales using morphological and molecular evidence. *Syst. Biol.* 50, 926–944.
- Kress, W.J., Prince, L.M., and Williams, K.J. (2002). The phylogeny and a new classification of the gingers (Zingiberaceae): evidence from molecular data. *Am. J. Bot.* 89, 1682–1696.
- Kubatko, L.S. (2009). Identifying Hybridization Events in the Presence of Coalescence via Model Selection. *Systematic Biology* 58, 478–488.
- Kubatko, L.S., Degnan, J.H., and Collins, T. (2007). Inconsistency of Phylogenetic Estimates from Concatenated Data under Coalescence. *Syst Biol* 56, 17–24.
- Kubatko, L.S., Carstens, B.C., and Knowles, L.L. (2009). STEM: species tree estimation using maximum likelihood for gene trees under coalescence. *Bioinformatics* 25, 971–973.
- Kupczok, A., Schmidt, H.A., and von Haeseler, A. (2010). Accuracy of phylogeny reconstruction methods combining overlapping gene data sets. *Algorithms Mol Biol* 5, 37.
- Lanfear, R., Calcott, B., Kainer, D., Mayer, C., and Stamatakis, A. (2014). Selecting optimal partitioning schemes for phylogenomic datasets. *BMC Evolutionary Biology* 14, 82.
- Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359.
- Lemmon, E.M., and Lemmon, A.R. (2013). High-Throughput Genomic Data in Systematics and Phylogenetics. *Annual Review of Ecology, Evolution, and Systematics* 44, 99–121.
- Leong-Škorničková, J., Šída, O., Jarolímová, V., Sabu, M., Fér, T., Trávníček, P., and Suda, J. (2007). Chromosome Numbers and Genome Size Variation in Indian Species of *Curcuma* (Zingiberaceae). *Ann Bot* 100, 505–526.

- Leong-Škorničková, J., Šída, O., Závěská, E., and Marhold, K. (2015). History of infrageneric classification, typification of supraspecific names and outstanding transfers in *Curcuma* (Zingiberaceae). *Taxon* *64*, 362–373.
- Léveillé-Bourret, É., Starr, J.R., Ford, B.A., Moriarty Lemmon, E., and Lemmon, A.R. (2017). Resolving Rapid Radiations within Angiosperm Families Using Anchored Phylogenomics. *Systematic Biology*.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* *25*, 2078–2079.
- Li, X., Acharya, A., Farmer, A.D., Crow, J.A., Bharti, A.K., Kramer, R.S., Wei, Y., Han, Y., Gou, J., May, G.D., et al. (2012). Prevalence of single nucleotide polymorphism among 27 diverse alfalfa genotypes as assessed by transcriptome sequencing. *BMC Genomics* *13*, 568.
- Linder, C.R., and Rieseberg, L.H. (2004). Reconstructing patterns of reticulate evolution in plants. *Am. J. Bot.* *91*, 1700–1708.
- Liu, L., and Yu, L. (2011). Estimating species trees from unrooted gene trees. *Syst. Biol.* *60*, 661–667.
- Maddison, W.P. (1997). Gene Trees in Species Trees. *Syst Biol* *46*, 523–536.
- Marioni, J.C., Mason, C.E., Mane, S.M., Stephens, M., and Gilad, Y. (2008). RNA-seq: An assessment of technical reproducibility and comparison with gene expression arrays. *Genome Research* *18*, 1509–1517.
- Martin, J.A., and Wang, Z. (2011). Next-generation transcriptome assembly. *Nature Reviews Genetics* *12*, 671–682.
- McCormack, J.E., Hird, S.M., Zellmer, A.J., Carstens, B.C., and Brumfield, R.T. (2013). Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution* *66*, 526–538.
- Meyer, M., Stenzel, U., and Hofreiter, M. (2008). Parallel tagged sequencing on the 454 platform. *Nat. Protocols* *3*, 267–278.
- Mirarab, S., Reaz, R., Bayzid, M.S., Zimmermann, T., Swenson, M.S., and Warnow, T. (2014). ASTRAL: genome-scale coalescent-based species tree estimation. *Bioinformatics* *30*, I541–I548.
- Newton, R.R., and Newton, I.L.G. (2013). PhyBin: binning trees by topology. *PeerJ* *1*, e187.
- Ngamriabsakul, C., Newman, M.F., and Cronk, Q.C.B. (2003). The phylogeny of tribe Zingibereae (Zingiberaceae) based on ITS (nr DNA) and trnL-F (cp DNA) sequences. *Edinburgh Journal of Botany* *60*.
- Nguyen, N., Mirarab, S., and Warnow, T. (2012). MRL and SuperFine plus MRL: new supertree methods. *Algorithms. Mol. Biol.* *7*, 3.

- Njuguna, W., Liston, A., Cronn, R., Ashman, T.-L., and Bassil, N. (2013). Insights into phylogeny, sex function and age of *Fragaria* based on whole chloroplast genome sequencing. *Molecular Phylogenetics and Evolution* 66, 17–29.
- Novak, P., Neumann, P., Pech, J., Steinhaisl, J., and Macas, J. (2013). RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. *Bioinformatics* 29, 792–793.
- Okou, D.T., Steinberg, K.M., Middle, C., Cutler, D.J., Albert, T.J., and Zwick, M.E. (2007). Microarray-based genomic selection for high-throughput resequencing. *Nature Methods* 4, 907–909.
- Paisooksantivatana, Y., Kako, S., and Seko, H. (2001). Genetic diversity of *Curcuma alismatifolia* Gagnep. (Zingiberaceae) in Thailand as revealed by allozyme polymorphism. *Genet. Resour. Crop Evol.* 48, 459–465.
- Pardo, C., Cubas, P., and Tahiri, H. (2004). Molecular phylogeny and systematics of *Genista* (Leguminosae) and related genera based on nucleotide sequences of nrDNA (ITS region) and cpDNA (trnL-trnF intergenic spacer). *Plant Syst. Evol.* 244, 93–119.
- Parks, M., Cronn, R., and Liston, A. (2009). Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. *BMC Biology* 7, 84.
- Pelser, P.B., Kennedy, A.H., Tepe, E.J., Shidler, J.B., Nordenstam, B., Kadereit, J.W., and Watson, L.E. (2010). Patterns and causes of incongruence between plastid and nuclear *Senecioneae* (Asteraceae) phylogenies. *Am. J. Bot.* 97, 856–873.
- Pickrell, J.K., and Pritchard, J.K. (2012). Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. *PLOS Genet* 8, e1002967.
- Price, M.N., Dehal, P.S., and Arkin, A.P. (2009). FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* 26, 1641–1650.
- Price, M.N., Dehal, P.S., and Arkin, A.P. (2010). FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLOS ONE* 5, e9490.
- Ramachandran, K. (1969). Chromosome Numbers in Zingiberaceae. *Cytologia* 34, 213–221.
- Renoult, J.P., Kjellberg, F., Grout, C., Santoni, S., and Khadari, B. (2009). Cyto-nuclear discordance in the phylogeny of *Ficus* section *Galoglychia* and host shifts in plant-pollinator associations. *BMC Evolutionary Biology* 9, 248.
- Robinson, D.F., and Foulds, L.R. (1981). Comparison of phylogenetic trees. *Mathematical Biosciences* 53, 131–147.
- Ruby, A.J., Kuttan, G., Babu, K.D., Rajasekharan, K.N., and Kuttan, R. (1995). Anti-tumour and antioxidant activity of natural curcuminoids. *Cancer Lett.* 94, 79–83.
- Sang, T. (2002). Utility of low-copy nuclear gene sequences in plant phylogenetics. *Crit. Rev. Biochem. Mol. Biol.* 37, 121–147.

- Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U.S.A.* *74*, 5463–5467.
- Schmickl, R., Liston, A., Zeisek, V., Oberlander, K., Weitemier, K., Straub, S.C.K., Cronn, R.C., Dreyer, L.L., and Suda, J. (2016). Phylogenetic marker development for target enrichment from transcriptome and genome skim data: the pipeline and its application in southern African *Oxalis* (Oxalidaceae). *Mol Ecol Resour* *16*, 1124–1135.
- Schwarzer, J., Misof, B., Tautz, D., and Schlieven, U.K. (2009). The root of the East African cichlid radiations. *BMC Evol Biol* *9*, 186.
- Škorničková, J., Sabu, M., and Prasanthkumar, M.G. (2004). *Curcuma mutabilis* (Zingiberaceae): a new species from South India. *Garden's Bulletin Singapore* 43–54.
- Škorničková, J. (2007). Taxonomic Studies in Indian *Curcuma* L. PhD. Thesis. Charles University, Prague.
- Small, R.L., Cronn, R.C., and Wendel, J.F. (2004). L. A. S. JOHNSON REVIEW No. 2. Use of nuclear genes for phylogeny reconstruction in plants. *Australian Systematic Botany* *17*, 145.
- Smith, S.A., and Dunn, C.W. (2008). Phyutility: a phyloinformatics tool for trees, alignments and molecular data. *Bioinformatics* *24*, 715–716.
- Sousa, F. de, Bertrand, Y.J.K., Nylinder, S., Oxelman, B., Eriksson, J.S., and Pfeil, B.E. (2014). Phylogenetic Properties of 50 Nuclear Loci in *Medicago* (Leguminosae) Generated Using Multiplexed Sequence Capture and Next-Generation Sequencing. *PLOS ONE* *9*, e109704.
- Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* *30*, 1312–1313.
- Stegemann, S., Keuthe, M., Greiner, S., and Bock, R. (2012). Horizontal transfer of chloroplast genomes between plant species. *PNAS* *109*, 2434–2438.
- Straub, S.C.K., Parks, M., Weitemier, K., Fishbein, M., Cronn, R.C., and Liston, A. (2012). Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. *Am. J. Bot.* *99*, 349–364.
- Syring, J., Farrell, K., Businský, R., Cronn, R., and Liston, A. (2007). Widespread genealogical nonmonophyly in species of *Pinus* subgenus *Strobus*. *Syst. Biol.* *56*, 163–181.
- Than, C., Ruths, D., and Nakhleh, L. (2008). PhyloNet: a software package for analyzing and reconstructing reticulate evolutionary relationships. *BMC Bioinformatics* *9*, 322.
- Tsitroni, A., Kirkpatrick, M., and Levin, D.A. (2003). A model for chloroplast capture. *Evolution* *57*, 1776–1782.
- Vachaspati, P., and Warnow, T. (2015). ASTRID: Accurate Species TREes from Internode Distances. *BMC Genomics* *16 Suppl 10*, S3.

- Valeton, T. (1918). New notes on the Zingiberaceae of Java and Malaya. *Bulletin Du Jardin Botanique de Buitenzorg* 27, 1–167.
- Weitemier, K., Straub, S.C.K., Cronn, R.C., Fishbein, M., Schmickl, R., McDonnell, A., and Liston, A. (2014). Hyb-Seq: Combining target enrichment and genome skimming for plant phylogenomics. *Appl Plant Sci* 2.
- Wen, D., Yu, Y., and Nakhleh, L. (2016a). Bayesian Inference of Reticulate Phylogenies under the Multispecies Network Coalescent. *PLOS Genetics* 12, e1006006.
- Wen, D., Yu, Y., Hahn, M.W., and Nakhleh, L. (2016b). Reticulate evolutionary history and extensive introgression in mosquito species revealed by phylogenetic network analysis. *Molecular Ecology* 25, 2361–2372.
- Wen, J., Xiong, Z., Nie, Z.-L., Mao, L., Zhu, Y., Kan, X.-Z., Ickert-Bond, S.M., Gerrath, J., Zimmer, E.A., and Fang, X.-D. (2013). Transcriptome Sequences Resolve Deep Relationships of the Grape Family. *PLoS ONE* 8, e74394.
- Wendel, J.F., Schnabel, A., and Seelanan, T. (1995). An Unusual Ribosomal DNA Sequence from *Gossypium gossypioides* Reveals Ancient, Cryptic, Intergenomic Introgression. *Molecular Phylogenetics and Evolution* 4, 298–313.
- Wiens, J.J., Kuczynski, C.A., Smith, S.A., Mulcahy, D.G., Sites, J.W., Townsend, T.M., and Reeder, T.W. (2008). Branch lengths, support, and congruence: testing the phylogenomic approach with 20 nuclear loci in snakes. *Syst. Biol.* 57, 420–431.
- Wyman, S.K., Jansen, R.K., and Boore, J.L. (2004). Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20, 3252–3255.
- Yasuda, K., Tsuda, T., Shimizu, H., and Sugaya, A. (1988). Multiplication of curcuma species by tissue culture. *Planta Med.* 54, 75–79.
- Yu, Y., and Nakhleh, L. (2015). A maximum pseudo-likelihood approach for phylogenetic networks. *BMC Genomics* 16, S10.
- Yu, Y., Barnett, R.M., and Nakhleh, L. (2013). Parsimonious Inference of Hybridization in the Presence of Incomplete Lineage Sorting. *Systematic Biology* 62, 738–751.
- Yu, Y., Dong, J., Liu, K.J., and Nakhleh, L. (2014). Maximum likelihood inference of reticulate evolutionary histories. *PNAS* 111, 16448–16453.
- Záveská, E., Fér, T., Šída, O., Leong-Škorničková, J., Sabu, M., and Marhold, K. (2011). Genetic diversity patterns in *Curcuma* reflect differences in genome size: GENETIC DIVERSITY PATTERNS IN CURCUMA. *Botanical Journal of the Linnean Society* 165, 388–401.
- Záveská, E., Fér, T., Šída, O., Krak, K., Marhold, K., and Leong-Škorničková, J. (2012). Phylogeny of *Curcuma* (Zingiberaceae) based on plastid and nuclear sequences: Proposal of the new subgenus *Ecomata*. *Taxon* 61, 747–763.

Záveská, E., Fér, T., Šída, O., Marhold, K., and Leong-Škorničková, J. (2016). Hybridization among distantly related species: Examples from the polyploid genus *Curcuma* (Zingiberaceae). *Molecular Phylogenetics and Evolution* 100, 303–321.

Záveská, E. (2014). *Phylogenetic Studies in the Polyploid Genus Curcuma L.* Ph.D. Thesis, Charles University, Prague.

Zeng, L., Zhang, Q., Sun, R., Kong, H., Zhang, N., and Ma, H. (2014). Resolution of deep angiosperm phylogeny using conserved nuclear genes and estimates of early divergence times. *Nature Communications* 5, 4956.

Zhou, P., Silverstein, K.A.T., Ramaraj, T., Guhlin, J., Denny, R., Liu, J., Farmer, A.D., Steele, K.P., Stupar, R.M., Miller, J.R., et al. (2017). Exploring structural variation and gene family architecture with De Novo assemblies of 15 *Medicago* genomes. *BMC Genomics* 18.

Zou, X.-H., Zhang, F.-M., Zhang, J.-G., Zang, L.-L., Tang, L., Wang, J., Sang, T., and Ge, S. (2008). Analysis of 142 genes resolves the rapid diversification of the rice genus. *Genome Biology* 9, R49.

INTERNETOVÉ ZDROJE:

1000 Plants [online], [cit. 2017-31-03]. Dostupné z: <http://www.onekp.com/samples/list.php>

Benchmarking seven most popular genome assemblers [online], [cit. 2017-24-03]. Dostupné z: <https://insidedna.me/tutorials/view/benchmark-seven-popular-genome-assemblers>

Fér, T. (2016). HybPhyloMaker – Pipeline for generating phylogenies based on target enriched libraries (Hyb-Seq data) [online], [cit. 2016-09-30]. Dostupné z: <https://github.com/tomas-fer/HybPhyloMaker>

Glenn, T. (2014). 2014 NGS Field Guide: Overview [online], [cit. 2016-22-08]. Dostupné z: <http://www.molecularecologist.com/next-gen-fieldguide-2014/>

Gordon A, Hannon GJ. (2010). Fastx-toolkit. FASTQ/A short-reads pre-processing tools [online], [cit. 2017-24-03]. Dostupné z: http://hannonlab.cshl.edu/fastx_toolkit

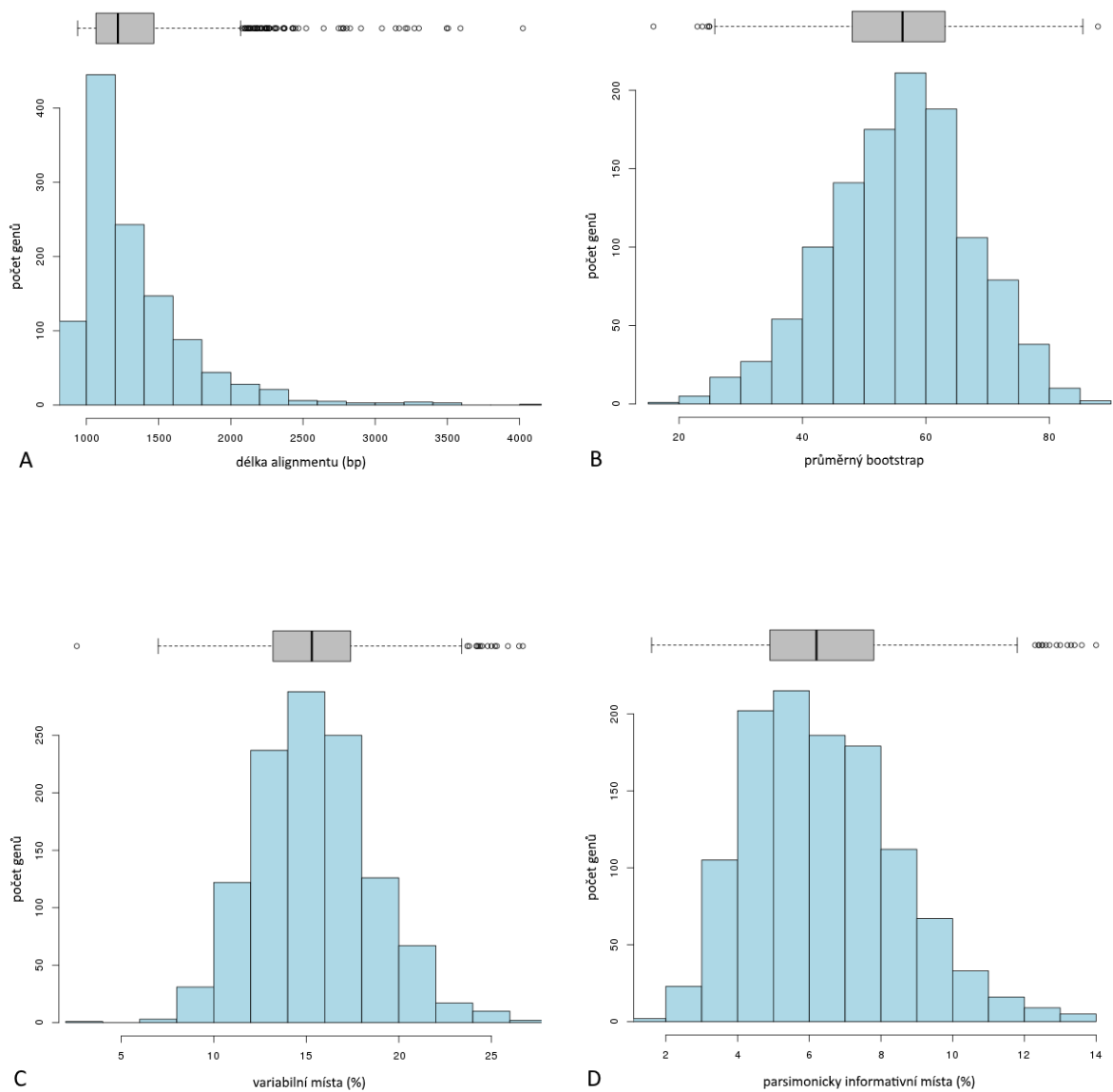
Hodcroft, E. (2016). TreeCollapserCL 4 [online], [cit. 2017-31-03]. Dostupné z: <http://emmahodcroft.com/TreeCollapseCL.html>

Illumina [online], [cit. 2016-3-29]. Dostupné z: http://www.illumina.com/documents/products/techspotlights/techspotlight_sequencing.pdf

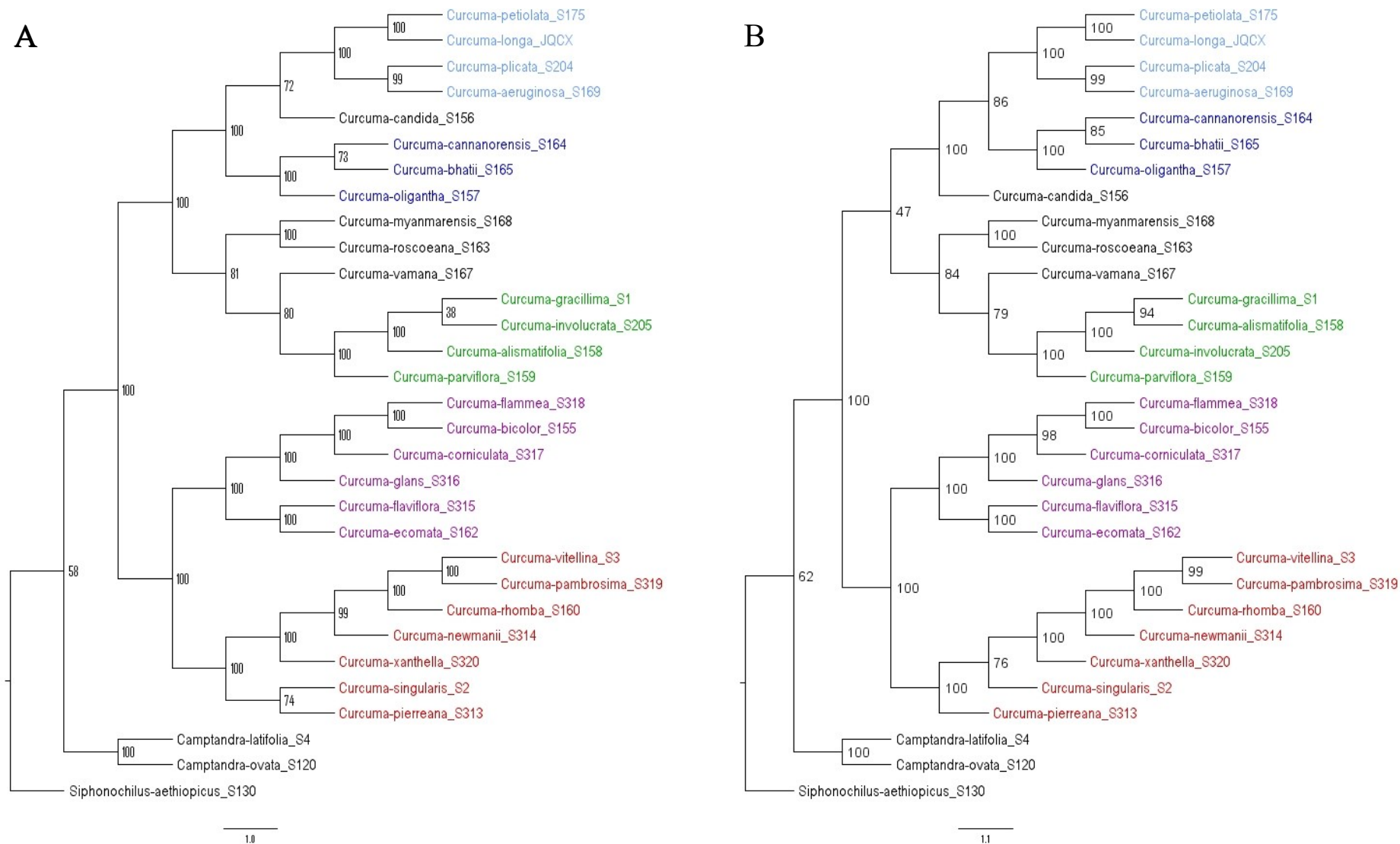
Mirarab, S. (2015). mrpmatrix [online], [cit. 2017-04-03]. Dostupné z: <https://github.com/smirarab/mrpmatrix>

R Core Team (2017). R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria [online], [cit. 2017-01-04]. Dostupné z: <https://www.R-project.org/>

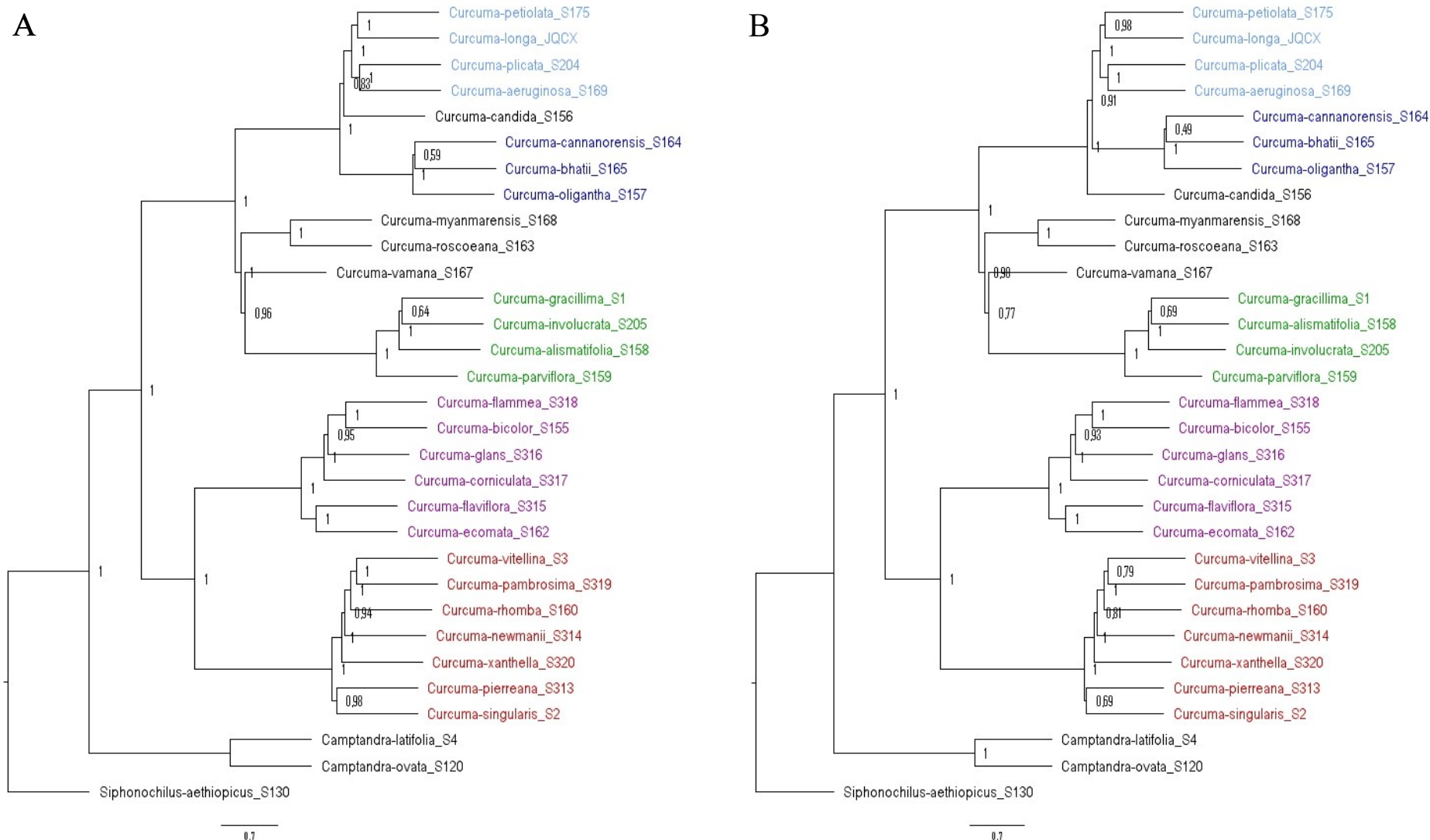
7 PŘÍLOHY



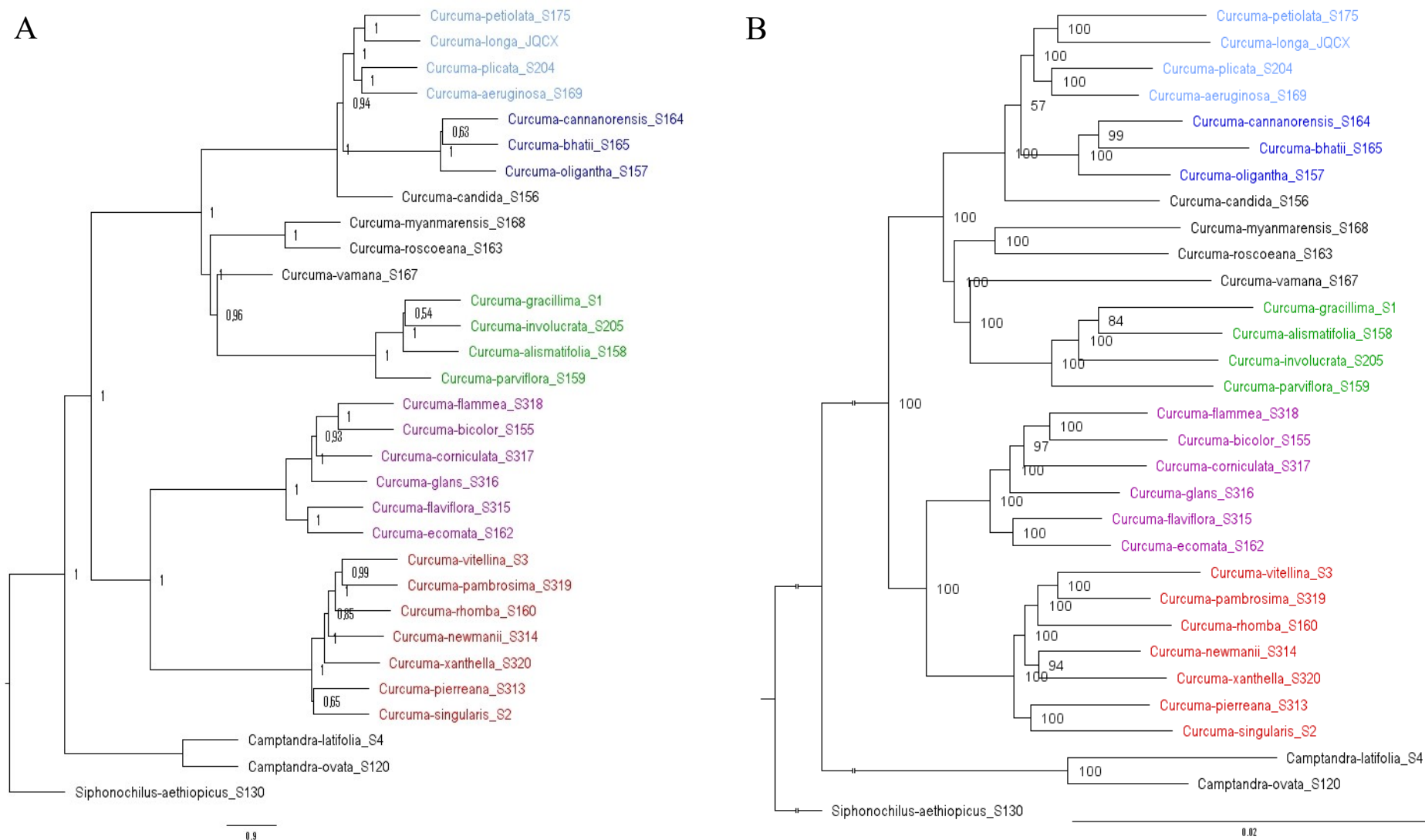
Příloha 1. Histogramy získané pomocí pipeline HybPhyloMaker při analýze 1 154 jaderných low-copy genů 28 druhů rodu *Curcuma* a 3 druhů tvořících outgroup. **A:** délka získaných alignmentů, **B:** průměrná bootstrapová podpora genových stromů vytvořených metodou Maximum Likelihood (RAxML), **C:** zastoupení variabilních pozic a **D:** množství parsimonicky informativních míst u jednotlivých genů.



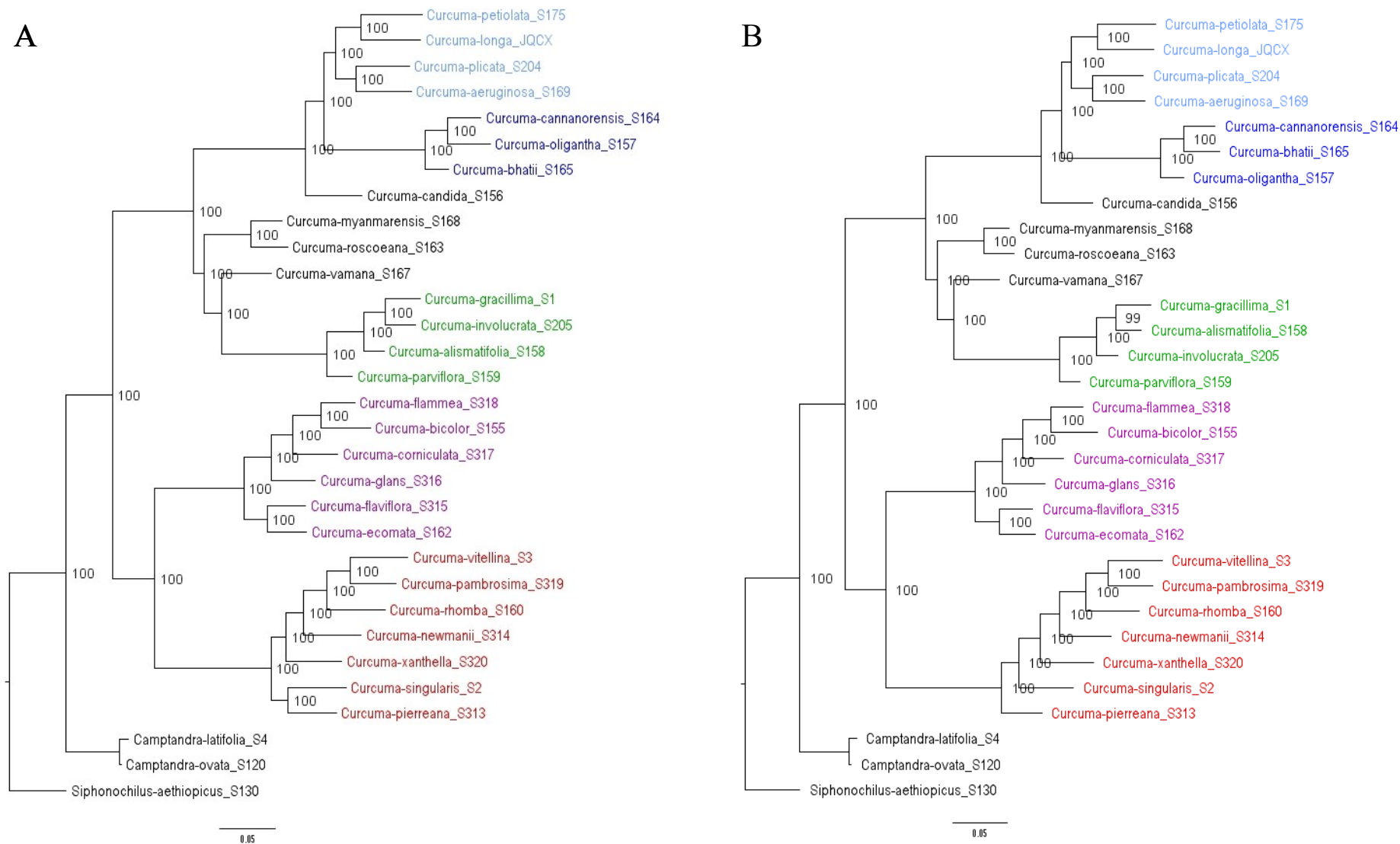
Příloha 2. Stromy druhů vytvořené metodou ASTRID s 1 000 multilokusovými bootstrapovými replikacemi. **A:** Původní species tree založený na 1 154 jaderných low-copy genech **B:** Update species tree založený na 811 jaderných genech. Hlavní linie označeny barevně – *Curcuma I* světle modrou, *Curcuma II* tmavě modrou, *Hitcheniopsis* zeleně, *Ecomata* fialově, *Pierreana* červeně a hybridní druhy černou barvou.



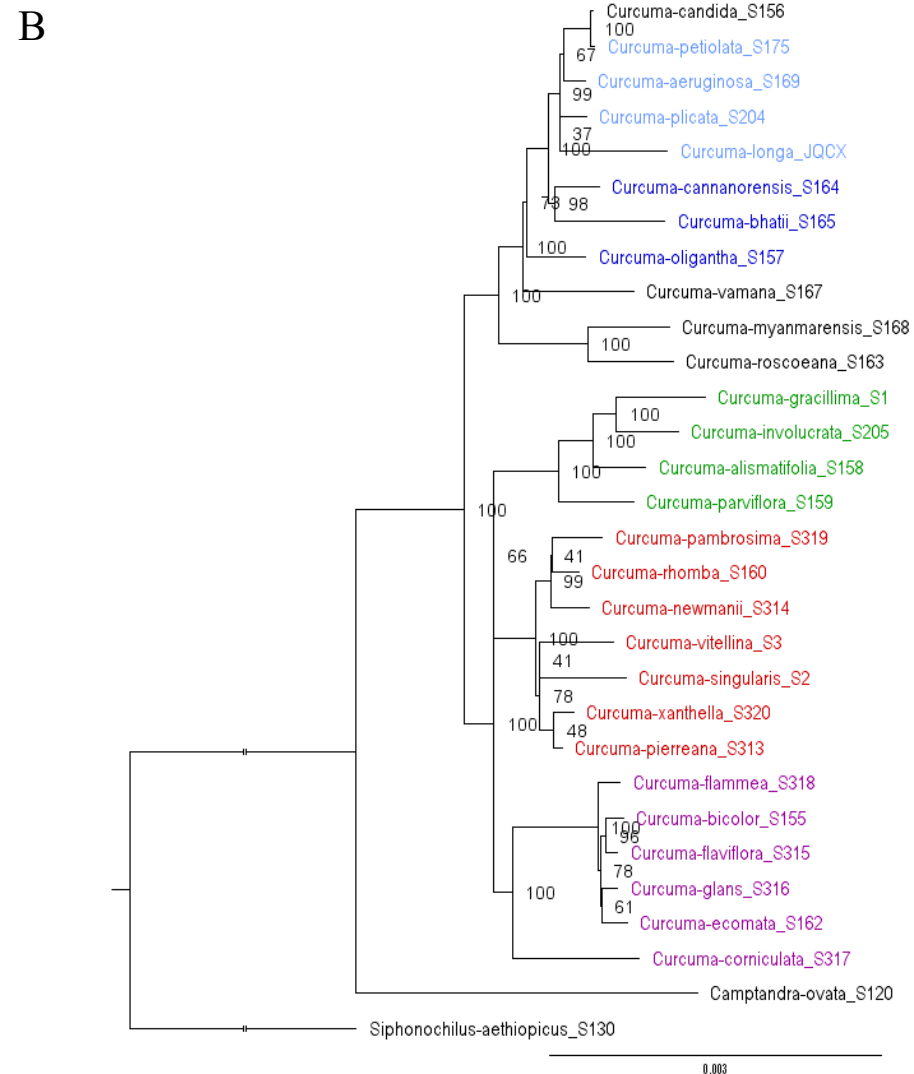
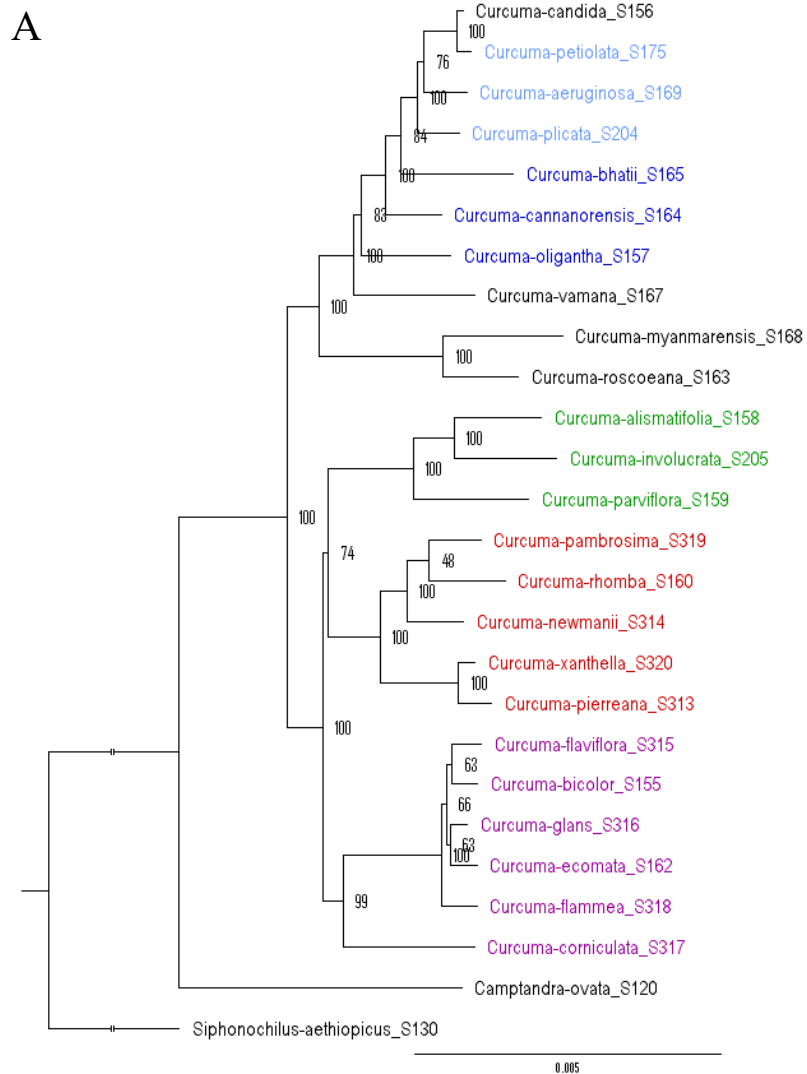
Příloha 3. Stromy druhů vytvořené metodou ASTRAL s hodnotami posterior probability. **A:** Původní species tree založený na 1 154 jaderných low-copy genech **B:** Update species tree založený na 811 jaderných genech. Hlavní linie označeny barevně – *Curcuma I* světle modrou, *Curcuma II* tmavě modrou, *Hitcheniopsis* zeleně, *Ecomata* fialově, *Pierreana* červeně a hybridní druhy černou barvou.



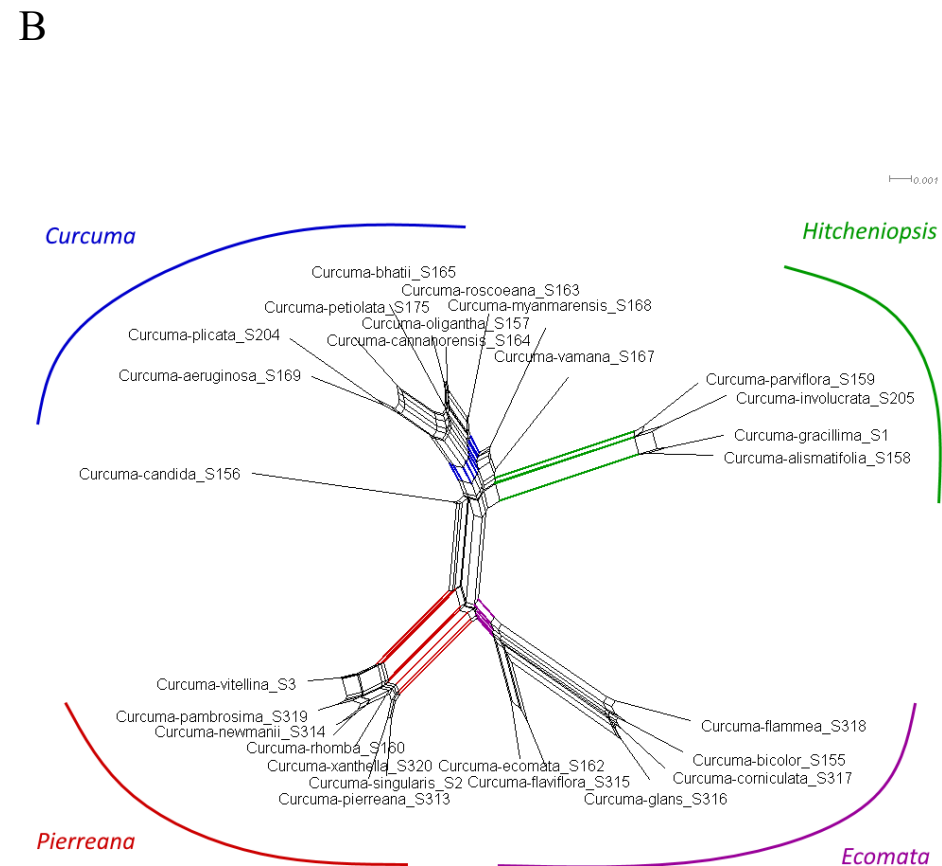
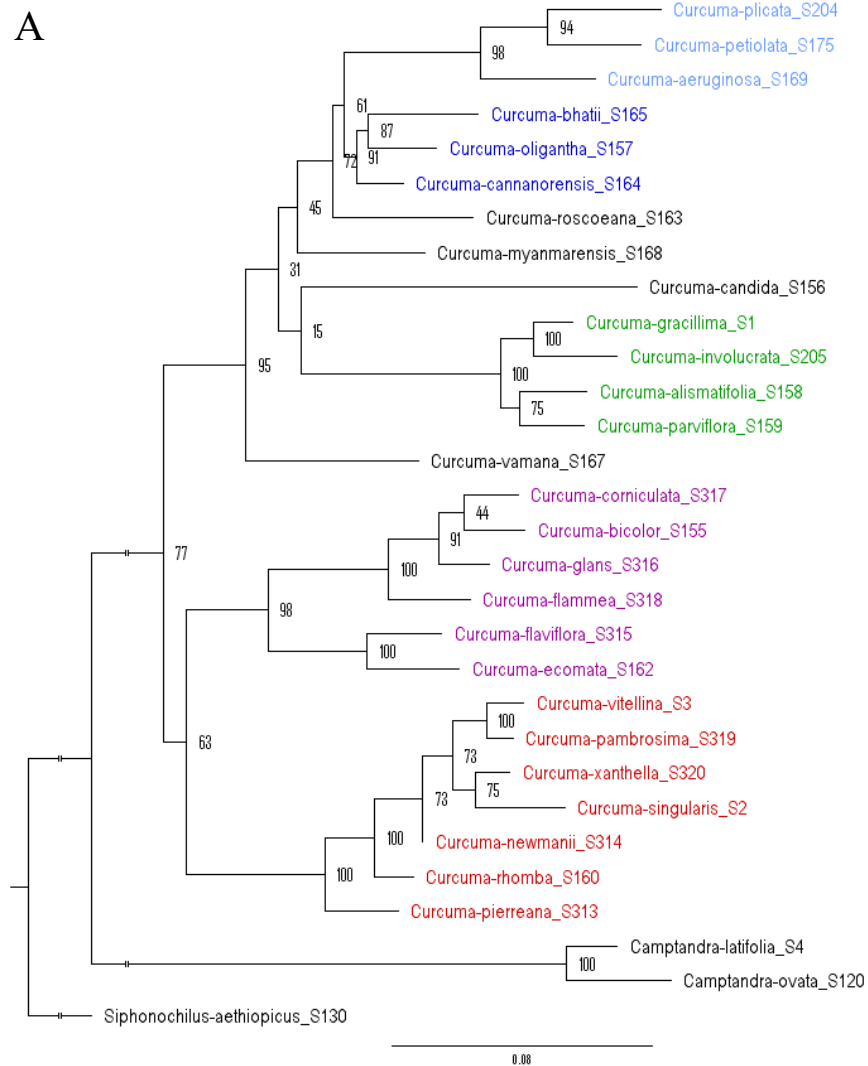
Příloha 4. A: Strom druhů vytvořený metodou ASTRAL s hodnotami posterior probability založený na 1 154 genových stromech se zkolabovanými nody s bootstrapem <70. **B:** Původní species tree založený na 1 154 konkatenovaných jaderných genech vytvořený metodou Maximum Likelihood se 100 bootstrapovými replikacemi. Hlavní linie označeny barevně – *Curcuma I* světle modrou, *Curcuma II* tmavě modrou, *Hitcheniopsis* zeleně, *Ecomata* fialově, *Pierreana* červeně a hybridní druhy černou barvou.



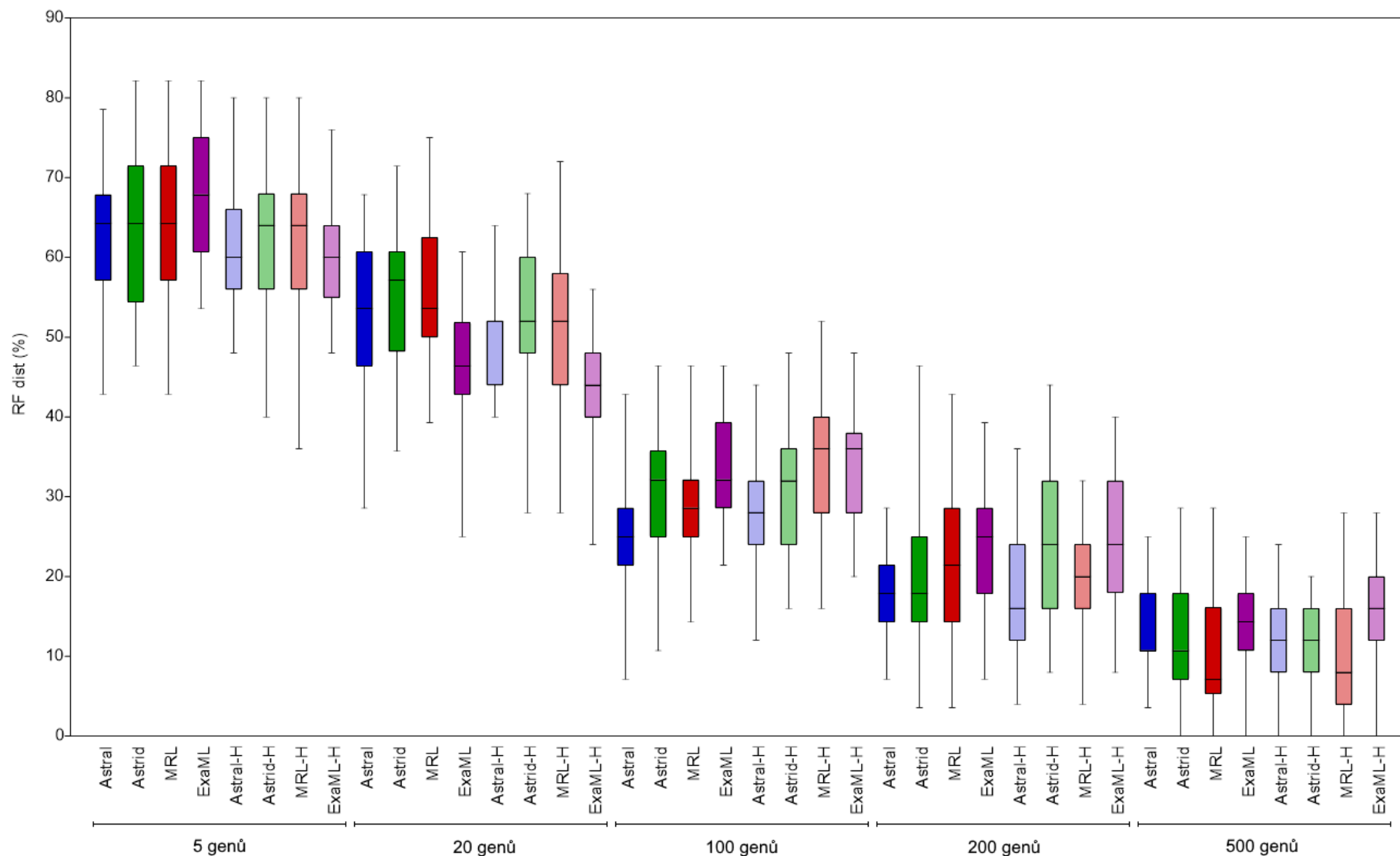
Příloha 5. Stromy druhů vytvořené metodou MRL s 1 000 bootstrapovými replikacemi. **A:** Původní species tree založený na 1 154 jaderných low-copy genech **B:** Update species tree založený na 811 jaderných genech. Hlavní linie označeny barevně – *Curcuma I* světle modrou, *Curcuma II* tmavě modrou, *Hitcheniopsis* zeleně, *Ecomata* fialově, *Pierreana* červeně a hybridní druhy černou barvou.



Příloha 6. A: Fylogenetický strom vytvořený z celého chloroplastového genomu pomocí *de-novo* assembly a metody Maximum Likelihood s 1 000 bootstrapovými replikacemi. **B:** Fylogenetický strom z 86 konkatenovaných chloroplastových genů získaný metodou Maximum Likelihood s 1 000 bootstrapovými replikacemi. Hlavní linie označeny barevně – *Curcuma I* světle modrou, *Curcuma II* tmavě modrou, *Hitcheniopsis* zeleně, *Ecomata* fialově, *Pierreana* červeně a hybridní druhy černou barvou.



Příloha 7. A: Fylogenetický strom z rDNA cistronu získaný mapováním na referenci a metodou Maximum Likelihood s 1 000 bootstrapovými replikacemi. Hlavní linie označeny barevně – *Curcuma I* světle modrou, *Curcuma II* tmavě modrou, *Hitcheniopsis* zeleně, *Ecomata* fialově, *Pierreana* červeně a hybridní druhy černou barvou. **B:** Neighbor-Net analýza stejného datasetu.



Příloha 8. Srovnání RF distancí v závislosti na použité metodě tvorby species tree. RF distance vyjádřené jako procento z maximální RF distance mezi 10 druhovými stromy získanými vždy z 5, 20, 100, 200 a 500 jaderných low-copy genů metodami ASTRAL, ASTRID, MRL a ExaML. „-H“ značí dataset bez přítomnosti meziliniových hybridů *C. myanmarensis*, *C. roscoeana* a *C. vamana*. Jednotlivé metody označeny stejnou barvou, dataset neobsahující hybridní druhy je ve světlejších odstínech.

Příloha 9. Průměrná Robinson-Fouldova vzdálenost mezi 10 stromy vytvořenými z náhodně vybraných 5, 20, 100, 200 a 500 low-copy nukleárních genů a mezi těmito stromy a fylogenetickým stromem založeným na 1 154 genech. RF distance uvedena jako procento z maximální vzdálenosti mezi stromy vypočítané podle vzorce $RF_{max} = 2(n-3)$, kde n je počet taxonů ve stromu (pro A: 31 druhů, pro B: 28 druhů). Metoda = metoda tvorby stromu (EML = ExaML), mezi stromy = průměrná hodnota RF distance mezi jednotlivými stromy, 1000 genů = průměrná hodnota vzdálenosti mezi příslušnými stromy a stromem z přibližně 1000 genů. **A:** srovnání RF distancí stromů obsahujících meziliniové hybridy (*C. roscoeana*, *C. vamana* a *C. myanmarensis*), **B:** bez meziliniových hybridů.

A: s hybridy	5 genů				20 genů				100 genů				200 genů				500 genů			
Metoda	Astral	Astrid	MRL	EML	Astral	Astrid	MRL	EML	Astral	Astrid	MRL	EML	Astral	Astrid	MRL	EML	Astral	Astrid	MRL	EML
Mezi stromy	62,3	63,8	65,2	67,0	53,4	54,6	55,4	47,1	25,2	30,7	29,6	32,3	18,0	20,0	22,1	23,6	12,7	12,1	10,9	14,8
1000 genů	51,8	53,2	54,3	57,1	43,2	43,9	43,2	36,4	17,9	21,4	20,0	24,3	13,9	13,2	14,3	17,5	9,6	8,6	6,4	10,7

B: bez hybr.	5 genů				20 genů				100 genů				200 genů				500 genů			
Metoda	Astral	Astrid	MRL	EML	Astral	Astrid	MRL	EML	Astral	Astrid	MRL	EML	Astral	Astrid	MRL	EML	Astral	Astrid	MRL	EML
Mezi stromy	62,0	60,8	60,8	59,4	49,6	52,2	51,6	43,6	27,6	29,8	34,8	34,0	18,6	23,8	19,8	24,6	11,4	11,2	10,0	15,0
1000 genů	52,8	48,6	50,4	52,0	42,0	42,4	42,4	34,8	20,0	22,4	24,4	25,2	13,6	15,2	12,4	17,6	8,4	7,2	5,6	11,2

Příloha 10. Celková tabulka výsledných hodnot lineage movement analýzy 28 druhů kurkum založené na 970 stomech z low-copy nukleárních genů. Druhy rozděleny do skupin podle příslušnosti k molekulární linii - *Curcuma* ještě rozdělena na *Curcuma I* a *II* podle skupin velikosti genomu (podle Leong-Škorničková et al., 2007) , poslední skupinu tvoří hybridní druhy. Hodnoty jsou uvedeny v procentech a zaokrouhleny.

Druh	<i>Curcuma I</i>	<i>Curcuma II</i>	<i>Curcuma</i>	<i>Hitchen.</i>	<i>Hybridi</i>	<i>Ecomata</i>	<i>Pierreana</i>
<i>Curcuma I</i>							
<i>C. aeruginosa</i>	31,1	13,9	45,1	0,3	2,1	0,2	0,2
<i>C. longa</i>	38,8	14,8	53,6	0,6	2,9	0,0	0,2
<i>C. petiolata</i>	42,8	13,9	56,7	1,1	2,1	0,3	0,1
<i>C. plicata</i>	36,3	13,9	50,2	0,4	1,8	0,2	0,2
<i>Curcuma II</i>							
<i>C. bhatii</i>	7,3	54,6	62,0	0,2	1,5	0,1	0,1
<i>C. kannanorensis</i>	7,2	59,0	66,2	0,5	1,4	0,1	0,1
<i>C. oligantha</i>	6,2	55,2	61,3	0,4	1,9	0,1	0,1
<i>C. candida</i>	35,2	4,3	39,5	0,6	4,4	0,0	0,2
<i>Hitcheniopsis</i>							
<i>C. alismatifolia</i>	0,1	0,6	0,7	61,6	1,4	0,3	0,0
<i>C. gracillima</i>	0,5	0,1	0,6	64,8	1,0	0,0	0,6
<i>C. involucrata</i>	1,1	0,2	1,3	60,4	1,2	0,2	0,0
<i>C. parviflora</i>	0,7	0,8	1,5	39,8	2,9	0,5	0,3
<i>Hybridi</i>							
<i>C. myanmarensis</i>	2,3	2,2	4,4	1,2	50,7	0,2	0,2
<i>C. roscoeana</i>	3,1	3,2	6,3	1,9	51,5	0,2	0,5
<i>C. vamana</i>	3,1	3,7	6,8	3,3	13,7	0,4	0,2
<i>Ecomata</i>							
<i>C. bicolor</i>	0,1	0,1	0,2	0,3	0,2	36,2	1,4
<i>C. corniculata</i>	0,2	0,0	0,2	0,2	0,1	32,0	0,4
<i>C. ecomata</i>	0,2	0,0	0,2	0,1	0,0	47,9	1,2
<i>C. flammea</i>	0,0	0,1	0,1	0,0	0,2	65,1	0,6
<i>C. flaviflora</i>	0,2	0,1	0,3	0,3	0,5	45,5	1,0
<i>C. glans</i>	0,0	0,1	0,1	0,1	0,0	41,3	1,5
<i>Pierreana</i>							
<i>C. newmanii</i>	0,2	0,0	0,2	0,0	0,0	0,8	56,2
<i>C. pambrosima</i>	0,0	0,0	0,0	0,0	0,0	1,0	60,8
<i>C. pierreana</i>	0,0	0,1	0,1	0,0	0,4	0,5	46,9
<i>C. rhomba</i>	0,1	0,0	0,1	0,3	0,1	0,4	61,6
<i>C. singularis</i>	0,2	0,2	0,4	0,2	0,2	1,2	51,4
<i>C. vitellina</i>	0,2	0,0	0,2	0,4	0,2	0,9	59,7
<i>C. xanthella</i>	0,0	0,2	0,2	0,0	0,0	0,7	52,3

Příloha 11. Počty genů obsahující jednotlivé monofyletické skupiny získané při testech monofylie 28 druhů rodu *Curcuma* L. založených na 1 103 - 1 154 low-copy nukleárních genech (vyřazeny z analýzy ty geny, které neobsahovaly testovaný druh) se zkolabovanými větvemi s bootstrapovou podporou <70. C/H/E/P - počet genů, kde je monofyletická skupina *Curcuma/Hitcheniopsis/Ecomata/Pierreana* bez testovaného druhu. Chyb/Hhyb/Ehyb/Phyb - počet genů, kde je monofyletická skupina *Curcuma/Hitcheniopsis/Ecomata/Pierreana* včetně testovaného druhu. CHhyb/EPhyb - monofyletičnost skupiny *Curcuma* a *Hitcheniopsis/Ecomata* a *Pierreana*. C1 - monofyletičnost skupiny *Curcuma* I bez testovaného druhu, C1hyb - monofyletičnost skupiny *Curcuma* I včetně testovaného druhu, C2hyb - monofyletičnost skupiny *Curcuma* II a testovaného druhu, báze C - testovaný druh stojící na bazální pozici skupiny *Curcuma* a bázeCH - testovaný druh stojící na bazální pozici skupin *Curcuma* a *Hitcheniopsis* po jejich spojení.

druh	C1	C1hyb	C2hyb	C	Chyb	bázeC	H	Hhyb	bázeCH	CHhyb	E	Ehyb	P	Phyb	EPhyb
<i>Curcuma I</i>															
<i>C. aeruginosa</i>	7	29	2	9	565	7	818	0	1	566	541	0	656	0	0
<i>C. longa</i>	13	29	6	24	565	17	818	1	2	566	541	0	656	1	3
<i>C. petiolata</i>	6	28	4	19	552	10	796	1	3	547	529	0	642	1	0
<i>C. plicata</i>	5	29	2	12	565	7	818	0	2	565	541	0	655	4	1
<i>Curcuma II</i>															
<i>C. bhatii</i>	29	2	19	4	563	3	813	0	0	562	539	0	652	0	0
<i>C. kannanorensis</i>	29	2	19	9	565	7	817	0	0	565	541	0	656	0	0
<i>C. oligantha</i>	29	2	16	11	564	7	816	1	1	565	539	0	654	0	0
<i>Hitcheniopsis</i>															
<i>C. alismatifolia</i>	29	0	0	559	0	0	70	808	5	560	538	0	651	0	1
<i>C. gracillima</i>	26	0	0	545	0	0	62	790	5	546	523	0	634	0	0
<i>C. involucrata</i>	29	0	0	556	0	0	74	811	7	561	534	0	651	0	0
<i>C. parviflora</i>	29	0	0	559	0	0	266	804	14	557	535	1	647	0	0
<i>Hybridi</i>															
<i>C. candida</i>	20	29	10	55	564	41	817	3	5	566	541	0	656	2	0
<i>C. myanmarensis</i>	27	1	0	461	108	66	758	138	148	507	541	2	656	0	4
<i>C. roscoeana</i>	25	2	0	448	112	60	755	138	156	519	545	0	650	0	3
<i>C. vamana</i>	28	1	4	427	135	58	737	199	131	540	538	0	647	0	3

Příloha 11. Pokračování.

druh	C1	C1hyb	C2hyb	C	Chyb	bázeC	H	Hhyb	bázeCH	CHhyb	E	Ehyb	P	Phyb	EPhyb
<i>Ecomata</i>															
<i>C. bicolor</i>	29	0	0	559	0	0	810	0	0	1	8	535	650	3	404
<i>C. corniculata</i>	29	0	0	563	0	0	816	1	0	0	39	540	656	2	407
<i>C. ecomata</i>	29	0	0	564	0	0	816	0	0	2	31	540	655	5	407
<i>C. flammea</i>	29	0	0	564	0	0	814	0	0	1	15	538	655	1	406
<i>C. flaviflora</i>	29	0	0	565	0	0	817	0	0	0	48	541	656	3	408
<i>C. glans</i>	28	0	0	564	0	0	816	0	0	0	14	541	655	2	407
<i>Pierreana</i>															
<i>C. newmanii</i>	29	0	0	565	0	0	818	0	0	0	540	1	10	655	407
<i>C. pambrosima</i>	29	0	0	564	0	0	817	0	0	0	540	0	4	654	407
<i>C. pierreana</i>	29	0	0	563	0	0	817	0	0	0	538	1	41	654	406
<i>C. rhomba</i>	29	0	0	563	0	0	816	0	0	0	541	0	14	655	407
<i>C. singularis</i>	29	0	0	553	0	0	806	0	1	1	531	2	33	644	403
<i>C. vitellina</i>	29	0	0	548	0	0	787	0	0	0	519	0	12	629	389
<i>C. xanthella</i>	29	0	0	564	0	0	815	0	0	1	539	0	16	655	405