The following article appeared in Molecular Phylogenetics and Evolution 117: 111-123 (2017); and may be found at: https://doi.org/10.1016/j.ympev.2017.05.014

Molecular Phylogenetics and Evolution 117 (2017) 111–123

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

Recalcitrant deep and shallow nodes in *Aristolochia* (Aristolochiaceae) illuminated using anchored hybrid enrichment

Stefan Wanke [a,*], Carolina Granados Mendoza [b,c,1], Sebastian Müller [a,1], Anna Paizanni Guillén [d,e], Christoph Neinhuis [a], Alan R. Lemmon [f], Emily Moriarty Lemmon [g], Marie-Stéphanie Samain [e]

[a] Technische Universität Dresden, Institut für Botanik, Zellescher Weg 20b, 01062 Dresden, Germany
[b] CONACYT División de Biología Molecular, Instituto Potosino de Investigación Científica y Tecnológica A.C., Camino a la Presa de San José 2055, Lomas 4a. sección, C.P. 78216 San Luis Potosí, San Luis Potosí, Mexico
[c] Departamento de Botánica, Instituto de Biología, Universidad Nacional Autónoma de México, Apartado Postal 70-367, 04510 Coyoacán, Distrito Federal, Mexico
[d] Universidad Michoacana de San Nicolás de Hidalgo, 58066 Morelia, Michoacán, Mexico
[e] Instituto de Ecología, A.C., Centro Regional del Bajío, 61600 Pátzcuaro, Michoacán, Mexico
[f] Department of Scientific Computing, Florida State University, Dirac Science Library, Tallahassee, FL 32306-4102, USA
[g] Department of Biological Science, Florida State University, 319 Stadium Drive, PO Box 3064295, Tallahassee, FL, 32306-4295, USA

ARTICLE INFO

Article history:
Received 29 September 2016
Revised 12 May 2017
Accepted 15 May 2017
Available online 20 May 2017

Keywords:
Enrichment strategy
Next generation sequencing
Phylogenomics
Polyploidy
Short internodes
Recent diversification
Universal nuclear probe set

ABSTRACT

Recalcitrant relationships are characterized by very short internodes that can be found among shallow and deep phylogenetic scales all over the tree of life. Adding large amounts of presumably informative sequences, while decreasing systematic error, has been suggested as a possible approach to increase phylogenetic resolution. The development of enrichment strategies, coupled with next generation sequencing, resulted in a cost-effective way to facilitate the reconstruction of recalcitrant relationships. By applying the anchored hybrid enrichment (AHE) genome partitioning strategy to *Aristolochia* using an universal angiosperm probe set, we obtained 231–233 out of 517 single or low copy nuclear loci originally contained in the enrichment kit, resulting in a total alignment length of 154,756 bp to 160,150 bp. Since *Aristolochia* (Piperales; magnoliids) is distantly related to any angiosperm species whose genome has been used for the plant AHE probe design (*Amborella trichopoda* being the closest), it serves as a proof of universality for this probe set. *Aristolochia* comprises approximately 500 species grouped in several clades (OTUs), whose relationships to each other are partially unknown. Previous phylogenetic studies have shown that these lineages branched deep in time and in quick succession, seen as short-deep internodes. Short-shallow internodes are also characteristic of some *Aristolochia* lineages such as *Aristolochia* subsection *Pentandrae*, a clade of presumably recent diversification. This subsection is here included to test the performance of AHE at species level. Filtering and subsampling loci using the phylogenetic informativeness method resolves several recalcitrant phylogenetic relationships within *Aristolochia*. By assuming different ploidy levels during bioinformatics processing of raw data, first hints are obtained that polyploidization contributed to the evolution of *Aristolochia*. Phylogenetic results are discussed in the light of current systematics and morphology.

1. Introduction

Hundreds of loci are potentially needed to resolve recalcitrant phylogenetic relationships (Leaché and Rannala, 2010; Wickett et al., 2014; Prum et al., 2015), which are very short internodes that are often present at shallow scales, as well as among deep level nodes. These so-called short branched clades are found in many plant lineages all over the tree of life. For many years, resolving such nodes in phylogenies has been recognized as most challenging (e.g. Richardson et al., 2004). An ongoing discussion surrounds the question of increasing taxonomic sampling versus incrementing character sampling for previously poorly resolved phylogenies (e.g. Goldman, 1998; Geuten et al., 2007; Leaché and Rannala, 2010; Townsend and López-Giráldez, 2010; San Mauro et al., 2012).

Sequencing costs per DNA base pair are continuously decreasing and at the same time accuracy of raw data is increasing (Cai et al., 2012; Lemmon and Lemmon, 2013). Despite reduced

* Corresponding author.
E-mail address: stefan.wanke@tu-dresden.de (S. Wanke).
[1] Authors contributed equally to this study.

http://dx.doi.org/10.1016/j.ympev.2017.05.014
1055-7903/

sequencing costs, whole genome sequencing is still very expensive if applied to a broad sampling and, in most cases, far beyond necessary, as only a fraction of the genome is potentially needed (Ruane et al., 2015). Furthermore, it requires extensive downstream (post-sequencing) processing with many custom scripts, whereas only a small amount of the obtained data may be well suited for phylogenetic questions (Carstens et al., 2012; Lemmon and Lemmon, 2013). Aiming at more cost-efficient results, only presumably informative loci can be sequenced. As such, it is avoided to produce a large amount of potentially uninformative molecular data that might demand high computational effort or present the difficulty of discerning between orthologs and paralogs when genes are multi-copy. Furthermore, the acquired data can be subsampled to increase the phylogenetic information while simultaneously decreasing systematic error (Lemmon and Lemmon, 2013). Consequently, the era of next generation sequencing used in phylogenomics contains a new challenge: applying new techniques to select and sequence only those loci that are informative and necessary for answering the particular phylogenetic question.

One recently developed technique is "Anchored Hybrid Enrichment" (AHE). In this approach selected loci are enriched with the help of enrichment probes, oligonucleotide sequences of ca. 60–120 bp that hybridize to moderately conserved target regions (compared to ultraconserved elements), followed by sequencing of the enriched target regions, and more variable flanking regions, on a high-throughput sequencing platform (Lemmon and Lemmon, 2013). Lemmon et al. (2012) designed a vertebrate-wide enrichment probe set and obtained a fully resolved and well-supported species tree for vertebrates. More recently, this methodology was applied to snakes (Pyron et al., 2014; Ruane et al., 2015; Pyron et al., 2016; Chen et al., 2017), lizards (Leaché et al., 2014; Brandley et al., 2015; Tucker et al., 2017; Domingos et al., 2017; Manthey et al., 2016), frogs (Peloso et al., 2015), fish (Eytan et al., 2015; Stout et al., 2016), birds (Prum et al., 2015), flies (Young et al., 2016), and spiders (Hamilton et al., 2016). Although solid evidence exists from diverse animal lineages, a probe set for angiosperms only recently became available (Buddenhagen et al., 2016), which since then has been applied at species level to the genera *Salvia* (Lamiaceae, Fragoso et al., 2017) and *Protea* (Proteaceae, Mitchel et al., 2017). Although a number of plant studies utilized enrichment strategies (Table 1), the majority applied lineage specific probes (e.g. genus or family) whereas our study uses probes potentially applicable to angiosperms as a whole. This angiosperm probe set is here implemented in a study of the genus *Aristolochia* (Aristolochiaceae, magnoliids), which is only distantly related (~140 Ma. of divergence time; Magallón et al., 2015) to any of the taxa used to design the probe set.

Another recent development addresses data subsampling and selecting loci according to potential phylogenetic information, because adding data itself will not necessarily provide accurate phylogenetic relationships if the rate of evolution is inappropriate (Townsend and López-Giráldez, 2010). Townsend (2007) developed a method called phylogenetic informativeness (PI) that provides a quantitative measure of the phylogenetic signal of a set of loci across a defined topology, by quantifying the probability of a character state change at a certain position of a tree, remaining subsequently unchanged. In order to estimate this, the ideal change rate based on an ultrametric tree, with branches being proportional to evolutionary units, is compared with the evolutionary changes across sites. Starting from a specific taxonomic sampling scheme, combining AHE and PI might allow selecting loci that match the requirements of the respective phylogenetic question.

Performance of these complementary methods is best explored in lineages where relationships have been difficult to resolve by using traditional molecular systematic approaches. In particular, we focus on two types of recalcitrant relationships: first, short and deep internodes where lineages split early in quick succession, and second, shallow level nodes at the species level (Townsend and Leuenberger, 2011). In order to be informative for a specific divergence event in a phylogenetic tree, a character must suffer a change in state in that specific time frame and remain unchanged along the complete branch length. Short and deep internodes are difficult to resolve for two reasons. Firstly, the probability that a character state change occurs in a short time period (short internode) is lower than after a long period (long internode). Secondly, a character has a higher probability to subsequently change its state in deep branches than in shallow ones. Therefore, a phylogenetic signal derived from a character state change during a short and deep internode is likely to be obscured by phylogenetic noise caused by subsequent character state changes along deep branches.

Table 1
List of recently published studies using enrichment strategies in plants.

Study	Studied Taxon	Taxon level	# of recovered loci
Fragoso-Martínez et al. (2017)	Salvia (Lamiaceae)	Genus	448 nuclear loci
Heyduk et al. (2016)	*Sabal* (Arecaceae)	Genus	133 nuclear loci
Mandel et al. (2015)	Asteraceae	Family	795 nuclear loci
Mitchell et al. (2017)	*Protea* (Proteaceae)	Genus	498 nuclear loci
Sass et al. (2016)	Zingiberales	Order	308 nuclear loci
Schmickl et al. (2015)	*Oxalis* (Oxalidaceae)	Genus	727 nuclear loci
Sousa et al. (2014)	*Medicago* (Fabaceae)	Genus	50 nuclear loci
Stephens et al. (2015a)	*Sarracenia* (Sarraceniaceae)	Genus	199 nuclear loci
Stephens et al. (2015b)	*Helianthus* (Asteraceae)	Genus	170 nuclear loci
Syring et al. (2016)	*Pinus albicaulis* (Pinaceae)	Species	4452 nuclear loci

Traditionally, plastid and nuclear ribosomal DNA markers are most commonly used in plant phylogenetics by applying a set of primers followed by PCR and Sanger sequencing (e.g. Small et al., 1998; Álvarez and Wendel, 2003; Shaw et al., 2005; 2007). Although the application of these traditional approaches revolutionized our current understanding of plant evolution, their phylogenetic informativeness is limited with respect to resolving difficult relationships. Plastid markers show a limited variability, as evolution of the plastome is relatively slow compared to that of the nuclear genome (Sang, 2002; Clegg et al., 1994). Consequently, plastid markers with limited variability are frequently unable to resolve nodes with very short branches (Granados Mendoza et al., 2015; Fragoso-Martínez et al., 2017; Mitchel et al., 2017). Furthermore, plastids are inherited maternally (with few exceptions) and therefore show only the maternal lineage when it comes to hybridization or introgression (Naumann et al., 2011; Zimmer and Wen, 2013). Nuclear ribosomal markers do not share the inheritance disadvantage, but potentially occur with a high copy number and might cause problems with regard to orthology establishment (Álvarez and Wendel, 2003; Song et al., 2012).

Comparatively little effort, measured in number of publications, has been spent on mining the information from the remaining nuclear genome. Nuclear single or low copy genes have been promoted as a solution to the problem of ambiguous relationships, as well as to overcome issues that might result from plastid loci or nuclear ribosomal loci. Nuclear single and low copy genes have mostly been used in studies were plastid and nuclear ribosomal DNA resulted in unresolved or unsupported relationships

(e.g. Désamoré et al., 2012; Guo et al., 2012; Marcussen et al., 2012; Zhang et al., 2012; Zimmer and Wen, 2013; Müller et al., 2015). However, it also became clear from these studies that the traditional approaches of primer design, testing and application require at least the modification of existing primers and settings, if not a reset.

Here we apply the presumably universal angiosperm enrichment probe set (Buddenhagen et al., 2016) on a small but taxonomically representative sampling of *Aristolochia* (Aristolochiaceae). This plant lineage is particularly suitable because it has been shown to have short-deep and shallow nodes that are difficult to resolve (Neinhuis et al., 2005; Wanke et al., 2006; Ohi-Toma et al., 2006). We consider Aristolochiaceae to consist of only two genera (*Thottea* and *Aristolochia*) and the families Asaraceae, Lactoridaceae and Hydnoraceae as successive sister groups forming the perianth-bearing Piperales (magnolliids; Naumann et al., 2013; Horner et al., 2015). This contrasts with the results of Christenhusz et al. (2015), which were based on an online survey. Unfortunately, the latter was followed by APG IV (The Angiosperm Phylogeny Group 2016).

Aristolochia comprises about 500 species (González et al., 2010; Wagner et al., 2014) from all continents except Antarctica. This genus is subdivided in three subgenera: *Siphisia*, *Pararistolochia* and *Aristolochia*, all of which are monophyletic according to morphological and molecular data (e.g. González and Stevenson, 2000, 2002; Wanke et al., 2006). *Aristolochia* subgenus *Aristolochia* is by far the most species-rich lineage and is further subdivided into five monophyletic groups. Relationships among these five lineages remain unknown, and current infra-subgeneric taxonomy is not reflecting natural relationships; therefore, the clades are treated here as operational taxonomic units (OTUs). The unresolved relationships between these OTUs potentially branched deep in time (Salomo et al., in review), and in quick succession, indicated by very short internodes (Neinhuis et al., 2005; Wanke et al., 2006; Ohi-Toma et al., 2006; Wanke et al., 2007; González et al., 2014). One of the five OTUs is predominantly distributed in Mexico plus a handful of South American species. Own unpublished plastid-based datasets provide little resolution and no support for species level relationships. The majority of the species of this OTU belong to subsection *Pentandrae* (Pfeifer, 1970; González et al., 2010; Paizanni Guillén et al., 2016). These species potentially diverged very recently, being a perfect case to test the potential of AHE on a shallow taxonomic level.

As a first test of the anchored hybrid enrichment (AHE) method in *Aristolochia*, we here focus on adding more data. However, future work will also test if taxon addition proximal to poorly resolved branches increases support without adding more data (Goldman, 1998; Geuten et al., 2007; Leaché and Rannala, 2010; Townsend and Lopez-Giraldez, 2010; San Mauro et al., 2012). The aims of this study are to (i) apply the presumably universal probe set for angiosperms to a lineage (*Aristolochia*) that is only distantly related to taxa used to design the probe set (monocots, eudicots, *Amborella*), (ii) evaluate the performance of AHE on recalcitrant deep and shallow nodes in *Aristolochia,* (iii) resolve the relationships between the five *Aristolochia* OTUs, as well as at species level within subsection *Pentandrae*, and (iv) discuss these relationships in the light of current systematics and biogeography.

2. Material and methods

2.1. Taxon sampling

Taxon sampling includes all three subgenera of *Aristolochia*: *Siphisia* with one species from each of its diversity centers in Asia (*A. hainanensis*) and Central and North America (*A. arborea*; González et al., 2014); the Old World subgenus *Pararistolochia* (*A. praevenosa*; Buchwalder et al., 2014); and at least one species from each of the five monophyletic OTUs previously identified for subgenus *Aristolochia* (Neinhuis et al., 2005; Wanke et al., 2006; Ohi-Toma et al., 2006). *Aristolochia labiata* belongs to a mostly South American clade (OTU 1). *Aristolochia grandiflora* represents a Neotropical clade formerly referred to as the "*A. grandiflora* complex" (OTU 2). *Aristolochia maxima* is one of a small number of species belonging to subseries *Thyrsicae* (OTU 3). *Aristolochia baetica* represents an Old World clade referred to as *Aristolochia* sensu strictu (OTU 4). Finally, the fifth clade consists of *A. lindneri* (*A. lindneri* complex) and subsection *Pentandrae* (OTU 5). The *A. lindneri* complex is a South American subtropical and temperate group, whereas species of subsection *Pentandrae* are virtually confined to Mexico with few species occurring in the Caribbean and the southern USA (González et al., 2014; Paizanni Guillén et al., 2016). Eleven species of this latter subsection are included here, representing about 23% of the total diversity. One species is represented by two accessions from different localities (*A. pentandra*). The genus *Thottea*, which is sister to *Aristolochia*, serves as outgroup (additional file 1).

2.2. Raw data acquisition

Genomic DNA was extracted using a CTAB isolation method with an additional RNase A treatment. DNA was extracted either from silica gel dried leaves collected in the field or from fresh leaves obtained from the Botanical Garden Dresden, Germany. Genomic DNA was loaded in 1.2% agarose gels, run in electrophoresis at 100V for 60 min and visualized under UV light. A minimum DNA concentration of 0.02 μg/μl and purity ca. 1.8 as measured by both 260/280 and 260/230 ratios was ensured and quantified with a Qubit® fluorometer.

The starting point to find suitable loci for AHE in flowering plants consisted of 959 nuclear genes present in single copy in each of *Arabidopsis thaliana*, *Populus trichocarpa*, *Vitis vinifera* and *Oryza sativa* genomes (Duarte et al., 2010). 3050 exons from these genes were found in *Arabidopsis thaliana* to fit a minimum threshold size needed for enrichment (Buddenhagen et al., 2016). 1721 of the 3050 *Arabidopsis thaliana* exons had a similarity $\geq$ 55% with those of *O. sativa* and orthologous regions of these 1721 exons were identified in 29 complete and additional nine low-coverage angiosperm genomes. Finally, 499 exons with $\leq$ 1.2 average copy number and $\geq$ 85% of occurrence among the genomes were further selected. A custom Agilent SureSelect Target Enrichment Kit (Angiosperm v1 kit; Agilent Technologies, XT) was designed based on these 499 exons plus 18 selenium-tolerance genes (Buddenhagen et al., 2016). This kit was successfully tested on the aforementioned species and additional angiosperms (Buddenhagen et al., 2016).

Anchored hybrid enrichment data were collected following the methods of Buddenhagen et al. (2016) through the Center for Anchored Phylogenomics at the Florida State University (www.anchoredphylogeny.com). Genomic DNA samples were sonicated to a fragment size of ~300–800 bp using a Covaris E220 Focused-ultrasonicator with Covaris microTUBES. Library preparation and indexing were performed on a Beckman-Coulter Biomek FXp liquid-handling robot following a modified version of Meyer and Kircher's (2010) protocol. One important modification to this protocol is a size-selection step after blunt-end repair using SPRIselect beads (Beckman-Coulter Inc.; 0.9x ratio of bead to sample volume). All libraries were then pooled at equal quantities, and enrichments were performed using the Angiosperm v1 kit (Buddenhagen et al., 2016). After enrichment, a post-enrichment PCR was performed to ensure sufficient sample volume for sequencing using IS5 and IS6 reamplification primers (Meyer and Kircher, 2010), followed by

pooling the samples and sequencing one lane PE150 on a Illumina HiSeq2000. Sequencing was performed in the Translational Science Laboratory in the College of Medicine at Florida State University.

2.3. Raw data processing

Sequencing outputs were processed using the CASAVA v1.8 pipeline provided with the Illumina HiSeq software (reads were quality filtered using the high-chastity setting). Quality filtered reads were demultiplexed using 21 out of 96 in-house-developed indexes (8 bp each; indexes at least 2 bases different). If index sequence did not match one of the 21 expected index sequences, the associated read was discarded. Pair-read merging was performed following Rokyta et al. (2012) in order to improve read accuracy and length. For read assembly, a divergent assembly followed by a de novo approach was applied as in Prum et al. (2015) and Buddenhagen et al. (2016). During divergent assembly, reads were mapped to probe region sequences using a selection of taxa employed for probe design, which included *Arabidopsis thaliana* (Brassicales: Brassicaceae), *Billbergia nutans* (Poales: Bromeliaceae), and *Carex lurida* (Poales: Cyperaceae). This approach involves applying divergent references to initiate the assembly of each locus in the conserved probe region, and then the assembled reads are used as references for extending assemblies into the flanking regions. Because all downstream analyses rely on accurate assemblies, several control mechanisms and refinement steps were included. For instance, sites with <10 called bases were marked as N in the final assembly and nucleotide ambiguity code was used for sites with unambiguous consensus bases. Additionally, consensus sequences produced with fewer than 30 reads were removed from further analysis. Furthermore, all assemblies were classified by quality, and if necessary, manually inspected and adjusted. After these adjustments, assemblies were considered as final and merged into alignments for each locus individually. Prior to further analyses, we generated phased alignments, as in Pyron et al. (2016), that are different with respect to the numbers of "assumed" ploidy level. Firstly, diploidy was assumed for all taxa as is usually done. Secondly, we tested "assumed" tetraploidy because we obtained odd phylogenetic results in initial analyses (non-monophyly of species, see results) for: (1) all accessions of this study; and (2) all accessions of OTU 5, while assuming diploidy for the remaining accessions.

2.4. Concatenated and coalescence-based phylogenetic analyses

Maximum Likelihood (ML) analyses were performed with a partitioned concatenated data matrix of all assembled loci using the "rapid Bootstrap and search for best-scoring ML tree" algorithm in RAxML HPC2 version 8.2.9 (Stamatakis, 2014) on XSEDE in the CIPRES Science Gateway (Miller et al., 2010). Gene partitions subsets and their best-fitting models of evolution were identified with PartitionFinder v. 1.1.1 (Lanfear et al., 2012); however, as recommended in the RAxML manual, the GTR + G model was implemented during the tree search. Bootstrap resampling of the data was applied to assess statistical support for nodes (1000 replicates). FigTree version 1.4.0 (Rambaut, 2009) was used for tree compilation and edition.

Coalescence-based methods were performed to compare the results to the concatenated analyses using ASTRAL (Mirarab et al., 2014), a program for species tree estimation that maximizes the number of quartet trees shared among source unrooted gene trees. ASTRAL analyses were run on an input file containing maximum likelihood (ML) gene trees for each individual locus. Gene trees were estimated using a workflow created in Geneious® 9.0.2 that runs in series ML analyses on all loci, applying the GTR + G model and the "rapid bootstrapping and search for the best-scoring ML tree" algorithm. Branch support was estimated through ASTRAL local posterior probabilities (LPP; main topology), which – when gene tree estimation error is low – has been proposed to be a more reliable measure for branch support compared to multi-locus bootstrapping (Sayyari and Mirarab, 2016).

2.5. Informativeness profiling and data filtering

As suggested by Fragoso-Martínez et al. (2017), net phylogenetic informativeness profiles were examined for the partitioned matrices and trees resulting from 1) the concatenated analysis of the matrix including all recovered loci (under partial tetraploidy), and 2) the concatenated analyses of four filtered matrices (assuming partial tetraploidy), excluding each site with substitution rates values higher than 20, 15, 10 and 5. Lists of characters to exclude for each of these four filtered matrices were produced using the script described in Fragoso-Martínez et al. (2017) and filtered matrices were written with the command -E of RAxML. Concatenated analyses of the filtered matrices were performed as described in the previous section for the complete data set.

Net phylogenetic informativeness profiles were calculated with the PhyDesign online program (López-Giráldez and Townsend, 2010). The trees were converted to ultrametric using the program TreeEdit v1.0a10 (Rambaut, 2002), applying the non-parametric rate smoothing method (Sanderson, 1997). Ultrametric trees were subsequently rescaled so that tips were assigned to time 0 and the root to time 1. The program HyPhy (Pond et al., 2005) implemented in PhyDesign was selected for calculating sites' substitution rates, applying a GTR model of substitution.

3. Results

3.1. Phylogenomic matrices

Important raw data parameters for all samples are provided in the additional file 2. In total 231 and 233 out of 517 loci originally contained in the enrichment kit were recovered for the complete diploid and partial tetraploid assembly schemes, respectively. Total aligned length of the complete diploid matrix and partial tetraploidy matrix was 154,756 bp and 160,150 bp, respectively. Length ranges of individual loci alignments were 158–1661 bp (average 670 bp) for the complete diploidy and 135–1681 bp (average 687 bp) for the partial tetraploidy assemblies. Nearly 38% of the alignments in both assembly schemes had full representation of the taxa, the remaining lacking one or more accessions. *Aristolochia pentandra* (accession I9826) was the species for which more loci could be retrieved (complete diploidy: 230 loci; partial tetraploidy: 232 loci), whereas *A. grandiflora* showed the lowest loci coverage (complete diploidy: 154; partial tetraploidy: 153).

3.2. Relationships and potential polyploidization

The concatenated analysis (Fig. 1A; diploidy) revealed a well-supported sister relationship of subgenus *Siphisia* (*A. hainanensis* and *A. arborea*) to the remaining two subgenera *Pararistolochia* and *Aristolochia* (BS $\geq$ 85). Two main lineages were recovered within subgenus *Aristolochia*. The first consists of a monophyletic OTU 5. Within the second lineage, a grade is recovered with *A. baetica* (OTU 4) branching first, followed by *A. labiata* (OTU 1), and *A. grandiflora* (OTU 2) and *A. maxima* (OTU 3) as sister to each other. All nodes, except the sister group relationship of OTU 1 to OTUs 2 and 3, receive bootstrap supports $\geq$85 (Fig. 1A). However, subsection *Pentandrae* is paraphyletic because *A. lindneri* is nested within it. Additionally, the majority of its species are recovered as non-monophyletic (Fig. 1A). These findings are similarly recovered in

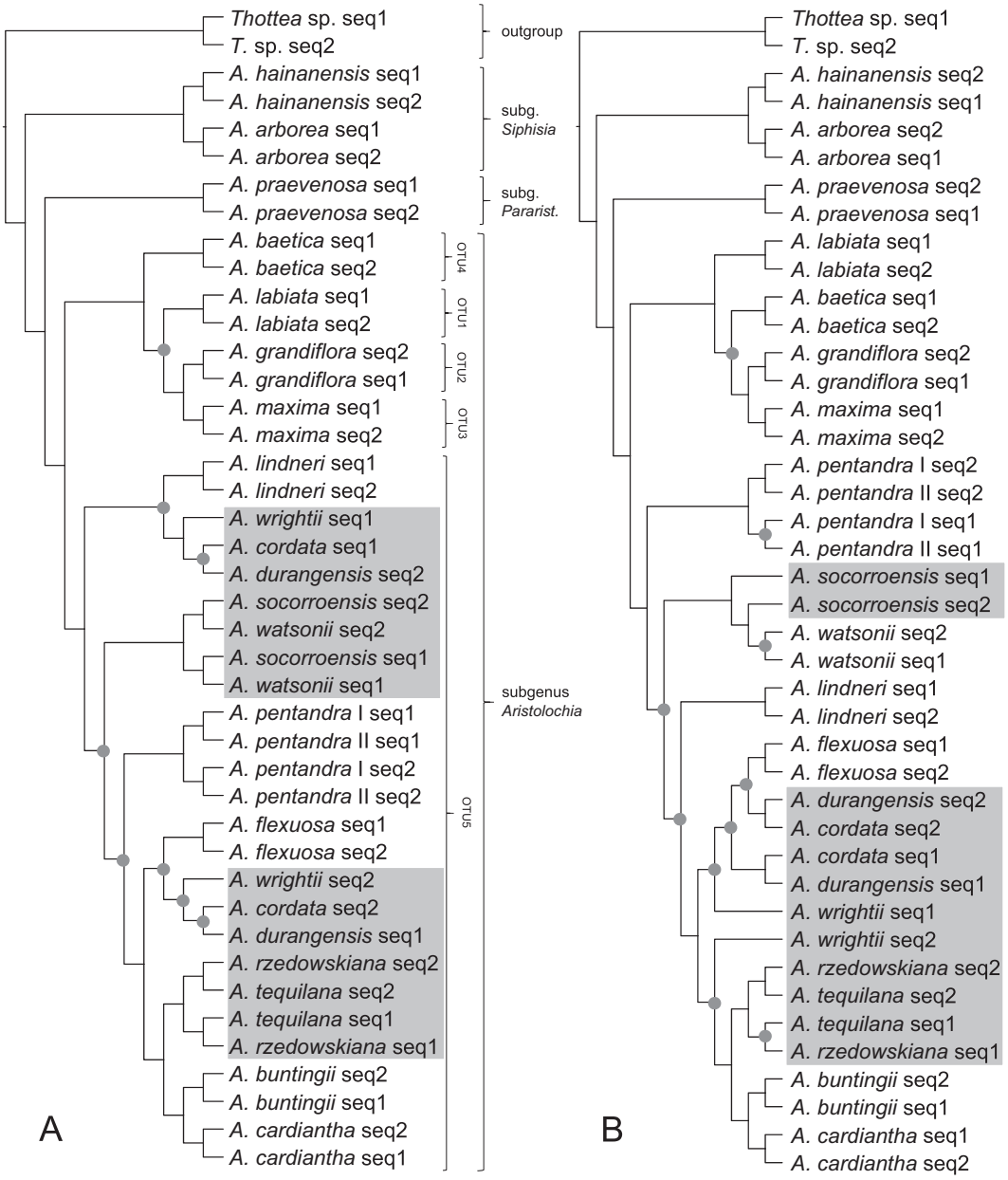

Fig. 1. Assumed diploidy during analyses results in non-monophyly of species. Trees resulting from the analysis of 231 loci recovered after the entire bioinformatic pipeline, assuming diploidy for all the species: (a) maximum likelihood trees obtained from the analysis of the concatenated dataset, (b) species tree resulting from the coalescence analysis approach. Subgeneric classification as well as affiliation of species to operational taxonomic units (OTUs) applies to both trees. Grey boxes indicate non-monophyletic species (considering their two alleles), whereas grey dots denote nodes with low support (bootstrap BS < 85; local posterior probabilities LPP < 0.85).

the coalescence-based approach (Fig. 1B), with differences present only at unsupported nodes in both analyses. Consistently, ASTRAL recovers nearly the same species of OTU 5 as non-monophyletic.

As a consequence of the congruence of both approaches with respect to the unexpected non-monophyly at species level, we hypothesized that the evolution of the organisms is different from the diploidy we assumed in the concatenated and coalescence-based analyses. As a result, we performed additional analyses assuming tetraploidy (Fig. 2). Independently from whether we assumed tetraploidy for all accessions (data not shown) or for OTU 5 only, identical results with respect to relationships and the monophyly of accessions were obtained. Both concatenated (Fig. 2A) and coalescence (Fig. 2B) analyses resulted in the species of OTU 5 being monophyletic. However, both approaches differ with respect to the position of *A. lindneri* and *A. pentandra,* with the first one being sister to all remaining species of OTU 5 in the concatenated analysis.

In general, very short branches are observed within OTU 5 (Fig. 3). Given the results of the concatenated analysis assuming tetraploidy (Fig. 2A), a grade is recovered, consisting of a clade formed by *A. watsonii* and *A. socorroensis,* a clade consisting of the two *A. pentandra* accessions, followed by *A. wrigthii,* and a clade containing all the remaining species. This last one consists of a clade with (*A. flexuosa* (*A. durangensis* + *A. cordata*) being sister to a clade of the following species ((*A. buntingii* + *A. cardiantha*) (*A. rzedowskiana* + *A. tequilana*)). In contrast to the concatenated analysis, the coalescence analysis flips the position of *A. pentandra* and *A. lindneri* (Fig. 2B). Within subsection *Pentandrae* and depending on the analyzed data set (Figs. 2–4), two to three nodes on species level remain unsupported (BS $\leq$ 0.85, LPP $\leq$ 0.85).

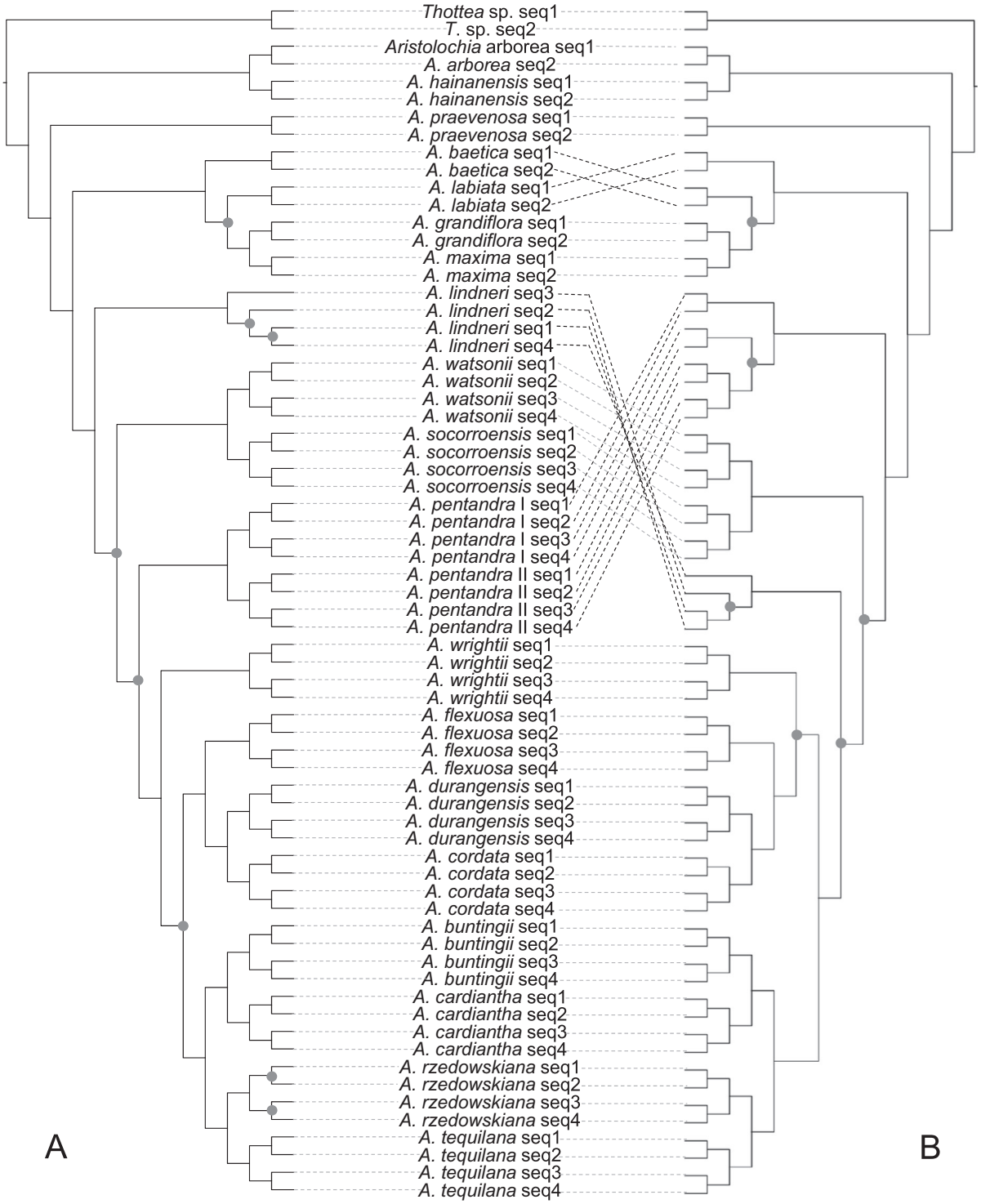

Fig. 2. Assumed tetraploidy during analyses results in monophyly of species. Trees (left: concatenated analysis, right: coalescence analysis) resulting from the analysis of 233 loci recovered after the entire bioinformatic pipeline assuming tetraploidy for a subset of species and diploidy for the rest. Dotted red lines indicate incongruence of species relationships between both approaches. Grey dots denote nodes with low support (bootstrap BS < 85; local posterior probabilities LPP < 0.85).

3.3. Phylogenetic informativeness and data filtering results

Given the aforementioned results, we assumed that some accessions are more likely tetraploid than diploid. Therefore, all subsequent PI analyses were performed on the partial tetraploid datasets and reference trees obtained from their analyses. When considering the complete (unfiltered) data set, a considerable variation in net PI across all loci, with maximum net PI of individual loci ranging from 10.86 to 668.37, was observed (Fig. 3). Maximum net PI of individual loci was reached at reference tree times (t) ranging from 0.01 to 0.57. The majority of the loci presented a steady increase in their net PI until reaching a maximum, followed by a subsequent gradual decrease. About 65 loci peaked before the time of divergence of *A. lindneri* and subsection *Pentandrae* ($t = 0.11$); among them 29 showed a sharp increase in net PI peaking at or more recently than the shallowest species pair divergence

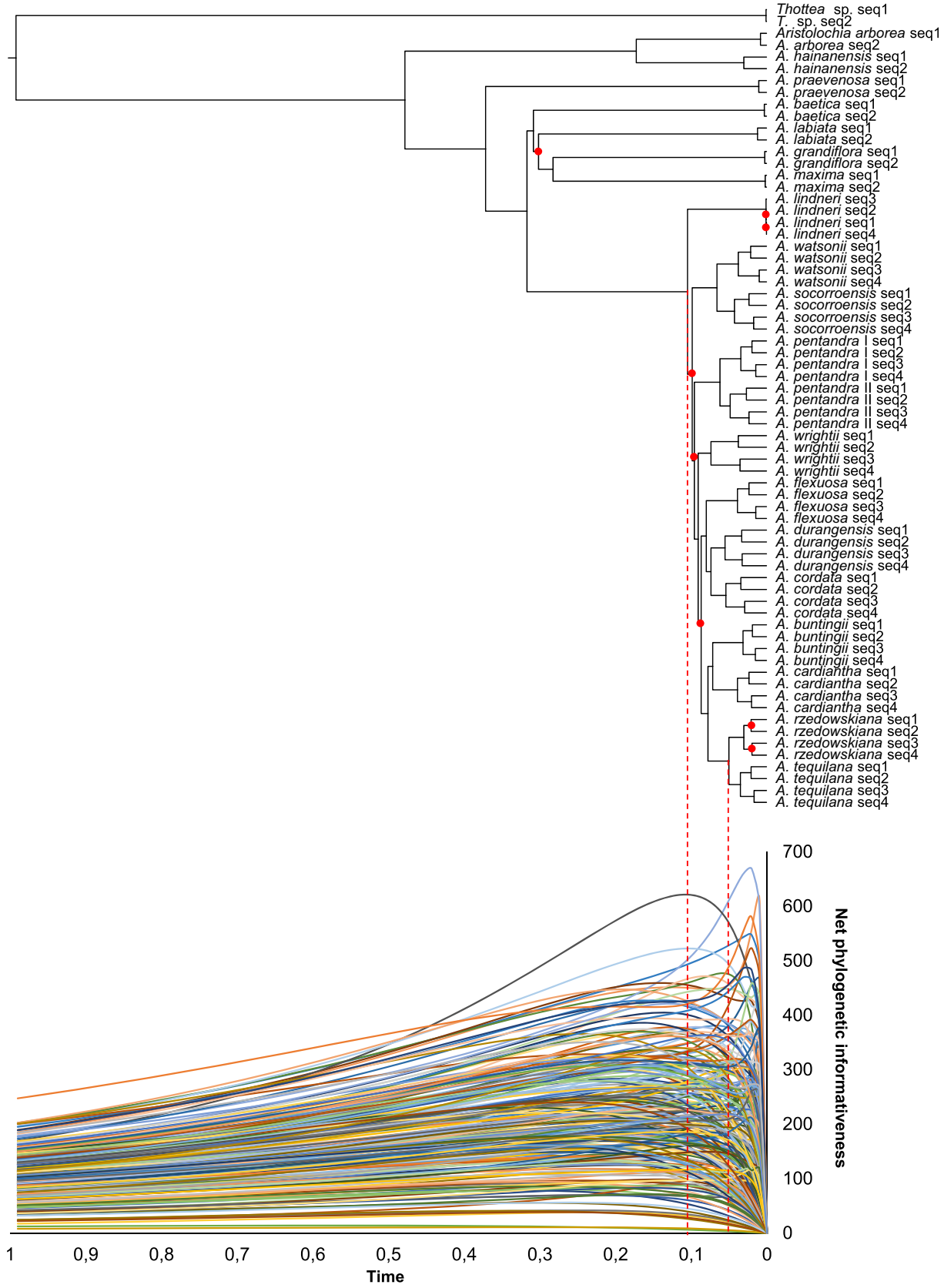

Fig. 3. Phylogenetic informativeness profiles identify sites with unusually high substitution rates. Net phylogenetic informativeness profiles of loci from the unfiltered dataset are shown at the bottom of the figure. An arbitrary timescale is shown on an ultrametric tree modified from the concatenated dataset analysis results, with the root at time 1 and the tips at time 0 (top of figure). Red dotted lines indicate times for the shallowest species divergence and the divergence of OTU 5 (*A. lindneri* and subsection *Pentandrae*). Red dots denote nodes with low bootstrap support (BS < 85).

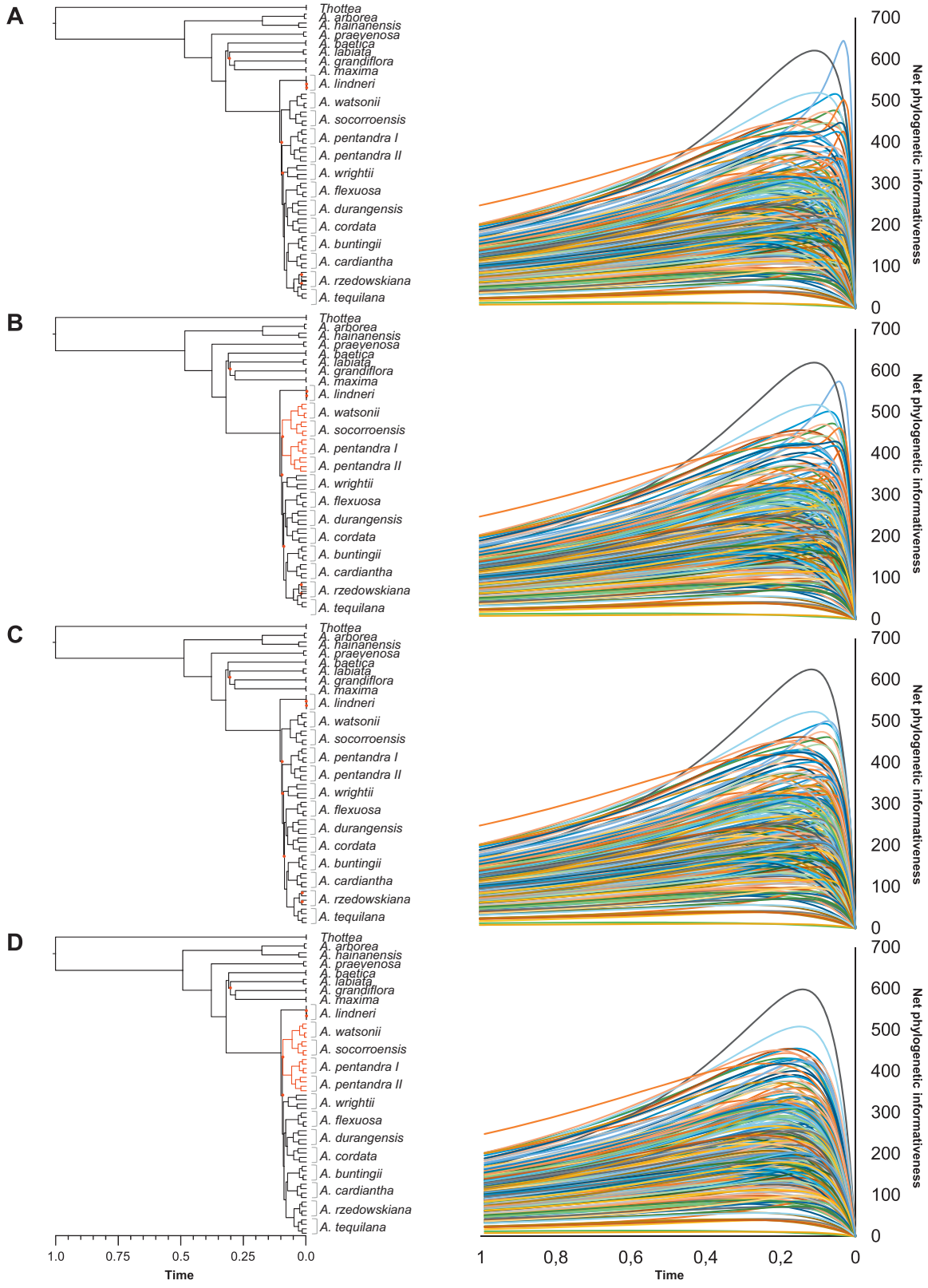

Fig. 4. Excluding sites with high substitution rates results in better supported relationships. Net phylogenetic informativeness profiles and corresponding trees from the four data filtering schemes are depicted, excluding sites with substitution rates higher than 20 (A), 15 (B), 10 (C) and 5 (D). Trees are ultrametric and reconstructed with RAxML using the concatenated dataset. Clades in red denote different topologies from that obtained from the unfiltered data matrix (Fig. 1). Red dots denote nodes with low bootstrap support (BS < 85).

(ca. $t = 0.05$; between *A. rzedowskiana* and *A. tequilana*), followed by a sudden decrease.

Analyses from the filtered matrices resulted in identical topologies to that of the non-filtered matrix, except for the filtering

schemes excluding sites with substitution rate values higher than 15 and 5, where the clade (*A. pentandra* (*A. watsonii* + *A. socorroensis*)) was recovered as sister to a clade containing the remaining species from subsection *Pentandrae*. The number of weakly supported nodes (eight) remained the same as that of the complete matrix for the filtering schemes excluding sites with substitution rate values higher than 15 and 10, whereas the filtering schemes excluding rates above 20 and 5 had only seven and five low supported nodes, respectively. Informativeness profiles showed a gradual attenuation of the curves' peaks as more data were excluded, so that maximum net PI values ranges narrowed to 10.85–643.66, 10.78–619.12, 10.99–624.54 and 11.08–598.16 for each of the filtering schemes from the more to the less inclusive, respectively. The number of loci peaking before the time of divergence of *A. lindneri* and subsection *Pentrandrae* ($t = 0.11$) decreased towards the strictest filtering schemes, with 66, 63, 46 and 1 loci, respectively. Similarly, the number of loci peaking close to zero and more recently than the divergence time between *A. rzedowskiana* and *A. tequilana* decreased as more data were excluded, with 23, 10, one and zero loci from the less to the more strict filtering scheme.

4. Discussion

4.1. Performance of the AHE approach

The results of AHE demonstrate the power of this strategy to unravel evolutionary relationships as well as processes that would have been difficult to resolve using traditional approaches. AHE allowed gathering simultaneously a tremendous amount of highly informative, orthologous single or low-copy nuclear markers (here up to 233 loci, minimum length 135 bp), providing sufficient phylogenetic information to resolve short branches that occurred early in the evolution of *Aristolochia*, as well as those of recent diversification. The AHE plant probe set avoids the problem of lacking large lineage-specific genome or transcriptome scale data for marker characterization or primer design. The probe set has been designed based on genomes of 31 plant genera; the *Amborella trichopoda* genome (Amborella Genome Project, 2013) was the only one representing the 'extant early diverging angiosperms', whereas the other genomes belonged to eudicot or monocot species. Consequently, a large evolutionary distance exists between the taxa used for probe design and our study group *Aristolochia* (~140 Ma., Magallón et al., 2015). Given this evolutionary distance and the resulting sequence divergence, this study acts as proof of principle that the probe set is most likely applicable to any angiosperm group, although the number of recovered loci might vary. For *Aristolochia* loci number was significantly lower than in *Salvia* (448 loci, Fragoso-Martínez et al., 2017) and *Protea* (498 loci, Mitchel et al., 2017). The potential to enrich hundreds of loci in almost every angiosperm group is based upon the probe design, which is chosen to enrich conserved regions that are likely available in all angiosperms (Buddenhagen et al., 2016). More variable regions, as well as conserved coding regions that flank the probe regions, provide in sum the needed genetic variability to resolve phylogenetic questions at a wide range of evolutionary time scales. Also, the locus length could be adjusted by selecting different size bands during size selection process (Lemmon et al., 2012), to optimize a priori the resulting variability. With an increase in locus length, more of the variable adjacent region of each locus would get sequenced (e.g. introns) at the cost of decreasing the number of recovered loci. However, our study used the same threshold as the *Salvia* and *Protea* studies (Fragoso-Martínez et al., 2017; Mitchel et al., 2017) and hence, the recovered loci number is exclusively resulting from the genetic distance.

Although PCR-based approaches can potentially be applied to deep and shallow nodes, they likely require individual primers when applied to other more distantly related lineages. Along with AHE, other recent approaches have resulted in large phylogenomic datasets for plants and contributed to significant progress in resolving difficult relationships. Using enrichment strategies, the number of recovered loci is largely variable, ranging from 50 (Sousa et al., 2014) to 4452 (Syring et al., 2016; Table 1). However, one should keep in mind that the probe sets for these studies have specifically been designed for the respective taxon. Eaton and Ree (2013) used a RAD approach to obtain 4837 loci for *Pedicularis* section *Cyathophora* (Orobanchaceae) obtaining a highly supported phylogenetic hypothesis. However, RAD sequencing strategies are limited to shallow level phylogenetics and thus become less useful with larger genetic distance of target lineages (e.g. Rubin et al., 2012). Gostel et al. (2015) used a microfluidic PCR-based target enrichment approach on the recently radiated Madagascan endemic genus *Commiphora* (Burseraceae) obtaining, based on 49 recovered loci, the first supported shallow level *Commiphora* phylogeny. Focusing on very deep level relationships and using non-enrichment approaches, Ruhfel et al. (2014) analyzed 78 plastid genes for 360 species of green plants (*Viridiplantae*) and Wickett et al. (2015) analyzed 852 nuclear genes (transcriptome) to reconstruct the origin and early diversification of land plants. Although these and other studies (e.g. Blaimer et al., 2015) have applied different high-throughput approaches, they all concluded that such strategies outperform traditional approaches with respect to cost-benefit ratio and the ability to answer the respective landmark questions in evolutionary biology.

The performance is influenced by the average locus length in addition to the number of recovered loci. Although the average loci length in this study (670–687 bp) is lower than what was recovered in previous tests performed by Buddenhagen et al. (2016), using an angiosperm wide sampling, it is significantly higher than in *Protea* (551 bp, Mitchel et al., 2017) and only slightly lower than in *Salvia* (704 bp, Fragoso-Martínez et al., 2017). When shallow-level taxon sets are aligned loci are expected to be longer compared to alignments of more distantly related taxa because alignments can be extended to non-codding flanking regions such as introns and integenic spacers. However, it is likely that the rather low percentage of reads mapping to the respective loci in *Aristolochia* (0.91–13.81%) also influenced the average loci length negatively. Buddenhagen et al. (2016) retrieved 20–75% reads mapping to individual loci.

Concisely, AHE for plants is still under development. The quantity of recovered loci for specific taxa might increase, potentially also resulting in longer loci contigs, when probes are designed based on newly sequenced genomes from critical angiosperm lineages and when pilot studies like this one are used to improve the probe set. A follow-up study, which will test taxon addition proximal to still ambiguous nodes, will also make use of a probe kit containing *Aristolochia* to better represent the magnoliids.

4.2. Phylogenomic hypotheses and current systematics

Filtering out sites with unusually high substitution rates decreased the number of loci peaking more recently than the shallowest species divergence or the divergence time of OTU 5 (*A. lindneri* plus subsection *Pentandrae*) to nearly zero, suggesting that little or no phylogenetic noise could be affecting our analysis from the strictest filtering scheme (Townsend, 2007; Granados Mendoza et al., 2013). Similarly, as more data were filtered (Fig. 4) the number of unsupported nodes decreased (from 8 to 5), indicating that some of the unsupported nodes can be attributed to potential phylogenetic noise introduced by a subset of "fast-evolving" sites. Conversely, maximum net phylogenetic informativeness ranges

narrowed as more data were excluded, which could also have a negative impact on the ability of the filtered datasets in supporting difficult relationships (e.g. short internodes). Specifically excluding unusually "fast-evolving" sites had a positive impact, increasing the number of highly-supported nodes and allowing us to produce a dataset with minimal phylogenetic noise. It can thus be concluded that the loci obtained via the AHE plant probe set perform best at subgeneric level for *Aristolochia* but reach their limit at species level, particularly for closely related and recently diverged species given the here applied taxon set. In future studies focusing on subgenus *Pentandrae* we will use outgroups that are closely related, allowing constructing alignments at shallow scales, and make use of more variable regions at the flanks of the loci.

Previous molecular phylogenetic relationships between subgenera *Siphisia*, *Pararistolochia* and *Aristolochia* (Neinhuis et al., 2005; Ohi-Toma et al., 2006; Wanke et al., 2006), which are also supported by morphological synapomorphies (González and Stevenson, 2002), are confirmed by our analyses. Given that earlier studies (e.g. Ohi-Toma et al., 2006; Wanke et al., 2006; González et al., 2010) showed the monophyly of all OTUs, the limited sampling in this pilot study is justified.

Although phylogenetic hypotheses from concatenated and coalescence approaches (Figs. 1 and 2) reconstruct slightly different relationships within subgenus *Aristolochia*, these differences are unsupported and do not alter the main result that section *Gymnolobus* Duchartre (OTUs 1, 2, 3, and 5) is paraphyletic with respect to OTU 4 (*Aristolochia baetica,* representative of the Old World section *Diplolobus* Duchartre). As OTUs 1, 2, 3, and 5 are all occurring in the New World, the subdivision of *Aristolochia* subgenus *Aristolochia* into one Old and one New World clade could not be confirmed, which is in contrast to an earlier phylogenetic analysis based on morphological data (González and Stevenson 2002). While a single subtending bract of the individual flowers and a complex syrinx support a closer relationship between the New World OTUs 2, 3 and 5 (*A. grandiflora* complex, subseries *Thyrsicae*, and subsection *Pentandrae* including *A. lindneri*), the lack of an abscission zone and pollen characters indicate a closer relationship of OTU 5 to OTU 1 (González and Stevenson, 2002). None of these relationships are supported by our results, because OTU 5 is always sister to a clade consisting of the remaining OTUs of *Aristolochia* subgenus *Aristolochia*. In depth morphological analyses, keeping the molecular results recovered here in mind, might lead to the discovery of new or differently interpreted morphological synapomorphies supporting these relationships.

OTU 5 consists of subsection *Pentandrae* and the "*Aristolochia lindneri* group", a South American subtropical and temperate lineage. Their sister group relationship that is only recovered in the tetraploid concatenated analysis is supported by morphological characters, e.g. loculicidal, basipetal capsules, bracteate flowers and dark brown to black triangular flattened seeds, which have a prominulous raphe and lack wings (González, 1999; González and Stevenson, 2002; González and Rudall, 2003). However, these morphological characters of OTU 5 are also present in OTU 2 (González 1999; González and Stevenson, 2002) and are thus not synapomorphies.

4.3. Species-level evolution within subsection Pentandrae

Currently, a published molecular phylogenetic hypothesis for subsection *Pentandrae* does not exist. Our own unpublished plastid-based datasets virtually provide very little resolution and no support for species level relationships. Short branch lengths indicate that this lineage, which is nearly completely endemic to Mexico, either diverged very recently or shows an extremely low mutational rate (Figs. 3 and 4). The monophyly of the subsection *Pentandrae* is supported based on the most comprehensive dataset to date (concatenate, Fig. 2A and 4), although the coalescence-based approach suggests the inclusion of *A. lindneri* in this subsection (Fig. 2B). The future inclusion of additional species from the *A. lindneri* complex, such as *A. burelae*, *A. urbaniana, A. lozaniana,* and *A. stuckertii* (González et al., 2010), might allow one to clarify whether the *A. lindneri* complex is monophyletic and sister to subsection *Pentandrae* or nested within.

The analyses resulting from assumed diploidy resulted in the non-monophyly of the majority of subsection *Pentandrae* species (Fig. 1), whereas all species were resolved as monophyletic when tetraploidy was assumed (Fig. 2). This finding was consistently recovered in concatenated and coalescence-based approaches and independent from the filtering threshold of the phylogenetic informativeness (Figs. 2 and 4), suggesting that polyploidization likely played a major role during the evolution of this lineage. However, *A. lindneri* has been reported to have the smallest genome of all *Aristolochia* species yet studied (0.67 pg/2C, Bliss et al., 2013). Subsection *Pentandrae* species have not yet been investigated for chromosome numbers or genome size and we might have assumed more species than necessary to be tetraploid as indicated by species of subsection *Pentandrae* that are also monophyletic when assuming diploidy for them. Additionally, it is yet impossible to judge if polyploidization took place only once or multiple times independently. However, it is important to state that the AHE approach allowed us to obtain an indication for polyploidization in subsection *Pentandrae,* which would have been more difficult using traditional sequencing methods.

In general, *Aristolochia pentandra* shows the broadest range of phenotypic variability with respect to vegetative characters, as well as the largest area of distribution of all *Pentandrae* species (Pfeifer, 1970). Possibly the most recent common ancestor of subsection *Pentandrae* also showed a large distribution area and featured a high degree of plasticity. The distribution area of extant taxa undoubtedly experienced fragmentation and isolation of habitats through the formation of larger mountain ranges that are well-known to promote diversification (Sierra Madre Occidental, trans-Mexican Volcanic Belt, Sierra Madre del Sur; Rzedowski, 1991; Ferrusquía-Villafranca, 1993; Delgadillo et al., 2003; Morrone, 2005; Torres and Luna, 2006). A similar pattern occurs within some of the most species-rich angiosperm families in Mexico, such as Asteraceae and Poaceae (Delgadillo et al., 2003), Fabaceae (Sousa and Delgado, 1993), and Lamiaceae (Ramamoorthy and Elliott, 1993). Leaf morphology within subsection *Pentandrae* supports the sister-group relationship of *A. socorroensis* and *A. watsonii*, both characterized by sagittate (arrow-shaped) leaves, whereas *A. pentandra* has a variety of leaves shapes (cordiform, hastate to trilobate) and the leaves of *A. wrightii* are hastate. The remaining species, *A. cordata*, *A. flexuosa*, *A. tequilana, A. cardiantha* and *A. buntingii*, all have small distribution areas except for *A. rzedowskiana,* being distributed in the Mexican states of Jalisco and Colima, and *A. durangensis* distributed in Durango and Nayarit (Paizanni Guillén et al., 2016).

5. Conclusion

We confirm the technical applicability of the plant AHE enrichment probe set to the angiosperm genus *Aristolochia* (magnoliids), although they are only distantly related to the eudicot and monocot taxa, as well as *Amborella trichopoda*, which were used to design this enrichment probe set. This study acts as proof of principle that the plant AHE method and the respective probe set is potentially useful for any angiosperm lineage. Given its universality, this approach could be highly cost-effective compared to non-universal probe sets specifically designed for individual genera or families. Furthermore, we showed that the enrichment probe set

provides enough informative loci to successfully resolve recalcitrant phylogenetic relationships within *Aristolochia*. However, the efficiency of the plant AHE method could further be improved by optimizing the probe set by adding taxa from yet underrepresented lineages splitting deep in time, i.e. lineages between *Amborella*, the eudicot and monocot radiation. Such efforts are already underway. Additionally, the AHE raw data provide a first hint of potential polyploidization involved in the diversification of closely related species in the absence of chromosome counts or flow cytometric results. However, there is still room for improvement for the analysis of datasets with phased alleles, under both concatenated and coalescence based approaches, but we hope that this pilot study will motivate further developments of this analytically complex area. For *Aristolochia,* and specifically for *Aristolochia* subgenus *Aristolochia*, this study resolves for the first time relationships between previously recovered monophyletic groups (OTUs) and renders section *Diplolobus* paraphyletic. However, the Neotropical groups of *Aristolochia* subgenus *Aristolochia* do not form a monophyletic entity. Increased sampling as well as an improved AHE probe set will allow addressing the remaining questions with respect to relationships prior to a taxonomic revision of the genus *Aristolochia*.

Acknowledgements

Financial support for this study comes from a bilateral cooperation project between DAAD and CONACYT (project number 204693), supporting the exchange of researchers and students between Mexico and Germany. APG is particularly grateful to the authorities of the Universidad de Guadalajara, Dept. de Ecología y Recursos Naturales del Centro Universitario de la Costa Sur. CGM thanks the Dirección General de Asuntos del Personal Académico (DGAPA-UNAM, 2014–2016) for two postdoctoral grants. We thank the Mexican authorities (Secretaría de Medio Ambiente y Recursos Naturales; SEMARNAT; permit numbers SGPA/DGGFS/712/1643/13, and SGPA/DGGFS/712/1613/14) for permission to collect material.

References

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.

Blaimer, B.B., Brady, S.G., Schultz, T.R., Lloyd, M.W., Fisher, B.L., Ward, P.S., 2015. Phylogenomic methods outperform traditional multi-locus approaches in resolving deep evolutionary history: a case study of formicine ants. BMC Evol. Biol. 15 (1), 271.

Bliss, B.J., Wanke, S., Barakat, A., Ayyampalayam, S., Wickett, N., Wall, P.K., Jiao, Y., Landherr, L., Ralph, P.E., Hu, Y., Neinhuis, C., Leebens-Mack, J., Arumuganathan, K., Clifton, S.W., Maximova, S.N., Ma, H., dePamphilis, C.W., 2013. Characterization of the basal angiosperm *Aristolochia fimbriata*: a potential experimental system for genetic studies. BMC Plant Biol. 13, 13.

Brandley, M.C., Bragg, J.G., Singhal, S., Chapple, D.G., Jennings, C.K., Lemmon, A.R., Lemmon, E.M., Thompson, M.B., Moritz, C., 2015. Evaluating the performance of anchored hybrid enrichment at the tips of the tree of life: a phylogenetic analysis of Australian *Eugongylus* group scincid lizards. BMC Evol. Biol. 15, 62. http://dx.doi.org/10.1186/s12862-015-0318-0.

Buddenhagen, C., Lemmon, A.R., Lemmon E.M., Bruhl, J., Cappa, J., Clement, W.L., Donoghue, M., Edwards, E.J., Hipp, A.L., Kortyna, M., Mitchell, N., Moore, A., Prychid, C.J., Segovia-Salcedo, M.C., Simmons, M.P., Soltis, P.S., Wanke, S., Mast, A., 2016. Anchored phylogenomics of angiosperms I: Assessing the robustness of phylogenetic estimates. http://dx.doi.org/10.1101/086298 (availble at bioRxiv.org).

Cai, G., Li, H., Lu, Y., Huang, X., Lee, J., Müller, P., Ji, Y., Liang, S., 2012. Accuracy of RNA-Seq and its dependence on sequencing depth. BMC Bioinformatics 13 (Suppl 13), S5. http://dx.doi.org/10.1186/1471-2105-13-S13-S5.

Carstens, B., Lemmon, A.R., Lemmon, E.M., 2012. The promises and pitfalls of next-generation sequencing data in phylogeography. Syst. Biol. 61, 713–715. http://dx.doi.org/10.1093/sysbio/sys050.

Chen, X., Lemmon, A.R., Lemmon, E.M., Pyron, R.A., Burbrink, F.T., 2017. Using phylogenomics to understand the link between biogeographic origins and regional diversification in ratsnakes. Mol. Phylogenet. Evol. 111, 206–218.

Christenhusz, M.J.M., Vorontsova, M.S., Fay, M.F., Chase, M.W., 2015. Results from an online survey of family delimitation in angiosperms and ferns: recommendations to the Angiosperm Phylogeny Group for thorny problems in plant classification. Bot. J. Linn. Soc. 178, 501–528.

Clegg, M.T., Gaut, B.S., Learn, G.H., Morton, B.R., 1994. Rates and patterns of chloroplast DNA evolution. PNAS 91, 6795–6801.

Delgadillo, C., Villaseñor, J.L., Dávila, P., 2003. Endemism in the Mexican flora: a comparative study in three plant groups. Ann. Mo. Bot. Gard. 90, 25–34.

Désamoré, A., Laenen, B., González-Mancebo, J.M., Jaén Molina, R., Bystriakova, N., Martinez-Klimova, E., Carine, M.A., Vanderpoorten, A., 2012. Inverted patterns of genetic diversity in continental and island populations of the heather *Erica scoparia* s.l. J. Biogeogr. 39, 574–584. http://dx.doi.org/10.1111/j.1365-2699.2011.02622.x.

Domingos, F.M., Colli, G.R., Lemmon, A., Lemmon, E.M., Beheregaray, L.B., 2017. In the shadows: phylogenomics and coalescent species delimitation unveil cryptic diversity in a Cerrado endemic lizard (Squamata: *Tropidurus*). Mol. Phylogenet. Evol. 107, 455–465.

Duarte, J.M., Wall, P.K., Edger, P.P., Landherr, L.L., Ma, H., Pires, J.C., Leebens-Mack, J., 2010. Identification of shared single copy nuclear genes in *Arabidopsis, Populus, Vitis* and *Oryza* and their phylogenetic utility across various taxonomic levels. BMC Evol. Biol. 10, 61.

Eaton, D.A.R., Ree, R.H., 2013. Inferring phylogeny and introgression using RADseq data: an example from flowering plants (Pedicularis: Orobanchaceae). Syst. Biol. 62, 689–706. http://dx.doi.org/10.1093/sysbio/syt032.

Eytan, R.I., Evans, B.R., Dornburg, A., Lemmon, A.R., Lemmon, E.M., Wainwright, P.C., Near, T.J., 2015. Are 100 enough? Inferring acanthomorph teleost phylogeny using Anchored Hybrid Enrichment. BMC Evol. Biol. 15, 113. http://dx.doi.org/10.1186/s12862-015-0415-0.

Ferrusquía-Villafranca, I., 1993. Geology of Mexico: A Synopsis. In: Ramamoorthy, T. P., Bye, R., Lot, A., Fa, J. (Eds.), Biological Diversity of Mexico. Origins and Distribution. Oxford University Press, UK, p. 812.

Fragoso-Martínez, I., Salazar, G.A., Martínez-Gordillo, M., Magallón, S., Sánchez-Reyes, L., Moriarty, Lemmon E., Lemmon, A., Sazatornil, F., Granados Mendoza, C., 2017. A pilot study applying the Plant Anchored Hybrid Enrichment method to New World sages (*Salvia* subgenus *Calosphace*; Lamiaceae). Mol. Phylogenet. Evol. 117, 124–134. http://dx.doi.org/10.1016/j.ympev.2017.02.006.

Geuten, K., Massingham, T., Darius, P., Smets, E., Goldman, N., 2007. Experimental design criteria in phylogenetics: where to add taxa. Syst. Biol. 56, 609–622.

Goldman, N., 1998. Phylogenetic information and experimental design in molecular systematics. Proc. Biol. Sci. 265, 1779–1786.

González, F., 1999. Inflorescence morphology and the systematics of Aristolochiaceae. System. Geograp. Plants 68, 159–172. http://dx.doi.org/10.2307/3668598.

González, F., Esquivel, H.E., Murcia, G.A., Pabón-Mora, N., 2010. *Aristolochia pentandra* (Aristolochiaceae) in Colombia: biogeographic implications and proposed synapomorphies between the pentandrous species of *Aristolochia* and its South American sister group. Rev. Acad. Colomb. Cienc 34, 467–478.

González, F., Rudall, P.J., 2003. Structure and development of the ovule and seed in Aristolochiaceae, with particular reference to *Saruma*. Plant Syst. Evol. 241, 223–244. http://dx.doi.org/10.1007/s00606-003-0050-x.

González, F., Stevenson, D., 2002. A phylogenetic analysis of the subfamily Aristolochioideae (Aristolochiaceae). Rev. Acad. Colomb. Cienc. Exact. Fis. Nat. 26, 25–57.

González, F., Stevenson, D.W., 2000. Perianth development and systematics of *Aristolochia*. Flora (Jena) 195, 370–391.

González, F., Wagner, S.T., Salomo, K., Symmank, L., Samain, M.-S., Isnard, S., Rowe, N.P., Neinhuis, C., Wanke, S., 2014. Present trans-Pacific disjunct distribution of *Aristolochia* subgenus *Isotrema* (Aristolochiaceae) was shaped by dispersal, vicariance and extinction. J. Biogeogr. 41, 380–391. http://dx.doi.org/10.1111/jbi.12198.

Gostel, M.R., Coy, K.A., Weeks, A., 2015. Microfluidic PCR-based target enrichment: A case study in two rapid radiations of *Commiphora* (Burseraceae) from Madagascar. J. Sytem. Evol. 53, 411–431. http://dx.doi.org/10.1111/jse.12173.

Granados Mendoza, C., Wanke, S., Salomo, K., Goetghebeur, P., Samain, M.-S., 2013. Application of the phylogenetic informativeness method to chloroplast markers: a test case of closely related species in tribe Hydrangeeae (Hydrangeaceae). Mol. Phylogenet. Evol. 66, 233–242. http://dx.doi.org/10.1016/j.ympev.2012.09.029.

Granados Mendoza, C.G., Naumann, J., Samain, M.-S., Goetghebeur, P., Smet, Y.D., Wanke, S., 2015. A genome-scale mining strategy for recovering novel rapidly-evolving nuclear single-copy genes for addressing shallow-scale phylogenetics in *Hydrangea*. BMC Evol. Biol. 15, 132. http://dx.doi.org/10.1186/s12862-015-0416-z.

Guo, Y.-Y., Luo, Y.-B., Liu, Z.-J., Wang, X.-Q., 2012. Evolution and biogeography of the slipper orchids: eocene vicariance of the conduplicate genera in the old and new world tropics. PLoS ONE 7, e38788. http://dx.doi.org/10.1371/journal.pone.0038788.

Hamilton, C.A., Lemmon, A.R., Lemmon, E.M., Bond, J.E., 2016. Expanding anchored hybrid enrichment to resolve both deep and shallow relationships within the spider tree of life. BMC Evol. Biol. 16 (1), 212.

Heyduk, K., Trapnell, D.W., Barrett, C.F., Leebens-Mack, J., 2016. Phylogenomic analyses of species relationships in the genus *Sabal* (Arecaceae) using targeted sequence capture. Biol. J. Linn. Soc. 117, 106–120. http://dx.doi.org/10.1111/bij.12551.

Horner, H.T., Samain, M.-S., Wagner, S.T., Wanke, S., 2015. Towards uncovering evolution of lineage-specific calcium oxalate crystal patterns in Piperales. Botany 93, 159–169. http://dx.doi.org/10.1139/cjb-2014-0191.

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. http://dx.doi.org/10.1093/molbev/mss020.

Leaché, A.D., Rannala, B., 2010. The accuracy of species tree estimation under simulation: a comparison of methods. Syst. Biol. syq073. http://dx.doi.org/10.1093/sysbio/syq073.

Leaché, A.D., Wagner, P., Linkem, C.W., Böhme, W., Papenfuss, T.J., Chong, R.A., Lavin, B.R., Bauer, A.M., Nielsen, S.V., Greenbaum, E., Rödel, M.-O., Schmitz, A., LeBreton, M., Ineich, I., Chirio, L., Ofori-Boateng, C., Eniang, E.A., Baha El Din, S., Lemmon, A.R., Burbrink, F.T., 2014. A hybrid phylogenetic–phylogenomic approach for species tree estimation in African *Agama* lizards with applications to biogeography, character evolution, and diversification. Mol. Phylogenet. Evol. 79, 215–230. http://dx.doi.org/10.1016/j.ympev.2014.06.013.

Lemmon, A.R., Emme, S.A., Lemmon, E.M., 2012. Anchored hybrid enrichment for massively high-throughput phylogenomics. Syst. Biol. sys049. http://dx.doi.org/10.1093/sysbio/sys049.

Lemmon, E.M., Lemmon, A.R., 2013. High-throughput genomic data in systematics and phylogenetics. Annu. Rev. Ecol. Evol. Syst. 44, 99–121. http://dx.doi.org/10.1146/annurev-ecolsys-110512-135822.

López-Giráldez, F., Townsend, J., 2011. PhyDesign: an online application for profiling phylogenetic informativeness. BMC Evol Biol. 11, 152.

Magallón, S., Gómez-Acevedo, S., Sánchez-Reyes, L., Hernández-Hernández, T., 2015. A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. New Phytol. 207 (2), 437–453.

Mandel, J.R., Dikow, R.B., Funk, V.A., 2015. Using phylogenomics to resolve mega-families: an example from Compositae. J. Sytem. Evol. 53, 391–402. http://dx.doi.org/10.1111/jse.12167.

Manthey, J.D., Tollis, M., Lemmon, A.R., Moriarty Lemmon, E., Boissinot, S., 2016. Diversification in wild populations of the model organism *Anolis carolinensis*: a genome-wide phylogeographic investigation. Ecol. Evol. 6 (22), 8115–8125.

Marcussen, T., Jakobsen, K.S., Danihelka, J., Ballard, H.E., Blaxland, K., Brysting, A.K., Oxelman, B., 2012. Inferring species networks from gene trees in high-polyploid North American and Hawaiian Violets (*Viola*, Violaceae). Syst. Biol. 61, 107–126. http://dx.doi.org/10.1093/sysbio/syr096.

Meyer, M., Kircher, M., 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb Protoc 2010, pdb.prot5448. http://dx.doi.org/10.1101/pdb.prot5448.

Miller, M.A., Pfeiffer, W., Schwartz, T., 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: Proceedings of the Gateway Computing Environments Workshop (GCE). New Orleans, pp. 1–8. http://dx.doi.org/10.1109/GCE.2010.5676129.

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.

Mitchell, N., Lewis, P.O., Lemmon, E.M., Lemmon, A.R., Holsinger, K.E., 2017. Anchored phylogenomics improves the resolution of evolutionary relationships in the rapid radiation of *Protea* L. Am. J. Bot. 104 (1), 102–115.

Morrone, J.J., 2005. Toward a synthesis of Mexican Biogeography. Revista Mexicana de Biodiversidad 76, 207–252.

Müller, S., Salomo, K., Salazar, J., Naumann, J., Jaramillo, M.A., Neinhuis, C., Feild, T.S., Wanke, S., 2015. Intercontinental long-distance dispersal of Canellaceae from the New to the Old World revealed by a nuclear single copy gene and chloroplast loci. Mol. Phylogenet. Evol. 84, 205–219. http://dx.doi.org/10.1016/j.ympev.2014.12.010.

Naumann, J., Salomo, K., Der, J.P., Wafula, E.K., Bolin, J.F., Maass, E., Frenzke, L., Samain, M.-S., Neinhuis, C., dePamphilis, C.W., Wanke, S., 2013. Single-copy nuclear genes place haustorial hydnoraceae within piperales and reveal a cretaceous origin of multiple parasitic Angiosperm lineages. PLoS ONE 8, e79204. http://dx.doi.org/10.1371/journal.pone.0079204.

Naumann, J., Symmank, L., Samain, M.-S., Müller, K.F., Neinhuis, C., dePamphilis, C. W., Wanke, S., 2011. Chasing the hare - evaluating the phylogenetic utility of a nuclear single copy gene region at and below species level within the species rich group *Peperomia* (Piperaceae). BMC Evol. Biol. 11, 357. http://dx.doi.org/10.1186/1471-2148-11-357.

Neinhuis, C., Wanke, S., Hilu, K.W., Müller, K., Borsch, T., 2005. Phylogeny of Aristolochiaceae based on parsimony, likelihood, and Bayesian analyses of *trn L-trn F* sequences. Plant Syst. Evol. 250, 7–26.

Ohi-Toma, T., Sugawara, T., Murata, H., Wanke, S., Neinhuis, C., Murata, J., 2006. Molecular phylogeny of *Aristolochia* sensu lato (Aristolochiaceae) based on sequences of *rbcL*, *matK*, and *phyA* genes, with special reference to differentiation of chromosome numbers. Syst. Bot. 31, 481–492. http://dx.doi.org/10.1043/05-38.1.

Paizanni Guillén, A., Santana Michel, F.J., Ramírez Amezcua, J.M., Wagner, S.T., Müller, S., Montero Castro, J.C., Wanke, S., Samain, M.-S., 2016. Four new species of *Aristolochia* subsection *Pentandrae* from western Mexico. Syst. Bot. 41, 128–141.

Peloso, P.L.V., Frost, D.R., Richards, S.J., Rodrigues, M.T., Donnellan, S., Matsui, M., Raxworthy, C.J., Biju, S.D., Lemmon, E.M., Lemmon, A.R., Wheeler, W.C., 2015. The impact of anchored phylogenomics and taxon sampling on phylogenetic inference in narrow-mouthed frogs (Anura, Microhylidae). Cladistics 32, 113–140. http://dx.doi.org/10.1111/cla.12118.

Pfeifer, H.W., 1970. A taxonomic revision of the Pentandrous species of *Aristolochia*.

Pond, S.L.K., Frost, S.D.W., Muse, S.V., 2005. HyPhy: hypothesis testing using phylogenies. Bioinformatics 21, 676–679. http://dx.doi.org/10.1093/bioinformatics/bti079.

Prum, R.O., Berv, J.S., Dornburg, A., Field, D.J., Townsend, J.P., Lemmon, E.M., Lemmon, A.R., 2015. A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. Nature 526, 569–573. http://dx.doi.org/10.1038/nature15697.

Pyron, R.A., Hsieh, F.W., Lemmon, A.R., Lemmon, E.M., Hendry, C.R., 2016. Integrating phylogenomic and morphological data to assess candidate species-delimitation models in brown and red-bellied snakes (*Storeria*). Zool. J. Linnean Soc. http://dx.doi.org/10.1111/zoj.12392.

Pyron, R.A., Hendry, C.R., Chou, V.M., Lemmon, E.M., Lemmon, A.R., Burbrink, F.T., 2014. Effectiveness of phylogenomic data and coalescent species-tree methods for resolving difficult nodes in the phylogeny of advanced snakes (Serpentes: Caenophidia). Mol. Phylogenet. Evol. 81, 221–231. http://dx.doi.org/10.1016/j.ympev.2014.08.023.

Ramamoorthy, T.P., Elliott, M., 1993. Mexican Lamiaceae: diversity, distribution, endemism and evolution. In: Ramamoorthy, T.P., Bye, R., Lot, A., Fa, J. (Eds.), Biological Diversity of Mexico. Origins and Distribution. Oxford University Press, UK, p. 812.

Rambaut, A., 2002. TreeEdit, Version 1.0a10. <http://tree.bio.ed.ac.uk/software/treeedit/>.

Rambaut, A., 2009. FigTree ver. 1.3.1. <http://tree.bio.ed.ac.uk/software/figtree/>.

Richardson, J.E., Chatrou, L.W., Mols, J.B., Erkens, R.H.J., Pirie, M.D., 2004. Historical biogeography of two cosmopolitan families of flowering plants: Annonaceae and Rhamnaceae. Phil. Trans. R. Soc. Lond. B 359, 1495–1508. http://dx.doi.org/10.1098/rstb.2004.1537.

Rokyta, D.R., Lemmon, A.R., Margres, M.J., Aronow, K., 2012. The venom-gland transcriptome of the eastern diamondback rattlesnake (*Crotalus adamanteus*). BMC Genom. 13, 1–23. http://dx.doi.org/10.1186/1471-2164-13-312.

Ruane, S., Raxworthy, C.J., Lemmon, A.R., Lemmon, E.M., Burbrink, F.T., 2015. Comparing species tree estimation with large anchored phylogenomic and small Sanger-sequenced molecular datasets: an empirical study on Malagasy pseudoxyrhophiine snakes. BMC Evol. Biol. 15, 221. http://dx.doi.org/10.1186/s12862-015-0503-1.

Rubin, B.E.R., Ree, R.H., Moreau, C.S., 2012. Inferring phylogenies from RAD sequence data. PLoS ONE 7, e33394. http://dx.doi.org/10.1371/journal.pone.0033394.

Ruhfel, B.R., Gitzendanner, M.A., Soltis, P.S., Soltis, D.E., Burleigh, J.G., 2014. From algae to angiosperms–inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. BMC Evol. Biol. 14, 23. http://dx.doi.org/10.1186/1471-2148-14-23.

Rzedowski, J., 1991. Diversidad y orígenes de la flora fanerogámica de México. Acta Botanica Mexicana 14, 3–21.

San Mauro, D., Gower, D.J., Cotton, J.A., Zardoya, R., Wilkinson, M., Massingham, T., 2012. Experimental design in phylogenetics: testing predictions from expected information. Syst. Biol. 61, 661–674.

Sanderson, M.J., 1997. A nonparametric approach to estimating divergence times in the absence of rate constancy. Mol. Biol. Evol. 14, 1218–1231.

Sang, T., 2002. Utility of low-copy nuclear gene sequences in plant phylogenetics. Crit. Rev. Biochem. Mol. Biol. 37, 121–147. http://dx.doi.org/10.1080/10409230290771474.

Sass, C., Iles, W.J.D., Barrett, C.F., Smith, S.Y., Specht, C.D., 2016. Revisiting the Zingiberales: using multiplexed exon capture to resolve ancient and recent phylogenetic splits in a charismatic plant lineage. PeerJ 4. http://dx.doi.org/10.7717/peerj.1584.

Sayyari, E., Mirarab, S., 2016. Fast coalescent-based computation of local branch support from quartet frequencies. Mol. Biol. Evol. 33, 1654–1668. http://dx.doi.org/10.1093/molbev/msw079.

Schmickl, R., Liston, A., Zeisek, V., Oberlander, K., Weitemier, K., Straub, S.C.K., Cronn, R.C., Dreyer, L.L., Suda, J., 2015. Phylogenetic marker development for target enrichment from transcriptome and genome skim data: the pipeline and its application in southern African *Oxalis* (Oxalidaceae). Mol. Ecol. Resour. http://dx.doi.org/10.1111/1755-0998.12487. n/a-n/a.

Shaw, J., Lickey, E.B., Beck, J.T., Farmer, S.B., Liu, W., Miller, J., Siripun, K.C., Winder, C. T., Schilling, E.E., Small, R.L., 2005. The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. Am. J. Bot. 92, 142–166. http://dx.doi.org/10.3732/ajb.92.1.142.

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.

Small, R.L., Ryburn, J.A., Cronn, R.C., Seelanan, T., Wendel, J.F., 1998. The tortoise and the hare: choosing between noncoding plastome and nuclear *Adh* sequences for phylogeny reconstruction in a recently diverged plant group. Am. J. Bot. 85, 1301–1315.

Song, J., Shi, L., Li, D., Sun, Y., Niu, Y., Chen, Z., Luo, H., Pang, X., Sun, Z., Liu, C., Lv, A., Deng, Y., Larson-Rabin, Z., Wilkinson, M., Chen, S., 2012. Extensive pyrosequencing reveals frequent intra-genomic variations of internal transcribed spacer regions of nuclear ribosomal DNA. PLoS ONE 7, e43971. http://dx.doi.org/10.1371/journal.pone.0043971.

Sousa, S.M., Delgado, S.A., 1993. Mexican Leguminosae: phytogeography, endemism and origins. In: Ramamoorthy, T.P., Bye, R., Lot, A., Fa, J. (Eds.), Biological Diversity of Mexico. Origins and Distribution. Oxford University Press, UK. 812 pp.

Sousa, F.de, Bertrand, Y.J.K., Nylinder, S., Oxelman, B., Eriksson, J.S., Pfeil, B.E., 2014. Phylogenetic properties of 50 *N*uclear *L*oci in *Medicago* (Leguminosae) generated using multiplexed sequence capture and next-generation

sequencing. PLoS ONE 9, e109704. http://dx.doi.org/10.1371/journal.pone.0109704.
Stamatakis, A., 2014. RAxML Version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics btu033. http://dx.doi.org/10.1093/bioinformatics/btu033.
Stephens, J.D., Rogers, W.L., Heyduk, K., Cruse-Sanders, J.M., Determann, R.O., Glenn, T.C., Malmberg, R.L., 2015a. Resolving phylogenetic relationships of the recently radiated carnivorous plant genus *Sarracenia* using target enrichment. Mol. Phylogenet. Evol. 85, 76–87. http://dx.doi.org/10.1016/j.ympev.2015.01.015.
Stephens, J.D., Rogers, W.L., Mason, C.M., Donovan, L.A., Malmberg, R.L., 2015b. Species tree estimation of diploid *Helianthus* (Asteraceae) using target enrichment. Am. J. Bot. 102, 910–920. http://dx.doi.org/10.3732/ajb.1500031.
Stout, C.C., Tan, M., Lemmon, A.R., Lemmon, E.M., Armbruster, J.W., 2016. Resolving Cypriniformes relationships using an anchored enrichment approach. BMC Evol. Biol. 16 (1), 244.
Syring, J.V., Tennessen, J.A., Jennings, T.N., Wegrzyn, J., Scelfo-Dalbey, C., Cronn, R., 2016. Targeted capture sequencing in whitebark pine reveals range-wide demographic and adaptive patterns despite challenges of a large, repetitive genome. Front. Plant Sci. 7. http://dx.doi.org/10.3389/fpls.2016.00484.
The Angiosperm Phylogeny Group, 2016. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. Bot. J. Linn. Soc. 181 (1), 1–20.
Torres, M.A.I., Luna, V., 2006. Análisis de trazos para establecer áreas de conservación en la Faja Volcánica Transmexicana. Interciencia 31, 849–855.
Townsend, J.P., 2007. Profiling Phylogenetic Informativeness. Syst. Biol. 56, 222–231. http://dx.doi.org/10.1080/10635150701311362.
Townsend, J.P., Leuenberger, C., 2011. Taxon sampling and the optimal rates of evolution for phylogenetic inference. Syst. Biol. 60, 358–365. http://dx.doi.org/10.1093/sysbio/syq097.
Townsend, J.P., Lopez-Giraldez, F., 2010. Optimal selection of gene and ingroup taxon sampling for resolving phylogenetic relationships. Syst. Biol. 59, 446–457.
Tucker, D.B., Colli, G.R., Giugliano, L.G., Hedges, S.B., Hendry, C.R., Lemmon, E.M., Lemmon, A.R., Sites Jr., J.W., Pyron, R.A., 2017. Methodological congruence in phylogenomic analyses with morphological support for teiid lizards (Sauria: Teiidae). Mol. Phylogenet. Evol. 103, 75–84. http://dx.doi.org/10.1016/j.ympev.2016.07.002.
Wagner, S.T., Hesse, L., Isnard, S., Samain, M.-S., Bolin, J., Maass, E., Neinhuis, C., Rowe, N.P., Wanke, S., 2014. Major trends in stem anatomy and growth forms in the perianth-bearing Piperales, with special focus on *Aristolochia*. Ann. Bot. mcu044. http://dx.doi.org/10.1093/aob/mcu044.
Wanke, S., Gonzalez, F., Neinhuis, C., 2006. Systematics of pipevines: combining morphological and fast-evolving molecular characters to investigate the relationships within subfamily Aristolochioideae (Aristolochiaceae). Int. J. Plant Sci. 167, 1215–1227. http://dx.doi.org/10.1086/508024.
Wanke, S., Jaramillo, M.A., Borsch, T., Samain, M.-S., Quandt, D., Neinhuis, C., 2007. Evolution of Piperales—*matK* gene and *trnK* intron sequence data reveal lineage specific resolution contrast. Mol. Phylogenet. Evol. 42, 477–497. http://dx.doi.org/10.1016/j.ympev.2006.07.007.
Wickett, N.J., Mirarab, S., Nguyen, N., Warnow, T., Carpenter, E., Matasci, N., Ayyampalayam, S., Barker, M.S., Burleigh, J.G., Gitzendanner, M.A., Ruhfel, B.R., Wafula, E., Der, J.P., Graham, S.W., Mathews, S., Melkonian, M., Soltis, D.E., Soltis, P.S., Miles, N.W., Rothfels, C.J., Pokorny, L., Shaw, A.J., DeGironimo, L., Stevenson, D.W., Surek, B., Villarreal, J.C., Roure, B., Philippe, H., dePamphilis, C.W., Chen, T., Deyholos, M.K., Baucom, R.S., Kutchan, T.M., Augustin, M.M., Wang, J., Zhang, Y., Tian, Z., Yan, Z., Wu, X., Sun, X., Wong, G.K.-S., Leebens-Mack, J., 2014. Phylotranscriptomic analysis of the origin and early diversification of land plants. PNAS 111, E4859–E4868. http://dx.doi.org/10.1073/pnas.1323926111.
Young, A.D., Lemmon, A.R., Skevington, J.H., Mengual, X., Ståhls, G., Reemer, M., Jordaens, K., Kelso, S., Lemmon, E.M., Hauser, M., De Meyer, M., Misof, B., Wiegmann, B.M., 2016. Anchored enrichment dataset for true flies (order Diptera) reveals insights into the phylogeny of flower flies (family Syrphidae). BMC Evol. Biol. 16, 143. http://dx.doi.org/10.1186/s12862-016-0714-0.
Zhang, N., Zeng, L., Shan, H., Ma, H., 2012. Highly conserved low-copy nuclear genes as effective markers for phylogenetic analyses in angiosperms. New Phytol. 195, 923–937. http://dx.doi.org/10.1111/j.1469-8137.2012.04212.x.
Zimmer, E.A., Wen, J., 2013. Using nuclear gene data for plant phylogenetics: Progress and prospects. Mol. Phylogenet. Evol. 66, 539–550. http://dx.doi.org/10.1016/j.ympev.2013.01.005.