Molecular Phylogenetics and Evolution 119 (2018) 13–24

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Transcriptome mining for phylogenetic markers in a recently radiated genus of tropical plants (*Renealmia* L.f., Zingiberaceae)

Eugenio Valderrama[a,b,c,*], James E. Richardson[a,d], Catherine A. Kidner[a,e], Santiago Madriñán[f], Graham N. Stone[a,b]

[a] Royal Botanic Garden Edinburgh, Edinburgh, Scotland, United Kingdom
[b] Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, Scotland, United Kingdom
[c] Laboratorio de Análisis del Comportamiento Humano, Facultad de Psicología y Grupo de Investigación de Biología (GRIB), Universidad El Bosque, Bogotá, Colombia
[d] Programa de Biología, Universidad del Rosario, Bogotá, Colombia
[e] Institute of Molecular Plant Sciences, University of Edinburgh, Edinburgh, Scotland, United Kingdom
[f] Laboratorio de Botánica y Sistemática, Universidad de Los Andes, Bogotá, Colombia

ARTICLE INFO

Keywords:
Phylogenetics
Transcriptomes
Markers
Diversification
Zingiberaceae
Neotropical

ABSTRACT

The reconstruction of relationships within species-rich groups that have recently evolved in biodiversity hotspots is hampered by a lack of phylogenetically informative markers. It is also made difficult by the lack of sampling necessary to reconstruct a species-level phylogeny. We use transcriptome mining to search for markers to reconstruct a phylogeny of the amphi-Atlantic genus *Renealmia* L. f. (Zingiberaceae). We recover seven introns from single copy genes and use them to reconstruct the phylogeny of the genus together with a commonly used phylogenetic marker, internal transcribed spacers of ribosomal DNA (ITS) that has previously been used to reconstruct the phylogeny of the genus. We targeted genes with low numbers of base pairs that improves sequencing success using highly degraded DNA from herbarium specimens. The use of herbarium specimens greatly increased the number of species in the study as these were readily available in historical collections. Data were obtained for 14 of the 17 African species and 54 of the 65 Neotropical species. The phylogeny was well-supported for a number of *Renealmia* subgroups although relationships among those clades remained poorly supported.

1. Introduction

Recent diversification of the species-rich Neotropical flora presents a challenge for studies attempting to elucidate evolutionary history at the species level. Widely used phylogenetic markers struggle to recover species level relationships within recent and species-rich lineages (Hughes and Eastwood, 2006; Nicholls et al., 2015; e.g. Richardson et al., 2001; Särkinen et al., 2007) a scenario that could be expected when different and non-mutually exclusive processes affect radiations. Recovering phylogenies when lineages are recent and so lack lineage specific mutations is challenging, also when influenced by incomplete lineage sorting because they are recent and/or comprise large population sizes and finally when introgression through hybridization has occurred within them (Maddison, 1997; Maddison and Knowles, 2006; Pamilo and Nei, 1988). Although this pattern points to the importance of these processes, it is necessary to improve our understanding of the evolutionary history of these lineages in order to elucidate the reasons for such diversifications and gain insights into the mechanisms that yielded the outstanding diversity of the Neotropics.

The traditional approach over the last two decades to obtain species-level phylogenies in plants has been to amplify a limited set of non-coding cpDNA loci (Shaw et al., 2007) and the nrDNA internal transcribed spacer (ITS) region (e.g. Armstrong et al., 2014; Álvarez and Wendel, 2003). These are expected to be variable at low taxonomic levels and are easy to sequence using universal PCR primers. This approach has shed light on many relationships among genera and even at the species level, but has also led to poorly resolved phylogenies that are not very informative for studies beyond classification purposes, even in large and morphologically diverse genera (Hughes et al., 2006). In recent radiations, incomplete lineage sorting and low differentiation among species are expected to cause incongruence between the phylogenetic signals of sampled loci, necessitating approaches that incorporate coalescent variation. During the last decade, several methods have been developed that accommodate incongruence between gene

* Corresponding author at: Facultad de Psicología y Programa de Biología, Universidad El Bosque, Bogotá, Colombia.
E-mail addresses: e.valderrama.e@gmail.com (E. Valderrama), jamese.richardson@urosario.edu.co (J.E. Richardson), c.kidner@rbge.ac.uk (C.A. Kidner), samadrin@uniandes.edu.co (S. Madriñán), graham.stone@ed.ac.uk (G.N. Stone).

http://dx.doi.org/10.1016/j.ympev.2017.10.001
Received 27 June 2016; Received in revised form 17 July 2017; Accepted 3 October 2017
Available online 04 October 2017
1055-7903/

trees and the underlying species tree (Heled and Drummond, 2010; Liu, 2008; Mirarab et al., 2014; Shaw et al., 2013). However, to estimate the species trees it is necessary to sample multiple low-copy number, phylogenetically-informative, orthologous and unlinked loci across the genome of the study lineage, a challenging task in some plant groups where traditional markers are not informative enough (Hughes et al., 2006).

Recent developments in DNA sequencing technology (Metzker, 2009) have facilitated the generation of genome-wide phylogenetic datasets, allowing resolution of the evolutionary history of challenging study systems (Brassac and Blattner, 2015; e.g. Nicholls et al., 2015; Yang et al., 2015). An alternative approach to genome-wide sampling of lineages with large genomes that avoids the costs of either generating high enough coverage to reliably sequence homologous regions in multiple accessions is to sequence transcribed sequences of the genome (the transcriptome) with RNA extracted from living specimens (Wang et al., 2009). Transcriptomes represent a substantial reduction of the genome that can be sequenced with a smaller budget. The development of specific software for assembling transcripts (Haas et al., 2013) and the growing number of lineages for which transcriptomes are becoming available in public repositories (Droc et al., 2013; Goodstein et al., 2012) make *de-novo* transcriptomes easier to assemble and compare. Transcriptomes can be used as a genome-wide sample to target useful markers to estimate phylogenies (Chamala et al., 2015; Rothfels et al., 2013) and could be used for selecting markers for targeted enrichment (e.g. Nicholls et al., 2015).

Besides using appropriate markers, a common problem in phylogenetic studies of widespread groups is that some species are only available for sampling as herbarium specimens. Sampling widespread species from throughout their distributions or local endemics in the field to acquire high quality tissue samples would require considerable research resources, intricate logistics and long-term studies to estimate species level phylogenies for species-rich groups. The issue is exacerbated in the Neotropics by the widespread habit of using alcohol to preserve specimens in difficult environments until drying facilities (*e.g.* the local herbarium) are reached by the botanists, reducing the already diminished quality of the DNA that can be extracted from herbarium specimens (Särkinen et al., 2012). However, the likelihood of successful PCR amplifications of DNA extracted from herbarium specimens can be increased by modifying the standard DNA extraction protocols and targeting short amplicons (Särkinen et al., 2012). Field collection of RNA samples for transcriptomic approaches is more challenging, requiring more complex preservatives (such as RNA-later) and cryostorage – and RNA preservation is usually very poor in herbarium material.

In this study we present an alternative method to generate species level phylogenies of recent radiations of plants with big genomes and for which several species are only available as herbarium specimens. We generated three transcriptomes and compared them to available public repositories to target short introns within potentially conserved, orthologous and low copy-number loci. Those loci were amplified with PCR's that can be used on low quality DNA extractions from herbarium specimens to obtain markers to cover as much species as possible of the genus *Renealmia* L.f. (Zingiberaceae) to estimate a species-level phylogeny. Finally, we use the resulting phylogenies to ask whether groups and subgroups defined by morphological characters (see Supplementary Notes 1) are supported by molecular data, and whether some widespread species are resolved as monophyletic clades.

2. Material and methods

2.1. de novo transcriptomes

2.1.1. Tissue

Three transcriptomes were generated from individuals held in the living collection of the Royal Botanic Garden Edinburgh under the accession numbers 20080229A, 19740104 and 19750180 corresponding to the Neotropical species *R. alpinia* (Rottb.) Maas, the African species *R. battenbergiana* H.A. Cummins ex Baker and a species of *Alpinia* Roxb., respectively (Supplementary Table 1).

2.1.2. RNA extraction

Young buds were sampled so that a variety of tissues including young leaves, rhizomes and meristems could be incorporated to harvest mRNAs for as many as possible of the expressed proteins in the vegetative parts. The buds were frozen immediately after collection in liquid nitrogen and kept frozen during the grinding process. PureLink Plant RNA Reagent (Life Technologies) was used following the manufacturer's protocol for small scale RNA isolation with two adjustments. 500 µL of acid phenol:chloroform (5:1, pH 4.7; SIGMA-ALDRICH) was added to the aqueous phase obtained after the protocol's first chloroform addition, then mixed and centrifuged at 12,000*g* for ten minutes at 4 °C. The aqueous phase to which isopropyl alcohol was added was left for at least an hour at −20 °C to improve precipitation. Integrity of the extracted RNA was assessed on 1% agarose gels stained with ethidium bromide. RNA concentrations were measured using a Qubit 2.0 Fluorometer (Life Technologies) after applying the RQ1 DNase (Promega) treatment; both procedures followed manufacturer's protocols. Pooling of multiple extractions (up to four) was necessary to increase the total quantity of RNA per species.

2.1.3. de novo transcriptomes

TruSeq RNA Libraries (Illumina) were prepared and sequenced in a MiSeq v2 (Illumina) lane (12×10^6 base pairs [bp] in 150 bp paired-end-reads) by Edinburgh Genomics at the University of Edinburgh. *Trimmomatic* 0.3 (Bolger et al., 2014) was used to remove adapters and bases with poor quality using a sliding window trimming approach with windows of 4 bp and a threshold value for the average Phred score of 15. Phred scores are logarithmically related to the base-calling error probability, a Phred score of 15 represents a probability of incorrect base call of 0.0316 (Ewing and Green, 1998) and represents a coarse filter to avoid losing too much information at this step in the assemblies. Bases with Phred scores below three at both ends of the reads and reads shorter than 36 bp were removed in *Trimmomatic*. *FastQC* (Andrews et al., 2015) was used to check quality before and after trimming. The remaining reads after trimming were assembled in *Trinity* with default settings (Haas et al., 2013).

2.1.4. Choosing markers for phylogenetic analyses

To choose appropriate genes in the transcripts for phylogenetic analysis it is necessary to distinguish which genes evolved from a common ancestral gene by speciation (orthologous genes) from those that are related by duplication within a genome (paralogous genes). To estimate which genes in the transcripts of the three species are potentially orthologous we generated a list of the Reciprocal Best Hits (RBH) using *BLAST* (Altschul et al., 1990). We assumed that the transcripts of orthologous genes present in two different genomes will identify each other as reciprocal best hits due to shared common ancestry (Tatusov et al., 1997; Ward and Moreno-Hagelsieb, 2014).

We checked if the potential orthologous genes found in the three species share the same best *BLAST* hit in the transcriptomes of *Arabidopsis thaliana* (TAIR10; Lamesch et al., 2012) and *Musa acuminata* (Droc et al., 2013). We used the unique locus identifier of *A. thaliana* genes that were best *BLAST* hits to identify the potential orthologs among the three transcriptomes included in the COSII list. The COSII lists 2869 genes selected within the coding sequences (CDS) of *A. thaliana* and representatives of the euasterids that, according to phylogenies and reciprocal *BLAST* methods, are potentially orthologs and single-copy genes (F. Wu et al., 2006). We also used the *A. thaliana* locus identifiers to visually inspect the gene model of each potential orthologous gene in all the available data in *Phytozome* 9.1 for model angiosperms (Goodstein et al., 2012). Loci that included introns of

100–800 bp, showed low copy number, and showed a conserved gene model across available angiosperms were selected. We confirmed the gene model in the annotated genome of the closest relative available, *i.e.* The Banana Genome Hub (Droc et al., 2013), for the *M. acuminata* best *BLAST* hit of each potential ortholog and aligned it to the transcripts of the *Renealmia* and *Alpinia* species we generated using *Geneious* 4.8.5 (Biomatters Development Team, 2010). We selected short introns to include non-coding regions expecting them to be more variable and less prone to selection and recombination. The short length of the markers seeks to improve the PCR success rate given that it is inversely proportional to the amplicon size, especially because DNA extracted from herbarium specimens is expected to be degraded and fragmented (Särkinen et al., 2012).

Sixteen introns flanked by CDS (183–893bp long) were selected and *Primer3* (Untergasser et al., 2012) was used to design primers for polymerase chain reactions (PCR). DNA was extracted from the individuals used to generate the transcriptomes using the Qiagen Plant DNeasy kit following the manufacturer's instructions. PCR reactions using the 16 sets of primers were carried out in 20 μL volume reactions containing 2 μL 2 mM dNTPs, 2 μL 10× NH_4 reaction buffer, 0.6 μL 50 mM $MgCl_2$, 1 μL 10 μM of each primer, 4 μL 5× combinatorial enhancer solution with 2.7 M betaine, 6.7% DMSO and 55 μg/ml BSA (Ralser et al., 2006), 0.2 μL of 5 u/μL Biotaq DNA polymerase buffer, 1 μL of genomic DNA and 8.2 μL of water. The thermal cycling consisted of 3 min denaturation at 94 °C, followed by 34 cycles of 1 min at 94 °C for denaturation, 30 s for annealing and 72 °C for 1 min and 30 s for extension with a final extension period of 5 min at 72 °C on a Tetrad2 BioRad DNA Engine. Annealing temperatures for each set of primers were estimated by subtracting 5 °C from the minimum melting temperature (estimated in *Primer3*) of each pair. The PCR products were run on 1% agarose gels and we discarded the sets of primers that yielded non-specific products (*i.e.* produced multiple bands). PCR products were purified using ExoSAP IT (GE Healthcare) and sequenced by Macrogen Europe in both directions using Big Dye Terminator chemistry on an ABI 3730XL sequencer. Forward and reverse chromatograms were assembled and visually inspected in *Geneious* Pro 4.8.5 and resulting contigs were aligned with the transcripts used to design the primers to confirm that the target region was amplified. We selected eight introns flanked by coding sequences (191–924 bp) for which the transcripts and contigs matched and for which there were at least three variable sites among the three species.

2.2. Obtaining markers for Renealmia

Sixty-eight (14 Africa, 54 Neotropics) of *c.* 82 (17 Africa, 65 Neotropics) *Renealmia* species were sampled, covering the morphological diversity and geographic range of the genus. Outgroups were *Aframomum limbatum* (Oliv. & D. Hanb) K.Schum., *Aframomum verrucosum* Lock, *Alpinia abundiflora* Burtt & R.M.Sm. (two individuals sequenced), *Alpinia fax* B.L.Burtt & R.M.Sm. and *Amomum glabrum* S.Q.Tong. The first four are the closest relatives to *Renealmia*, while the latter is a more distantly related lineage within subfamily Alpinioideae (Auvray et al., 2010; Kress et al., 2005, 2007). The available material included high quality leaf samples collected in the field or in living collections and dried with silica gel and fragments of herbarium specimens encompassing a broad range of qualities for DNA extractions. The Qiagen Plant DNeasy kit was used to extract DNA from the high quality leaf samples and CTAB (Doyle, 1990) for the samples from herbarium specimens to increase the DNA yield (Särkinen et al., 2012). A protocol combining CTAB and Qiagen columns was used following Särkinen et al. (2012) to obtain a high yield and purity in a subset of problematic extractions but didn't result in any major improvement. The same PCR constituents and thermal cycling described above were used with adjustments in the extension times according to the efficiency of the DNA polymerase of *c.* 60 s per 1000 bp. PCR products were checked on 1% agarose gels. In addition to the eight introns, ITS 1 and ITS 2 were amplified in separate reactions with primers designed according to the available sequences in *GenBank* (Särkinen et al., 2007).

When amplifications for some samples were unsuccessful, reactions were performed using nested primers designed in *Primer3* and the PCR products of previous attempts as template instead of genomic DNA. As nested PCRs are more prone to contamination and PCR artefacts, some individuals were amplified and sequenced more than once and the negative controls of the previous attempts were also used as templates. Additionally, Platinum Taq DNA Polymerase (Life Technologies) was used with the same protocols for problematic accessions. Purification, sequencing and chromatogram assembly was performed as described above. Apparent heterozygotes were scarce and coded as ambiguities using IUPAC codes. Alignments of each locus were performed using *MUSCLE* (Edgar, 2004) and inspected visually in *Geneious Pro 4.8.5*. GenBank numbers for all sequences used are detailed in Supplementary Table 1.

2.3. Phylogenetic analyses

DNA substitution models for each locus were estimated using *jModelTest 2*.1.6 (Darriba et al., 2012) and with *PartitionFinder* v1.1.1 (Lanfear et al., 2012), using Bayesian information criterion (BIC) and maximum likelihood trees to select the best models. *PartitionFinder* was also used to determine the best partition schemes for each locus and for the concatenated matrix. For introns we considered the two flanking CDS and the intron sequence as three candidate partitions and five candidate sections for the ITS region, the 18S ribosomal RNA (rRNA), ITS1, 5.8S rRNA, ITS2 and 26S rRNA.

Because schemes with separate partitions for the two flanking CDS of each intron and separate partitions within the rRNA and introns of the ITS region were not selected, just two candidate partitions per locus (CDS or rRNA and intron sequences) were included in the analysis to determine the best scheme for the concatenated matrix.

Maximum likelihood trees for the nine loci and the concatenated matrix were estimated with *GARLI* 2.01 (Zwickl, 2006) using the substitution models selected for the best partition schemes, with 100 bootstrap replicates to obtain support values. Support values from the bootstrap analysis were estimated for the maximum likelihood tree using *SumTrees* 4.0.0 in *Dendropy* 4.0.2 (Sukumaran and Holder, 2010). Bayesian inference analyses were performed for the concatenated matrix in *MrBayes* 3.2.6 (Ronquist et al., 2012) using the best partition schemes and respective substitution models. Each analysis comprised two independent runs, each with one cold chain and three heated chains with incremental temperature of 0.10, run for 150 million generations, sampling every 1000 generations. When models included more than one substitution type, we allowed *MrBayes* to explore different substitution schemes using reversible-jump sampling. To confirm convergence of independent runs and appropriate mixing we used statistics from *MrBayes* and inspected parameters in *Tracer* v 1.6.1 (Rambaut et al., 2014). At least 25% of the first samples of each run were discarded as burn-in. Whenever possible analyses were run in the *CIPRES* portal (Miller et al., 2011). Only the *Alpinia* sp. accession was defined as outgroup in all the analyses to avoid assuming *Renealmia* spp. monophyly.

We estimated species trees to incorporate incongruence among gene trees using methods statistically consistent under a multi-species coalescent model. *ASTRAL*-II (Mirarab and Warnow, 2015), *STAR* (Liu et al., 2009), *MP_EST* (Liu et al., 2010) and *NJst* (Liu and Yu, 2011) using the maximum likelihood gene trees and their bootstrap replicates was used to estimate a species tree with 100 replicates of multi-locus bootstrapping (Seo, 2008). The Species Tree Analysis Web server (STRAW; T. I. Shaw et al., 2013) performed the analyses of the latter three methods. Additionally, we simultaneously estimated the species and gene trees with ∗*BEAST* (Heled and Drummond, 2010) as implemented in *BEAST* 2.1.3 (Bouckaert et al., 2014). The substitution model for each gene was specified assuming a lognormal relaxed clock

prior for each gene, a Yule process for the species tree prior and a constant population size model. Two independent runs of 300 million generations each were performed sampling every 10000 steps and discarding the first 25% as burn-in. Results of separate runs were combined using *LogCombiner* 2.3.0 (Bouckaert et al., 2014) and *Tracer* 1.6.1 was used to check effective sample sizes (ESS) of parameter estimates, convergence and mixing of runs. *TreeAnnotator* v2.3.0 (Rambaut and Drummond, 2002) was used to obtain the 95% highest posterior density of estimates of node ages, and used *FigTree* v1.4.2 (Rambaut, 2014) to examine trees. Finally *BEST* 2.3 (Liu, 2008) was used, which simultaneously estimates gene and species trees, specifying the substitution model for each gene and adjusting the Θ prior to the average across loci of the number of variable sites per base pair within species (Thetapr = invgamma [3,0.008] as the average was three variable sites per pair per thousand sites within species) and allowing substitution rates across loci to show a three-fold variation among them. Six independent runs were performed without metropolis coupling (following Castillo-Ramírez et al., 2011) for 20 million generations, sampling each 1000 steps and discarding 5000 samples as burn-in. We used *DensiTree* (Bouckaert, 2010) to jointly visualize all topologies included in the posterior sample.

3. Results

3.1. de novo transcriptomes and phylogenetic markers

Phred scores for all bases in the reads corresponding to the three transcriptomes after quality control and trimming were above 32, meaning that base calling-error probabilities were below 0.000631 and sequences are reliable. The summary statistics of the assemblies of *R. alpinia*, *R. battenbergiana* and *Alpinia* are presented in Table 1. We obtained 13,392 transcripts of potential gene orthologs among the three transcriptomes with the RBH analysis. 502 potential gene orthologs shared the same *A. thaliana* and *M. acuminata* best *BLAST* hit among the three taxa and were included in the COSII list. After inspecting the gene model of those 502 genes in *M. acuminata* and the available data in *Phytozome*, 16 introns (164–893 bp) flanked by CDS of low copy number and conserved proteins across representatives of the angiosperms were picked.

Of those 16 introns that were amplified and sequenced for the three individuals used to obtain the transcriptomes, one apparently has several copies and PCR products were non-specific, three produced sequences that didn't match the target regions when compared with the transcripts used to design the primers, and four didn't have three clear variable sites among the three species. The remaining eight introns (Table 2 and Fig. 1a–h) and ITS 1 & 2 were amplified and sequenced for as many accessions as possible. A summary of the workflow from the transcriptomes to the selected markers is shown in Fig. 2. We obtained seven or more markers for 40 of the 64 sampled species, for some of which multiple individuals were included for a total of 71 individuals (Supplementary Table 1). All the concatenated and species tree analyses were performed with this dataset. Herbarium specimens comprised 39.4% of the individuals included in that dataset with seven or more markers (and 67.9% of the individuals for which we were able to obtain at least one marker). PCR was most successful for ITS and the shorter markers (Fig. 3). Table 3 indicates the total number of sites, constant sites, parsimony informative sites, parsimony uninformative sites and the percentage of parsimony informative sites.

Table 1
Summary statistics of the assemblies obtained in *Trinity*.

Taxon	Transcripts	GC%	N50	Total assembled bases
R. alpinia	46,111	46.89	945	30,589,619
R. battenbergiana	46,549	47.34	882	29,154,199
Alpinia sp.	49,159	46.81	1017	34,640,218

3.2. Renealmia phylogeny

Gene trees obtained with Bayesian and likelihood methods confirmed that *Renealmia* is monophyletic with the outgroup included in this study. Overall the individual gene trees resolved species level relationships poorly (Supplementary Fig. 1a–i). Some recovered a clade including all the Neotropical species and gave moderate support to some relationships within the African and Neotropical lineages. The ITS region was among the least informative markers at species level (Table 3). The introns obtained showed a similar number of informative sites and the individual locus phylogenies recovered more clades with moderate support.

The Bayesian Inference analysis had convergence problems in the concatenated matrix runs that were solved after the removal of the iH1 region, which was subsequently excluded from the multi-locus analyses. The phylogenies resulting from the concatenation confirmed *Renealmia* monophyly with high support. The Neotropical species were recovered in a single monophyletic clade nested within the African species (Fig. 4). The relationships among African species are moderately supported with a few highly supported. Some clades are recovered within the Neotropical *Renealmia* with moderate to high support but the relationships among these clades are not resolved by the analyses and the branches among them are very short. The supported clades are broadly consistent with existing taxonomic subgroups (Maas, 1977). All the species of the *R. chalcochlora* and *R. breviscapa* subgroups are placed in their respective monophyletic clades with moderate support but the *R. aromatica* subgroup members are scattered in different clades, some of which include species of the same subgroup, taxa with undefined affinities or belonging to different groups and subgroups. The *R. ligulata* subgroup includes *R. cuatrecasasii* from the *R. aromatica* subgroup and all of the newly described species. The *R. thyrsoidea* subgroup is monophyletic with the species included in this analysis. A clade with species belonging to the *R. alpinia* subgroup includes *R. nicolaioides* from the related *R. ligulata* subgroup and an accession of *R. aromatica*; *R. petasites* appears unrelated to the other species of the subgroup. The two species of the *R. cernua* group are not related to each other. All the major groups (*R. cernua, R. alpinia, R. aromatica)* are polyphyletic. In widespread species like *R. cernua* and *R. ligulata,* where more than one individual per species was included, the conspecifics were closely related to each other, except in *R. aromatica* and to a lesser extent *R. lucida* and *R. polypus*.

Interestingly, in a locality in the Western ridge of the Colombian Andes called Cerro el Inglés in the Serranía de los Paragüas where several species are sympatric, some individuals show intermediate morphologies with the new species and are closely related in the phylogenies, pointing to the possibility of hybridization and introgression among *Renealmia* species. An individual of *R. lucida* ('R. lucida2' in Fig. 4) appears related to *R. sn. verrucosa* accessions ('R. sn ver') instead of the other *R. lucida* individuals and shows a pale green ovary that is more similar to *R. sn. verrucosa* morphologically. Also an individual morphologically similar to *R. sn. flaviflora* ('R. sn fla2') is nested in the *R. sn. verrucosa* clade and its closest relative is an individual of *R. sn. verrucosa* ('R. sn ver3') that shows leaf sheaths with an indumentum not found in other collections. Finally *R. sn. margaritae* ('R. sn mar') is related to two accessions of the sympatric species *R. sn. flaviflora*, one of which shows a white corolla and round labellum shape ('R. sn fla3') more similar to *R. sn. margaritae* than to other accessions identified as *R. sn. flaviflora*. Similarly, an individual ('R. aromatica1') identified as *R. aromatica* (a species with a paniculate thyrse inflorescence) is the closest relative of an individual ('R. alpinia1') identified as *R. alpinia*, a species commonly with a raceme inflorescence (*e.g.*'R. alpinia2') that also presents a paniculate thyrse.

The species tree methods show phylogenies very similar to the ones obtained with the concatenated analysis but in most cases the support values are lower, especially when compared to the Bayesian Inference concatenated analysis. The incongruence among the species trees is

Table 2
Identifiers of the eight selected transcripts of the three species, best *BLAST* hits in *A. thaliana* and *M. acuminata* for each and the names of the corresponding primers.

Primer name	*R. alpinia*	*R. battenbergiana*	*Alpinia* sp.	*A. thaliana*	*M. acuminata*	Product size (bp)
iH1_Re	Acomp12717_c0_seq1	Bcomp12102_c0_seq1	Ccomp10756_c0_seq1	At4g01940.1	GSMUA_Achr11T20820_001	215
iH2_Re	Acomp1360_c0_seq1	Bcomp13515_c0_seq1	Ccomp22414_c0_seq1	At2g02880.1	GSMUA_Achr8T30210_001	575
iH3_Re	Acomp18101_c0_seq1	Bcomp1799_c0_seq1	Ccomp12521_c0_seq1	At5g38880.1	GSMUA_Achr4T09790_001	735
iH4_Re	Acomp28884_c0_seq1	Bcomp15937_c0_seq1	Ccomp2075_c0_seq1	At3g01060.1	GSMUA_Achr9T26800_001	341
iH5_Re	Acomp35199_c0_seq1	Bcomp48055_c0_seq1	Ccomp35950_c0_seq1	At1g15980.1	GSMUA_Achr1T10090_001	403
iN_Re	Acomp14191_c0_seq1	Bcomp13503_c0_seq1	Ccomp5795_c0_seq1	At5g41190.1	GSMUA_Achr9T23650_001	183
iPG_Re	Acomp7115_c0_seq1	Bcomp9019_c0_seq1	Ccomp18349_c0_seq1	At2g16860.1	GSMUA_Achr1T15550_001	892
iPS_Re	Acomp15617_c0_seq1	Bcomp19624_c0_seq1	Ccomp20501_c0_seq1	At1g73820.1	GSMUA_Achr8T30330_001	341

considerable. The *ASTRAL* analysis score of the bootstrap replicates are around 0.6 implying that only 60% of the quartet trees included in the bootstrap replicates are included in the species trees suggesting high levels of incomplete lineage sorting (Fig. 5). The *BEST* analysis failed to converge, showing a decrease in the time-series plot of the log-likelihood estimates (*i.e.* "burnout") in all the independent runs. The **BEAST* analysis recovered a very similar topology to the one obtained by the concatenated Bayesian analysis, showing the same moderately and well supported groups with some differences like *R. battenbergiana* nested within a well-supported African clade.

4. Discussion

4.1. Transcriptome mining for phylogenetic markers

The number of informative sites per amplicon length in the introns we target using the transcriptomes is similar to the levels of variation found in ITS (Table 3). In some study cases the variation in ITS plus a chloroplast marker are good enough to recover species level relationships in recently diversified genera in the Neotropics (e.g. Armstrong et al., 2014; Martínez et al., 2015). Given that ITS is a marker variable enough to be informative of species level relationships in several groups (Ennos et al., 1999) the approach of targeting introns is a promising one to obtain several loci for the phylogenetic study of species-rich lineages. One of the advantages of the approach presented here is that Sanger sequencing allows some degree of success when sampling herbarium specimens (Särkinen et al., 2012). Although a considerable proportion of the individuals included in this analysis were sampled from herbarium specimens, the degraded and fragmented DNA in these extractions was difficult to amplify in the longer amplicons indicating shorter markers should be considered to be more likely to yield sequence data. However short markers could have insufficient informative sites and thus be unsuccessful in resolving phylogenies (Mirarab et al., 2014) and obtaining longer markers by combining multiple amplicons could be a useful approach.

Alternative approaches involving high throughput next generation sequencing platforms are appealing and studies using these approaches with historical samples are becoming frequent (Dodsworth, 2015; Bakker et al., 2015; Staats et al., 2013). Also, if the number of loci required to estimate species level phylogenies is much higher than the number of loci obtained here, as is the case with *Renealmia* species (see Section 4.2 below), targeted enrichment could be very cost effective (Bi et al., 2012; Cronn et al., 2012; Peñalba et al., 2014). Targeted enrichment has been used successfully to resolve the phylogenetic history of *Inga*, another recent South American radiation (Nicholls et al., 2015). Targeting regions with higher copy number, such as plastid genomes, could lead to better results than with nuclear genomes (Straub et al., 2012). Obtaining chloroplast genomes might be useful to acquire a high number of informative sites, but because the plastid genome represents a single large non-recombining and maternally inherited locus, sequence variation within it cannot provide insights into lineage sorting, hybridization and introgression, which require the exploration of multiple unlinked loci. These processes could play an important role in the evolutionary history of Neotropical plant lineages like *Renealmia* and are worth addressing. A combination of the approach given here to target some regions scattered through the genome with targeted enrichment approaches could yield very robust results and make good use of the newly available bioinformatics resources (Chamala et al., 2015) and public repositories (Droc et al., 2013; Goodstein et al., 2012).

4.2. Renealmia phylogeny

The possibility of hybridization events violates the assumptions of both the multispecies coalescent and concatenation methods and hence it is difficult to choose one for the dataset presented here. Because species tree estimation could be problematic when using loci with minimal phylogenetic signal (Xi et al., 2015), the debate of the performance of current multispecies coalescent methods over concatenation approaches is still open (Tonini et al., 2015) and since there is no hard incongruence in the results presented here, both methods are considered in the discussion of *Renealmia* evolutionary history.

In the specific case of *Renealmia* the gene trees obtained with individual markers are not informative enough to recover the relationships of the species, especially in the Neotropical lineages. This failure reflects the low numbers of phylogenetically informative sites available in the sequenced regions of each locus. This could result from a combination of recent origin and low mutation rates for the introns, despite being less prone to selection and hence more variable. However, mutation rates of nuclear genes are higher than chloroplast and mitochondrial genes in plants (Ennos et al., 1999). In *Renealmia* the phylogenies estimated with the ITS region are very poorly resolved and even with the addition of the seven introns in the concatenated analysis, although some clades are recovered, the relationships among those clades are uncertain.

As an alternative to low mutation rates, the lack of genetic differentiation in such a morphologically diverse and widely distributed group could be related to high effective population sizes and recent radiation, combining to generate contrasting patterns of sorting of ancestral polymorphism among loci. Population sizes are probably very large in common and widely distributed species (Steege et al., 2013) like *R. alpinia* that is distributed throughout Tropical South America, the lesser Antilles, Central America and Mexico (Maas, 1977). However, there are also narrow endemics, in Africa and in the Neotropics (*e.g. R. sancti-thomae* from São Thomé Island and *R. helenae* endemic to the cloud forest of Cerro Jefe, Panamá, respectively), for which lower population sizes would be expected. An additional factor making *Renealmia* phylogeny difficult to estimate could be that diversification occurred rapidly. Rapid diversifications have been reported in several lineages of plants (Drummond et al., 2012; Madriñán et al., 2013; Martínez et al., 2015; Nicholls et al., 2015; Richardson et al., 2001) in the Neotropics, suggesting that it plausibly would have shaped the evolutionary history of *Renealmia* (Valderrama et al. in preparation).

Finally, the possibility of hybridization and subsequent introgression could be a good explanation for the patterns obtained. Extensive hybridization and introgression could homogenise genetic differentiation among *Renealmia* species (Comes and Abbott, 2001). The scenario

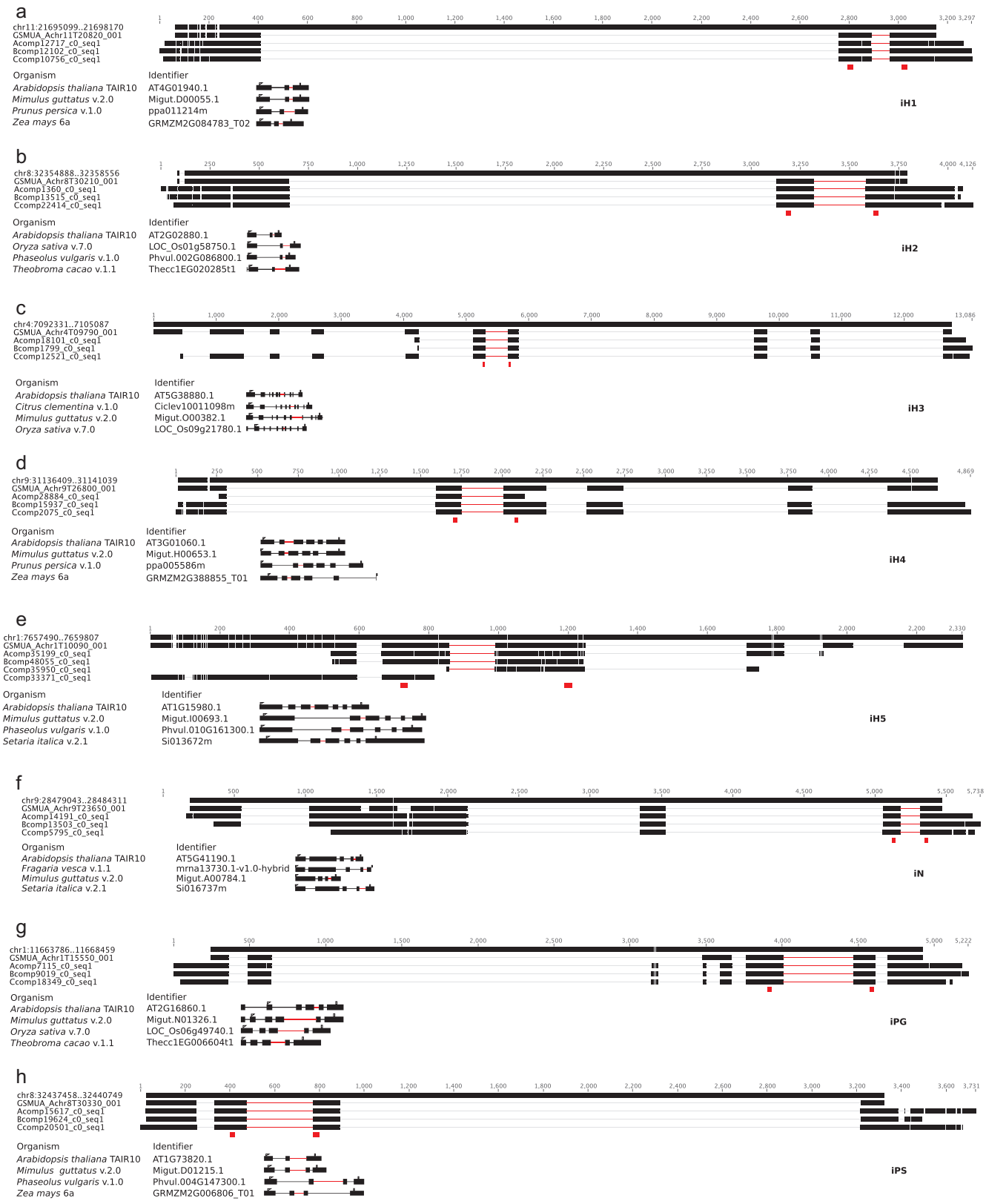

Fig. 1. (a–h) Alignments of *M. acuminata* chromosome and transcript followed by the generated *Renealmia* L.f. and *Alpinia* sp. transcripts (shown in that order) for the chosen eight loci; the target intron and the respective primers are shown in red. Below, the gene models of *A. thaliana* and additional examples throughout the angiosperms are shown with the target intron highlighted in red. Names given to the introns are shown in the bottom right hand corner. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

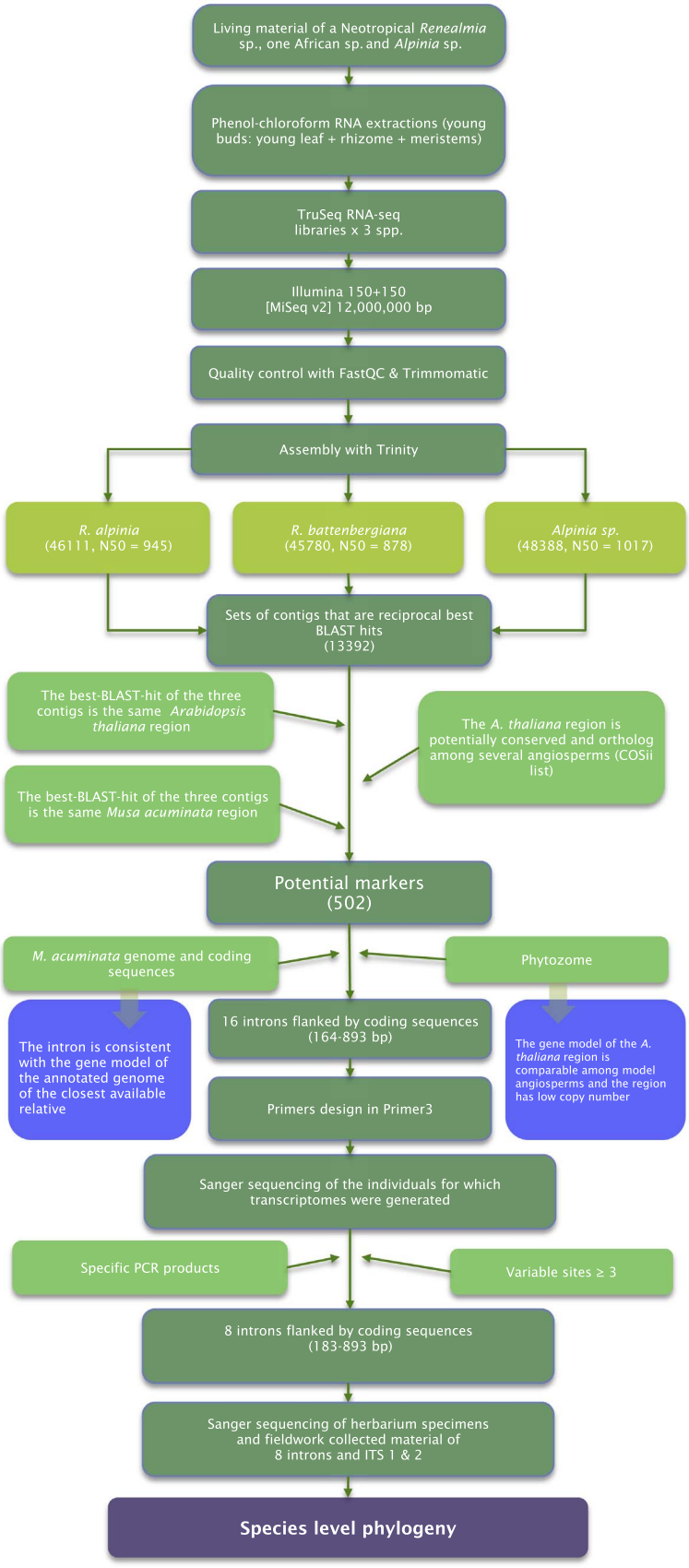

Fig. 2. Summary of the workflow to select the introns used as molecular markers to estimate the *Renealmia* species level phylogeny.

suggested is supported by morphology. For example, the occasional occurrence of a paniculate thyrse inflorescence in *Renealmia alpinia* (a species commonly with a raceme inflorescence) could be attributable to hybridization and introgression with *R. aromatica* that also has a paniculate thyrse (Maas, 1977). In addition, some species have individuals that show intermediate morphologies with other sympatric species and

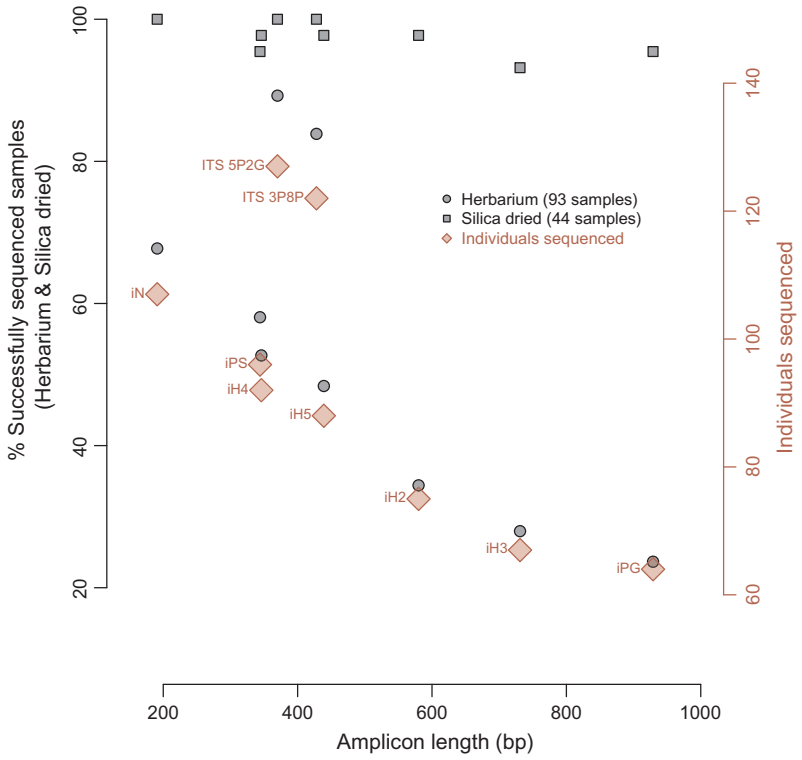

Fig. 3. Percent of successfully sequenced individuals sampled from silica dried and herbarium specimens against amplicon length for each locus (left axis) and the total number of individuals sequenced per locus (right axis and red symbols). Ribosomal RNA marker and shorter introns were obtained for more individuals, especially when extractions were made with herbarium material.

Table 3
Number of informative sites per locus.

Locus	ITS	iH1	iH2	iH3	iH4	iH5	iN	iPG	iPS
Constant sites	572	171	413	606	291	326	129	742	249
Parsimony informative sites	89	26	103	46	20	36	33	67	37
Parsimony uninformative sites	63	22	64	79	35	53	29	115	58
Total sites	724	219	580	731	346	415	191	924	344
% Parsimony informative sites	12.3	11.9	17.8	6.3	5.8	8.7	17.3	7.3	10.8

are more closely related to them than to conspecifics from other regions, *e.g. R. lucida* with *R. verrucosa, R. flaviflora* with *R. verrucosa* and *R. margaritae* with *R. flaviflora* from la Serranía de los Paragüas in Colombia, consistently with hybridization and probably introgression. This pattern contrasts with other Neotropical diversifications where species in sympatry seldom hybridize or introgress (Richardson et al., 2001; Twyford et al., 2015) and impels us to consider hybridization as a driver of *Renealmia* evolutionary history (Särkinen et al., 2007). Admixture is widespread in some plant lineages (Abbott et al., 2003; Comes and Abbott, 2001; Linder and Rieseberg, 2004; Slovák et al., 2014; Wu, 2015; Vargas et al., 2017) including *Renealmia* relatives like *Curcuma* L. (Zingiberaceae; Záveská et al., 2012), and this scenario of reticulate evolution has also been documented in corals (Hatta et al., 1999; Veron, 1995) and gallwasps (Nicholls et al., 2012). Interestingly the traditional and species tree based phylogenetic methods don't explicitly model hybridization and introgression events (Linder and Rieseberg, 2004; Xu, 2000) and hence the uncertainty of our estimated phylogenies could be higher than reported here. Methods such as the species tree approach in *BEST* assume that there is no extensive hybridization in the species analysed and violations of this assumption could explain the difficulties observed in our *BEST* analysis.

Alternatively, the accidental inclusion of non-homologous loci in the analysis could also be a source of disagreement between the gene trees (Nicholls et al., 2015). This scenario is difficult to reject because of the low signal in some loci. However, considering that the methods to pick markers were designed to avoid multiple copy regions and that all the gene trees (except the one obtained for iH1 locus) include close relatives of *Renealmia* and recover the genus as a monophyletic clade, we consider it unlikely that the loci obtained were involved in duplication events.

5. Conclusions

Using transcriptomes to target genes to estimate phylogenies is a promising approach and the genes used here are probably useful for estimating phylogenies at low taxonomic levels in the family and probably in the order. Using herbarium specimens as the source for DNA extractions for Sanger sequencing remains challenging even when targeting short amplicons. Although building a dataset when only herbarium specimens are available is feasible with the methods presented here, the possibility of a combination of recent radiations, hybridization and incomplete lineage sorting suggests that obtaining samples in the field and obtaining a high number of markers using next generation sequencing approaches covering plastid and nuclear regions could be the safest way to succeed when estimating species level phylogenies for problematic groups like *Renealmia*. Some of the subgroups within the genus defined by morphology are supported by the molecular data but the relationships among them are unresolved. Incomplete lineage sorting that might be related to rapid diversifications and large population sizes in some taxa added to the examples of possible hybridization could explain the difficulties evidenced here to estimate a species level phylogeny for the genus.

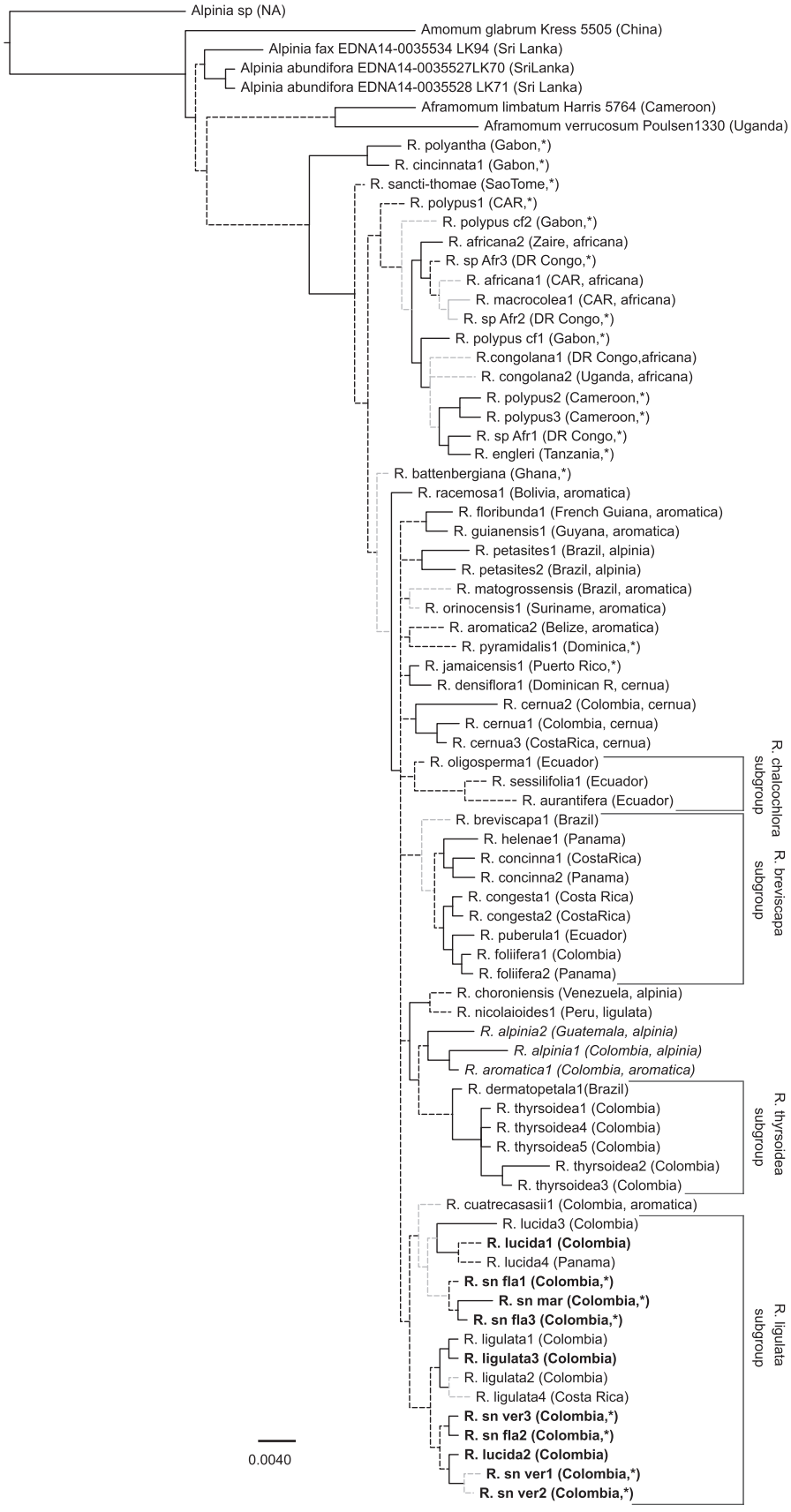

Fig. 4. Majority rule consensus tree of the Bayesian Inference analysis including the individuals with seven or more markers in a concatenated matrix. Branches with posterior probability values below 0.9 are shown in grey and branches with bootstrap support values of the Maximum Likelihood analysis (made in *GARLI* with the same dataset) below 80% are shown with dashed lines. The Neotropical species are a monophyletic clade and within it some clades match the taxonomic subgroups, however the relationships among those clades are not resolved. The country of each accession and the taxonomic subgroup is indicated for each accession (* indicate taxa that are not included in a taxonomic subgroup). The clade with the accessions of *R. alpinia* and *R. aromatica* mentioned in the discussion are shown in italics and the individuals from Serranía de los Paragüas in the Western ridge of the Colombian Andes are shown in bold type.

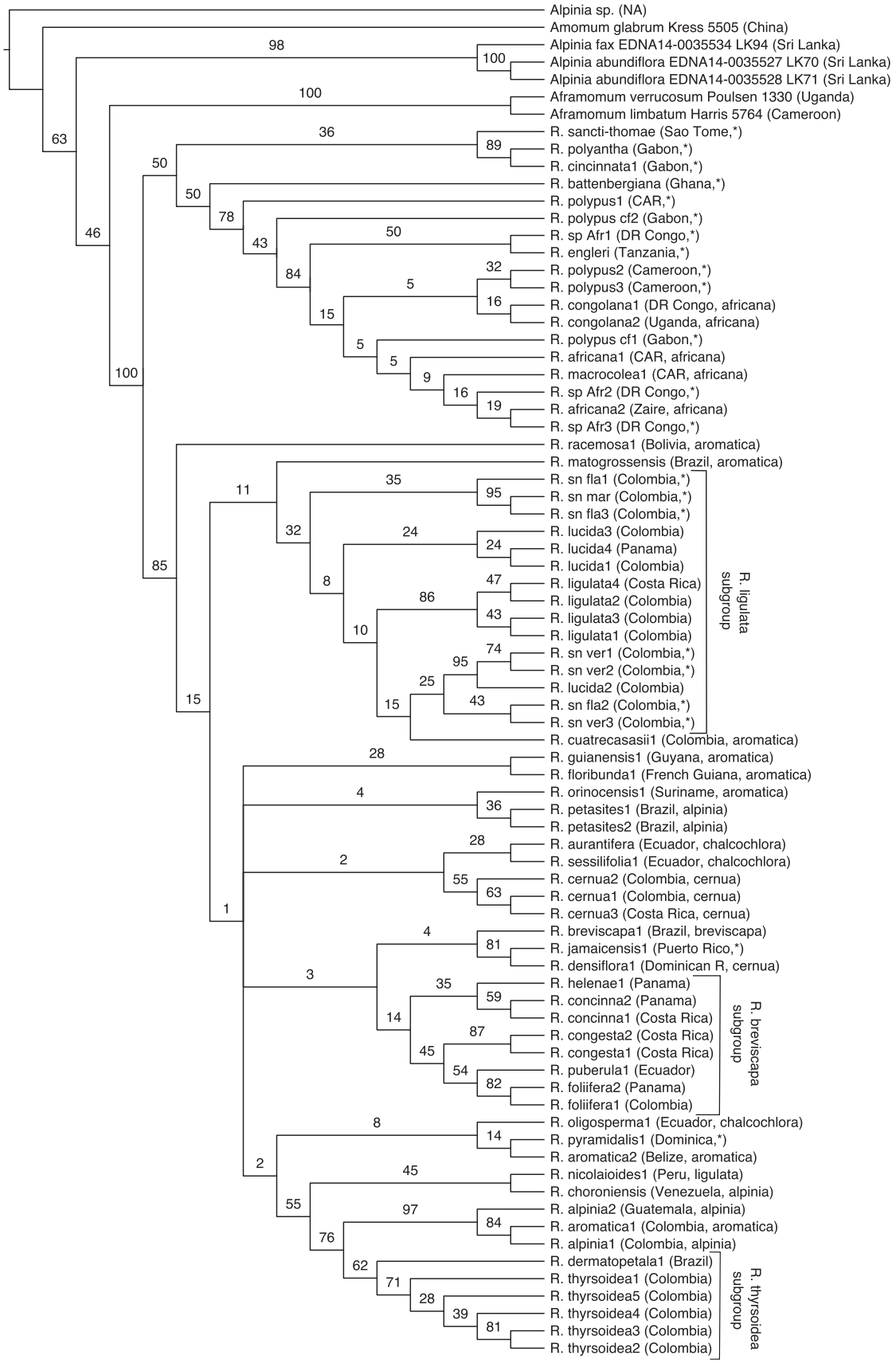

Fig. 5. Species trees generated with *ASTRAL* with multi-locus bootstrapping. A greedy consensus of the 100 replicates, support values are shown above the branches. The species tree shows a similar topology to the one obtained with the concatenated analyses but with lower support values. The amount of incongruence among the gene trees is evident from the low support values of the consensus. The country and the taxonomic subgroup is indicated for each accession (* indicate taxa that are not included in a taxonomic subgroup).

Acknowledgements

We acknowledge the Darwin Trust who provided a scholarship for the first author. Fieldwork was supported on two occasions by the University of Edinburgh Davis Expedition Fund. The Synthesys program of the European Commission funded a visit of the first author to the National Herbarium Nederland, Herbaria that provided samples of *Renealmia* species including Naturalis Biodiversity Center in the Netherlands (U, L and WAG collections) and the Missouri Botanical Garden's Herbarium (MO). Paul and Hiltje Maas for their advice and support. Fabiola Areces of the University of Puerto Rico-Río Piedras Herbarium (UPRRP) and David Harris of the Royal Botanic Garden Edinburgh (RBGE) for sending samples of their collection numbers. Ruth Hollands, Michelle Hart and Laura Forrest of RBGE for their help in the laboratory. To Toby Pennington and Tiina Särkinen of the RBGE for providing advice throughout the project. To Elspeth Haston (RBGE) and Félix Forest (Royal Botanic Gardens, Kew) for comments on the manuscript.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ympev.2017.10.001.

References

Abbott, R.J., James, J.K., Milne, R.I., Gillies, A.C.M., 2003. Plant introductions, hybridization and gene flow. Philos. Trans. R. Soc. B: Biol. Sci. 358, 1123–1132. http://dx.doi.org/10.1098/rstb.2003.1289.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410. http://dx.doi.org/10.1016/S0022-2836(05)80360-2.

Andrews, S., Lindenbaum, P., Howard, B., Ewels, P., 2015. FastQC.

Armstrong, K.E., Stone, G.N., Nicholls, J.A., Valderrama, E., Anderberg, A.A., Smedmark, J., Gautier, L., Naciri, Y., Milne, R., Richardson, J.E., 2014. Patterns of diversification amongst tropical regions compared: a case study in Sapotaceae. Front. Genet. 5, 1–13. http://dx.doi.org/10.3389/fgene.2014.00362.

Auvray, G., Harris, D.J., Richardson, J.E., Newman, M.F., Särkinen, T.E., 2010. Phylogeny and Dating of *Aframomum* (Zingiberaceae). In: Seberg, O., Sánchez Barfod, A., Petersen, G., Davis, J.I. (Eds.), Diversity, Phylogeny and Evolution in the Monocotyledons. Aarhus, Denmark, pp. 287–305.

Álvarez, I., Wendel, J.F., 2003. Ribosomal ITS sequences and plant phylogenetic inference. Mol. Phylogenet. Evol. 29, 417–434. http://dx.doi.org/10.1016/s1055-7903(03)00208-2.

Bakker, F.T., Lei, D., Yu, J., Mohammadin, S., Wei, Z., van de Kerke, S., Gravendeel, B., Nieuwenhuis, M., Staats, M., Alquezar-Planas, D.E., Holmer, R., 2015. Herbarium genomics: plastome sequence assembly from a range of herbarium specimens using an Iterative Organelle Genome Assembly pipeline. Biol. J. Lin. Soc. 117, 33–43.

Bi, K., Vanderpool, D., Singhal, S., Linderoth, T., Moritz, C., Good, J.M., 2012. Transcriptome-based exon capture enables highly cost-effective comparative genomic data collection at moderate evolutionary scales. BMC Genomics 13, 403.

Biomatters Development Team, 2010. Geneious Pro.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120. http://dx.doi.org/10.1093/bioinformatics/btu170.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M.A., Rambaut, A., Drummond, A.J., 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. 10http://dx.doi.org/10.1371/journal.pcbi.1003537. e1003537–6.

Bouckaert, R.R., 2010. DensiTree: making sense of sets of phylogenetic trees. Bioinformatics 26, 1372–1373. http://dx.doi.org/10.1093/bioinformatics/btq110.

Brassac, J., Blattner, F.R., 2015. Species-Level Phylogeny and Polyploid Relationships in *Hordeum* (Poaceae) Inferred by Next-Generation Sequencing and In Silico Cloning of Multiple Nuclear Loci. Syst. Biol. 64, 792–808. http://dx.doi.org/10.1093/sysbio/syv035.

Castillo-Ramírez, S., Liu, L., Pearl, D., Edwards, S.V., 2011. Bayesian estimation of species trees: a practical guide to optimal sampling and analysis. In: Knowles, L.L., Kubatko, L.S. (Eds.), Estimating Species Trees. John Wiley & Sons, Hoboken, New Jersey, pp. 15–34.

Chamala, S., García, N., Godden, G.T., Krishnakumar, V., Jordon-Thaden, I.E., Smet, R.D., Barbazuk, W.B., Soltis, D.E., Soltis, P.S., 2015. MarkerMiner 1.0: a new application for phylogenetic marker development using angiosperm transcriptomes. Appl. Plant Sci. 3, 1400115–1400118. http://dx.doi.org/10.3732/apps.1400115.

Comes, H.P., Abbott, R.J., 2001. Molecular phylogeography, reticulation, and lineage sorting in Mediterranean *Senecio* sect. *Senecio* (Asteraceae). Evolution 55, 1943. http://dx.doi.org/10.1554/0014-3820(2001) 055[1943:mprals]2.0.co;2.

Cronn, R., Knaus, B.J., Liston, A., Maughan, P.J., Parks, M., Syring, J.V., Udall, J., 2012. Targeted enrichment strategies for next-generation plant biology. Am. J. Bot. 99, 291–311. http://dx.doi.org/10.3732/ajb.1100356.

Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2012. JModelTest 2: more models, new heuristics and parallel computing. Nat. Meth. 9http://dx.doi.org/10.1038/nmeth.2109. 772-772.

Dodsworth, S., 2015. Genome skimming for next-generation biodiversity analysis. Trends Plant Sci. 20, 525–527.

Doyle, J.J., 1990. Isolation of plant DNA from fresh tissue. Focus 12, 13–15.

Droc, G., Larivière, D., Guignon, V., Yahiaoui, N., This, D., Garsmeur, O., Dereeper, A., Hamelin, C., Argout, X., Dufayard, J.-F., Lengelle, J., Baurens, F.-C., Cenci, A., Pitollat, B., D'Hont, A., Ruiz, M., Rouard, M., Bocs, S., 2013. The banana genome hub. Database (Oxford) 2013, bat035. http://dx.doi.org/10.1093/database/bat035.

Drummond, C.S., Eastwood, R.J., Miotto, S.T.S., Hughes, C.E., 2012. Multiple continental radiations and correlates of diversification in *Lupinus* (Leguminosae): testing for key innovation with incomplete taxon sampling. Syst. Biol. 61, 443–460. http://dx.doi.org/10.1093/sysbio/syr126.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797. http://dx.doi.org/10.1093/nar/gkh340.

Ennos, R.A., Sinclair, W.T., Hu, X.-S., Langdon, A., 1999. Using organelle markers to elucidate the history, ecology and evolution of plant populations. In: Hollingsworth, P.M., Bateman, R.M., Gornall, R.J. (Eds.), Molecular Systematics and Plant Evolution. pp. 1–19.

Ewing, B., Green, P., 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res. 8, 186–194.

Goodstein, D.M., Shu, S., Howson, R., Neupane, R., Hayes, R.D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., Rokhsar, D.S., 2012. Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res. 40, D1178–D1186. http://dx.doi.org/10.1093/nar/gkr944.

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., MacManes, M.D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C.N., Henschel, R., LeDuc, R.D., Friedman, N., Regev, A., 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. 8, 1494–1512. http://dx.doi.org/10.1038/nprot.2013.084.

Hatta, M., Fukami, H., Wang, W., Omori, M., Shimoike, K., Hayashibara, T., Ina, Y., Sugiyama, T., 1999. Reproductive and genetic evidence for a reticulate evolutionary history of mass-spawning corals. Mol. Biol. Evol. 16, 1607–1613.

Heled, J., Drummond, A.J., 2010. Bayesian inference of species trees from multilocus data. Mol. Biol. Evol. 27, 570–580. http://dx.doi.org/10.1093/molbev/msp274.

Hughes, C., Eastwood, R., 2006. Island radiation on a continental scale: exceptional rates of plant diversification after uplift of the Andes. Proc. Natl. Acad. Sci. U.S.A. 103, 10334–10339. http://dx.doi.org/10.1073/pnas.0601928103.

Hughes, C.E., Eastwood, R.J., Donovan Bailey, C., 2006. From famine to feast? Selecting nuclear DNA sequence loci for plant species-level phylogeny reconstruction. Philos. Trans. R. Soc. B: Biol. Sci. 361, 211–225. http://dx.doi.org/10.1098/rstb.2005.1735.

Kress, W.J., Liu, A.-Z., Newman, M., Li, Q.-J., 2005. The molecular phylogeny of *Alpinia* (Zingiberaceae): a complex and polyphyletic genus of gingers. Am. J. Bot. 92, 167–178. http://dx.doi.org/10.3732/ajb.92.1.167.

Kress, W.J., Newman, M.F., Poulsen, A.D., Specht, C., 2007. An analysis of generic circumscriptions in tribe Alpinieae (Alpinioideae: Zingiberaceae). Gard. Bull. Singapore 59, 113–128.

Lamesch, P., Berardini, T.Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., Muller, R., Dreher, K., Alexander, D.L., Garcia-Hernandez, M., Karthikeyan, A.S., Lee, C.H., Nelson, W.D., Ploetz, L., Singh, S., Wensel, A., Huala, E., 2012. The *Arabidopsis* Information Resource (TAIR): improved gene annotation and new tools. Nucleic Acids Res. 40, D1202–D1210. http://dx.doi.org/10.1093/nar/gkr1090.

Lanfear, R., Calcott, B., Ho, S.Y.W., Guindon, S., 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol. Biol. Evol. 29, 1695–1701. http://dx.doi.org/10.1093/molbev/mss020.

Linder, C.R., Rieseberg, L.H., 2004. Reconstructing patterns of reticulate evolution in plants. Am. J. Bot. 91, 1700–1708. http://dx.doi.org/10.3732/ajb.91.10.1700.

Liu, L., 2008. BEST: Bayesian estimation of species trees under the coalescent model. Bioinformatics 24, 2542–2543. http://dx.doi.org/10.1093/bioinformatics/btn484.

Liu, L., Yu, L., 2011. Estimating species trees from unrooted gene trees. Syst. Biol. 60, 661–667. http://dx.doi.org/10.1093/sysbio/syr027.

Liu, L., Yu, L., Edwards, S.V., 2010. A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. BMC Evol. Biol. 10, 302. http://dx.doi.org/10.1186/1471-2148-10-302.

Liu, L., Yu, L., Pearl, D.K., Edwards, S.V., 2009. Estimating species phylogenies using coalescence times among sequences. Syst. Biol. 58, 468–477. http://dx.doi.org/10.1093/sysbio/syp031.

Maas, P.J.M., 1977. Renealmia (Zingiberaceae - Zingiberoideae) Costoideae (Additions) (Zingiberaceae): Flora Neotropica 18. New York Botanical Garden Press, New York.

Maddison, W.P., 1997. Gene trees in species trees. Syst. Biol. 46, 523–536.

Maddison, W.P., Knowles, L.L., 2006. Inferring phylogeny despite incomplete lineage sorting. Syst. Biol. 55, 21–30. http://dx.doi.org/10.1080/10635150500354928.

Madriñán, S., Cortés, A.J., Richardson, J.E., 2013. Páramo is the world's fastest evolving and coolest biodiversity hotspot. Front. Genet. 4, 1–7. http://dx.doi.org/10.3389/fgene.2013.00192/abstract.

Martínez, C., Carvalho, M.R., Madriñán, S., Jaramillo, C.A., 2015. A Late Cretaceous *Piper* (Piperaceae) from Colombia and diversification patterns for the genus. Am. J. Bot. 102, 273–289. http://dx.doi.org/10.3732/ajb.1400427.

Metzker, M.L., 2009. Sequencing technologies—the next generation. Nat. Rev. Genet. 11, 31–46. http://dx.doi.org/10.1038/nrg2626.

Miller, M.A., Pfeiffer, W., Schwartz, T., 2011. The CIPRES science gateway: a community resource for phylogenetic analyses 1–8. http://dx.doi.org/10.1145/2016741.

2016785.
Mirarab, S., Reaz, R., Bayzid, M.S., Zimmermann, T., Swenson, M.S., Warnow, T., 2014. ASTRAL: genome-scale coalescent-based species tree estimation. Bioinformatics 30, i541–i548. http://dx.doi.org/10.1093/bioinformatics/btu462.
Mirarab, S., Warnow, T., 2015. ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. Bioinformatics 31, 44–52. http://dx.doi.org/10.1093/bioinformatics/btv234.
Nicholls, J.A., CHALLIS, RJ., MUTUN, S., Stone, GN., 2012. Mitochondrial barcodes are diagnostic of shared refugia but not species in hybridizing oak gallwasps. Mol. Ecol. 21, 4051–4062. http://dx.doi.org/10.1111/j.1365-294X.2012.05683.x.
Nicholls, J.A., Pennington, R.T., Koenen, E.J.M., Hughes, C.E., Hearn, J., Bunnefeld, L., Dexter, K.G., Stone, G.N., Kidner, C.A., 2015. Using targeted enrichment of nuclear genes to increase phylogenetic resolution in the neotropical rain forest genus *Inga* (Leguminosae: Mimosoideae). Front. Plant. Sci. 6, 710. http://dx.doi.org/10.3389/fpls.2015.00710.
Pamilo, P., Nei, M., 1988. Relationships between gene trees and species trees. Mol. Biol. Evol. 5, 568–583.
Peñalba, J.V., Smith, L.L., Tonione, M.A., Sass, C., Hykin, S.M., Skipwith, P.L., McGuire, J.A., Bowie, R.C.K., Moritz, C., 2014. Sequence capture using PCR-generated probes: a cost-effective method of targeted high-throughput sequencing for nonmodel organisms. Mol. Ecol. Resourc. 14, 1000–1010.
Ralser, M., Querfurth, R., Warnatz, H.-J., Lehrach, H., Yaspo, M.-L., Krobitsch, S., 2006. An efficient and economic enhancer mix for PCR. Biochem. Biophys. Res. Commun. 347, 747–751. http://dx.doi.org/10.1016/j.bbrc.2006.06.151.
Rambaut, A., 2014. FigTree.
Rambaut, A., Drummond, A.J., 2002. TreeAnnotator.
Rambaut, A., Suchard, M.A., Xie, W., Drummond, A.J., 2014. Tracer.
Richardson, J.E., Pennington, R.T., Pennington, T.D., Hollingsworth, P.M., 2001. Rapid diversification of a species-rich genus of neotropical rain forest trees. Science 293, 2242–2245. http://dx.doi.org/10.1126/science.1061421.
Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Hohna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61, 539–542. http://dx.doi.org/10.1093/sysbio/sys029.
Rothfels, C.J., Larsson, A., Li, F.-W., Sigel, E.M., Huiet, L., Burge, D.O., Ruhsam, M., Graham, S.W., Stevenson, D.W., Wong, G.K.-S., Korall, P., Pryer, K.M., 2013. Transcriptome-mining for single-copy nuclear markers in ferns. PLoS ONE 8http://dx.doi.org/10.1371/journal.pone.0076957. e76957-18.
Särkinen, T.E., Staats, M., Richardson, J.E., Cowan, R.S., Bakker, F.T., 2012. How to open the treasure chest? Optimising DNA extraction from herbarium specimens. PLoS ONE 7, e43808–e43809. http://dx.doi.org/10.1371/journal.pone.0043808.
Särkinen, T.E., Newman, M.F., Maas, P.J.M., Maas, H., Poulsen, A.D., Harris, D.J., Richardson, J.E., Clark, A., Hollingsworth, M., Toby Pennington, R., 2007. Recent oceanic long-distance dispersal and divergence in the amphi-Atlantic rain forest genus *Renealmia* L.f. (Zingiberaceae). Mol. Phylogenet. Evol. 44, 968–980. http://dx.doi.org/10.1016/j.ympev.2007.06.007.
Seo, T.-K., 2008. Calculating bootstrap probabilities of phylogeny using multilocus sequence data. Mol. Biol. Evol. 25, 960–971. http://dx.doi.org/10.1093/molbev/msn043.
Shaw, J., Lickey, E.B., Schilling, E.E., Small, R.L., 2007. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. Am. J. Bot. 94, 275–288. http://dx.doi.org/10.3732/ajb.94.3.275.
Shaw, T.I., Ruan, Z., Glenn, T.C., Liu, L., 2013. STRAW: species TRee analysis web server. Nucleic Acids Res. 41, W238–W241. http://dx.doi.org/10.1093/nar/gkt377.
Slovák, M., Kučera, J., Záveská, E., Vd'ačný, P., 2014. Dealing with discordant genetic signal caused by hybridisation, incomplete lineage sorting and paucity of primary nucleotide homologies: a case study of closely related members of the genus *Picris* Subsection *Hieracioides* (Compositae). PLoS ONE 9http://dx.doi.org/10.1371/journal.pone.0104929. e104929-19.
Staats, M., Erkens, R.H.J., van de Vossenberg, B., Wieringa, J.J., Kraaijeveld, K., Stielow, B., Geml J., Richardson, J.E., Bakker, F.T., 2013. Genomic treasure troves: complete genome sequencing of herbarium and insect museum specimens. PLoS ONE, 8, e69189.
Steege, ter, H., Pitman, N.C.A., Sabatier, D., Baraloto, C., Salomão, R.P., Guevara, J.E., Phillips, O.L., Castilho, C.V., Magnusson, W.E., Molino, J.-F., Monteagudo, A., Núñez Vargas, P., Montero, J.C., Feldpausch, T.R., Coronado, E.N.H., Killeen, T.J., Mostacedo, B., Vasquez, R., Assis, R.L., Terborgh, J., Wittmann, F., Andrade, A., Laurance, W.F., Laurance, S.G.W., Marimon, B.S., Marimon, B.-H., Guimarães Vieira, I.C., Amaral, I.L., Brienen, R., Castellanos, H., Cárdenas López, D., Duivenvoorden, J. F., Mogollón, H.F., Matos, F.D. de A., Dávila, N., García-Villacorta, R., Stevenson Diaz, P.R., Costa, F., Emilio, T., Levis, C., Schietti, J., Souza, P., Alonso, A., Dallmeier, F., Montoya, A.J.D., Fernandez Piedade, M.T., Araujo-Murakami, A., Arroyo, L., Gribel, R., Fine, P.V.A., Peres, C.A., Toledo, M., Aymard C, G.A., Baker, T.R., Cerón, C., Engel, J., Henkel, T.W., Maas, P., Petronelli, P., Stropp, J., Zartman, C.E., Daly, D., Neill, D., Silveira, M., Paredes, M.R., Chave, J., Lima Filho, D. de A., Jørgensen, P.M., Fuentes, A., Schöngart, J., Cornejo Valverde, F., Di Fiore, A., Jimenez, E.M., Peñuela Mora, M.C., Phillips, J.F., Rivas, G., van Andel, T.R., Hildebrand, von, P., Hoffman, B., Zent, E.L., Malhi, Y., Prieto, A., Rudas, A., Ruschell, A.R., Silva, N., Vos, V., Zent, S., Oliveira, A.A., Schutz, A.C., Gonzales, T., Trindade Nascimento, M., Ramirez-Angulo, H., Sierra, R., Tirado, M., Umaña Medina, M.N., van der Heijden, G., Vela, C. I.A., Vilanova Torre, E., Vriesendorp, C., Wang, O., Young, K.R., Baider, C., Balslev, H., Ferreira, C., Mesones, I., Torres-Lezama, A., Urrego Giraldo, L.E., Zagt, R., Alexiades, M.N., Hernandez, L., Huamantupa-Chuquimaco, I., Milliken, W., Palacios Cuenca, W., Pauletto, D., Valderrama Sandoval, E., Valenzuela Gamarra, L., Dexter, K.G., Feeley, K., Lopez-Gonzalez, G., Silman, M.R., 2013. Hyperdominance in the Amazonian tree flora. Science 342, 1243092. http://dx.doi.org/10.1126/science.1243092.
Straub, S.C.K., Parks, M., Weitemier, K., Fishbein, M., Cronn, R.C., Liston, A., 2012. Navigating the tip of the genomic iceberg: next-generation sequencing for plant systematics. Am. J. Bot. 99, 349–364. http://dx.doi.org/10.3732/ajb.1100335.
Sukumaran, J., Holder, M.T., 2010. DendroPy: a Python library for phylogenetic computing. Bioinformatics 26, 1569–1571. http://dx.doi.org/10.1093/bioinformatics/btq228.
Tatusov, R.L., Koonin, E.V., Lipman, D.J., 1997. A genomic perspective on protein families. Science 278, 631–637.
Tonini, J., Moore, A., Stern, D., Shcheglovitova, M., Ortí, G., 2015. Concatenation and species tree methods exhibit statistically indistinguishable accuracy under a range of simulated conditions. PLoS Curr 7, 1–16. http://dx.doi.org/10.1371/currents.tol.34260cc27551a527b124ec5f6334b6be.
Twyford, A.D., Kidner, C.A., Ennos, R.A., 2015. Maintenance of species boundaries in a Neotropical radiation of *Begonia*. Mol. Ecol. 24, 4982–4993. http://dx.doi.org/10.1111/mec.13355.
Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., Rozen, S.G., 2012. Primer3–new capabilities and interfaces. Nucleic Acids Res. 40http://dx.doi.org/10.1093/nar/gks596. e115-e115.
Vargas, O.M., Ortiz, E.M., Simpson, B.B., 2017. Conflicting phylogenomic signals reveal a pattern of reticulate evolution in a recent high-Andean diversification (Asteraceae: Astereae: *Diplostephium*). New Phytol. 214, 1736–1750. http://dx.doi.org/10.1111/nph.14530.
Veron, J.E.N., 1995. Corals in Space and Time. Cornell University Press, Ithaca, NY.
Wang, Z., Gerstein, M., Snyder, M., 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. 10, 57–63. http://dx.doi.org/10.1038/nrg2484.
Ward, N., Moreno-Hagelsieb, G., 2014. Quickly finding orthologs as reciprocal best hits with BLAT, LAST, and UBLAST: how much do we miss? PLoS ONE 9, e101850–e101856. http://dx.doi.org/10.1371/journal.pone.0101850.
Wu, F., Mueller, L.A., Crouzillat, D., Pétiard, V., Tanksley, S.D., 2006. Combining bioinformatics and phylogenetics to identify large sets of single-copy orthologous genes (COSII) for comparative, evolutionary and systematic studies: a test case in the euasterid plant clade. Genetics 174, 1407–1420. http://dx.doi.org/10.1534/genetics.106.062455.
Wu, J., 2015. Phylogeny of Salix subgenus *Salix* s.l. (Salicaceae): delimitation, biogeography, and reticulate evolution. BMC Evol. Biol. 15, 1–13. http://dx.doi.org/10.1186/s12862-015-0311-7.
Xi, Z., Liu, L., Davis, C.C., 2015. Genes with minimal phylogenetic information are problematic for coalescent analyses when gene tree estimation is biased. Mol. Phylogenet. Evol. 92, 63–71. http://dx.doi.org/10.1016/j.ympev.2015.06.009.
Xu, S., 2000. Phylogenetic analysis under reticulate evolution. Mol. Biol. Evol. 17, 897–907.
Yang, Y., Moore, M.J., Brockington, S.F., Soltis, D.E., Wong, G.K.-S., Carpenter, E.J., Zhang, Y., Chen, L., Yan, Z., Xie, Y., Sage, R.F., Covshoff, S., Hibberd, J.M., Nelson, M.N., Smith, S.A., 2015. Dissecting molecular evolution in the highly diverse plant clade Caryophyllales using transcriptome sequencing. Mol. Biol. Evol. 32, 2001–2014. http://dx.doi.org/10.1093/molbev/msv081.
Záveská, E., Fér, T., Šída, O., Krak, K., Marhold, K., 2012. Phylogeny of *Curcuma* (Zingiberaceae) based on plastid and nuclear sequences: proposal of the new subgenus Ecomata. Taxon 61, 747–763.
Zwickl, D.J., 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Austin, TX.