Molecular Phylogenetics and Evolution 75 (2014) 91–102

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

Phylogenomics of the plant family Araceae

Claudia L. Henriquez [a,b,*], Tatiana Arias [d], J. Chris Pires [c], Thomas B. Croat [b], Barbara A. Schaal [a]

[a] Evolution, Ecology & Population Biology, Division of Biology, Washington University in St. Louis, One Brookings Drive, St. Louis, MO 63130, USA
[b] Missouri Botanical Garden, 4344 Shaw Blvd., St. Louis, MO 63110, USA
[c] Department of Biological Sciences, University of Missouri, Columbia, MO 65211, USA
[d] School of Biological Sciences, The University of Hong Kong, Pokfulam Road, Hong Kong

ARTICLE INFO

Article history:
Received 26 July 2013
Revised 15 February 2014
Accepted 18 February 2014
Available online 1 March 2014

Keywords:
Araceae
Massively parallel sequencing
Clade
Phylogeny

ABSTRACT

The biogeography, chromosome number evolution, pollination biology and evolutionary history of the plant family Araceae have recently become much clearer (Cabrera et al., 2008; Chartier et al., 2013; Cusimano et al., 2011, 2012; Nauheimer et al., 2012). However, phylogenetic ambiguity near the root of the tree precludes answering questions about the early evolution of the family. We use Illumina sequencing technology and reference based assembly to resolve the remaining questions in the deep phylogeny of Araceae. We sampled 32 genera and obtained 7 from GenBank (including an outgroup), representing 42 of 44 major clades described in Cusimano et al. (2011). A subsequent phylogenomic analysis based on mitochondrial data was performed to test congruence between plastid and mitochondrial data for phylogenetic inference. Plastid sequences produced strongly supported phylogenies. In contrast, mitochondrial phylogenies were weakly supported and incongruent with chloroplast data (Templeton test, $p \leqslant 0.0001$), although several smaller clades were recovered. New strongly-supported clades seen here are: (1) *Anubias* and *Montrichardia*, excluding *Calla*, form a clade that is sister to the *Zantedeschia* clade; (2) the South African genus *Zantedeschia* is sister to the Old World *Anchomanes* clade; and (3) within the *Zantedeschia* clade, *Philodendron* is sister to the rest. *Calla* and *Schismatoglottis* form a clade at the base of one of two major clades in Aroideae based on complete chloroplast sequences. Although statistical support is weak, morphological and cytological features support this topology.

1. Introduction

Araceae, or the Arum family, is a large and ancient monocot plant family most notable for its impressive morphological diversity, including the smallest known angiosperm and some of the largest vegetative and reproductive structures in the world (Simpson, 2006). The family consists of c. 3800 species in 118 genera, distributed mostly in the tropics but can range into temperate and, in the case of *Calla palustris*, circumboreal regions (Boyce and Croat, 2013; Ulrich et al., 2013). Members of Araceae occupy a wide array of ecological habitats from sea level to above 3000 m and range from submerged, emergent or free-floating aquatics, to epiphytic, climbing and terrestrial plants (Bown, 2000; Cabrera et al., 2008; Croat, 1988; Gonçalves, 2004; Gonçalves et al., 2007). Stems can be rhizomatous, cormose, tuberous or reduced to a thallus-like structure and leaves can be simple, highly divided or fenestrate (Mayo et al., 1997; Simpson, 2006). Araceae are distinguished from closely related families in having a great diversity of calcium oxalate crystals (raphides, druses, crystal sand, styloids and prismatics), possessing a spadix of small, bisexual or unisexual flowers, subtended by a spathe, and they lack ethereal oil cells (Grayum, 1990; Keating, 2003; Stevens, 2001 onwards).

Detailed classification of Araceae, established as a family in 1789 (Jussieu, 1789), began in the nineteenth century with the work of Heinrich Wilhelm Schott (1794–1865) and Adolf Gustav Engler (1844–1930). Schott's pre-Darwinian classification grouped genera based on inflorescences, flowers and fruits (Mayo et al., 1997; Nicolson, 1987). A modified version of this classification was used by Hooker (1883) who divided Araceae into 11 tribes, and later by Hutchinson (1973) who divided the family into 18 tribes (Grayum, 1990; Hooker, 1883; Hutchinson, 1973). Engler's new system of classification, which included hypotheses of evolutionary transitions of not only floral, but also of vegetative morphological and anatomical characters (Engler, 1920, Mayo et al., 1997, Nicolson, 1987), has been the framework for much subsequent work (Bogner, 1978; Bogner and Nicolson, 1991; Grayum, 1990; Hotta, 1970, Mayo et al., 1997, Nakai, 1943). Grayum's

* Corresponding author at: Evolution, Ecology & Population Biology, Division of Biology, Washington University in St. Louis, One Brookings Drive, Campus Box 1137, St. Louis, MO 63130, USA.
E-mail address: chenriquez@wustl.edu (C.L. Henriquez).

http://dx.doi.org/10.1016/j.ympev.2014.02.017
1055-7903/

1990 revision, based on a large survey of palynological characters, is notable in recognizing *Acorus* as separate from all other Araceae (Grayum, 1987, 1990).

Since the chloroplast restriction site data of French et al. (1995), molecular data have been used to infer evolutionary relationships at all levels in the family (Barabé et al., 2002; Cabrera et al., 2008; Chartier et al., 2013; Cusimano et al., 2011; Gauthier et al., 2008, Gonçalves et al., 2007, Nauheimer et al., 2012; Renner et al., 2004; Renner and Zhang, 2004; Rothwell et al., 2004; Tam et al., 2004; Wong et al., 2010). To date, the most comprehensive family-wide molecular data set consists of six chloroplast (*rbcL*, *matK*, partial *trnK* intron, partial *tRNA-Leu* gene, *trnL-trnF* spacer, and partial *tRNA-Phe* gene) and one nuclear (*PhyC*) markers (Cabrera et al., 2008; Chartier et al., 2013), and has been used to clarify the evolutionary history, biogeography, pollination biology and chromosomal evolution of Araceae (see also Cusimano et al., 2011; Nauheimer et al., 2012). Araceae has an inferred ancestral haploid chromosome number of $n = 16$ or $n = 18$, and began to diversify in the Early Cretaceous, approximately 122 Mya, as the breakup of Pangea was finalizing (Cusimano et al., 2012; Nauheimer et al., 2012). By the Cretaceous/Paleogene boundary all eight of the currently recognized subfamilies, including the duckweed subfamily Lemnoideae, were present and form a clade that is sister to a clade comprising all other members of the order Alismatales (Cabrera et al., 2008; Nauheimer et al., 2012; Tobe and Kadokawa, 2010). Evolutionary relationships among six of the eight subfamilies (Gymnostachydoideae, Orontioideae, Lemnoideae, Pothoideae, Monsteroideae and Lasioideae), all of which contain bisexually-flowered members, are well-supported (Cabrera et al., 2008; Cusimano et al., 2011; Nauheimer et al., 2012). The Unisexual Flowers clade, containing subfamilies Zamioculcadoideae and the highly diverse Aroideae (1573 species, 75 genera), diverged from the bisexual-flowered lineage during the Late Cretaceous approximately 90 Mya (Nauheimer et al., 2012). Low resolution of several deep nodes in the phylogeny of the Unisexual Flowers clade leaves open several important questions, including the position of the highly autapomorphic, bisexually-flowered genus *Calla* (Cabrera et al., 2008; Chartier et al., 2013; Cusimano et al., 2011; Ulrich et al., 2013). Although *Calla* is well-supported in a clade with two unisexually-flowered genera (*Montrichardia* and *Anubias*) in the nuclear tree from Chartier et al. (2013), the position of that clade at the base of Aroideae is not strongly supported and biogeographical and morphological features make this grouping dubious. *Calla* has spirally arranged perfect flowers that emerge acropetally, disulcate pollen, an inferred ancestral haploid chromosome number of $n = 18$ and a circumboreal, mainly European geographical distribution (Chartier et al., 2013; Stevens, 2001 onwards, Ulrich et al., 2013). *Anubias* is an African genus and *Montrichardia* is South American, but both share an inferred ancestral haploid chromosome number of $n = 12$. The only feature shared by all three is a helophytic habit, which occurs elsewhere in the family (Chartier et al., 2013; Cusimano et al., 2011; Grayum, 1990). Another generic placement that warrants further investigation is the weakly-supported sister relationship of the South African genus *Zantedeschia* ($n = 16$) with the strictly South American tribe Spathicarpeae ($n = 17$) (Cabrera et al., 2008; Chartier et al., 2013; Cusimano et al., 2011; Nauheimer et al., 2012). In addition, weakly-supported relationships among the smaller clades within the *Zantedeschia* clade are in need of further clarification.

With the advent of massively parallel sequencing, phylogenetic analyses can now be based on tens of thousands of nucleotides, which can greatly enhance our confidence in the resulting evolutionary hypotheses (Givnish et al., 2010; Steele et al., 2012; Xi et al., 2012). Sequencing of plastomes and mitogenomes includes de novo assemblies using complete genomic DNA and reference-based assemblies using DNA enriched for chloroplasts, and combinations thereof (Givnish et al., 2010; Steele et al., 2012). In addition to variation in the proportion of organellar and nuclear DNA used in creating libraries for sequencing, the suite of genomic tools now available to process and analyze the resulting deluge of genomic data permits the use of multiple software programs to corroborate results.

Phylogenomic studies in plants have generally focused on the chloroplast genome, whereas the mitochondrial genome, due to its complicated mutational dynamics, has been more commonly used in studies of structural variation, nucleotide substitution rates and horizontal gene transfer (Knoop et al., 2011; Mower et al., 2007; Palmer et al., 2000; Richards et al., 2009; Richardson et al., 2013; Xi et al., 2012). The low silent-site substitution rate of plant mitochondrial DNA, which has been shown to be one-third less than that of plant chloroplast DNA, plus the extensive RNA-editing and retroprocessing that occurs in this genome perhaps explain why, in plant phylogenetic studies, mitochondrial regions have typically been used in combination with plastid regions (Renner and Zhang, 2004; Seberg and Petersen, 2006; Seberg et al., 2012; Steele et al., 2012; Wolfe et al., 1987). In addition, previous studies have shown that phylogenies reconstructed from mitochondrial data are less resolved and incongruent with plastid data (Petersen et al., 2006, 2013). However, slow silent substitution rates are not consistent across the entire mitochondrial genome or among all plant lineages; mitochondrial genes from highly divergent plant genera have been shown to have substitution rates similar to that of the rapidly evolving mammalian mitochondrial genome (Mower et al., 2007; Palmer et al., 2000). The question remains whether large-scale datasets based on tens to hundreds of thousands of aligned nucleotides, representing both coding and non-coding regions, from the mitochondrial genome possess enough phylogenetic signal to resolve evolutionary relationships in plants at the family level.

Here we use Illumina sequencing technology with total genomic DNA and reference-based assembly of the chloroplast, using the programs Geneious 6.0.3 and Bowtie2, to resolve some of the major remaining questions in the current phylogeny of Araceae. A subsequent phylogenomic analysis of mitochondrial sequences obtained from reference-based assembly was performed to compare congruence of the mitochondrial phylogeny with the plastid phylogeny.

2. Materials and methods

2.1. Taxon sampling

For the chloroplast analysis, we sampled 32 genera of Araceae and obtained from GenBank the complete, annotated chloroplast genomes of 5 additional genera: *Colocasia esculenta* (Ahmed et al., 2012), *Lemna minor* (Mardanov et al., 2008), *Wolffiella lingulata*, *Wolffia australiana* and *Spirodela polyrhiza* (Wang and Messing, 2011). We included at least one representative from 42 of the 44 clades of Araceae named in Cusimano et al. (2011). For a list of genera included in this study and the higher taxa they represent refer to Table 1. The two taxa not sampled were Cryptocoryneae and Culcasieae, although larger clades within which they are nested were sampled; these are the Rheophytes clade and the *Homalomena* clade, respectively. Of the 11 phylogenetically isolated genera in Cusimano et al. (2011) (*Calla*, *Callopsis*, *Montrichardia*, *Anubias*, *Zantedeschia*, *Philonotion*, *Protarum*, *Pistia*, *Alocasia*, *Pinellia* and *Arisaema*), 4 genera (*Callopsis*, *Philonotion*, *Protarum*, *Pistia*) were not sampled. *Gymnostachys anceps*, the sole member of subfamily Gymnostachydoideae, was not sampled but its sister relationship with subfamily Orontioideae, here represented by *Orontium*, has been strongly supported in other studies (Cabrera et al., 2008;

Table 1
List of genera, taxa represented from Cusimano et al. (2011) and putative synapomorphic indels.

Clade	Clade name	Representative genera and clades	Synapomorphic indels in plastid genes		
			Type	bp	gene
1	Orontioideae	*Orontium*			
2	Lemnoideae	*Spirodela, Lemna, Wolffia, Wolfiella*			
3	Potheae	*Pothos*			
4	*Heteropsis* clade	*Stenospermation*			
5	Spathiphylleae	*Spathiphyllum*			
6	*Rhaphidophora* clade	*Monstera, Rhaphidophora*			
7	Lasioideae	*Lasia*	Deletion	2145	*rpoC2*
8	Zamioculcadoideae	*Zamioculcas*			
9	Aglaonemateae	*Aglaonema*			
10	Nephthytideae	*Anchomanes*			
11	Culcasieae				
12	*Philodendron* clade	*Philodendron*	Insertion	14	*matK*
13	Spathicarpeae	*Dieffenbachia, Taccarum*	Insertion	15	*matK*
14	Cryptocoryneae				
15	Schismatoglottideae	*Schismatoglottis*			
16	Thomsonieae	*Amorphophallus*			
17	Caladieae	*Syngonium, Ulearum, Xanthosoma, Zomicarpella*			
18	Arisareae	*Arisarum*			
19	Arophyteae	*Carlephyton*			
20	*Colocasia* clade	*Colocasia, Steudnera*			
21	Areae	*Typhonium*			
22	Proto-Araceae	*Orontium*			
23	Pothoideae	*Anthurium*, clade 3			
24	Monsteroideae	Clades 4,5,6			
25	*Stylochaeton* clade	*Stylochaeton*, clade 8			
26	*Anchomanes* clade	Clades 9, 10	Insertion	6	*petA*
27	*Homalomena* clade	Clades 11, 12			
28	Rheophytes clade	Clades 14, 15			
29	*Typhonodorum* clade	Clade 19			
30	*Alocasia* clade	*Alocasia, Arisaema, Pinellia,* clade 21			
31	Bisexual Climbers *clade*	Clades 23, 24			
32	*Zantedeschia* clade	*Zantedeschia*, clades 13, 26, 27			
33	*Colletogyne* clade	Clades 18, 29			
34	*Pistia* clade	Clades 20, 30	Insertion	3	*atpE*
35	*Amorphophallus* clade	Clades 16, 17			
36	*Ambrosina* clade	Clades 33, 34			
37	*Dracunculus* clade	Clades 35, 36			
38	*Philonotion* clade	Clades 28, 37			
39	Aroideae	*Anubias, Montrichardia* clades 32, 38			
40	Unisexual Flowers clade	Clades 25, 39			
41	*Podolasia* clade	Clades 7, 40			
42	True Araceae *clade*	Clades 31, 41			
43	*Spirodela* clade	Clades 2, 42	Deletion	9	*atpF*
44	Araceae	Clades 22, 43	Insertion	22	*rpoC1*
			Deletion	36	*rpoB*
			Insertion	69	*ndhK*
			Deletion	214	*cemA*

Cusimano et al., 2011; Nauheimer et al., 2012). Two species of *Acorus* were used as an outgroup: *Acorus americanus* (Peery et al., 2007) and *Acorus calamus* (Goremykin et al., 2005). For the mitochondrial analysis, the complete, annotated mitochondrial genome of *Spirodela polyrhiza* (Wang et al., 2012) was taken from GenBank and the 32 genera sampled (above) were included. Silica samples of *Calla palustris* were obtained from the Nancy Botanical Garden in France. All remaining genera were collected as fresh samples from the Araceae Greenhouse and Temperate Greenhouse at the Missouri Botanical Garden in St. Louis, Missouri. Although 31.4% of the total genera in the family were sampled (37 of 118), they represent 95.5% of the major named taxa (42 of 44) in Araceae. The list of species used in this study with GenBank accession numbers and herbarium voucher numbers appears in Appendix 1.

2.2. Illumina sequencing

Total genomic DNA was extracted from 100 mg tissue for all fresh samples, 20 mg for the silica sample of *Calla palustris*, using Qiagen DNeasy Minikit (Qiagen, Germantown, Maryland, USA). Two extractions per taxon were performed eluting DNA with 125 μL elution buffer for each extraction, which were then combined for a total of 250 μL, or alternatively each extraction was eluted with 75 μL for a total of 150 μL. The quality and concentration of DNA samples were quantified with Nanodrop (ThermoScientific, Delaware, USA) and gel electrophoresis. Illumina TruSeq kits recommend 55 μL of DNA at a concentration of 20 ng/μL; samples that were below this concentration threshold were concentrated using ethanol precipitation, while those that were overconcentrated were diluted with autoclaved H_2O or elution buffer from Qiagen DNeasy Minikit. Library preparation in the Pires lab at the University of Missouri, Columbia, followed the TruSeq DNA Sample Preparation Guide protocol (Illumina, Inc., 2010), except where noted. Sonication to shear total genomic DNA was performed for a total of 15–24 min. using a Bioruptor (Diagenode, Inc., New Jersey, USA). Gel extractions of size-selected samples (200–400 bp) were performed using x-tracta disposable gel extraction tools (USA Scientific, Ocala, Florida, USA) and purified with the Gel Extraction kit (Qiagen) for the end repair, adenylation of 3′ ends, ligation, and enrichment steps. All gels for electrophoresis were 2% low-melt agarose stained with ethidium bromide and run at 120 V for 1 h with a 100 bp ladder to visualize sheared

DNA for size. Prepared DNA libraries were sent to the University of Missouri DNA Core for quantification and fragment-size verification with an ABI 3730xl DNA Analyzer (Life Technologies Corp., Carlsbad, California). Sequencing was performed with the Illumina HiSeq 2000 (Illumina, Inc., San Diego, California) for single-end reads of a length of 101 bps. For the first sequencing run, we multiplexed 8 samples per lane using adapters 1–8. For the second and final sequencing run we multiplexed 12 samples using adapters 1–12. Illumina HiSeq 2000 automatically removes adapter ends and parses reads based on adapter ends into separate files. Raw fastq reads for each taxon were concatenated when presented in multiple files. The total number of reads generated for each taxon is listed in Table 2.

2.3. Data quality-trimming and filtering

Raw fastq reads were quality trimmed using DynamicTrim (Cox et al., 2010), which uses the Burrows-Wheeler Alignment trimming algorithm to crop reads that are below a quality cutoff ($p = 0.05$, Phred score $Q = 13$). Based on the increased mutational complexities of mitochondrial sequences, an additional filtering step was performed for the mitochondrial analysis using Prinseq-lite-0.20.3 (Schmieder and Edwards, 2011), in which sequences under a length of 40 base pairs with a quality score lower than 30, and having more than 1% Ns were removed. For the number of filtered reads passing quality control for each taxon, refer to Table 2.

2.4. Sequence assembly, validation and alignment of chloroplast coding sequences

Quality-trimmed reads for each taxon were assembled to the chloroplast genome of *Lemna minor* (165,955 bases), used here as the reference sequence. Assembly was performed using Bowtie2-2.0.0-beta6 (Langmead and Salzberg, 2012) with parameters set to default. Assembly was also performed using Geneious versions 5.6–6.0.3 (created by Biomatters) using the Custom Sensitivity setting. The Custom sensitivity values were chosen to have a more stringent minimum overlap length than Bowtie2 and the maximum gap size was changed from 15 to 3 based on the need to minimize computing requirements. Fine-tuning was iterated up to five times. The consensus sequences from Bowtie2 and Geneious for each taxon were extracted and aligned to the reference genome using the mauveAligner algorithm plugin in Geneious. All discrepancies in consensus sequences from Bowtie2, Geneious and the chloroplast reference genome, excluding the second inverted repeat, were viewed with the 'highlight disagreements to the reference' setting in Geneious.

Table 2
Raw data information for each of the alignments used in phylogenetic analysis.

Taxa	Raw reads	Filtered reads	Mean coverage plastid protein-coding genes	Mean coverage mitochondrial genes, tRNAs, rRNAs	Mean coverage entire chloroplast sequence	Mean coverage entire mitochondrial sequence
Aglaonema costatum	19,455,075	16,097,991	89.4			
Aglaonema modestum	13,771,620		61.8			
Aglaonema nitidum	23,037,033	17,535,145	103.0	3247.9	88.1	1643.0
Alocasia fornicata	23,100,492		136.1			
Alocasia navicularis	12,580,339	10,496,587	75.6	381.1	64.9	197.6
Amorphophallus titanum	18,021,330	14,882,635	146.6	210.4	123.5	528.5
Anchomanes hookeri	28,743,540	22,307,306	1303.2	521.0	1123.8	389.1
Anthurium huixtlense	14,129,239	11,677,966	108.9	23.3	97.6	1612.3
Anubias heterophylla	24,011,002	20,266,562	191.8	362.3	180.8	346.7
Arisaema franchetianum	24,341,409	17,693,614	387.1	469.4	329.3	701.1
Arisarum simorrhinum	14,646,430	12,811,273	228.5	300.4	202.7	320.2
Calla palustris	16,715,852	14,608,316	938.3	275.9	832.1	469.4
Carlephyton glaucophyllum	17,921,059	15,258,012	410.2	49.5	367.1	566.3
Dieffenbachia parlatorei	11,306,711	9,358,638	173.4	539.1	162.5	273.1
Lasia spinosa	38,056,690	31,722,071	973.6	1515.0	816.4	1353.1
Monstera adansonii	36,278,235	27,002,232	774.9	761.1	727.0	1000.6
Montrichardia arborescens	12,493,655	11,108,076	192.7	204.9	183.8	220.6
Orontium aquaticum	20,860,738	18,136,283	99.2	145.1	81.8	299.2
Philodendron lanceolatum	21,202,424	17,651,991	291.0	404.3	276.6	339.2
Pinellia pedatisecta	8,856,861	7,309,388	1150.3	594.6	1064.5	482.0
Pinellia tripartita	20,898,685	15,048,528	526.7			
Pothos scandens	22,207,465	18,417,735	353.2	56.6	328.0	305.0
Rhaphidophora amplissima	12,870,626	11,122,937	115.8	136.6	130.2	213.8
Schismatoglottis calyptrata	13,869,294	12,054,112	62.2	77.4	53.3	95.4
Spathiphyllum patulinervum	13,898,684	12,465,700	34.4	534.0	29.7	544.8
Stenospermation multiovulatum	42,788,539	35,975,487	343.8	2388.5	321.0	2258.7
Steudnera colocasiifolia	13,038,295	11,332,048	100.5	116.0	93.2	233.0
Stylochaeton bognerii	12,709,376	11,211,182	67.7	477.6	65.7	2418.4
Syngonium angustatum	24,674,957	18,265,606	483.6	595.5	450.6	486.5
Taccarum caudatum	18,998,555	15,611,970	449.1	114.7	417.5	361.7
Typhonium blumei	21,868,376	15,754,513	955.2	1188.1	849.3	604.5
Ulearum donburnsii	24,551,954	20,392,535	159.7	808.7	151.0	834.4
Xanthosoma helleborifolium	11,667,350	9,823,510	492.6	145.8	458.0	303.7
Zamioculcas zamiifolia	43,288,898	32,819,219	540.0	2883.0	496.3	3808.8
Zantedeschia aethiopica	14,776,054	12,956,453	1553.4	221.4	1021.4	295.2
Zomicarpella amazonica	17,421,774	15,066,253	101.9	31.8	92.6	125.2
Mean coverage total			393.8	618.2	365.0	738.5
Total bases			61,716	113,181	211,614	318,210
Constant bases			45,615	95,153		
Parsimony uninformative bases			6335	11,871		
Parsimony informative bases			9766	6157		

Blank boxes denote genera that were excluded from the alignment due to redundancy.

Validation of sequences for each taxon, based on highlighted differences, was performed using the assemblies of raw mapped reads from both Bowtie2 and Geneious to ensure appropriate SNP calling and indel mapping. This combinatorial approach for validation using the consensus sequences and mapped reads from both programs was used only on those regions of the genome that were homologous with the annotated protein-coding sequences of *Lemna minor*. The restriction of sequence validation to these regions was due to the high level of conservation among plastid protein-coding sequences. Intergenic regions were too many and too variable to validate manually. This first round of assembly and validation permitted incorporation of many SNPs and indels that were lost in each single assembly, but ambiguities in more variable genes such as *ndhF* still remained. Therefore, the consensus sequences from the first round of assembly and validation for each taxon were then used as reference sequences in a subsequent round. After this second round, taxa still containing ambiguous sequences were assembled and validated reiteratively only in Geneious. Protein-coding sequences for each taxon were extracted and concatenated using the 'extract annotations' tool and were checked for start and stop codons using the 'translation' tool in Geneious. The concatenated protein-coding sequences for all taxa were aligned using the mauveAligner algorithm plugin in Geneious.

The focus on protein-coding sequences notwithstanding, certain genes proved to be too variable and labor intensive to validate even after several rounds (up to 10) of assembly and validation and were discarded from all taxa in the final alignment. Problematic genes, the species in which they occur, and a description of the issue are listed in Table 3. In total, 10 protein-coding genes (*infA*, *ycf68*, *rpl20*, *rps12*, *accD*, *clpP*, *rps19*, *rpl23*, *ycf1* and *rps15*) were removed from the final alignment, which consisted of 70 plastid protein-coding genes for 37 genera of Araceae and two species of *Acorus*. The final alignment can be accessed in TreeBASE at http://purl.org/phylo/treebase/phylows/study/TB2:S15395.

2.5. Sequence assembly and alignment of the entire chloroplast

We were interested in comparing the phylogenetic potential of a shotgun approach to obtaining complete chloroplast sequences versus the detailed validation approach (above). For this, the consensus sequences spanning the entire chloroplast from the second round of assembly and validation of coding sequences (above) were extracted for each taxon. Intergenic regions and introns spanning the entire genome, and coding sequences from the second inverted repeat were not validated. All consensus sequences were aligned using the mauveAligner algorithm plugin in Geneious.

2.6. Sequence assembly and alignment of mitochondrial sequences

Based on the well-supported phylogenies resulting from the shotgun approach to obtaining complete chloroplast sequences (refer to Section 3), we wanted to test whether large mitochondrial datasets obtained using this method could produce well-resolved phylogenies at the family level that are congruent with plastid data. For this, quality-trimmed and filtered reads were assembled to the *Spirodela polyrhiza* mitochondrial genome, which is the most compact monocot mitochondrial genome known to date (228,493 bases) (Wang et al., 2012). It contains a total of 57 genes encoding 35 proteins, 3 ribosomal RNAs and 19 transfer RNAs (Wang et al., 2012). Assembly was performed in Geneious with the custom sensitivity settings similar to those above except that the maximum gap size allowed was increased to 70 base pairs. We assembled reads to the mitochondrial genome three times reiteratively, taking the consensus sequence from each assembly as the reference for the subsequent assembly.

Assembly of raw reads to the mitochondrial reference genome averaged over all genera was 2.11% compared to 3.58% for the chloroplast. In general, assembly to the mitochondrial genome was notably more sporadic than assembly to the chloroplast genome even though mean coverage values are on par with those in the chloroplast genome (Table 2). A consistent theme among most genera was two islands of extreme depth coverage in the *nad4* and *nad2* genes (i.e. up to 8000× in a non-coding region of *nad2* in *Amorphophallus*), which may explain the high average coverage among mitochondrial assemblies. All genes encoding proteins, rRNAs and tRNAs for each taxon were extracted and concatenated using the 'extract annotations' tool in Geneious. Concatenated genes were aligned using the mauveAligner algorithm plugin in Geneious. For complete mitochondrial sequences, consensus sequences spanning the entire mitochondria were extracted for each taxon and aligned using the mauveAligner algorithm plugin in Geneious.

Concatenated plastid and mitochondrial gene alignments were analyzed using PAUPrat (Sikes and Lewis, 2001) in CIPRES to obtain the number of constant, parsimony-uninformative and parsimony-informative bases. For the mean coverage of raw reads for each

Table 3
Genes removed from the alignment of plastid protein-coding sequences.

Plastid gene	Species	Problem
accD	*Anchomanes hookeri, Anthurium huixtlense, Zantedeschia aethiopica*	Assembly toward beginning, no start/stop codon
clpP	*Stylochaeton bognerii*	Assembly throughout
infA	*Acorus americanus, Acorus calamus*	Present only in these species, a pseudo-copy in *Colocasia esculenta* (Ahmed et al., 2012)
rpl20	*Spirodela polyrhiza*	Absent from this genus
rpl23	*Anchomanes hookeri*	No start/stop codon
rps12	*Spirodela polyrhiza, Wolffia australiana, Wolfiella lingulata*	Absent from these genera
rps15	*Colocasia esculenta*, duckweeds	In IR[a] region in duckweed, in SSC[b] in *Colocasia esculenta* (Ahmed et al., 2012)
rps19	Anchomanes hookeri, Zantedeschia aethiopica	No start/stop codon
ycf1	*Aglaonema costatum, Aglaonema modestum, Aglaonema nitidum, Alocasia fornicata, Alocasia navicularis, Amorphophallus titanum, Anchomanes hookeri, Anthurium huixtlense, Calla palustris, Dieffenbachia parlatorei, Orontium aquaticum, Pinellia tripartita, Pothos scandens, Rhaphidophora amplissima, Stenospermation multiovulatum, Stylochaeton bognerii, Taccarum caudatum, Typhonium blumei, Ulearum donburnsii, Zantedeschia aethiopica, Zomicarpella amazonica*	Assembly, indels, no start/stop codons, in IR region in duckweed, in SSC in *Colocasia esculenta* (Ahmed et al., 2012)
ycf68	*Colocasia esculenta*	A pseudo-copy of *ycf68* reported in duckweed (Ahmed et al., 2012)

[a] Inverted repeat.
[b] Small single copy region.

taxon in each alignment, the mean coverage of raw reads for all taxa in each alignment and the number of total bases in each alignment refer to Table 2. For the number of constant bases, parsimony uninformative bases and parsimony informative bases in the concatenated gene alignments refer to Table 2.

2.7. Phylogenomic analyses

Phylogenetic analyses consisted of Maximum Likelihood and Bayesian analytical methods using Mr. Bayes version 3.1.2 (Huelsenbeck and Ronquist, 2001) and PhyloBench (Stamatakis et al., 2008). Maximum likelihood analysis, performed using RAxML HPC Black Box (Stamatakis, 2006a) consisted of 1000 rapid bootstrap inferences and a thorough ML search thereafter. The likelihood of the final tree was evaluated and optimized under the General Time Reversible substitution model with gamma-distributed rate variation across sites, and a proportion of invariant sites GTR + Γ + I (Stamatakis, 2006a; Yang, 1993). RAxML HPC Black Box uses the GTRCAT approximation of the GTR + Γ model with 25 per-site rate categories (Stamatakis, 2006b). The congruence between the ML phylogenetic trees based on complete plastid and complete mitochondrial sequences was compared using the Templeton test in PAUP (Swofford, 1991; Templeton, 1983).

Bayesian analysis was performed on plastid and mitochondrial concatenated gene sequences using Mr. Bayes version 3.1.2 using the GTR + Γ substitution model with the number of gamma categories set to 4. The GTR + Γ substitution model was chosen using the Akaike information criterion (Akaike, 1974) in jmodeltest2 (Darriba et al., 2012). The prior probability distribution for the substitution rates of the GTR model and the state frequencies was a flat Dirichlet. The prior for the shape parameter of the gamma distribution of rate variation was set to uniform. The prior for branch lengths was unconstrained and exponential. The analyses were run two times independently for 10,000,000 generations, sampling trees every 1000 generations. Three heated chains (temp = 0.200) and one cold chain were used. The first 25% of samples were discarded from the cold chain as burnin. Graphical exploration of MCMC convergence was performed using AWTY (Nylander et al., 2008; Wilgenbusch et al., 2004). The Bayesian consensus phylogenetic tree inferred from concatenated plastid protein-coding sequences can be accessed in TreeBASE at http://purl.org/phylo/treebase/phylows/study/TB2:S15395.

3. Results

3.1. Phylogenomic analyses of chloroplast sequences

Phylogenomic analyses based on chloroplast sequences, both complete and concatenated protein-coding, yielded similar strongly-supported family-wide phylogenies except for the placement of *Calla* and *Schismatoglottis* (Figs. 1 and 2). The following are novel strongly-supported evolutionary relationships. *Anubias* and *Montrichardia* form a clade (BS = 99%, PP = 0.99) that is the sister group to the *Zantedeschia* clade (BS = 100%, PP = 0.99) based on concatenated protein-coding (PC) sequences (Fig. 1). Although this topology does not change in the ML phylogeny inferred from complete chloroplast (C) sequences, the addition of the (*Calla*, *Schismatoglottis*) clade at its base decreases bootstrap support to 63% and 77%, respectively (Fig. 2). The *Zantedeschia* clade consists of a grade, with *Philodendron* (representing the *Homalomena* clade) as sister to the rest (PC: BS = 100%, PP = 0.99; C: BS = 100%), followed by Spathicarpeae (PC: BS = 67%, PP = 0.89; C: BS = 93%), then followed by the South African genus *Zantedeschia* as the sister taxon to the Old World *Anchomanes* clade (PC: BS = 73%, PP = 0.99; C: BS = 81% (Figs. 1 and 2).

The placement of *Calla* is consistently the only node in all phylogenies with bootstrap support less than 85% and a posterior probability less than 0.95 (Figs. 1 and 2). *Calla* is seen as either the sister taxon to the *Philonotion* clade (PC: BS = 58%, PP = 0.94) (Table 1), which includes *Schismatoglottis* (PC: BS = 64%, PP = 0.98) (Fig. 1), or forms a clade with *Schismatoglottis* (C: BS = 40%) that is sister to the ((*Montrichardia*, *Anubias*)*Zantedeschia* clade) (C: BS = 58%) (Fig. 2). The position of *Schismatoglottis* as sister to the other members of the *Philonotion* clade, seen in the phylogenies based on concatenated protein-coding sequences, is weakly supported in the ML analysis (BS = 64%) but strongly supported in the Bayesian analysis (PP = 0.98).

One strongly-supported clade presented here that was not seen in previous studies based on chloroplast sequences is tribe Spathiphylleae forms the sister taxon to the rest of subfamily Monsteroideae (PC: BS = 100%, PP = 0.99; C: BS = 100%) (Figs. 1 and 2). The Unisexual Flowers clade, containing the bisexually-flowered genus *Calla*, is also strongly supported for the first time using plastid data alone (PC: BS = 100%, PP = 0.99; C: 100%) (Fig. 2).

3.2. Comparison of chloroplast and mitochondrial phylogenies

In contrast to the strongly supported phylogenies obtained from chloroplast data, phylogenies based on mitochondrial sequences did not have strong statistical support, with the exception of several small clades (Fig. 2). In the phylogenies inferred from concatenated mitochondrial genes (data not shown), strongly-supported clades were limited to the *Rhaphidophora* clade (BS = 100%, PP = 1.0), Spathicarpeae (BS = 100, PP = 1.0), Pothoideae (BS = 33%, PP = 1.0), the (*Pinellia*(*Arisaema*, *Typhonium*)) clade (BS = 100%, 100%, PP = 1.0, 1.0, respectively) and a highly morphologically incongruent clade composed of *Aglaonema* and *Spirodela* (BS = 100%, PP = 1.0). The Maximum Likelihood analysis based on complete mitochondrial sequences produced a greater number of clades with strong bootstrap support than did the analysis based on mitochondrial genes, but still not many (Fig. 2). These include Pothoideae, Monsteroideae, Spathicarpeae, Caladieae, the *Pistia* clade, and the (*Pinellia*(*Arisaema*, *Typhonium*)) clade. All have a bootstrap support value of 100%, except Pothoideae with BS = 85%. No other node in the phylogeny is statistically well supported, but the relationships among the seven subfamilies mirror those in the chloroplast phylogeny with one exception. Lasioideae, containing bisexually-flowered taxa, is sister to the Bisexual Climbers clade, composed of Pothoideae and Monsteroideae, instead of sister to the Unisexual Flowers clade. This placement of Lasioideae is interesting in that it makes all bisexually-flowered taxa (with the exception of *Calla*) within True Araceae monophyletic.

Within subfamily Aroideae, the *Dracunculus* clade remains wholly intact except for the rearrangement of *Xanthosoma* and *Zomicarpella*. Relationships among the *Zantedeschia* clade, *Montrichardia*, *Anubias*, *Calla* and *Schismatoglottis* do not match those in either of the chloroplast phylogenies (Figs. 1 and 2). *Calla* and *Schismatoglottis* (with the addition of *Montrichardia*) form a clade that is sister to the rest of Aroideae. Although smaller clades found in the complete chloroplast phylogeny are recovered in the complete mitochondrial phylogeny, and many relationships among genera and clades appear superficially similar, the results of the Templeton test reject congruency between the two topologies ($p \leqslant 0.0001$).

3.3. Synapomorphic indels in chloroplast protein-coding genes for major named taxa

Various indels notable among the clades in the alignment of chloroplast protein-coding sequences of Araceae and *Acorus* are listed in Table 1. As more genera are sequenced and added to the

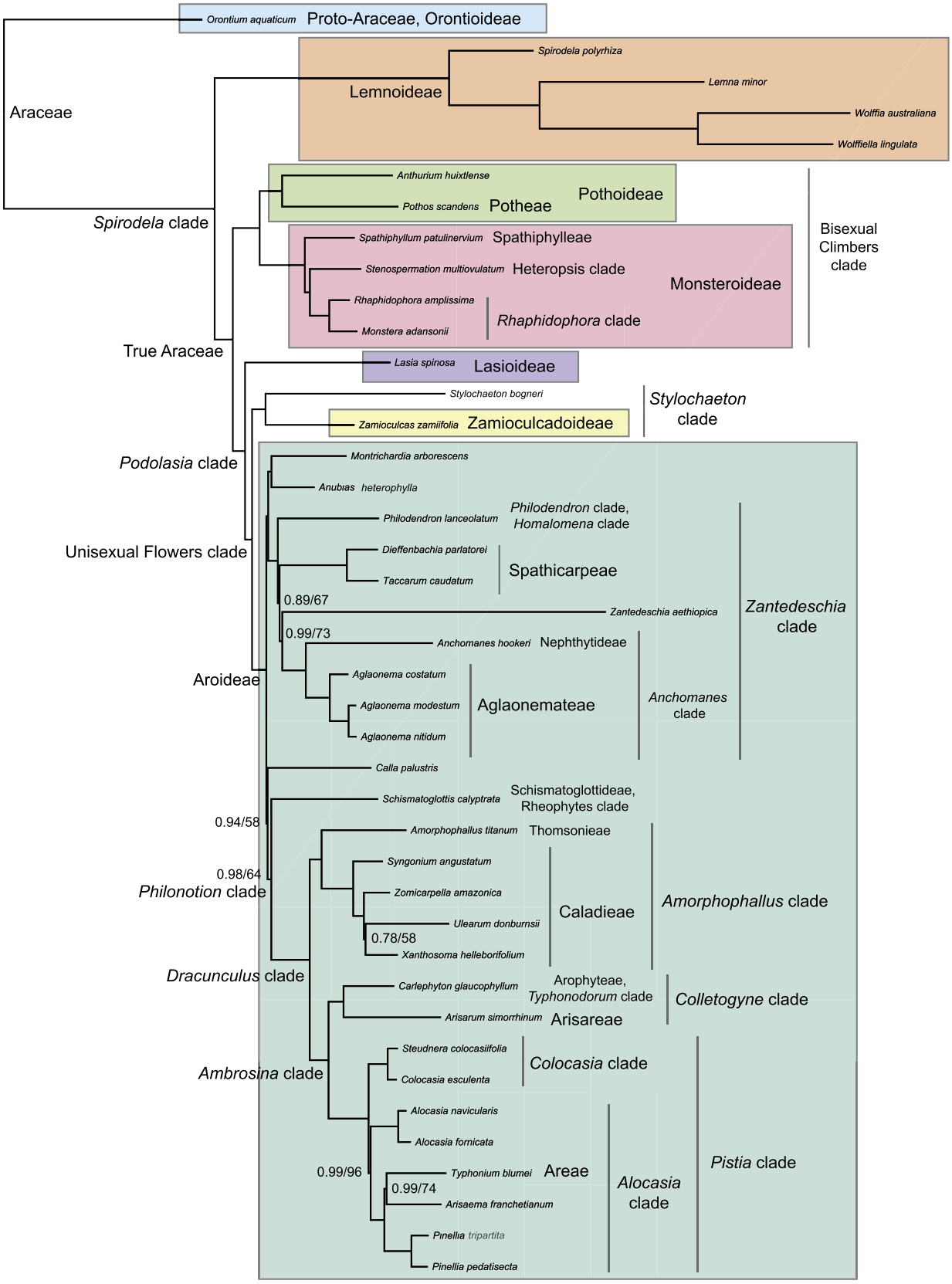

Fig. 1. Phylogenetic tree obtained from ML and Bayesian analysis of 70 plastid protein-coding genes for 37 genera of Araceae and two species of *Acorus*, used as the outgroup (not shown). Subfamilies are boxed (excluding Gymnostachydoideae). Nodes with no values have a posterior probability $\geqslant$0.98 and bootstrap support $\geqslant$99%.

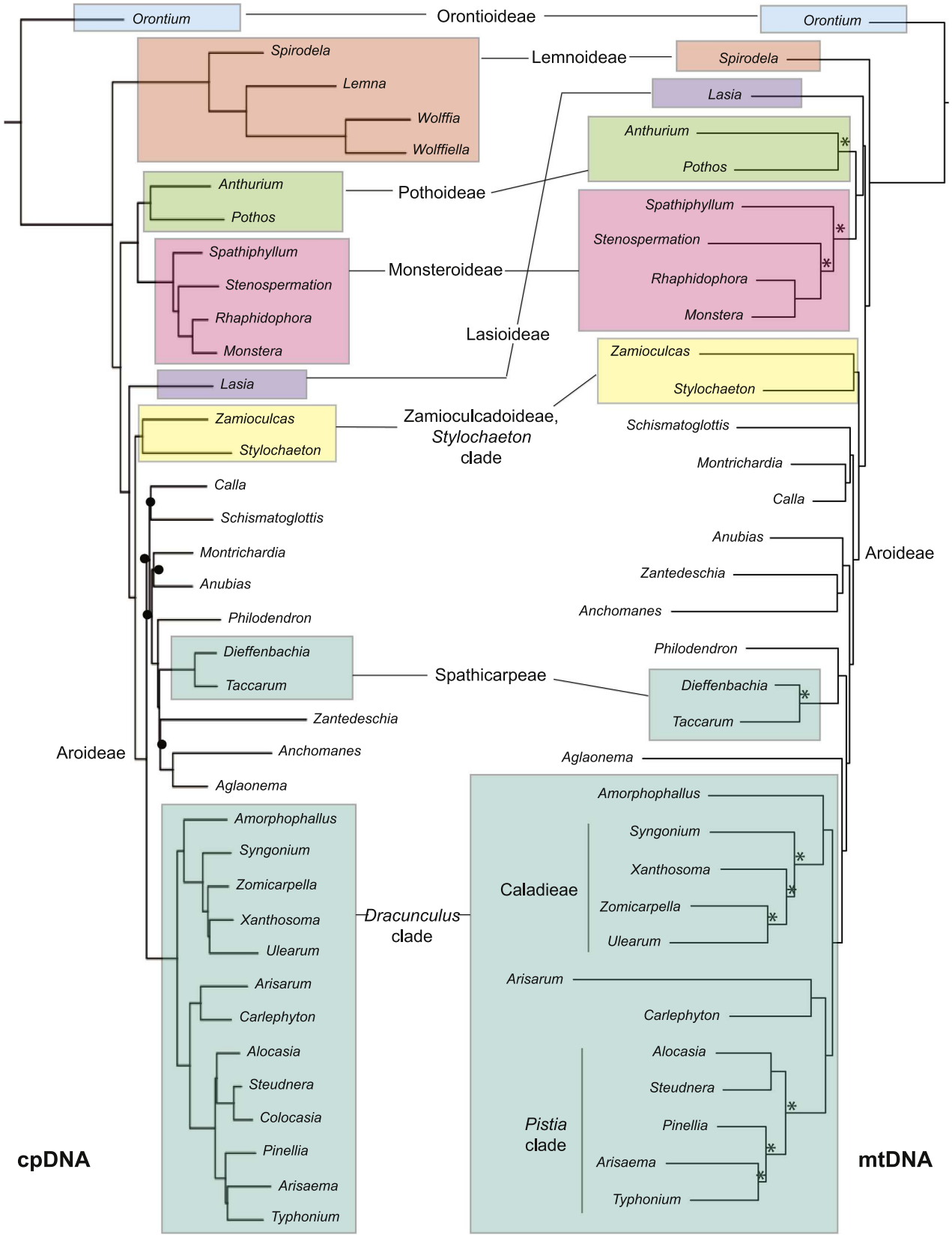

Fig. 2. Best-scoring RAxML phylogenies based on complete chloroplast and mitochondrial sequences. Closed circles mark nodes with <85% bootstrap support in the plastid tree, asterisks nodes with ⩾85% bootstrap support in the mitochondrial tree. Boxes mark clades recovered in both analyses, colors correspond to subfamilies.

phylogeny, it will be interesting to discover if these indels hold as synapomorphies.

Synapomorphic indels for Araceae include a 22 bp insertion at the beginning of *rpoC1* (*Anchomanes hookeri* has an additional 5 bp insertion), a 36 bp deletion at the beginning of *rpoB*, a 69 bp insertion at the beginning of *ndhK* (*Stylochaeton bogneri* has an additional 3 bp insertion) and a 214 bp deletion at the beginning of *cemA*. Members of the *Spirodela* clade, formed by Lemnoideae and the True Araceae clade, have no morphological synapomorphies, but all have a 9 bp deletion at the end of the second exon of *atpF*. In Lasioideae, *Lasia spinosa* has a 2145 bp or 715 amino acid deletion at the end of *rpoC2*. The *Anchomanes* clade, formed by *Anchomanes hookeri* and three species of *Aglaonema*, has a 6 bp insertion toward the beginning of *petA*. The *Pistia* clade, containing *Colocasia*, *Steudnera*, *Alocasia*, *Typhonium*, *Pinellia* and *Arisaema*, has a 3 bp insertion in *atpE* followed closely by a transition SNP from A to G. Tribe Spathicarpeae, formed by *Dieffenbachia* and *Taccarum*, has a 15 bp insertion at the end of *matK*. *Philodendron*, representing the *Philodendron* and *Homalomena* clades, has a 14 bp insertion toward the end of *matK*.

4. Discussion

4.1. Evolutionary relationships of Araceae

This study provides the first well-supported phylogeny based on chloroplast sequences for the early evolution of Araceae, particularly the early evolution of the generically rich subfamily Aroideae. It is also the first glimpse at a mitochondrial phylogeny for the family. Most of our results corroborate previously established phylogenetic relationships, however, several key findings pertaining to the early evolution of Aroideae differ greatly from previous studies.

Our results support the current circumscription of Araceae into eight subfamilies: Gymnostachydoideae, Orotioideae, Lemnoideae, Pothoideae, Monsteroideae, Lasioideae, Zamioculcadoideae and Aroideae (for Gymnostachydoideae refer to Cabrera et al., 2008; Chartier et al., 2013). Relationships among the subfamilies containing bisexually-flowered genera, Orontioideae, Lemnoideae, Pothoideae and Monsteroideae are strongly supported. Although there are no morphological synapomorphies for the *Spirodela* clade, this study suggests a 9 base pair deletion in the plastid gene *atpF* may be a diagnostic synapomorphic indel for the group. The sister relationship of Lemnoideae and True Araceae (Table 1) within the *Spirodela* clade is well established in this and previous studies, as is the sister relationship of the Bisexual Climbers and *Podolasia* clades (Fig. 1).

Within the Bisexual Climbers clade, the placement of Spathiphylleae in subfamily Monsteroideae as sister to a clade containing the *Rhaphidorphora* and *Heteropsis* clades has not been observed previously in chloroplast phylogenies (Figs. 1 and 2). This result, distancing Spathiphylleae from other members of Monsteroideae evolutionarily, is not surprising given the tribe's unique morphological features such as lacking calcium oxalate prisms, vessels, a stem endodermis and cortical vascular system, and in having pollen and trichosclereids unlike any others in Araceae (Grayum, 1990).

The Unisexual Flowers clade, containing a sister group relationship between the *Stylochaeton* clade and subfamily Aroideae, is here strongly supported. In our chloroplast phylogeny (Fig. 1), *Stylochaeton bogneri* is the only taxon not included in one of the eight subfamilies. We agree with previous workers that subfamily Zamioculcadoideae should be expanded to include *Stylochaeton*, thus characterizing the former as consisting of geophytic, sub-saharan African plants that have perigoniate, unisexual flowers and lack laticifers.

In Aroideae, within the *Zantedeschia* clade, *Philodendron* is sister to all other genera. *Philodendron* here represents the *Homalomena* clade, which have the morphological synapomorphies of the occurrence of sclerotic hypodermis and resin canals in the roots and absence of endothecial thickenings in the anthers (Cusimano et al., 2011). The South African genus *Zantedeschia* is sister to the *Anchomanes* clade formed by African Nephthytideae (tuberous or rhizomatous, seasonally dormant to evergreen) and Asian Agalonemateae (entirely evergreen), which share an inferred ancestral haploid chromosome number of twenty (Cusimano et al., 2012; Mayo et al., 1997). *Zantedeschia* has an inferred ancestral haploid chromosome number of $n = 16$, and the species are seasonally dormant, occasionally evergreen, tuberous herbs. This arrangement, seen here for the first time, makes the whole group strictly Old World and is biogeographically more congruent than previous studies in which the genus *Zantedeschia* is sister to the strictly South American Spathicarpeae. All members of Spathicarpeae have an inferred ancestral haploid chromosome number of seventeen (Cusimano et al., 2012).

The phylogenetic position of *Calla*, a genus with a unique character combination in the family, has shifted dramatically among the various studies attempting to resolve its evolutionary history (Ulrich et al., 2013). Morpho-anatomical and palynological data suggest that *Calla* belongs in a lineage by itself in a transition zone between bisexually- and unisexually-flowered clades, while previous molecular data suggests that *Calla* is embedded in the Unisexual-Flowers clade but its placement therein is unresolved (Ulrich et al., 2013). This study unequivocally supports the inclusion of *Calla* in the Unisexual Flowers clade, but presents yet another hypothesis, albeit poorly-supported, of its evolutionary relationship to other unisexually-flowered genera (Fig. 2). Interestingly, the phylogeny based on complete chloroplast sequences in this study is similar to the strict consensus tree from the combined parsimony analysis of chloroplast data by Cabrera et al. (2008) in that *Calla* and the Rheophytes clade (represented by *Schismatoglottis*) form a sister relationship. In Cabrera et al. (2008) the (*Calla,* Rheophytes clade) clade is at the base of what is now accepted as Aroideae, whereas in this study that clade is at the base of one of the two major clades forming Aroideae. The sister relationship between *Anubias*, *Montrichardia* and *Calla* at the base of Aroideae, as seen in the nuclear data (Chartier et al., 2013), was not recovered in either of our chloroplast phylogenies, but a variation of it was seen with low statistical support in the mitochondrial phylogeny (Fig. 2). Although the addition of the (*Calla*, Rheophytes clade) clade at the base of the ((*Montrichardia*, *Anubias*) *Zantedeschia* clade)) clade weakens the statistical support for the latter, this topology is supported by morphological and cytological features. Rare modifications of the leaf sheath into "ligule-" or "stipulelike" structures, in which the leaf sheath is free at the tips, are shared by *Calla*, several Schismatoglottideae (Table 1), and some *Philodendron* species (Grayum, 1990). *Calla* also shares the morphological characters of unilocular ovules and basal placentation with *Nephthytis* (Nephthytideae, here represented by *Anchomanes*) and *Callopsis* (Grayum, 1990). Furthermore, *Calla*, *Philodendron*, and members of the Rheophytes clade (*Lagenandra* and *Cryptocoryne*) share an inferred ancestral haploid chromosome number of $n = 18$. In fact, the larger ((*Calla*, Rheophytes clade) ((*Montrichardia*, *Anubias*) *Zantedeschia* clade)) clade is reminiscent of Grayum's morphology-based circumscription of subfamily Calloideae Schott, including 14 of his 17 tribes (only tribes Peltandreae, Arophyteae and Callopsideae fall out elsewhere).

The inclusion of *Calla* in the Unisexual Flowers clade implies a return from unisexual to bisexual flowers – a transition that is exceedingly rare (Barrett, 2013). However, the multiple developmental pathways leading to unisexual flowers and the retention of sexual bipotency in many unisexual flower primordia suggest

that sex determination in floral organs is a much more labile process than previously recognized (Dellaporta and Calderon-Urrea, 1993; Mitchell and Diggle, 2005). The presence of male flowers at the tips of spadices above the normal bisexual flowers in *Calla* and *Orontium* (Grayum, 1990) attest to this notion of 'sex flexibility' and make the placement of *Calla* within the Unisexual Flowers clade more tenable. A closer look into the developmental pathways of floral organ evolution in Araceae is highly desired.

4.2. Phylogenomics: chloroplasts vs. mitochondria

The structural and mutational complexities of land-plant mitochondrial genomes and the difficulties they present for phylogenetic analyses have long been recognized (Petersen et al., 2006; Seberg and Petersen, 2006; Seberg et al., 2012). The results of this study show that indeed, even when using tens of millions of Illumina sequencing reads from total genomic DNA, the shotgun approach has vastly different potential for phylogenomics in plastid versus mitochondrial genomes. The shotgun approach, using the entire chloroplast genome as a reference, mainly had results as strongly supported as those of the carefully validated concatenated protein-coding sequences, with few inconsistencies between the two. In contrast, although many of the clades seen in the chloroplast phylogeny were recovered in the phylogeny based on complete mitochondrial sequences, the statistical support was too low in the latter to draw any meaningful conclusions about most generic relationships in the family. In addition, what appears at face value to be considerable similarity between two organellar phylogenies, is strongly rejected when scrutinized methodically.

In this study, the whole suite of mitochondrial genes was, *on average*, less informative than chloroplast protein-coding genes for a family-level phylogeny in Araceae. It is interesting to note that in spite of the questionable homology of intergenic regions in plant mitochondria the alignment including these genomic regions yielded a better-supported topology than did the combined-gene alignment alone; a possible explanation could be the small size of the *Spirodela* mitochondrion. Reference assemblies based on mitochondria are, in general, more problematic than chloroplast reference assemblies (Argelia Cuenca pers. comm.). Previous workers have shown that excluding predicted RNA-edited sites in mitochondrial genes increases congruence between mitochondrial and plastid phylogenies (Petersen et al., 2013). Comparison of reference-based and de novo assemblies, and a look into the role of RNA-edited sites and processed paralogs in the mitochondrial phylogeny of Araceae are yet to be studied.

4.3. Problematic genes and Zantedeschia

Ten genes (*infA*, *ycf68*, *rpl20*, *rps12*, *accD*, *clpP*, *rps19*, *rpl23*, *ycf1* and *rps15*) were removed from the final concatenated plastid protein-coding sequence alignment because of either problematic assembly of raw reads and/or uncertainty of their presence or absence. In reference-based assembly, structural changes between the reference and target genomes are not captured and regions of exceptionally high variation are difficult to assemble. Several genes listed above have already been confirmed as pseudogenes in other angiosperms (*rpl23* in spinach, *ycf1* in rice and maize, *infA* in tobacco, *Arabidopsis* and *Oenothera*, *accD* in grasses) (Millen et al., 2001). In the case of *accD*, a study of mutation rates in eudicot legume chloroplast genomes showed that the *accD-psal-ycf4-cemA* region was hypermutable and that *accD* was transferred to the nucleus in *Trifolium* (Magee et al., 2010). Here, the hypermutability of the gene *accD* was observed in phylogenetically independent genera of Araceae (*Anthurium*, *Zantedeschia* and *Anchomanes*) and based on evidence in legumes, the transfer of *accD* to the nucleus in these genera may be a possibility. A similar pattern is seen in the *clpP* gene of *Stylochaeton*. Interestingly, *clpP* and *accD* were found to be essential for shoot and leaf development, respectively, in the eudicot tobacco, yet *accD* is unnecessary during development in grasses, a highly derived monocot group (Kode et al., 2005). These contradictory roles make the study of *accD* in Araceae especially intriguing, considering the leaf developmental patterns in the family that are transitional between 'dicots' and monocots (Bharathan, 1996).

The gene *infA* is among the most easily lost in the chloroplast genomes of land plants (Millen et al., 2001). Both species of *Acorus* used in this study possess the gene but not one of the duckweed genera obtained from GenBank has it. Using the sequence of *infA* from *Acorus americanus* as a reference, we found varying levels of coverage of the gene in phylogenetically disparate genera of Araceae. Coverage ranged from complete (*Orontium*), intermediate (*Calla*, *Lasia*, and *Anubias*) to absent (*Anthurium* and *Spathiphyllum*) and suggests that the mutational dynamics of *infA* have not stabilized within Araceae. The list of ten genes omitted from the final concatenated plastid protein-coding sequence alignment serves as a basis for future work into the structural and functional properties of these plastid genes in Araceae.

Of all the chloroplast genomes assembled in this study, *Anchomanes* and *Zantedeschia* proved to be the most problematic. The amount of autapomorphic substitutions in the chloroplast genome of *Zantedeschia* was surprisingly high, with the associated branch length almost as long as those in the highly morphologically derived duckweed subfamily Lemnoideae. Interestingly, the species of *Zantedeschia* used in this study (*Z. aethiopica*) is morphologically distinct from the other seven species in the genus (Mayo et al., 1997).

5. Conclusions

This study presents the first well-supported phylogeny for deep branches of the plant family Araceae using strictly chloroplast data, and the first glimpse at a family-wide phylogeny based on mitochondrial sequences. New evolutionary relationships seen in this study, the mutational dynamics of several plastid protein-coding genes and a comparison of chloroplast and mitochondrial sequences for phylogenomics are discussed. Although sampling was sufficient to resolve the relationships between the major clades in the family, the lack of sampling of several key genera including *Pistia*, *Protarum*, *Philonotion* and especially *Callopsis* leaves room for future work. As more nuclear data become available for the family, it will be interesting to see the ultimate placement of *Calla* and *Schismatoglottis* (Rheophytes clade).

Acknowledgments

We would like to thank Geneviève Ferry for sending the silica sample of *Calla palustris* from the Nancy Botanical Garden. We would like to thank Pat Edgar, Dustin Mayfield, Sidonie Bellot and Pu Huang for helpful conversations on bioinformatics. We thank Susanne Renner, Lars Nauheimer, Josef Bogner and Peter Boyce for helpful comments on Araceae taxonomy and systematics. We thank Susanne Renner for helpful suggestions on mitochondrial data. We thank Toby Kellogg, Peter Stevens, Kenneth Olsen, Kate Waselkov, Allan Larson and Alan Templeton for helpful comments on systematics. We thank Monica Carlsen for help with Mesquite and TreeBASE. We also thank Emily Colletti in the Aroid Greenhouse for help with living material. Funding for this study was provided by the GAANN fellowship, the Rettner B. Morris Scholarship, Washington University in St. Louis, J. Chris Pires Lab (NSF DEB 1146603).

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. 2014.02.017.

References

Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. Automat. Contr. 19, 716–723.

Ahmed, I., Biggs, P.J., Matthews, P.J., Collins, L.J., Hendy, M.D., Lockhart, P.J., 2012. Mutational dynamics of aroid chloroplast genomes. Genome Biol. Evol. 4 (12), 1316–1323.

Barabé, D., Bruneau, A., Forest, F., Lacroix, C., 2002. The correlation between development of atypical bisexual flowers and phylogeny in the Aroideae (Araceae). Plant. Syst. Evol. 232, 1–19.

Barrett, S.C.H., 2013. The evolution of plant reproductive systems: how often are transitions irreversible? Proc. R. Soc. London B 280, 20130913.

Bharathan, G., 1996. Does the monocot mode of leaf development characterize all monocots? Aliso 14 (4), 271–279.

Bogner, J., 1978. A critical list of the aroid genera. Aroideana 1 (3), 63–73.

Bogner, J., Nicolson, D.H., 1991. A revised classification of Araceae with dichotomous keys. Willdenowia 21, 35–50.

Bown, D., 2000. Aroids: Plants of the *Arum* family, second ed. Timber Press, Portland, Oregon, USA.

Boyce, P.C., Croat, T.B., 2013. The Überlist of Araceae, totals for published and estimated number of species in aroid genera. <http://www.aroid.org/genera/130307uberlist.pdf> (accessed 01.14).

Cabrera, L.I., Salazar, G.A., Chase, M.W., Mayo, S.J., Bogner, J., Dávila, P., 2008. Phylogenetic relationships of aroids and duckweeds (Araceae) inferred from coding and noncoding plastid DNA. Am. J. Bot. 95 (9), 1153–1165.

Chartier, M., Gibernau, M., Renner, S.S., 2013. The evolution of pollinator-plant interaction types in the Araceae. Evolution. http://dx.doi.org/10.1111/evo.12318.

Cox, M.P., Peterson, D.A., Biggs, P.J., 2010. SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. BMC Bioinform. 11, 485.

Croat, T.B., 1988. Ecology and life-forms of Araceae. Aroideana 11, 4–56.

Cusimano, N., Bogner, J., Mayo, S.J., Boyce, P.C., Wong, S.Y., Hesse, M., Hetterscheid, W.L.A., Keating, R.C., French, J.C., 2011. Relationships within the Araceae: comparison of morphological patterns with molecular phylogenies. Am. J. Bot. 98 (4), 654–668.

Cusimano, N., Sousa, A., Renner, S.S., 2012. Maximum likelihood inference implies a high, not a low, ancestral haploid chromosome number in Araceae, with a critique of the bias introduced by ‘‘x’’. Ann. Bot. 109 (4), 681–692.

Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2012. JModelTest 2: more models, new heuristics and parallel computing. Nat. Meth. 9 (8), 772.

Dellaporta, S.L., Calderon-Urrea, A., 1993. Sex determination in flowering plants. Plant Cell 5, 1241–1251.

Engler, A., 1920. Araceae—Pars generalis et index familiae generalis. In: Engler, A. (Ed.), Das Pflanzenreich 74 (IV.23A), pp. 1–71.

French, J.C., Chung, M.G., Hur, Y.K., 1995. Chloroplast DNA phylogeny of the Ariflorae. In: Rudall, P.J., Cribb, P.J., Cutler, D.F., Humphries, C.J. [Eds.], Monocotyledons: Systematics and Evolution, vol. 1. Royal Botanic Gardens, Kew, UK. pp. 255–275.

Gauthier, M.-P.L., Barabe, D., Bruneau, A., 2008. Molecular phylogeny of the genus Philodendron (Araceae): delimitation and infrageneric classification. Bot. J. Linn. Soc. 156 (1), 13–27.

Geneious version (5.6–6.0.3) created by Biomatters. <http://www.geneious.com/>.

Givnish, T.J., Ames, M., McNeal, J.R., McKain, M.R., Steele, P.R., dePamphilis, C.W., Graham, S.W., Pires, J.C., Stevenson, D.W., Zomlefer, W.B., Briggs, B.G., Duvall, M.R.Moore, M.J., Heaney, J.M., Soltis, D.E., Soltis, P.S., Thiele, K., Lebbens-Mack, J.H., 2010. Assembling the tree of the monocotyledons: plastome sequence phylogeny and evolution of Poales 1. Ann. Mo. Bot. Gard 97 (4), 584–616.

Gonçalves, E.G., 2004. Araceae from Central Brazil: Comments on Their Diversity and Biogeography. Ann. Missouri Bot. Gard. 91, 457–463.

Gonçalves, E.G., Mayo, S.J., Van Sluys, M.-A., Salatino, A., 2007. Combined genotypic-phenotypic phylogeny of the tribe Spathicarpeae (Araceae) with reference to independent events of invasion to Andean regions. Mol. Phylogenet. Evol. 43 (3), 1023–1039.

Goremykin, V.V., Holland, B., Hirsch-Ernst, K.I., Hellwig, F.H., 2005. Analysis of *Acorus calamus* chloroplast genome and its phylogenetic implications. Mol. Biol. Evol. 22 (9), 1813–1822.

Grayum, M.H., 1987. A summary of evidence and arguments supporting the removal of *Acorus* from Araceae. Taxon 36, 723–729.

Grayum, M.H., 1990. Evolution and phylogeny of the Araceae. Ann. Mo. Bot. Gard. 77 (4), 628.

Hooker, J.F., 1883. Aroideae. In: Bentham, G., Hooker, J.D. (Eds.), Genera Plantarum, vol. 3, Reeve, London, UK. pp. 955–1000.

Hotta, M., 1970. A system of the family Araceae in Japan and adjacent areas. Memoirs of the Faculty of Science of Kyoto Imperial University. Series Biol. 4, 72–96.

Huelsenbeck, J.P., Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics (Oxford, England) 17 (8), 754–755.

Hutchinson, J., 1973. The Families of Flowering Plants. Clarendon Press, Oxford, UK.

Jussieu, A.L. de (1789). Genera plantarum. Paris (Araceae on pp. 23–25) 498 pp.

Keating, R.C., 2003. Leaf anatomical characters and their value in understanding morphoclines in the araceae. Bot. Rev. 68, 510–523.

Knoop, V., Volkmar, U., Hecht, J., Grewe, F., 2011. Mitochondrial Genome Evolution in the Plant Lineage. In: Kempken, F., (Ed.), Plant Mitochondria, Advances in Plant Biology 1.

Kode, V., Mudd, E.A., Iamtham, S., Day, A., 2005. The tobacco plastid accD gene is essential and is required for leaf development. Plant J. 44 (2), 237–244.

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Meth. 9 (4), 357–359.

Magee, A.M., Aspinall, S., Rice, D.W., Cusack, B.P., Sémon, M., Perry, A.S., Stefanovic, S., Milbourne, D., Barth, S., Palmer, J.D., Gray, J.C., Kavanagh, T.A., Wolfe, K.H., 2010. Localized hypermutation and associated gene losses in legume chloroplast genomes. Genom. Res. 20 (12), 1700–1710.

Mardanov, A.V., Ravin, N.V., Kuznetsov, B.B., Samigullin, T.H., Antonov, A.S., Kolganova, T.V., Skyabin, K.G., 2008. Complete sequence of the duckweed (Lemna minor) chloroplast genome: structural organization and phylogenetic relationships to other angiosperms. J. Mol. Evol. 66 (6), 555–564.

Mayo, S.J., Bogner, J., Boyce, P.C., 1997. The Genera of Araceae. Royal Botanic Gardens, Kew, UK.

Millen, R.S., Olmstead, R.G., Adams, K.L., Palmer, J.D., Lao, N.T., Heggie, L., Kavanagh, T.A., Hibberd, J.M., Gray, J.C., Morden, C.W., Calie, P.J., Jermiin, L.S., Wolfe, K.H., 2001. Many parallel losses of infA from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. Plant Cell 13 (3), 645–658.

Mitchell, C.H., Diggle, P.K., 2005. The evolution of unisexual flowers: morphological and functional convergence results from diverse developmental transitions. Am. J. Bot. 92 (7), 1068–1076.

Mower, J.P., Touzet, P., Gummow, J.S., Delph, L.F., Palmer, J.D., 2007. Extensive variation in synonymous substitution rates in mitochondrial genes of seed plants. BMC Evol. Biol. 7, 135.

Nakai, T., 1943. Araceae. In: Nakai, T., (Ed.), Ordines, familiae, tribi, genera, sectiones, species, varietates, formae et combinations novae a Prof. Nakai-Takenosin adhuc ut novis edita. Tokyo, Japan. pp. 214–221.

Nauheimer, L., Metzler, D., Renner, S.S., 2012. Global history of the ancient monocot family Araceae inferred with models accounting for past continental positions and previous ranges based on fossils. New Phytol. 195, 938–950.

Nicolson, D.H., 1987. History of Aroid Systematics. Aroideana 10 (4), 23–30.

Nylander, J.A.A., Wilgenbusch, J.C., Warren, D.L., Swofford, D.L., 2008. AWTY (are we there yet?): a system for graphical exploration of MCMC convergence in Bayesian phylogenetic inference. Bioinformatics 24 (4), 581–583.

Palmer, J.D., Adams, K.L., Cho, Y., Parkinson, C.L., Qui, Y.-L., Song, K., 2000. Dynamic evolution of plant mitochondrial genomes: mobile genes and introns and highly variable mutation rates. Proc. Natl. Acad. Sci. USA 97 (13), 6960–6966.

Petersen, G., Seberg, O., Davis, J.I., Goldman, D.H., Stevenson, D.W., Campbell, L.M., Michelangeli, F.A., Specht, C.D., Chase, M.W., Fay, M.F., Pires, J.C., Freudenstein, J.V., Hardy, C.R., Simmons, M.P., 2006. Mitochondrial data in monocot phylogenetics. Aliso 22, 52–62.

Petersen, G., Seberg, O., Davis, J.I., 2013. Phylogeny of the Liliales (Monocotyledons) with special emphasis on data partition congruence and RNA editing. Cladistics 29, 274–295.

Peery, R.M., Chumley, T.W., Kuehl, J.V., Boore, J.L., Raubeson, L.A., 2007. The complete chloroplast genome of *Acorus americanus*. NCBI.

Renner, S.S., Zhang, L.-B., 2004. Biogeography of the Pistia Clade (Araceae): based on chloroplast and mitochondrial DNA sequences and bayesian divergence time inference. Syst. Biol. 53 (3), 422–432.

Renner, S.S., Zhang, L.-B., Murata, J., 2004. A chloroplast phylogeny of *Arisaema* (Araceae) illustrates tertiary floristic links between Asia, North America, and East Africa. Am. J. Bot. 91 (6), 881–888.

Richards, T.A., Soanes, D.M., Foster, P.G., Leonard, G., Thornton, C.R., Talbot, N.J., 2009. Phylogenomic analysis demonstrates a pattern of rare and ancient horizontal gene transfer between plants and fungi. Plant Cell 21 (7), 1897–1911.

Richardson, A.O., Rice, D.W., Young, G.J., Alverson, A.J., Palmer, J.D., 2013. The ‘‘fossilized’’ mitochondrial genome of Liriodendron tulipifera: ancestral gene content and order, ancestral editing sites, and extraordinarily low mutation rate. BMC Evol. Biol. 11, 29.

Rothwell, G.W., Van Atta, M.R., Ballard Jr., H.E., Stockey, R.A., 2004. Molecular phylogenetic relationships among Lemnaceae and Araceae using the chloroplast trnL–trnF intergenic spacer. Mol. Phylogenet. Evol. 30 (2), 378–385.

Schmieder, R., Edwards, R., 2011. Quality control and preprocessing of metagenomic datasets. Bioinformatics (Oxford, England) 27 (6), 863–864.

Seberg, O., Petersen, G., 2006. Mitochondrial DNA Sequences in Plant Phylogenetics and Evolution-symposium at XVII IBC, Vienna, Austria. Taxon 55 (4), 833–835.

Seberg, O., Petersen, G., Davis, J.I., Pires, J.C., Stevenson, D.W., Chase, M.W., Fay, M.F., Devey, D.S., Jorgensen, T., Sytsma, K.J., Pillon, Y., 2012. Phylogeny of the Asparagales based on three plastid and two mitochondrial genes. Am. J. Bot. 99 (5), 875–889.

Sikes, D.S., Lewis, P.O., 2001. Beta software, version 1. PAUPRat: PAUP* implementation of the parsimony ratchet. Distributed by the authors. Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, USA.

Simpson, M.G., 2006. Plant Systematics. Elsevier Inc., Burlington, MA, USA.

Stamatakis, A., 2006a. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics (Oxford, England) 22 (21), 2688–2690.

Stamatakis, A., 2006b. Phylogenetic models of rate heterogeneity: a high performance computing perspective. In: Proceedings of the 20th IEEE/ACM International Parallel and Distributed Processing Symposium (IPDPS2006), Rhodos, Greece.

Stamatakis, A., Hoover, P., Rougemont, J., 2008. A rapid bootstrap algorithm for the RAxML web servers. Syst. Biol. 57 (5), 758–771.

Steele, P.R., Hertweck, K.L., Mayfield, D., McKain, M.R., Leebens-Mack, J., Pires, J.C., 2012. Quality and quantity of data recovered from massively parallel sequencing: examples in Asparagales and Poaceae. Am. J. Bot. 99 (2), 330–348.

Stevens, P.F. (2001 onwards). Angiosperm Phylogeny Website. Version 12, July 2012 [and more or less continuously updated since]. <http://www.mobot.org/MOBOT/research/APweb/>.

Swofford, D.L., 1991. PAUP: Phylogenetic Analysis Using Parsimony, Version 3.1 Computer program distributed by the Illinois Natural History Survey, Champaign, Illinois.

Tam, S.-M., Boyce, P.C., Upson, T.M., Barabé, D., Bruneau, A., Forest, F., Parker, J.S., 2004. Intergeneric and infrafamilial phylogeny of subfamily Monsteroideae (Araceae) revealed by chloroplast *trnL–F* sequences. Am. J. Bot. 91, 490–498.

Templeton, A.R., 1983. Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and the apes. Evolution 37 (2), 221–244.

Tobe, H., Kadokawa, T., 2010. Endosperm development in the Araceae (Alismatales) and evolution of developmental modes in monocots. J. Plant Res. 123 (6), 731–739.

Ulrich, S., Hesse, M., Bröderbauer, D., Bogner, J., Weber, M., Halbritter, H., 2013. *Calla palustris* (Araceae): new palynological insights with special regard to its controversial systematic position and to closely related genera. Taxon 62 (4), 701–712.

Wang, W., Messing, J., 2011. High-throughput sequencing of three lemnoideae (duckweeds) chloroplast genomes from total DNA. PLoS One 6 (9), E24670.

Wang, W., Wu, Y., Messing, J., 2012. The mitochondrial genome of an aquatic plant, Spirodela polyrhiza. PloS One 7 (10), e46747.

Wilgenbusch, J.C., Warren, D.L., Swofford, D.L., 2004. AWTY: A system for graphical exploration of MCMC convergence in Bayesian phylogenetic inference. <http://ceb.csit.fsu.edu/awty>.

Wolfe, K.H., Li, W.H., Sharp, P.M., 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc. Natl. Acad. Sci. USA 84 (24), 9054–9058.

Wong, S.Y., Boyce, P.C., Sofiman bin Othman, A., Pin, L.C., 2010. Molecular phylogeny of tribe Schismatoglottideae (Araceae) based on two plastid markers and recognition of a new tribe, Philonotieae, from the neotropics. Taxon 59 (1), 117–124.

Xi, Z., Ruhfel, B.R., Schaefer, H., Amorim, A.M., Sugumaran, M., Wurdack, K.J., Endress, P.K., Matthews, M.L., Stevens, P.F., Mathews, S., Davis, C.C., 2012. Phylogenomics and a posteriori data partitioning resolve the Cretaceous angiosperm radiation Malpighiales. Proc. Natl. Acad. Sci. USA 109 (43), 17519–17524.

Yang, Z., 1993. Maximum-likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. Mol. Biol. Evol. 10 (6), 1396–1401.