Syst. Biol. 63(6):933–950, 2014

DOI:10.1093/sysbio/syu054
Advance Access publication August 4, 2014

Chloroplast Phylogenomic Analyses Resolve Deep-Level Relationships of an Intractable Bamboo Tribe Arundinarieae (Poaceae)

PENG-FEI MA[1,2,3], YU-XIAO ZHANG[1,2], CHUN-XIA ZENG[2], ZHEN-HUA GUO[2], AND DE-ZHU LI[1,2,*]

[1]*Key Laboratory for Plant Diversity and Biogeography of East Asia, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, Yunnan 650201, China;* [2]*Plant Germplasm and Genomics Center, Germplasm Bank of Wild Species, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, Yunnan 650201, China; and* [3]*Kunming College of Life Sciences, University of Chinese Academy of Sciences, Kunming, Yunnan 650201, China*
**Correspondence to be sent to: Kunming Institute of Botany, Chinese Academy of Sciences, 132 Lanhei Road, Kunming, Yunnan 650201, China; E-mail: dzl@mail.kib.ac.cn.*

Received 14 April 2013; reviews returned 27 June 2013; accepted 24 July 2014
Associate Editor: Roberta Mason-Gamer

Abstract.—The temperate woody bamboos constitute a distinct tribe Arundinarieae (Poaceae: Bambusoideae) with high species diversity. Estimating phylogenetic relationships among the 11 major lineages of Arundinarieae has been particularly difficult, owing to a possible rapid radiation and the extremely low rate of sequence divergence. Here, we explore the use of chloroplast genome sequencing for phylogenetic inference. We sampled 25 species (22 temperate bamboos and 3 outgroups) for the complete genome representing eight major lineages of Arundinarieae in an attempt to resolve backbone relationships. Phylogenetic analyses of coding versus noncoding sequences, and of different regions of the genome (large single copy and small single copy, and inverted repeat regions) yielded no well-supported contradicting topologies but potential incongruence was found between the coding and noncoding sequences. The use of various data partitioning schemes in analysis of the complete sequences resulted in nearly identical topologies and node support values, although the partitioning schemes were decisively different from each other as to the fit to the data. Our full genomic data set substantially increased resolution along the backbone and provided strong support for most relationships despite the very short internodes and long branches in the tree. The inferred relationships were also robust to potential confounding factors (e.g., long-branch attraction) and received support from independent indels in the genome. We then added taxa from the three Arundinarieae lineages that were not included in the full-genome data set; each of these were sampled for more than 50% genome sequences. The resulting trees not only corroborated the reconstructed deep-level relationships but also largely resolved the phylogenetic placements of these three additional lineages. Furthermore, adding 129 additional taxa sampled for only eight chloroplast loci to the combined data set yielded almost identical relationships, albeit with low support values. We believe that the inferred phylogeny is robust to taxon sampling. Having resolved the deep-level relationships of Arundinarieae, we illuminate how chloroplast phylogenomics can be used for elucidating difficult phylogeny at low taxonomic levels in intractable plant groups. [Branch length; chloroplast phylogenomics; data partitioning; low sequence divergence; sampling; temperate woody bamboos.]

The estimation of phylogenetic relationships plays a key role in understanding evolution and has been an essential component of evolutionary biology. In plants, much effort in reconstructing the Tree of Life has focused on the relationships of major clades and significant advances have been made above the order or family levels (e.g., APG III 2009; Burlerigh et al. 2011; Soltis et al. 2011). However, progress in inferring phylogenetic relationships at lower taxonomic levels and among recently diverged species is less encouraging, especially for species-rich, morphologically diverse lineages (e.g., Waterway et al. 2009). The difficulty lies in the lack of phylogenetic signal due to the low sequence divergence of commonly used genetic markers and the rapid radiations documented in many recently diverged plant lineages (Valente et al. 2010; Schnitzler et al. 2011; Drummond et al. 2012). The phylogenies of rapidly radiating lineages are usually characterized by short internal branches associated with long external branches, posing a particularly difficult challenge for phylogenetic resolution (Rokas and Carroll 2006; Whitfield and Lockhart 2007; Wiens et al. 2008; Philippe et al. 2011). First, the time between successive speciation events is usually too short for species to accumulate sufficient evolutionary changes at the molecular level for phylogenetic inference. Second, more random mutations are expected to occur along the long external branches, generating homoplastic characters that obscure the true phylogenetic signal; therefore, the long branches may erroneously cluster together, that is, long-branch attraction (LBA) (Felsenstein 1978; Bergsten 2005).

The temperate woody bamboo clade, or tribe Arundinarieae (Poaceae: Bambusoideae), exemplifies these challenges. With approximately 533 extant species, the clade exhibits a wide range of morphological diversity, with mature plants ranging from 10-cm shrubs to 20-m treelike forms (Li et al. 2006). It is distributed primarily in montane forests of the northern temperate zone and at high elevations in the tropics and subtropics (Ohrnberger 1999; Li et al. 2006; Bamboo Phylogeny Group 2012). Over 95% of its species are endemic to East Asia, whereas only about 20 species are distributed among the montane areas of Sri Lanka, southern tip of India, Madagascar, and Africa; three species are confined to eastern North America (Li et al. 2006; Bamboo Phylogeny Group 2012). Molecular dating analyses support the tribe's origin between 23.02 and 40.33 Ma; major diversification may have occurred from 8.96 to 10.6 Ma (Bouchenak-Khelladi et al. 2010; Ruiz-Sanchez 2011) or 18.81 Ma (Christin et al. 2008a), coupled with a possible rapid radiation (Hodkinson et al. 2010).

Bamboos are one of the most important non-timber forest products and many species of temperate woody bamboo are of great value economically and ecologically (Li 1999; Bystriakova et al. 2003; Bamboo Phylogeny Group 2012). Among their many uses are in housing construction (as the "timber of the poor"), food, pulp manufacture, and material for weaving and furniture (McClure 1973; Li et al. 2006). Moreover, as a major lineage of the grass family diversifying in forests, these bamboos also play critical roles in ecological communities, providing shelter and food for many rare and endangered animals (e.g., the giant panda) (Li 1999; Ohrnberger 1999; Bystriakova et al. 2003). This group is of great interest from an evolutionary perspective for being one of the fastest-growing plants (e.g., moso bamboo can grow up to 119 cm in 24 h and 24 m high in 40–50 days [Fu 2001]), and with flowering intervals as long as 40–120 years (Janzen 1976).

Despite their importance, the bamboos are understudied relative to other grasses, especially with respect to a phylogeny-based classification. The Arundinarieae, together with Bambuseae (the tropical woody bamboo clade) and Olyreae (the herbaceous bamboo clade), comprise the Bambusoideae, all of which are strongly supported as monophyletic, with Arundinarieae as sister to the Bambuseae + Olyreae (Bouchenak-Khelladi et al. 2008; Sungkaew et al. 2009; Kelchner et al. 2013). The taxonomy of Arundinarieae is highly controversial, with many very different classifications based on morphological and anatomical characters (Li 1999; Bamboo Phylogeny Group 2012). The application of molecular phylogenetics has not solved these controversies, instead departing markedly from previous morphologically based classifications. Nearly all of the commonly recognized genera (21 based on Li [1999]; 28 based on Bamboo Phylogeny Group [2012]) and all two or three subtribes have been revealed as paraphyletic or polyphyletic (Peng et al. 2008; Triplett and Clark 2010; Zeng et al. 2010). Based on broad taxon sampling and approximately 10 kb chloroplast (cp) DNA sequence data, Zeng et al. (2010) divided the Arundinarieae into 10 major lineages for which no morphological synapomorphies can be easily identified. An additional lineage was recovered in Yang et al. (2013), and thus 11 major lineages in all are currently recognized within the tribe. The species-level diversity of the tribe is unevenly distributed among these lineages, with the overwhelming majority of species concentrated in just three lineages, and the others represented only by anywhere from 1 to 10 taxa (Triplett and Clark 2010; Zeng et al. 2010; Yang et al. 2013). The evolutionary relationships within and among these lineages are unclear (Triplett and Clark 2010; Zeng et al. 2010; Yang et al. 2013), due primarily to the extremely low rate of DNA sequence evolution (Zeng et al. 2010) associated with the long generation times (Janzen 1976).

Phylogenomics has proven useful for tackling difficult phylogenies (Hackett et al. 2008; Jian et al. 2008; Bewick et al. 2012; Zhou et al. 2012) and great progress has been achieved in resolving enigmatic relationships of plants with cp genome-scale data (Jansen et al. 2007; Moore et al. 2007, 2010; Parks et al. 2009; Xi et al. 2012; Barrett et al. 2013). Chloroplast DNA sequences have been extensively used for inferring plant phylogeny for several reasons. Chloroplast genomes are essentially recombination-free and have a smaller effective population size and thus shorter coalescent time when compared with nuclear genomes (Birky et al. 1983). They are structurally highly conserved without a complex gene structure, facilitating PCR primer design and sequencing. Their use furthermore obviates the need to assess orthology, a problem associated with biparentally inherited nuclear DNA sequences, especially in polyploid plants, as is the case in the tetraploid Arundinarieae (Chen et al. 2003; Zhang et al. 2012; Yang et al. 2013; $2n = 48$). Furthermore, cp phylogenomics has the potential to resolve relationships within Arundinarieae, as shown by an improvement in resolution from early multigene data sets (Triplett and Clark 2010; Zeng et al. 2010) to a more recent limited data set of whole cp genomes (Zhang et al. 2011), although to date, such studies have mainly focused on relationships at higher taxonomical levels (Jansen et al. 2007; Moore et al. 2007, 2010; Xi et al. 2012; Barrett et al. 2013).

The most common approach in cp phylogenomics is to analyze the concatenated data as a whole (Jansen et al. 2007; Moore et al. 2007; Parks et al. 2009) or only partitioning the data by gene (Moore et al. 2010; Barrett et al. 2013). A thorough exploration of possible partitioning schemes has not been applied to cp genome data, although different partitioning schemes can have substantial effects on phylogenetic reconstruction (Brown and Lemmon 2007; Li et al. 2008; Ward et al. 2010; Xi et al. 2012). In addition, the performances of noncoding versus coding sequences are usually not evaluated in cp phylogenomics and only the coding sequences are used (Jansen et al. 2007; Moore et al. 2007, 2010; Xi et al. 2012; Barrett et al. 2013). Nevertheless, the noncoding sequences should be especially considered when inferring phylogenies at lower taxonomic levels, in order to generate maximum phylogenetic signal (Parks et al. 2009; Zhang et al. 2011).

Here, we attempt to estimate deep-level phylogenetic relationships, namely those among major lineages, for Arundinarieae with cp phylogenomics. Using next-generation sequencing, 15 complete cp genome sequences were newly determined. We assembled a data set of whole genomes from 25 bamboos, 22 of which were sampled from representatives of eight major lineages of Arundinarieae. In addition, the remaining three lineages all had a taxon with more than 50% of cp genome sequenced, and we added these, to generate a second data set. Multiple phylogenetic reconstruction methods in combination with complex models of sequence evolution and a series of partitioning schemes were used to estimate phylogenetic relationships. This study has two major objectives. First, we evaluate the potential of whole cp genomes to estimate deep-level relationships within Arundinarieae, as a test case for resolving

TABLE 1. List of 28 taxa sampled in this study, for which complete or a majority of the chloroplast genome sequences are available

Tribe and species[a]	Lineage	Voucher specimen	Genome size (bp)[b]	Accession numbers
Arundinarieae				
Bergbambos tessellata	I	KMBG1301	110,634 (79.2%)	KJ522748
Oldeania alpina	II	KMBG1302	86,769 (62.1%)	KJ531443
Chimonocalamus longiusculus	III	MPF10182	139,821	JX513415
Ferrocalamus rimosivaginus	IV	—	139,467	HQ337794
Gelidocalamus tessellatus	V	MPF10049	139,712	JX513420
Arundinaria faberi	V	CZM025	139,629	JX513414
Arundinaria fargesii	V	MPF10139	139,696	JX513413
Fargesia nitida	V	Zhang08017	139,535	JX513416
Fargesia spathacea	V	MPF10141	139,767	JX513417
Fargesia yunnanensis	V	MPF10162	139,609	JX513418
Indocalamus longiauritus	V	—	139,668	HQ337795
Phyllostachys edulis	V	—	139,679	HQ337796
Phyllostachys nigra	V	—	139,839	HQ154129
Phyllostachys propinqua	V	—	139,704	JN415113
Yushania levigata	V	YD02	139,633	JX513426
Acidosasa purpurea	VI	—	139,697	HQ337793
Arundinaria gigantea	VI	—	138,935	JX235347
Indosasa sinica	VI	MPF10034	139,660	JX513422
Oligostachyum shiuyingianum	VI	DZL09122	139,647	JX513423
Pleioblastus maculatus	VI	MPF10161	139,720	JX513424
Thamnocalamus spathiflorus	VII	MPF10065	139,778	JX513425
Indocalamus wilsonii	VIII	MPF10146	139,962	JX513421
Gaoligongshania megalothyrsa	IX	MPF10056	140,064	JX513419
Indocalamus sinicus	X	Zeng&Zhang06081	71,694 (51.3%)	KJ531442
Ampelocalamus calcareus	XI	MPF10050	139,689	KJ496369
Outgroups				
Bambusa emeiensis	—	—	139,493	HQ337797
Bambusa oldhamii	—	—	139,350	FJ970915
Dendrocalamus latiflorus	—	—	139,394	FJ970916

[a]These newly sequenced taxa are indicated in bold. [b]For the three taxa with incomplete chloroplast genomes, the numbers in parentheses indicate the approximate percentage of the genome obtained.

phylogeny of intractable plants at lower taxonomic levels. Second, we aim for novel inferences about the performances of different sequences, analytical methods, and partitioning schemes in cp phylogenomic analysis from this data set.

MATERIALS AND METHODS

Taxon Sampling

From the 11 recognized lineages of Arundinarieae (Triplett and Clark 2010; Zeng et al. 2010; Yang et al. 2013), we chose 18 exemplar taxa to represent the diversity of the tribe. All 11 of the major lineages were sampled, but three (I, II, and X) were represented by partial cp genomes (see below). We completed our sampling with an additional seven previously published cp genomes of Arundinarieae in GenBank (Zhang et al. 2011; Burke et al. 2012; Wu and Ge 2012), such that 25 Arundinarieae species were sampled in all (Table 1). The three lineages (IV, V, and VI) which have the greatest species-level diversity of the tribe (Triplett and Clark 2010; Zeng et al. 2010) were represented by multiple (2–10) species spanning morphological and phylogenetic diversity within each lineage, whereas the other lineages were represented by a single species each. All three available cp genomes from Bambuseae (Wu et al. 2009; Zhang et al. 2011), thought to be most closely related to Arundinarieae (Bouchenak-Khelladi et al. 2008; Sungkaew et al. 2009; Kelchner et al. 2013), comprised the outgroups (Table 1). To assess the impact of taxon sampling, we also combined our data set with the previously described data set of Zeng et al. (2010), which included 141 Arundinarieae species sampled for eight cpDNA loci.

We assembled three data sets—(i) a core data set consisting of 22 Arundinarieae taxa representing eight major lineages of the tribe with complete cp genome sequences (*data-complete*); (ii) a data set of 25 Arundinarieae taxa with all 11 lineages sampled and some missing data in three samples (*lineage-complete*) (see Table 1); and (iii) a data set of 154 Arundinarieae taxa with extensive missing data (*combined-incomplete*). Taxon names generally followed those of Li et al. (2006), and *Bergbambos tessellata* and *Oldeania alpina* were used for the two African taxa following the recent suggestion of Stapleton (2013). Voucher specimens for the newly sequenced taxa were deposited at the Herbarium of Kunming Institute of Botany, Chinese Academy of Sciences.

Data Collection

The complete cp genomes were sequenced with Illumina sequencing technology (Shendure and Ji 2008), and two different methods were used in obtaining

cpDNA in accordance with the availability of fresh leaf material. For the taxa with more than 40 g fresh leaf material obtained from specimens in the field or those planted at the Kunming Botanical Garden of Kunming Institute of Botany, the protocol for isolating cpDNA is as described in Zhang et al. (2011). For *Ampelocalamus calcareus* and *B. tessellata*, total genomic DNA was first isolated from fresh leaf material with the CTAB method (Doyle and Doyle 1987). Subsequently, we tried to amplify the cpDNA using long-range PCR with nine primer pairs (Yang et al. 2014). However, we failed to amplify two cp genomic regions in these taxa. Amplicon DNA concentrations were determined by visual approximation with gel electrophoresis and amplicons were pooled in roughly equal mass mixtures for subsequent sequencing.

In addition to the isolated and amplified cpDNA, total genomic DNA of *A. calcareus* was also sheared and prepared for Illumina sequencing. The protocol for sequencing is as described in Zhang et al. (2011) with the following exceptions. In addition to a paired-end sequencing library with insert size of approximately 500 bp, a second paired-end sequencing library with insert size of approximately 170 bp was also constructed for accessions having more than or equal to 3 μg isolated cpDNA. Paired-end sequencing (75, 85, or 90 bp) was performed on the Illumina HiSeq 2000 instrument at BGI-Shenzhen, resulting in more than 200 Mb of DNA sequence data for each sample.

The raw sequence reads were assembled with SOAPdenovo v1.04 (Li et al. 2010) as in Zhang et al. (2011). Small gaps in the assemblies were bridged by designing custom primers (Supplementary Table S1; available from http://www.sysbio.oxfordjournals.org/, doi:10.5061/dryad.d5h1n.2) based on their flanking sequences for PCR and conventional Sanger sequencing. The four junctions between the single copy and inverted repeat (IR) regions of the cp genome were verified by Sanger sequencing (Zhang et al. 2011). A few possible sequencing and/or assembly errors in the assembled genomes were inspected and corrected by designed PCR (Supplementary Table S1) and Sanger sequencing. The complete genome sequences were imported into the program DOGMA (Wyman et al. 2004) for annotation, coupled with manual investigation of the positions of start and stop codons and boundaries between introns and exons.

Only silica-dried material was available for *Indocalamus sinicus* and *O. alpina*, the sole representative of lineages X and II (Triplett and Clark 2010; Zeng et al. 2010), respectively. Total genomic DNA was extracted from these samples with the CTAB method (Doyle and Doyle 1987). The DNA was degraded into short fragments in the dried material and could not be used for long-range PCR amplification. We thus designed an additional 90 primer pairs (Supplementary Table S2) in combination with those in Supplementary Table S1 and published previously (Zhang et al. 2011) to Sanger sequence the two taxa. The sequences of *I. sinicus* and *O. alpina* (as *Yushania alpina*) from the data set of eight cpDNA loci assembled in Zeng et al. (2010) were used here. In total, we collected 71,694 bp and 86,769 bp for *I. sinicus* and *O. alpina*, respectively, and we discontinued our attempts to sequence the remaining genomic regions after 1 month of focused efforts failed. All newly generated sequences have been deposited in GenBank (Table 1).

Sequence Alignment

The 25 cp genome sequences from the *data-complete* data set were aligned with MAFFT v6.833 (Katoh et al. 2005) by using default settings. The alignment was straightforward, and a few poorly aligned regions were manually adjusted in MEGA 4.0 (Tamura et al. 2007). During this process, we recorded the occurrences of more than 1 bp indels that are not associated with mononucleotide repeats and that can be unambiguously determined. There were 20 small inversions identified during the process of manual inspection of the alignment, which varied in size from 3 to 26 bp and were flanked by IRs between 8 and 24 bp in length. Small inversions that are prone to homoplasy can mislead phylogenetic inference (Zhang et al. 2011), and therefore we deleted them from the alignment. For the other two data sets, the sequences from the remaining taxa were added to the above alignment followed by manual alignment in MEGA.

Phylogenetic Analyses

Data subsets.—The angiosperm cp genome can be divided into large single copy (LSC) and small single copy (SSC) regions separated by the IRs regions; the three regions have very different molecular evolutionary rates (Palmer 1985). To evaluate convergence among these regions, we extracted three subsets (LSC, SSC, and IRs) from the *data-complete* data set. We also concatenated all the coding (exons of protein-coding genes, tRNAs, and rRNAs) and noncoding (intergenic regions and introns) sequences from this data set into another two subsets, coding and noncoding, respectively. For each data set and subset, the homogeneity of base frequencies across taxa was examined using the chi-square test implemented in PAUP* 4.0b10 (Swofford 2002).

Data partitioning.—To investigate the issue of data partitioning in cp phylogenomics, we employed a series of data partitions based on the *data-complete* data set in Bayesian and maximum-likelihood (ML) analyses. Six partitioning schemes were explored: (i) Partition2, in which the data were divided into coding and noncoding sequences; (ii) Partition3, in which the data were divided into three partitions corresponding to the LSC, SSC, and IRs regions of the cp genome; (iii) Partition6, in which tRNAs, rRNAs, and the noncoding sequences each formed a partition and the protein-coding sequences were divided into another three partitions by codon

position (Supplementary Table S3); (iv) Partition63, in which the protein-coding sequences were divided into 60 partitions by gene (77 unique genes, excluding 17 which have no variable nucleotide sites and/or are <200 bp and thus were separately combined with this from the same functional group; see details in Supplementary Table S4), and the other three partitions (tRNAs, rRNAs, and noncoding sequences) were the same as in Partition6; (v) Partition77, in which the coding sequences formed one partition and the noncoding sequences were divided into 76 partitions (each unique noncoding region except for those with no variable sites and/or a size of <200 bp was considered as a separate partition; see details in Supplementary Table S5); and (vi) PartitionS8, which contained eight partitions (see details in Supplementary Table S6) determined by PartitionFinder v1.1.1 (Lanfear et al. 2012), a program which compares numerous partitioning schemes and selects the optimal one, as described below. The alignment was first divided into 258 subsets: tRNAs, rRNAs, the noncoding sequences being divided into 76 subsets as in Partition77, and the protein-coding sequences being initially divided into 60 subsets, as in Partition63, and then by codon position. These 258 subsets were subsequently used in a heuristic search (search = rcluster) to identify a partitioning scheme according to the Bayesian information criterion (BIC) by running PartitionFinder.

For the *lineage-complete* data set, both unpartitioned and partitioned analyses were conducted. However, only the partitioning scheme PartitionS5 (see details in Supplementary Table S7) identified with PartitionFinder as described above was implemented in subsequent phylogenetic analyses. For the *combined-incomplete* data set, PartitionFinder identified the entire alignment as a single partition, so only unpartitioned analyses were conducted.

Model selection.—The choice of substitution model for each partition was primarily determined by using Modeltest 3.7 (Posada and Crandall 1998) with the Akaike information criterion (AIC) (Posada and Buckley 2004). However, for the partitions defined by PartitionFinder above, the substitution models determined simultaneously for them in the process of selecting partitioning schemes were used.

Maximum parsimony.—Maximum parsimony (MP) analyses were conducted with PAUP* 4.0b10 (Swofford 2002). All changes were considered as unordered and equally weighted. The gaps were treated as missing data, which was also applied to the following Bayesian and ML analyses. Heuristic searches for the most parsimonious trees were carried out with 1000 replicates of random taxon addition and 100 trees held at each step during stepwise addition, and tree-bisection-reconstruction (TBR) branch swapping, with all the trees saved per replicate. We assessed branch support by performing 200 bootstrap replicates using 100 random taxon addition replicates with 10 trees held at each step and TBR swapping, saving all the trees per replicate. However, for the data sets with broad taxon sampling (*combined-incomplete*) or a few variable characters (subsets SSC and IRs), MP analyses were conducted with one tree saved per taxon addition replicate with the MulTrees option off due to computation time.

Maximum likelihood.—ML analyses were performed with RAxML v7.2.6 (Stamatakis 2006). The ML tree was inferred with the combined rapid bootstrap (1000 replicates) and search for ML tree (the "-f a" option). The GTR + G model was used in all the analyses as suggested (see RAxML manual), whereas in unpartitioned ML analysis of the *data-complete* data set the GTR + I + G model was also used to assess the impact of different models (Table 2).

Bayesian analyses.—Bayesian inference (BI) was performed with MrBayes v3.1.2 (Ronquist and Huelsenbeck 2003). Two independent Markov chain Monte Carlo (MCMC) chains were run, each with three heated and one cold chain for 2 million generations. Each chain started with a random tree, default priors, and sampling trees every 100 generations. Because not all models applied by Modeltest are implemented in MrBayes, a more parameterized model was used when the selected model was not available in the program. In unpartitioned analyses, the GTR + I + G or GTR + I model was used depending on the selected model for each data set (Table 2). A recent empirical study has demonstrated that the use of both parameters I and G to accommodate rate variation across sites could produce multiple regions of high posterior probability (PP) in certain situations (Moyle et al. 2012); we thus also analyzed the *data-complete* data set with the reduced GTR + G model. In partitioned analyses, each partition was given its own model (Table 2 and Supplementary Tables S3–S7). Additionally, all parameters were unlinked across partitions except those for branch lengths and topology. Branch length rate multipliers were unlinked (prset applyto = (all) ratepr = variable).

Convergence was determined by examining the average standard deviation of split frequencies (ASDSFs < 0.01) and by using Are We There Yet graphical analysis (Nylander et al. 2008). Preliminary runs of MrBayes with the settings above converged in the majority of analyses, and reanalysis with the temperature parameter set to 0.05 reached convergence for all (ASDSF fell to 0.0165 at the end of the run of 4 million generations for the *combined-incomplete* data set), with the exception of analyses with Partition63 and Partition77. New analyses with the number of generations increased to 50 million and sampling every 1000 generations did not produce convergence. The failure of the MCMC searches to converge may be due to the parameter-rich models used—with so few substitutions in the data, model parameters could not be estimated accurately (Marshall 2010). Considering

TABLE 2. Data characteristics with models selected and used in unpartitioned ML and BI analyses for each data set

Data set	Taxa	Number of sites	Missing data (%)	Number of variable/informative sites[a]	GC (%)	Best fit model	ML	BI
LSC	25	86,580	—	1702/347 (2995/1467)	36.9	TVM+I+G	GTR+G	GTR+I+G
SSC	25	13,305	—	394/85 (641/281)	33.2	K81uf+I+G	GTR+G	GTR+I+G
IRs	25	44,098	—	172/22 (302/122)	44.2	K81uf+I	GTR+G	GTR+I
Coding	25	72,993	—	807/157 (1425/681)	41.9	GTR+I+G	GTR+G	GTR+I+G
Noncoding	25	70,990	—	1461/297 (2513/1189)	35.6	TVM+I+G	GTR+G	GTR+I+G
Data-complete	25	143,983	—	2268/454 (3938/1870)	38.9	TVM+I+G	GTR+I+G/GTR+G	GTR+I+G/GTR+G
Lineage-complete	28	144,369	3.8	2591/508 (4254/1918)	38.8	TVM+I+G	GTR+G	GTR+I+G
Combined-incomplete	157	144,805	77.5	2843/645 (4499/2052)	37.5	TVM+I+G	GTR+G	GTR+I+G

[a]Ingroup only (all taxa).

that the topology was our priority, and considering that it appeared that a stable topology was achieved in these analyses, we did not try further analyses. Twenty-five percent of trees (50% for the two highly partitioned schemes) were discarded as burn-in. The remaining trees were used to build a 50% majority-rule consensus tree.

Comparisons of partitioning strategies.—We used Bayes factors (BFs, defined as the ratio of marginal likelihoods from two competing models) to evaluate the various partitioning strategies (Nylander et al. 2004; Brandley et al. 2005). Marginal likelihoods were estimated from the harmonic means after discarding the burn-in from BI analyses with Tracer v1.5 (Rambaut and Drummond 2007). As these likelihoods could not be accurately estimated in analyses with Partition63 and Partition77 due to the lack of convergence, we also implemented PartitionFinder to test partitioning schemes according to AIC and BIC under the GTR + G model that was used in the ML analyses.

Indel analysis.—We took advantage of indels in the whole cp genomes to estimate phylogeny independently of the nucleotide sequences. All indels identified during manual investigation above were coded as binary characters that were concatenated into a matrix for MP analysis. Best tree searches were performed with 100 replicates of random taxon addition and one tree held at each step in PAUP.

Phylogeny Evaluation

The Arundinarieae phylogeny posed several key challenges for accurate phylogenetic inference (see details in Results). Meanwhile, the abundance of data in phylogenomics can lead to a highly supported but incorrect topology due to systematic errors (Delsuc et al. 2005; Jeffroy et al. 2006; Philippe et al. 2011). It is therefore necessary to evaluate the robustness of the inferred phylogeny. Phylogenetic analyses for this purpose were all conducted in an unpartitioned way on the *data-complete* data set if not specified using MP, ML, and BI methods as described above.

To evaluate the effect of distant outgroups on the ingroup phylogeny, we performed analyses in which outgroup taxa were removed. We also analyzed a data set of eight cpDNA loci used in Zeng et al. (2010) with our taxon sampling to assess the relative importance of large sequence data versus small taxon samples.

The artifact of LBA can be difficult to detect and different methods were applied to our data set. Long branch extraction is an approach suggested for cases where LBA is suspected (Bergsten 2005). The test for LBA described in Huelsenbeck (1998) was also implemented by using Seq-Gen v.1.3.2 (Rambaut and Grassly 1997) to simulate 100 data sets. The simulation used the tree shown in Figure 1e as a model tree with ML estimates of the branch lengths and the parameters of a nucleotide substitution model determined by Modeltest for the actual data set. Parsimony trees for these data sets were then constructed with PAUP as described above with the exception of 100 replicates of random taxon addition and 10 trees held at each step.

Assessments of competing hypotheses of phylogeny were conducted with the approximately unbiased (AU) test (Shimodaira 2002) as implemented in CONSEL v0.1i (Shimodaira and Hasewaga 2001). The site-wise log likelihoods for the trees were calculated in PAUP.

Network Analysis

We used a network analysis on the *lineage-complete* data set to investigate whether conflicting phylogenetic signal might explain the partial lack of resolution in Arundinarieae phylogeny (see Results) (Kennedy et al. 2005; Huson and Bryant 2006). SplitsTree4 (Huson and Bryant 2006) was used with the Neighbor-Net algorithm and uncorrected *P*-distances.

RESULTS

Characteristics of Chloroplast Genomes and Data Sets

Fifteen complete temperate woody bamboo cp genomes were newly determined in the present study (Table 1). Fourteen were obtained from isolated cpDNA (Zhang et al. 2011) and the other is from a combination

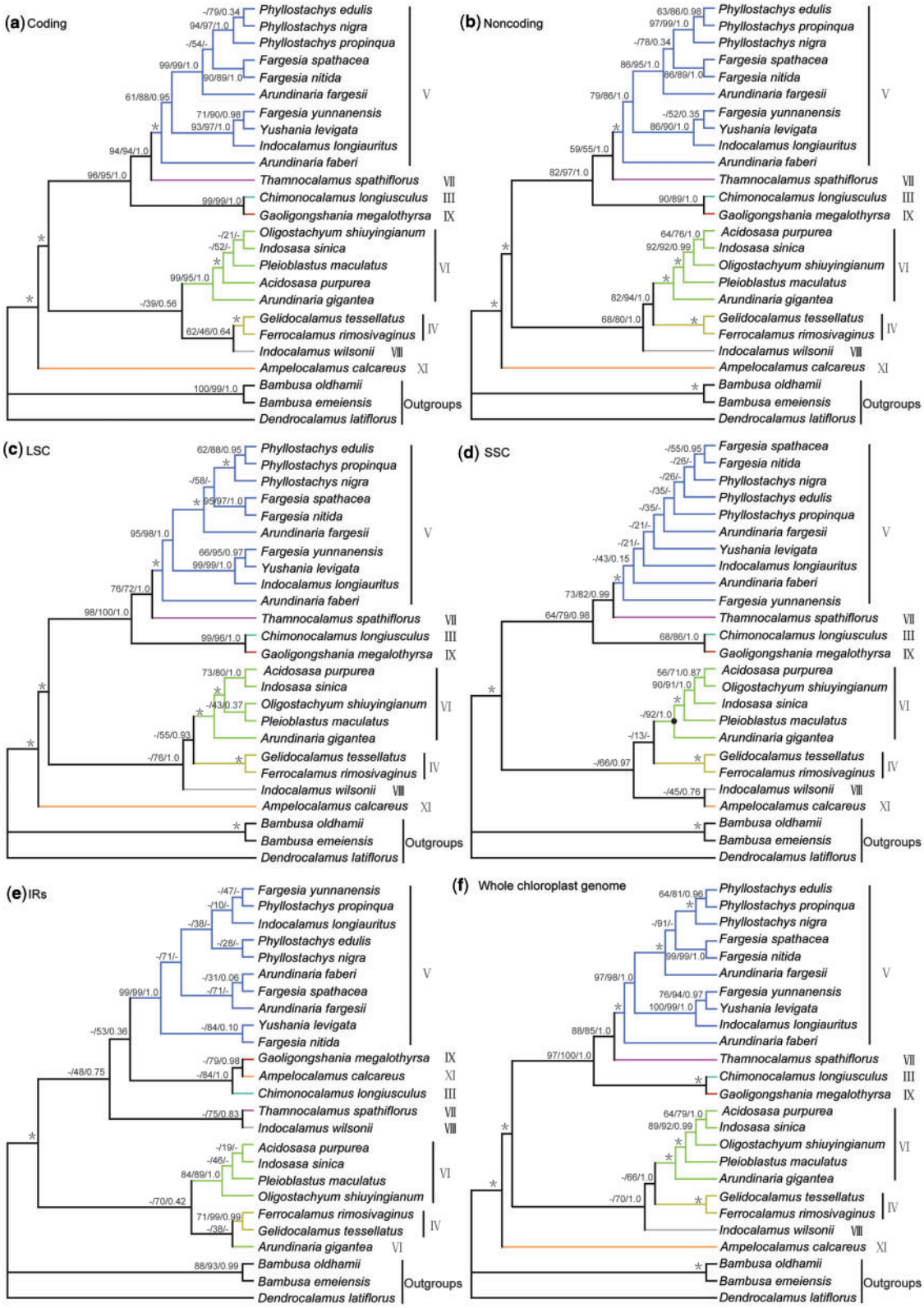

FIGURE 1. Phylogeny of 22 Arundinarieae species inferred from MP, ML, and BI analyses of different chloroplast genome sequences and total evidence. ML topology shown with MP bootstrap support values/ML bootstrap support values/Bayesian PPs listed at each node. Asterisks represent nodes with maximal support values in all analyses. Dash denotes nodes unresolved or without bootstrap support in the MP trees or contradicted by the BI trees with PPs < 0.50, see Supplementary Figures S2 and S3 for detail. Filled circle represents node that is contradicted by the MP topology. Colored branches in the online version of this figure indicate the eight Arundinarieae lineages.

of sequences from long-range cpDNA PCR products and from total genomic DNA. All sequenced genomes are very similar to previously published cp genomes of bamboos (Zhang et al. 2011; Burke et al. 2012; Wu and Ge 2012) and rather conservative in evolution of nearly every aspect of genome characters (e.g., genome structure and gene content). Their genome size ranges narrowly from 139,535 to 140,064 bp (Table 1). A gene map of the cp genome from a representative, *Arundinaria faberi*, is presented in Supplementary Figure S1.

The *data-complete* data set has an aligned length of 143,983 bp (Table 2). The alignment is characterized by extremely low sequence diversity, with only 2268 variable sites (1.58%) and 454 parsimony-informative sites (0.32%) across ingroup taxa (Table 2). Among regions of the cp genome, the lowest level of sequence divergence (0.39% variable sites for the ingroup) occurs in IRs, as expected (Palmer 1985). The GC content varies among the data subsets (Table 2), but none of the data subsets exhibit heterogeneous base composition across taxa. In the *lineage-complete* data set, 28 taxa were sampled and 3.8% of the cells in the matrix contain missing values (Table 2). An average of 64.4% of the complete genome sequences was assembled for the three taxa without complete genomic data; all had more than 50% of their genome sequenced. However, in the *combined-incomplete* data set, 77.5% of the cells in the matrix contain missing values and 129 of all 157 taxa were more than 90% incomplete (Table 2).

Phylogenetic Methods and Data Subsets

Across the five data subsets, the BI and ML analyses produced nearly identical phylogenetic trees, with the only differences being poorly supported (BS < 50%, PP < 0.50; Fig. 1a–e; the BI tree in Supplementary Fig. S2a–e). The MP analysis yielded a largely congruent but less resolved version of the BI/ML topology (Fig. 1a–e and Supplementary Fig. S3a–e). Furthermore, the well-supported relationships inferred from any one of the three tree-building methods received high support in the other two methods (Fig. 1a–e) with one exception that occurred in the SSC subset (indicated by a filled circle). Both the BI and ML analyses strongly supported the monophyly of the VI lineage (BS = 92%, PP = 1.0; Fig. 1d); however, the MP analysis placed the rapidly evolving *Arundinaria gigantea* (Supplementary Fig. S2d) as sister to all the other ingroup taxa with BS = 68% (Supplementary Fig. S3d). Phylogenetic analysis of the slowly evolving coding sequences from the SSC subset resolved this disagreement, with the monophyly of this lineage supported by all methods (MPBS = 59%, MLBS = 71%, and PP = 0.93; Supplementary Fig. S4).

Comparable phylogenetic resolution was recovered between the coding and noncoding data sets (Fig. 1a,b). The phylogenetic relationships derived from the LSC subset were largely resolved (Fig. 1c), whereas only some relationships were recovered by the IRs subset (Fig. 1e) and the SSC subset provided moderate phylogenetic resolution (Fig. 1d). Overall, phylogenetic analyses of the five different data subsets did not generate strongly supported topological conflict. Discrepancies mainly involved the placements of the IV, VI, and VIII lineages. In most analyses, their placements were either unresolved or poorly supported, although in the analysis of the noncoding subset the inferred relationships received moderate-to-high support (Fig. 1b).

Complete Chloroplast Genome and Data Partitioning

The topologies estimated from MP and the unpartitioned BI and ML analyses of the complete cp genome sequences were broadly similar to each other (Fig. 1f and Supplementary Figs S2f and S3f). Moreover, nearly identical results were obtained irrespective of the models (GTR + G and GTR + I + G) used in the BI and ML analyses (data not shown). In general, the recovered relationships received much higher support values than those inferred from the five data subsets (Fig. 1a–e).

In addition to the unpartitioned analysis, we also performed analyses with the entire data set divided into different partitions (Table 3) to better accommodate heterogeneity in the processes of molecular evolution across different sites. Overall, more highly partitioned models were favored over less partitioned ones (Table 3). However, the two most highly partitioned models, Partition63 and Partition77, were worse than the unpartitioned model (Table 3). The tree topologies inferred from the ML analyses under various partitioning strategies remained the same as in the unpartitioned analysis, and for the BI method only the unresolved relationships within the V lineage (Fig. 1f) changed but none of the possible topologies were well supported (PP < 0.40; data not shown). Moreover, the support values for the backbone phylogeny of Arundinarieae (Table 4) were nearly identical in these analyses. Thus, we focus on results mainly from the unpartitioned analyses.

Increased Resolution and Phylogeny Evaluation

Although the inferred tree has successive short internodes connected by long branches (Fig. 2), which can pose problems in phylogenetic reconstruction (Rokas and Carroll 2006; Whitfield and Lockhart 2007; Wiens et al. 2008), the phylogeny of Arundinarieae was largely resolved (Fig. 1f). Furthermore, the concerns that the long branch connecting the ingroup with outgroups (Fig. 2) would affect ingroup phylogeny were largely unfounded, because ingroup topologies recovered from the rooted and unrooted analyses were identical (Fig. 1f and Supplementary Fig. S5). Twelve of a total of the 20 ingroup nodes received high support in all three tree-building methods (MPBS ≥ 90%, MLBS ≥ 90%, and PP = 1.0) and another five received moderate-to-high support in different methods (Fig. 1f and Table 4). The XI lineage was recovered as the sister to the remainder of the Arundinarieae. The III, V, VII, and IX lineages formed

TABLE 3. Comparison of partitioning strategies used for the data sets of *data-complete* and *lineage-complete*

Data set	Partitioning strategy	Description	Comparison with PartitionFinder				Comparison with BFs	
			Parameters	ln *L*	AIC	BIC	Harmonic means	2ln (BFs)[a]
Data-complete	All	All together	56	−231216.40	462544.79	463097.93	−231070.83±0.12	—
	Partition2	Coding, noncoding	112	−230486.61	461197.22	462303.49	−230413.30±0.12	1315.06
	Partition3	LSC, SSC, IRs	168	−229773.23	459882.46	461541.87	−229835.14±0.15	1156.32
	Partition6	Noncoding, tRNAs, rRNAs, three codon positions	336	−229039.67	458751.34	462070.16	−229135.44±0.14	1399.40
	PartitionS8	Identified by PartitionFinder	448	−228439.36	457774.72	462199.82	−228604.99±0.2	1060.90
	Partition63	Noncoding, tRNAs, rRNAs, 60 protein genes	3528	−227917.60	462891.21	497738.86	—	—
	Partition77	Coding, 76 noncoding regions	4312	−227087.22	462798.43	505390.00	—	—
Lineage-complete	All	All together	62	−235507.06	471138.12	471750.69	−235225.36±0.14	—
	PartitionS5	Identified by PartitionFinder	310	−233137.80	466895.61	469958.45	−233084.63±0.16	4281.46

[a]Values for 2ln(BFs) are twice the difference in the –ln of the BF from the partitions model below compared with that above.

TABLE 4. Support values of Bayesian PPs and ML bootstraps for backbone phylogeny of Arundinarieae from different partitioned analyses of the data set of *data-complete*

Lineage	ML[a]							BI[b]						
	All	P2	P3	P6	P63	P77	PS8	All	P2	P3	P6	P63	P77	PS8
Arundinarieae	100	100	100	100	100	100	100	1.0	1.0	1.0	1.0	1.0	1.0	1.0
IV	100	100	100	100	100	100	100	1.0	1.0	1.0	1.0	1.0	1.0	1.0
V	100	100	100	100	100	100	100	1.0	1.0	1.0	1.0	1.0	1.0	1.0
VI	100	100	100	100	100	100	100	1.0	1.0	1.0	1.0	1.0	1.0	1.0
V + VII	85	85	84	85	87	89	84	1.0	1.0	1.0	1.0	1.0	1.0	1.0
III + IX	100	100	100	100	100	100	100	1.0	1.0	1.0	1.0	1.0	1.0	1.0
(V, VII) + (III, IX)	100	100	100	100	100	100	100	1.0	1.0	1.0	1.0	1.0	1.0	1.0
IV + VI	66	63	73	64	73	65	77	1.0	1.0	1.0	1.0	1.0	1.0	1.0
VIII + (IV, VI)	70	74	75	70	70	78	69	1.0	1.0	1.0	1.0	1.0	1.0	1.0
((V, VII) (III, IX)) + (VIII (IV, VI))	100	100	100	100	100	100	100	1.0	1.0	1.0	1.0	1.0	1.0	1.0

[a,b]P2–PS8 separately represents the six partitioning schemes: Partition2, Partition3, Partition6, Partition63, Partition77, and PartitionS8 in the main text.

a monophyletic group with a sister relationship among III and IX, V, and VII lineages (Fig. 1f). The phylogenetic relationships involving the IV, VI, and VIII lineages were not fully resolved (Fig. 1f and Table 4). High resolution was achieved for phylogenetic relationships within the major clades (Fig. 1f).

A total of 282 indels (Supplementary Table S8) were identified across the complete cp genomes according to the criteria described above. More than half of these occur in a single taxon (Supplementary Table S8), especially in those with a relatively high rate of molecular evolution (Fig. 2), and only 107 indels are shared by two or more taxa. Unordered parsimony analysis of the indels data set revealed relationships (Fig. 3) broadly similar to those in Figure 1f. The majority of indels in the ingroup were congruent with the optimal tree and five indels were inferred to be homoplasious (Fig. 3). The IV, V, and VI lineages were found to share 15, 8, and 5 indels, respectively, whereas only one to three indels supported other nodes (Fig. 3).

Although all the analyses recovered the sister relationship of the III and IX lineages with maximum support (Fig. 1f and Table 4), both lineages possess a very long branch (Fig. 2) and the LBA artifact should be suspected (Felsenstein 1978; Bergsten 2005). The exclusion of the IX lineage did not alter the phylogenetic placement of the III lineage (data not shown). Conversely, the IX lineage grouped with the VII lineage with weak support in the ML and MP analyses (MPBS = 65%, MLBS = 60%; Supplementary Fig. S6a) after excluding the III lineage whereas its placement remained unchanged in the BI tree (PP = 0.64; Supplementary Fig. S6b). When we repeated the analyses using only the coding sequences, the IX lineage was always in the original placement on the phylogenetic trees with strong support (MPBS = 86%, MLBS = 90%, PP = 1.0; Supplementary Fig. S7). We also applied a Monte Carlo simulation to test for LBA (Huelsenbeck 1998). The MP analysis did not recover a sister relationship between the XI and IX lineages as in the model tree (Fig. 1e) in 2 of 100 simulated data sets, and only one of these two trees yielded a sister relationship between the III and IX lineages. In sum, these results demonstrated that this sister relationship was not likely to be an LBA artifact.

In contrast to the poor phylogenetic resolution for Arundinarieae in previous studies (Triplett and Clark 2010; Zeng et al. 2010; Yang et al. 2013), substantial

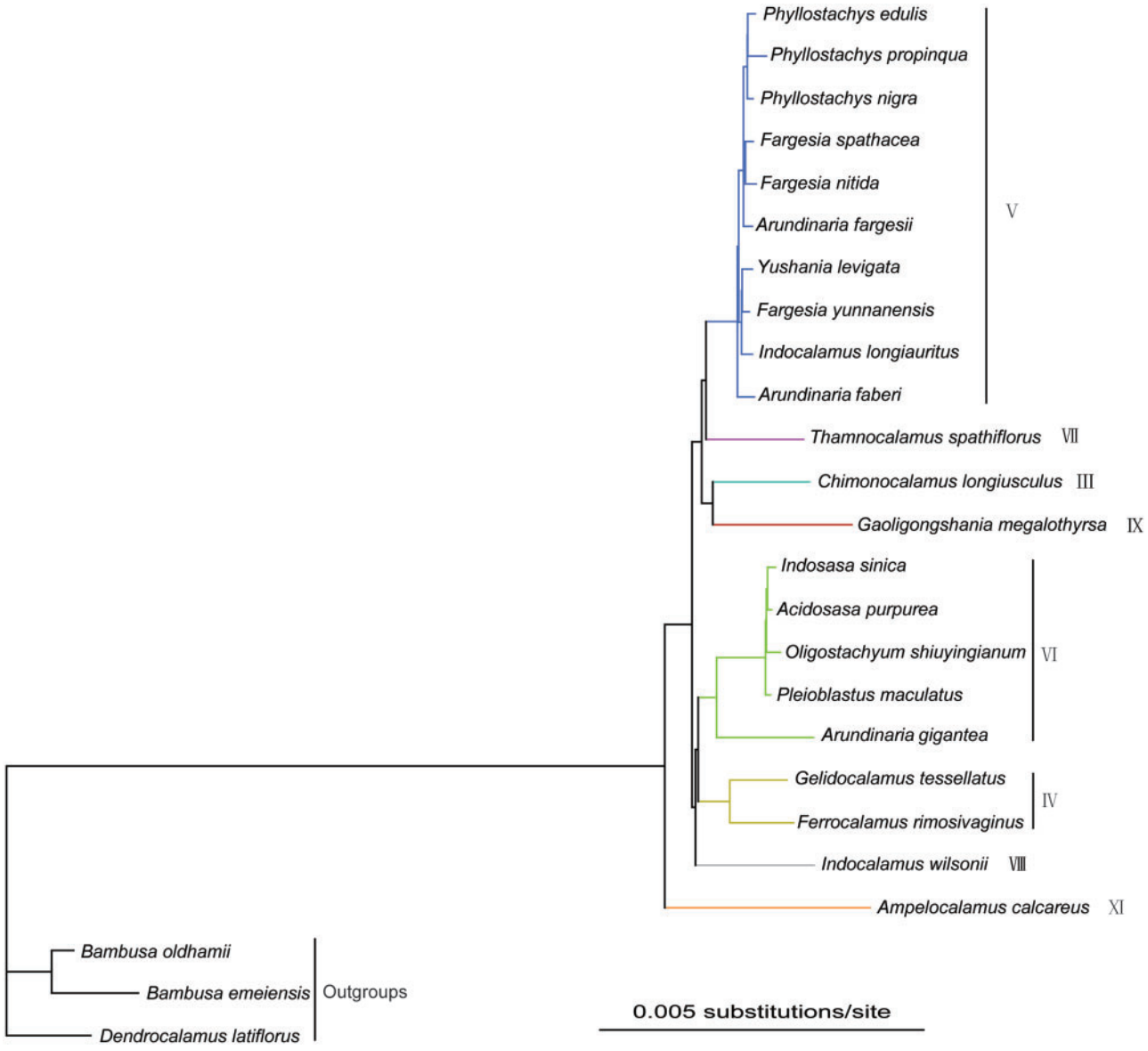

FIGURE 2. ML phylogram of 22 Arundinarieae species based on an unpartitioned analysis of the whole chloroplast genome sequences. This corresponds to the cladogram in Figure 1f. Colored branches in the online version of this figure indicate the eight Arundinarieae lineages.

improvement in resolution was achieved with our cp phylogenomic approach (Fig. 1f and Table 4). However, these studies and ours differed markedly in taxa and molecular data sampled, and the increased phylogenetic resolution could be due to much more data or less densely sampled taxa. Phylogenetic analyses based on the same eight cpDNA loci in Zeng et al. (2010) with our taxon sampling only resolved the monophyly of the major lineages, as in their study (Supplementary Fig. S8). We further performed phylogenetic analyses using the *lineage-complete* data set with all the 11 major lineages sampled (Table 1). The topologies generated under unpartitioned and partitioned BI and ML analyses of this data set were in full agreement with each other (Fig. 4 and Supplementary Fig. S9) as in the analysis of the *data-complete* data set. However, a notable change in BS values was observed for the grouping of the IV, VI, and VIII lineages, decreasing from 71% in unpartitioned to 61% in partitioned ML analysis (unresolved in MP and PP = 1.0 for BI in both models) (Fig. 4 and Supplementary Fig. S9). Moreover, the unpartitioned model was strongly rejected in favor of the partitioned model (Table 3), and thus we limit discussion to the phylogenetic trees estimated from partitioned analyses (Fig. 4). The phylogenetic relationships were generally congruent with those inferred from the *data-complete* data set, with only slight difference in support values (Figs. 1f and 4). The MLBS values for the nodes involving the IV, VI, and VIII lineages showed a decrease of 12–18% relative to the *data-complete* data set, whereas in the MP analysis a striking decrease in the BS values for the grouping of the V and VII lineages from 88% to 61% was

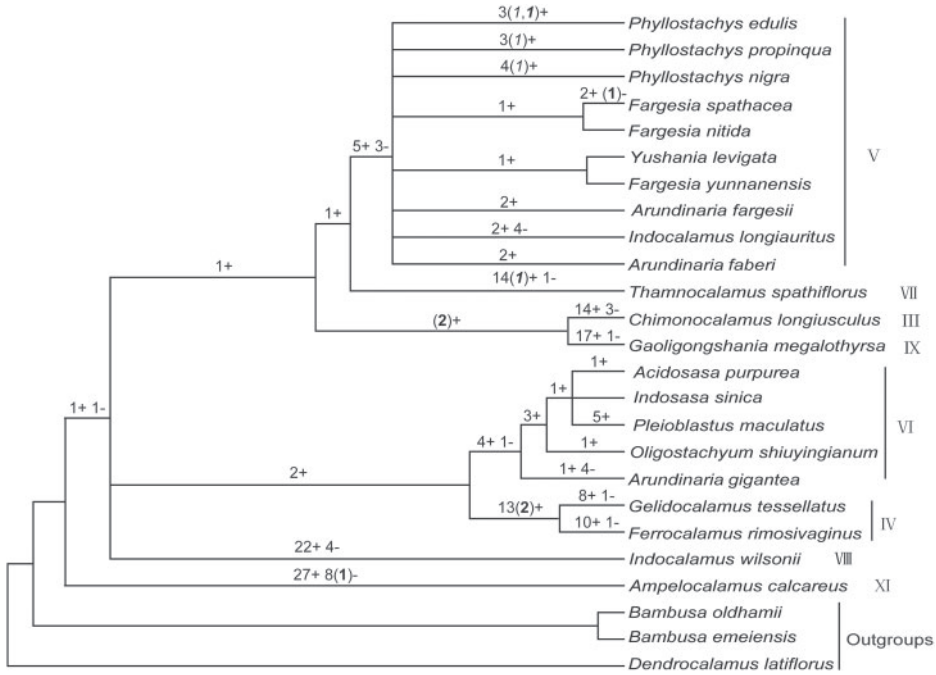

FIGURE 3. Phylogeny estimated using parsimony on 282 identified indels, with the 200 indels only involving the ingroup taxa mapped. Numbers above branch denote the number of unambiguously reconstructed insertions (+) and deletions (−), and those in the brackets indicate potentially homoplasious indels. The five insertions mapped to the V lineage include one shared by all the species except for *Phyllostachys edulis* within this lineage.

observed. Finally, we performed phylogenetic analyses using the *combined-incomplete* data set with a high percentage of missing data (Table 2). The phylogenetic placement of the VIII lineage in the strict consensus topology in MP analysis (Supplementary Fig. S10) differed from those in the ML and BI analyses (Fig. 5 and Supplementary Figs. S11 and S12) but without support. The phylogenetic relationships among major lineages were identical to those in Figure 4, although a further decrease in support values especially in MP analysis was revealed (Fig. 5). Resolution was poor within major lineages (Supplementary Figs. S10–S12).

Unresolved Phylogeny and Conflicting Phylogenetic Signals

Unresolved or contrasting branching order among the IV, VI, and VIII lineages were revealed in the analyses above (Figs. 1, 4, and 5). The sister-group relationship between lineages IV and VI was most frequently recovered and only this relationship received moderate-to-strong support (the noncoding subset; Fig. 1b). However, the AU test with the complete cp genome sequences failed to reject two alternative topologies involving these three clades, whereas it strongly rejected the hypothesis of an unresolved polytomy ($P < 0.001$) (Table 5). A similar result was obtained with the coding sequences, although different topologies were favored by the two data sets, whereas in the use of the noncoding sequences three of the four hypotheses were rejected ($P < 0.01$) in favor of the topology containing the grouping of the IV and VI lineages (Table 5). On the other hand, the phylogenetic relationships involving the II, III, IX, and X lineages did not receive strong support when all the major lineages were sampled (Fig. 4). A sister relationship between the II and IX lineages was resolved with weak-to-moderate support (MPBS = 58%, MLBS = 73%, BI = 0.98), and this grouping in turn was sister to the III lineage (Fig. 4). However, reanalysis with coding sequences revealed the grouping of the III and IX lineages again with slightly increased support values (MPBS = 74%, MLBS = 83%, BI = 0.89) whereas the other relationships remained unchanged (Supplementary Fig. S13).

In general, the network visualization of splits revealed support for the major groupings found by the phylogenetic analyses, as well as a number of alternative topologies not evident from phylogenetic analyses (Supplementary Fig. S14). Moreover, the region of the network concentrating around these incompletely resolved relationships described above showed substantial conflicting signal (Supplementary Fig. S14).

DISCUSSION

Inferring Phylogeny with Chloroplast Phylogenomics: An Example from Bamboos

To our knowledge, only three nuclear loci (ITS, GBSSI, and LEAFY) have been used for Arundinarieae phylogeny (Guo and Li 2004; Peng et al. 2008; Zhang

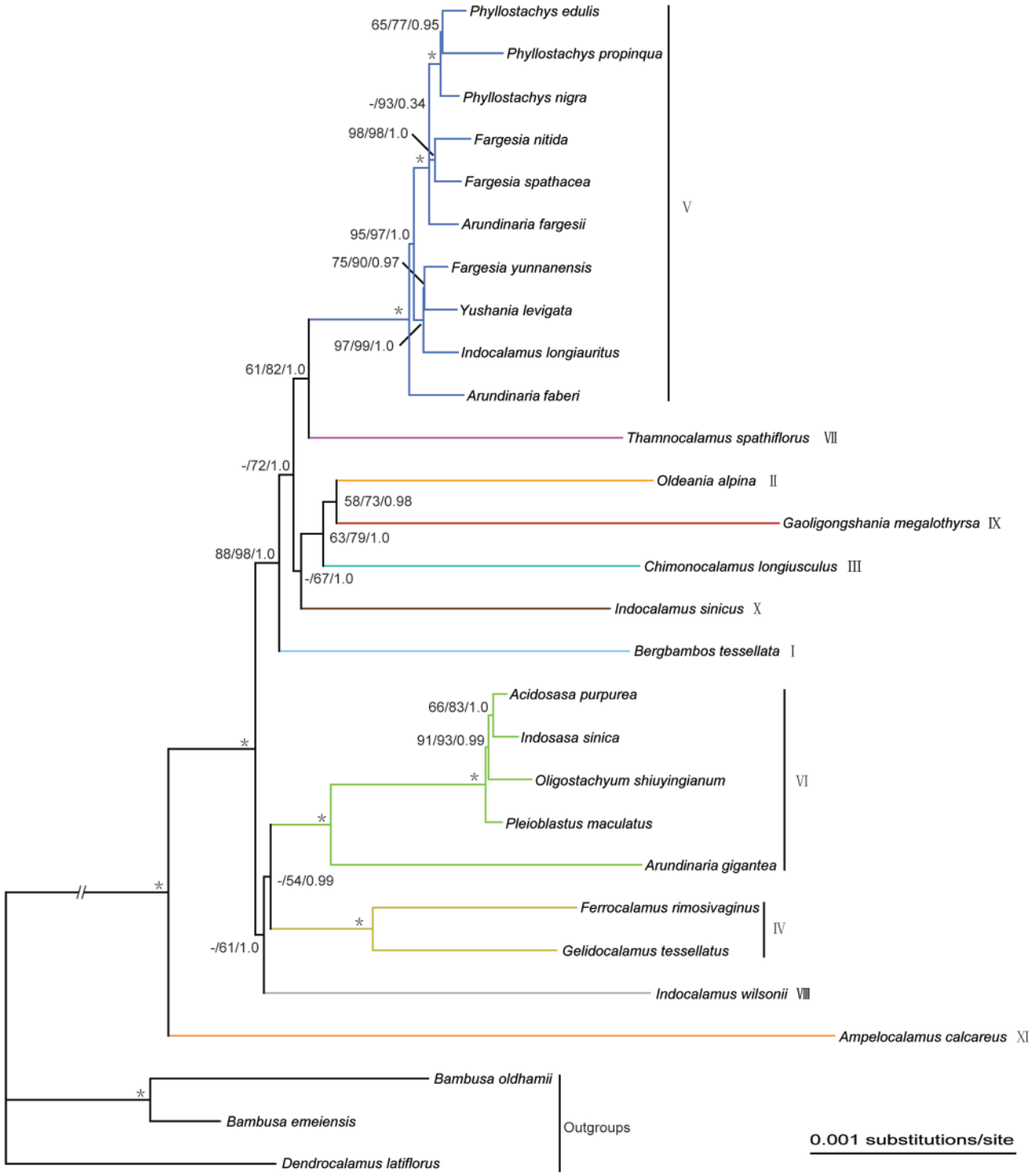

FIGURE 4. Phylogenetic relationships for Arundinarieae inferred from the 28-taxa data set with all the 11 major lineages of the tribe sampled. The complete chloroplast genome sequences were ML and BI analyzed with five partitions identified by software PartitionFinder. Numbers associated with nodes refer to MP bootstrap support values/ML bootstrap support values/Bayesian PPs. Asterisks represent nodes with maximal support values in all analyses. Dash denotes nodes unresolved in the 50% majority-rule topology in MP analysis. Colored branches in the online version of this figure indicate the 11 Arundinarieae lineages.

et al. 2012; Yang et al. 2013), largely due to the difficulties of identifying unambiguous single-copy nuclear loci in these tetraploid plants (Triplett and Clark 2010). The use of GBSSI and LEAFY, both of which can be considered as single copy in this group, also posed challenges for sequencing and identifying orthologous relationships (Zhang et al. 2012; Yang et al. 2013). Thus, most previous studies have focused on cpDNA sequences

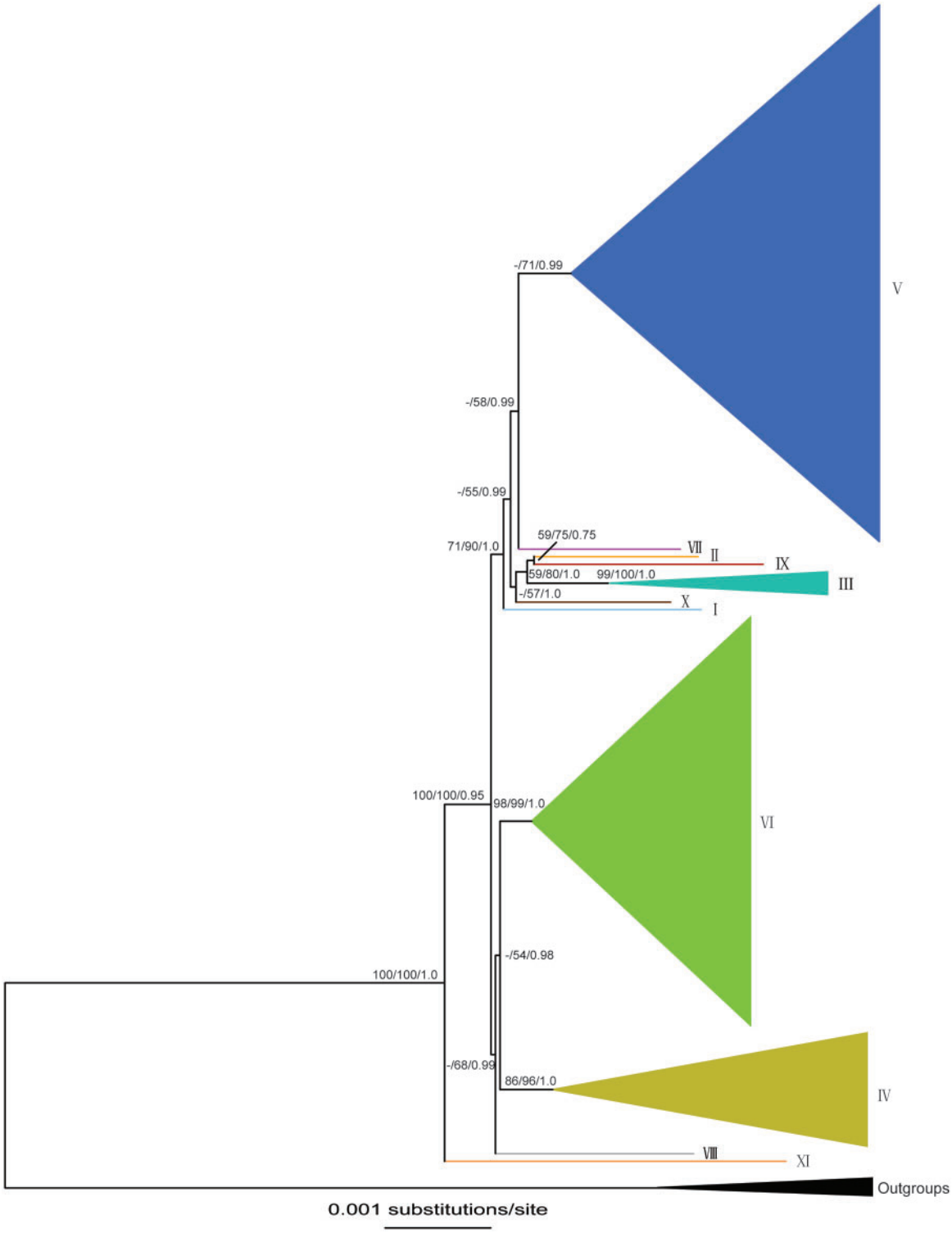

FIGURE 5. Summary of deep-level phylogenetic relationships for Arundinarieae based on ML analysis of the combined data set of 157 taxa with high percentage of missing data. The sizes of the triangles correspond to the number of taxa sampled in the lineages. Numbers associated with nodes refer to MP bootstrap support values/ML bootstrap support values/Bayesian PPs. Dash denotes nodes unresolved in the 50% majority-rule topology in MP analysis. See detailed phylogeny in Supplementary Figures S10–S12. Colored branches in the online version of this figure indicate the 11 Arundinarieae lineages.

TABLE 5. Likelihood tests of alternative topologies in CONSEL

Sequence data	Topology[a]	Δ ln *L*	AU probability
Complete genome sequence	(Outgroup(XI (VIII(IV, VI))((III, IX)(V, VII))))	Best	0.859
	(Outgroup(XI (VI(IV, VIII))((III, IX)(V, VII))))	13.2	0.222
	(Outgroup(XI (IV(VI, VIII))((III, IX)(V, VII))))	19.1	0.076
	(Outgroup(XI (IV(IV, IV, VIII))((III, IX)(V, VII))))	22.0	<0.001
Coding sequence	(Outgroup(XI (VI(IV, VIII))((III, IX)(V, VII))))	Best	0.773
	(Outgroup(XI (VIII(IV, VI))((III, IX)(V, VII))))	6.2	0.388
	(Outgroup(XI (IV(IV, VIII))((III, IX)(V, VII))))	8.2	0.243
	(Outgroup(XI (IV(IV, VI, VIII))((III, IX)(V, VII))))	12.7	<0.001
Noncoding sequence	(Outgroup(XI (VIII(IV, VI))((III, IX)(V, VII))))	Best	0.997
	(Outgroup(XI (VI(IV, VIII))((III, IX)(V, VII))))	15.5	0.003
	(Outgroup(XI (IV(VI, VIII))((III, IX)(V, VII))))	15.5	0.003
	(Outgroup(XI (IV, VI, VI))((III, IX)(V, VII))))	15.5	0.003

[a]The relationships within major lineages are constrained by the ML topology in Figure 1f.

but suffered from a lack of resolution, largely due to insufficient phylogenetic signal (Guo et al. 2001; Triplett and Clark 2010; Zeng et al. 2010). Moreover, sparse taxon sampling had been another problem for constructing Arundinarieae phylogeny until the studies of Triplett and Clark (2010) and Zeng et al. (2010). They both had a broad taxon sampling and identified the major lineages for Arundinarieae, providing a basis for building its backbone phylogeny. Here, we used the whole cp genomes from representatives of major lineages to infer the deep-level phylogeny of Arundinarieae.

We divided the complete cp genomes into three or two data subsets based on different genome regions or coding and noncoding sequences, respectively, to evaluate their performances on estimating phylogeny. As expected, no well-supported contradiction was recovered among phylogenies derived from the three different genome regions (Fig. 1c–e), because they are genetically linked. Moreover, phylogenetic resolution in analyses of these subsets is generally correlated with the number of variable sites they contain (Table 2). By contrast, and somewhat unexpectedly, phylogenetic resolution was comparable between the noncoding and coding subsets (Fig. 1a,b), despite the former containing nearly twice as many variable sites as the latter (1461 vs. 807; Table 2). However, these data subsets appear to support conflicting topologies involving the IV, VI, and VIII lineages as the AU test with the noncoding subset strongly rejects the topology inferred from the coding subset (Table 5). The main difference between these sequences is the rate of evolution, with the coding sequences more evolutionarily constrained. In addition, the coding sequences might be affected by natural selection, as has been documented in other grasses (Christin et al. 2008b). Although slowly and rapidly evolving sequences can have different phylogenetic utilities at various taxonomic levels (Jian et al. 2008) and only the coding genes were used in cp phylogenomics at high taxonomic levels previously (Jansen et al. 2007; Moore et al. 2007, 2010; Xi et al. 2012; Barrett et al. 2013), we advocate a total evidence approach to obtain a comprehensive view, especially at low taxonomic levels. In addition, the highest phylogenetic resolution was achieved in analysis of the complete genome sequences (Fig. 1).

Because of the differing rates and patterns of nucleotide substitution among the cp genome sequences (Palmer 1985), it may be inappropriate to analyze the genome as a whole, rather, data partitioning should be applied (Brown and Lemmon 2007; Li et al. 2008; Ward et al. 2010). Indeed, the more highly partitioned models usually have a much better fit to the data set than the less partitioned ones. However, the two most highly partitioned models that include significantly increased parameters both fit the data set more poorly than an unpartitioned model. Moreover, BI analysis using parameter-rich models often fail to converge (Marshall 2010), as was the case here. Our results also suggest that many commonly used partitioning schemes (e.g., by gene and codon position) cannot be guaranteed to be the best (Lanfear et al. 2012; Xi et al. 2012), because these are all inferior to the scheme identified by the software PartitionFinder (Table 3). It appears that the inferred phylogenetic trees from the *data-complete* data set are not sensitive to partitioning strategies (Table 4); however, this may be because our data set contains strong phylogenetic signal for robustly resolving relationships even without the use of an optimal partitioning scheme. In addition, a notable change in MLBS values for a key node between partitioned and unpartitioned analyses of the *lineage-complete* data set, which included a low percentage of missing data, was observed (Fig. 4 and Supplementary Fig. S9). Thus, it is always worth trying to analyze the data set partitioned in cp phylogenomic studies, even at lower taxonomic levels where low sequence divergence and homogeneity of molecular evolution are more likely to occur, whereas it is not necessarily better to consider each gene or noncoding region as a separate partition.

It is well known that both molecular characters and taxa sampled can affect the accuracy of phylogenetic estimates and there is long-standing debate about whether more characters or more taxa should be sampled (e.g., Rosenberg and Kumar 2001; Zwickl and Hillis 2002; Hedtke et al. 2006). The large amount of data used here is absolutely necessary for the substantial

improvement in resolution along the Arundinarieae backbone phylogeny (Fig. 1f) relative to previous studies (Triplett and Clark 2010; Zeng et al. 2010; Yang et al. 2013) as demonstrated by the analysis based on eight cpDNA loci used before combined with our taxon sampling (Supplementary Fig. S8). Our deep-level relationships may suffer from sparse taxon sampling (Fig. 1f); nevertheless, the phylogenetic trees with all the lineages sampled broadly corroborated these relationships (Figs. 1f and 4). Furthermore, even allowing for substantial difference in taxon sampling (28 vs. 157), the deep-level phylogenies inferred from the *lineage-complete* and *combined-incomplete* data sets were also in full agreement with each other (Figs. 4 and 5). The monophyly of major lineages is always strongly supported as demonstrated in Triplett and Clark (2010) and Zeng et al. (2010) and confirmed here, and it seems that more taxa sampled from them are likely not to alter the deep-level relationships recovered here. Meanwhile, the support values for these relationships of the major lineages of Arundinarieae tend to decrease as the number of sampled taxa increased (Figs. 1f, 4, and 5). We suspect that the 77.5% of missing data in the *combined-incomplete* data set, in which many taxa are more than 90% incomplete, is responsible for this. The MP method is more susceptible to the effects of missing data than the BI method (Flynn et al. 2005) and indeed the decrease of support values was more evident in the MP analyses than in the BI analyses (Figs. 1f, 4, and 5). However, we could not exclude the possibility that increased taxon sampling may be involved. Overall, it appears that the large molecular data set developed here played a fundamental role in recovering Arundinarieae backbone phylogeny, and more importantly, we can expect that the inferred phylogeny is robust to taxon sampling.

Although cp phylogenomics is usually employed at high taxonomic levels (Jansen et al. 2007; Moore et al. 2007, 2010; Xi et al. 2012; Barrett et al. 2013) and the cpDNA sequence divergence is very low among Arundinarieae species (Table 2), the relationships were still well resolved. More fundamentally, the resolved backbone phylogeny proved robust against sparse taxon sampling, the artifact of LBA, different tree-building methods and molecular evolutionary models, and partitioning strategies. The topology was further corroborated by the independent evidence from indels (Fig. 3). In conclusion, our results suggest that cp phylogenomics can be used to tackle other tough problems in plant phylogeny, taking full advantage of next-generation sequencing (Shendure and Ji 2008).

Short Internodes, Long Branches, and Phylogenetic Resolution

Our inferred phylogenetic trees possess a series of short internodes connected by long branches (Figs. 2, 4, and 5), indicating rapid radiation (Rokas and Carroll 2006; Whitfield and Lockhart 2007). It was thus not surprising that we could not completely resolve Arundinarieae backbone phylogeny despite the use of cp genome-scale data. The ambiguous relationships involving the IV, VI, and VIII lineages and partially resolved relationships among the II, III, IX, and X lineages were all associated with extremely short internodes combined with long branches (Figs. 2, 4, and 5). Topologies with short internodes can be resolved, as indicated by the resolution of relationships within major lineages (Figs. 1f and 4), but those that combine short and long internodes are more difficult to resolve. Conflicting phylogenetic signals present along the long branches, as revealed by the network analysis (Supplementary Fig. S14), hinders phylogenetic resolution. The incomplete data in the taxa from the I, II, and X lineages may also contribute to this problem; however, Wiens et al. (2005) found no correlation between the incompleteness of a taxon and the support for its phylogenetic placement. Moreover, all incomplete taxa had a large amount of data (Table 1). We thus conclude that the inclusion of taxa with missing data is not likely the underlying reason for the lack of phylogenetic resolution.

Incomplete lineage sorting is usually proposed as a potential explanation for incongruence among characters (Rokas and Carroll 2006; Whitfield and Lockhart 2007). Homoplasy is a better explanation in this case, since the cp genome is inherited as a single unit. Low phylogenetic signal along short branches, combined with the accumulation of numerous changes on long branches, can lead to a high level of homoplasy which appears as conflicting signal in the network analysis (Supplementary Fig. S14). Given this result, the slowly evolving coding sequences may be more reliable in inferring phylogenetic relationships among the II, III, IX, and X lineages (Supplementary Fig. S13) and thus the topology ((II, X), (III, IX)) is favored over competing alternatives.

Implications for Arundinarieae Phylogeny

The cpDNA-based phylogeny of the Arundinarieae represents one aspect of the overall evolutionary history of the group. In the future, additional evidence from the nuclear DNA sequences is needed in light of the potential discordance between phylogenies derived from cp and nuclear DNA sequences in Arundinarieae (Zhang et al. 2012; Yang et al. 2013). Nevertheless, our phylogeny provides many new and important insights into the evolution of Arundinarieae.

In our phylogeny, none of the eight lineages which are monotypic or represented by only a few taxa nest within the other three lineages (Triplett and Clark 2010; Zeng et al. 2010; Yang et al. 2013), validating their status as separate lineages. The XI lineage is the earliest diverging lineage within Arundinarieae and the remaining 10 lineages can be largely divided into two groups, with the IV, VI, and VIII lineages comprising one and the other seven lineages comprising

the other (Figs. 4 and 5). Within the second group, the I lineage is sister to all the others and the VII has close affinities to the V lineage, whereas the remaining four lineages form a monophyletic group with a topology ((II, X), (III, IX)). Most of the inferred relationships differ strongly from hypotheses based on morphology (Li 1999; Bamboo Phylogeny Group 2012); however, the apparent morphological disparity among closely related taxa is somewhat expected if the group has evolved through rapid radiation (e.g., Lerner et al. 2011; Drummond et al. 2012). Finally, a surprising finding is that the only two lineages (I and II) from the African mainland do not form a clade in the phylogenetic tree, and whether they have different geographical origins deserves further biogeographic work.

Supplementary Material

Supplementary material, including data files and online-only appendices, can be found in the Dryad data repository at http://dx.doi.org/10.5061/dryad.d5h1n.2.

Funding

This work was supported by the National Natural Science Foundation of China [grants U1136603, 31170204, and 30990244], and the Yunnan Province Government through an innovation team program [2009CI011].

Acknowledgements

The authors thank Lian-Ming Gao, Zhao-Ming Cai, and Hong-Mei Yang for assistance with field work; Lu Lu for help with obtaining two bamboo samples; Jun-Bo Yang, Hong-Tao Li, and Xian-Zhi Zhang for laboratory support; and Peter Fritsch for critically reading the article. They are also grateful to Frank Anderson, Roberta Mason-Gamer, and two anonymous reviewers for their valuable comments on an earlier version of this article.

References

APG III. 2009. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. Bot. J. Linn. Soc. 161:105–121.

Bamboo Phylogeny Group. 2012. An updated tribal and subtribal classification for the Bambusoideae (Poaceae). In: Gielis J., Potters G., editors. Proceedings of the 9th World Bamboo Congress, Antwerp, Belgium: World Bamboo Organization. p. 3–27.

Barrett C.F., Davis J.I., Leebens-Mack J., Conran J.G., Stevenson D.W. 2013. Plastid genomes and deep relationships among the commelinid monocot angiosperms. Cladistics 29:65–87.

Bergsten J. 2005. A review of long-branch attraction. Cladistics 21:163–193.

Bewick A.J., Chain F.J.J., Heled J., Evans B.J. 2012. The pipid root. Syst. Biol. 61:913–926.

Birky C.W. Jr., Maruyama T., Fuerst P. 1983. An approach to population and evolutionary genetic theory for genes in mitochondria and chloroplasts, and some results. Genetics 103:513–527.

Bouchenak-Khelladi Y., Salamin N., Savolainen V., Forest F., vander Bank M., Chase M.W., Hodkinson T.R. 2008. Large multi-gene phylogenetic trees of the grasses (Poaceae): progress towards complete tribal and generic level sampling. Mol. Phylogenet. Evol. 47:488–505.

Bouchenak-Khelladi Y., Verboom G.A., Savolainen V., Hodkinson T.R. 2010. Biogeography of the grasses (Poaceae): a phylogenetic approach to reveal evolutionary history in geographical space and ecological time. Bot. J. Linn. Soc. 162:543–557.

Brandley M.C., Schmitz A., Reeder T.W. 2005. Partitioned Bayesian analyses, partition choice, and the phylogenetic relationships of scincid lizards. Syst. Biol. 54:373–390.

Brown J.M., Lemmon A.R. 2007. The importance of data partitioning and the utility of Bayes factors in Bayesian phylogenetics. Syst. Biol. 56:643–655.

Burke S.V., Grennan C.P., Duvall M.R. 2012. Plastome sequences of two new world bamboos-*Arundinaria gigantea* and *Cryptochloa strictiflora* (Poaceae)-extended phylogenomic understanding of Bambusoideae. Am. J. Bot. 99:1951–1961.

Burlerigh J.G., Bansal M.S., Eulenstein O., Hartmann S., Wehe A., Vision T.J. 2011. Genome-scale phylogenetics: inferring the plant Tree of Life from 18,896 gene trees. Syst. Biol. 60:117–125.

Bystriakova N., Kapos V., Stapleton C., Lysenko I. 2003. Bamboo biodiversity: information for planning conservation and management in the Asia-Pacific region. UNEP-WCMC/INBAR.

Chen R.Y., Li X.L., Song W.Q., Liang G.L., Zhang P.X., Lin R.S., Zong W.X., Chen C.B., Feng X.L. 2003. Chromosome atlas of various bamboo species. Chromosome atlas of major economic plants genome in China IV. Beijing: Science Press.

Christin P.A., Besnard G., Samaritani E., Duvall M.R., Hodkinson T.R., Savolainen V., Salamin N. 2008a. Oligocene CO_2 decline promoted C_4 photosynthesis in grasses. Curr. Biol. 18:37–43.

Christin P.A., Salamin N., Muasya A.M., Roalson E.H., Russier F., Besnard G. 2008b. Evolutionary switch and genetic convergence on *rbcL* following the evolution of C_4 photosynthesis. Mol. Biol. Evol. 25:2361–2368.

Delsuc F., Brinkmann H., Philippe H. 2005. Phylogenomics and the reconstruction of the tree of life. Nat. Rev. Genet. 6:361–375.

Doyle J.J., Doyle J.L. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem. Bull. 19:11–15.

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

Felsenstein J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. Syst. Zool. 27:401–410.

Flynn J.J., Finarelli J.A., Zehr S., Hsu J., Nedbal M.A. 2005. Molecular phylogeny of the Carnivora (Mammalia): assessing the impact of increased sampling on resolving enigmatic relationships. Syst. Biol. 54:317–337.

Fu J.H. 2001. Chinese moso bamboo: its importance. Bamboo 22:5–7.

Guo Z.H., Chen Y.Y., Li D.Z., Yang J.B. 2001. Genetic variation and evolution of the alpine bamboos (Poaceae: Bambusoideae) using DNA sequence data. J. Plant Res. 114:315–322.

Guo Z.H., Li D.Z. 2004. Phylogenetics of the *Thamnocalamus* group and its allies (Gramineae: Bambusoideae): inference from the sequence of GBSSI gene and ITS spacer. Mol. Phylogenet. Evol. 30:1–12.

Hackett S.J., Kimball R.T., Reddy S., Bowie R.C., Braun E.L., Braun M.J., Chojnowski J.L., Cox W.A., Han K.L., Harshman J., Huddleston C.J., Marks B.D., Miglia K.J., Moore W.S., Sheldon F.H., Steadman D.W., Wirr C.C., Yuri T. 2008. A phylogenomic study of birds reveals their evolutionary history. Science 320:1763–1768.

Hedtke S.M., Townsend T.M., Hillis D.M. 2006. Resolution of phylogenetic conflict in large data sets by increased taxon sampling. Syst. Biol. 55:522–529.

Hodkinson T.R., Chonghaile G.N., Sungkaew S., Chase M.W., Salamin N., Stapleton C.M.A. 2010. Phylogenetic analyses of plastid and nuclear DNA sequences indicate a rapid late Miocene radiation of the temperate bamboo tribe Arundinarieae (Poaceae, Bambusoideae). Plant Ecol. Divers. 3:109–120.

Huelsenbeck J.P. 1998. Systematic bias in phylogenetic analysis: is the Strepsiptera problem solved? Syst. Biol. 47:519–537.

Huson D.H., Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23:254–267.

Jansen R.K., Cai Z., Raubeson L.A., Daniell H., dePamphilis C.W., Leebens-Mack J., MÜller K.F., Guisinger-Bellian M., Haberle R.C., Hansen A.K., Chumley T.W., Lee S., Peery R., McNeal J.R., Kuehl J.V., Boore J.L. 2007. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. Proc. Natl Acad. Sci. U. S. A. 104:19369–19374.

Janzen D.H. 1976. Why bamboos wait so long to flower. Ann. Rev. Ecol. Syst. 7:347–391.

Jeffroy O., Brinkmann H., Delsuc F., Philippe H. 2006. Phylogenomics: the beginning of incongruence? Trends Genet. 22:225–231.

Jian S., Soltis P.S., Gitzendanner M.A., Moore M.J., Li R., Hendry T.A., Qiu Y.L., Dhingra A., Bell C.D., Soltis D.E. 2008. Resolving an ancient, rapid radiation in Saxifragales. Syst. Biol. 57:38–57.

Katoh K., Kuma K., Toh H., Miyata T. 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res. 33:511–518.

Kelchner S.A., Bamboo Phylogeny Group. 2013. Higher level phylogenetic relationships within the bamboos (Poaceae: Bambusoideae) based on five plastid markers. Mol. Phylogenet. Evol. 67:404–413.

Kennedy M., Holland B.R., Gray R.D., Spencer H.G. 2005. Untangling long branches: identifying conflicting phylogenetic signals using spectral analysis, neighbor-net, and consensus networks. Syst. Biol. 54:620–633.

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

Lerner H.R.L., Meyer M., James H.F., Hofreiter M., Fleischer R.C. 2011. Multilocus resolution of phylogeny and timescale in the extant adaptive radiation of Hawaiian honeycreepers. Curr. Biol. 21:1838–1844.

Li C., Lu G., Ortí G. 2008. Optimal data partitioning and a test case for ray-finned fishes (Actinopterygii) based on ten nuclear loci. Syst. Biol. 57:519–539.

Li D.Z. 1999. Taxonomy and biogeography of the Bambuseae (Gramineae: Bambusoideae). In: Rao A.N., Rao V.R., editors. Bamboo—conservation, diversity, ecogeography, germplasm, resource utilization and taxonomy. Proceeding of training course cum workshop, Kunming and Xishuangbanna, Yunnan, China. Serdang, Malaysia: IPGRI-APO. p. 14–23.

Li D.Z., Wang Z.P., Zhu Z.D., Xia N.H., Jia L.Z., Guo Z.H., Yang G.Y., Stapleton C. 2006. Bambuseae in Flora of China (Wu Z.Y., Raven P.H., editors.). Vol. 22. Beijing & St. Louis: Science Press & Missouri Botanical Garden Press.

Li R., Zhu H., Ruan J., Qian W., Fang X., Shi Z., Li Y., Li S., Shan G., Kristiansen K., Li S., Yang H., Wang J., Wang J. 2010. De novo assembly of human genomes with massively parallel short read sequencing. Genome Res. 20:265–272.

Marshall D.C. 2010. Cryptic failure if partitioned Bayesian phylogenetic analyses: lost in the land of long trees. Syst. Biol. 59:108–117.

McClure F.A. 1973. The Bamboos: a fresh perspective. Cambridge: Harvard University Press. [Reprint 1993, Washington DC: Smithsonian Institution Press.]

Moore M.J., Bell C.D., Soltis P.S., Soltis D.E. 2007. Using plastid genome-scale data to resolve enigmatic relationships among the basal angiosperms. Proc. Natl Acad. Sci. U. S. A. 104:19363–19368.

Moore M.J., Soltis P.S., Bell C.D., Burleigh J.G., Soltis D.E. 2010. Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. Proc. Natl Acad. Sci. U. S. A. 107:4623–4628.

Moyle R.G., Andersen M.J., Oliveros C.H., Steinheimer F.D., Reddy S. 2012. Phylogeny and biogeography of the core babblers (Aves: Timaliidae). Syst. Biol. 61:631–651.

Nylander J.A.A., Ronquist F., Huelsenbeck J.P., Nieves-Aldrey J.L. 2004. Bayesian phylogenetic analysis of combined data. Syst. Biol. 53:47–67.

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 phylogenetics. Bioinformatics 24:581–583.

Ohrnberger D. 1999. The bamboos of the world: annotated nomenclature and literature of the species and the higher and lower taxa. Amsterdam: Elsevier Science.

Palmer J.D. 1985. Comparative organization of chloroplast genomes. Annu. Rev. Genet. 19:325–354.

Parks M., Cronn R., Liston A. 2009. Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. BMC Biol. 7:84.

Peng S., Yang H.Q., Li D.Z. 2008. Highly heterogeneous generic delimitation within the temperate bamboo clade (Poaceae: Bambusoideae): evidence from GBSSI and ITS sequences. Taxon 57:799–810.

Philippe H., Brinkman H., Lavrov D.V., Littlewood D.T.J., Manuel M., Wörheide G., Baurain D. 2011. Resolving difficult phylogenetic questions: why more sequences are not enough. PLoS Biol. 9:e1000602.

Posada D., Buckley T.R. 2004. Model selection and model averaging in phylogenetics: advantages of Akaike Information Criterion and Bayesian approaches over likelihood ratio tests. Syst. Biol. 53:793–808.

Posada D., Crandall K.A. 1998. Modeltest: testing the model of DNA substitution. Bioinformatics 14:817–818.

Rambaut A., Drummond A.J. 2007. Tracer v1.5 [Internet]. Available from: URL http://beast.bio.ed.ac.uk/Tracer.

Rambaut A., Grassly N.C. 1997. Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. Comput. Appl. Biosci. 13:235–238.

Rokas A., Carroll S.B. 2006. Bushes in the tree of life. PLoS Biol. 4:e352.

Ronquist F., Huelsenbeck J.P. 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574.

Rosenberg M.S., Kumar S. 2001. Incomplete taxon sampling is not a problem for phylogenetic inference. Proc. Natl Acad. Sci. U. S. A. 98:10751–10756.

Ruiz-Sanchez E. 2011. Biogeography and divergence time estimates of woody bamboos: insights in the evolution of neotropical bamboos. Bol. Soc. Bot. Mex. 88:67–75.

Schnitzler J., Barraclough T.G., Boatwright J.S., Goldblatt P., Manning J.C., Powell M.P., Rebelo T., Savolainen V. 2011. Causes of plant diversification in the cape biodiversity hotspot of South Africa. Syst. Biol. 60:343–357.

Shendure J., Ji H. 2008. Next-generation DNA sequencing. Nat. Biotechnol. 26:1135–1145.

Shimodaira H. 2002. An approximately unbiased test of phylogenetic tree selection. Syst. Biol. 51:492–508.

Shimodaira H., Hasewaga M. 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics 17:1246–1247.

Soltis D.E., Smith S.A., Cellinese N., Wurdack K.J., Tank D.C., Brockington S.F., Refulio-Rodriguez N.F., Walker J.B., Moore M.J., Carlsward B.S., Bell C.D., Latvis M., Crawley S., Black C., Diouf D., Xi Z., Rushworth C.A., Gitzendanner M.A., Sytsma K.J., Qiu Y.L., Hilu K.W., Davis C.C., Sanderson M.J., Beaman R.S., Olmstead R.G., Judd W.S., Donoghue M.J., Soltis P.S. 2011. Angiosperm phylogeny: 17 genes, 640 taxa. Am. J. Bot. 98:704–730.

Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688–2690.

Stapleton C.M.A. 2013. *Bergbambos* and *Oldeania*, new genera of African bamboos (Poaceae, Bambusoideae). PhytoKeys 25:87–103.

Sungkaew S., Stapleton C.M.A., Salamin N., Hodkinson T.R. 2009. Non-monophyly of the woody bamboos (Bambuseae; Poaceae): a multi-gene region phylogenetic analysis of Bambusoideae s.s.. J. Plant Res. 122:95–108.

Swofford D.L. 2002. PAUP*. Phylogenetic analysis using parsimony (* and other methods), version 4. Sunderland (MA): Sinauer Associates.

Tamura K., Dudley J., Nei M., Kumar S. 2007. MEGA 4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol. Biol. Evol. 24:1596–1599.

Triplett J.K., Clark L.G. 2010. Phylogeny of the temperate bamboos (Poaceae: Bambusoideae: Bambuseae) with an emphasis on *Arundinaria* and allies. Syst. Bot. 35:102–120.

Valente L.M., Savolainen V., Vargas P. 2010. Unparalleled rates of species diversification in Europe. Proc. R. Soc. B 277:1489–1496.

Ward P.S., Brady S.G., Fisher B.L., Schultz T.R. 2010. Phylogeny and biogeography of Dolichoderine ants: effects of data partitioning and relict taxa on historical inference. Syst. Biol. 59:342–362.

Waterway M., Hoshino T., Masaki T. 2009. Phylogeny, species richness, and ecological specialization in Cyperaceae tribe Cariceae. Bot. Rev. 75:138–159.

Whitfield J.B., Lockhart P.J. 2007. Deciphering ancient rapid radiations. Trends Ecol. Evol. 22:258–265.

Wiens J.J., Fetzner J.W., Parkinson C.L., Reeder T.W. 2005. Hylid frog phylogeny and sampling strategies for speciose clades. Syst. Biol. 54:719–748.

Wiens J.J., Kuczynski C.A., Smith S.A., Mulcahy D.G., Sites J.W., Townsend T.M., Reeder T.W. 2008. Branch length, support, and congruence: testing the phylogenomic approach with 20 nuclear loci in snakes. Syst. Biol. 57:420–431.

Wu F.H., Kan D.P., Lee S.B., Daniell H., Lee Y.W., Lin C.C., Lin N.S., Lin C.S. 2009. Complete nucleotide sequence of *Dendrocalamus latiflorus* and *Bambusa oldhamii* chloroplast genomes. Tree Physiol. 29:847–856.

Wu Z.Q., Ge S. 2012. The phylogeny of the BEP clade in grasses revisited: evidence from the whole-genome sequences of chloroplast. Mol. Phylogenet. Evol. 62:573–578.

Wyman S.K., Jansen R.K., Boore J.L. 2004. Automatic annotation of organellar genomes with DOGMA. Bioinformatics 20: 3252–3255.

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. U. S. A. 109:17519–17524.

Yang H.M., Zhang Y.X., Yang J.B., Li D.Z. 2013. The monophyly of *Chimonocalamus* and conflicting gene trees in Arundinarieae (Poaceae: Bambusoideae) inferred from four plastid and two nuclear markers. Mol. Phylogenet. Evol. 68:340–356.

Yang J.B., Li D.Z., Li H.T. 2014. Highly effective sequencing whole chloroplast genomes of angiosperms by nine novel universal primer pairs. Mol. Ecol. Resour. 14:1024–1031.

Zeng C.X., Zhang Y.X., Triplett J.K., Yang J.B., Li D.Z. 2010. Large multi-locus plastid phylogeny of the tribe Arundinarieae (Poaceae: Bambusoideae) reveals ten major lineages and low rate of molecular divergence. Mol. Phylogenet. Evol. 56:821–839.

Zhang Y.J., Ma P.F., Li D.Z. 2011. High-throughput sequencing of six bamboo chloroplast genomes: phylogenetic implications for temperate woody bamboos (Poaceae: Bambusoideae). PLoS One 6:e20596.

Zhang Y.X., Zeng C.X., Li D.Z. 2012. Complex evolution in Arundinarieae (Poaceae: Bambusoideae): incongruence between plastid and nuclear GBSSI gene phylogenies. Mol. Phylogenet. Evol. 63:777–797.

Zhou X., Xu S., Xu J., Chen B., Zhou K., Yang G. 2012. Phylogenomic analysis resolves the interordinal relationships and rapid diversification of the Laurasiatherian mammals. Syst. Biol. 61:150–164.

Zwickl D.J., Hillis D.M. 2002. Increased taxon sampling greatly reduces phylogenetic error. Syst. Biol. 51:588–598.