MOLECULAR ECOLOGY
RESOURCES

Molecular Ecology Resources (2014) **14**, 988–999 doi: 10.1111/1755-0998.12248

Identification of putative orthologous genes for the phylogenetic reconstruction of temperate woody bamboos (Poaceae: Bambusoideae)

LI-NA ZHANG,*†‡[1] XIAN-ZHI ZHANG,*†‡[1] YU-XIAO ZHANG,*† CHUN-XIA ZENG,† PENG-FEI MA,*† LEI ZHAO,† ZHEN-HUA GUO† and DE-ZHU LI*†

**Key Laboratory for Plant Diversity and Biogeography of East Asia, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, Yunnan 650201, China, †Plant Germplasm and Genomics Center, Germplasm Bank of Wild Species, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, Yunnan 650201, China, ‡Kunming College of Life Sciences, University of Chinese Academy of Sciences, Kunming, Yunnan 650201, China*

Abstract

The temperate woody bamboos (Arundinarieae) are highly diverse in morphology but lack a substantial amount of genetic variation. The taxonomy of this lineage is intractable, and the relationships within the tribe have not been well resolved. Recent studies indicated that this tribe could have a complex evolutionary history. Although phylogenetic studies of the tribe have been carried out, most of these phylogenetic reconstructions were based on plastid data, which provide lower phylogenetic resolution compared with nuclear data. In this study, we intended to identify a set of desirable nuclear genes for resolving the phylogeny of the temperate woody bamboos. Using two different methodologies, we identified 209 and 916 genes, respectively, as putative single copy orthologous genes. A total of 112 genes was successfully amplified and sequenced by next-generation sequencing technologies in five species sampled from the tribe. As most of the genes exhibited intra-individual allele heterozygotes, we investigated phylogenetic utility by reconstructing the phylogeny based on individual genes. Discordance among gene trees was observed and, to resolve the conflict, we performed a range of analyses using BUCKy and HybTree. While caution should be taken when inferring a phylogeny from multiple conflicting genes, our analysis indicated that 74 of the 112 investigated genes are potential markers for resolving the phylogeny of the temperate woody bamboos.

Keywords: Arundinarieae, *Fargesia*, orthologous genes, *Phyllostachys*, phylogeny

Received 5 August 2013; revision received 2 March 2014; accepted 4 March 2014

Introduction

The bamboos (Poaceae: Bambusoideae) consist of about 1500 species in 116 genera, and, together with Ehrhartoideae (e.g. rice) and Pooideae (e.g. wheat), are grouped into the BEP clade (Bouchenak-Khelladi *et al.* 2008; Bamboo Phylogeny Group 2012). They are very important economically, ecologically and culturally in Asia, Africa and Latin America (Soderstrom 1981; Li 1999). Most members of the Bambusoideae are woody species and have a long interval of flowering, which can be used as food and materials for buildings and furniture while some species are herbaceous and produce flowers annually. Recent molecular phylogenetic studies have indicated that Bambusoideae could be classified into three tribes: Arundinarieae (temperate woody bamboos), Bambuseae (tropical woody bamboos) and Olyreae (herbaceous bamboos) (Bouchenak-Khelladi *et al.* 2008; Sungkaew *et al.* 2009). The temperate woody bamboos are mainly distributed either in the North Temperate Zone or at high elevations in tropical regions, consist of more than 500 species in 28 genera and are recognized as tetraploids with a chromosome complement of $2n = 48$ (Soderstrom 1981; Orhrberger 1999; Li *et al*. 2006; Bamboo Phylogeny Group 2012). Although the temperate woody bamboos are well resolved as monophyletic, the taxonomy within this group is extraordinarily troublesome. Recent studies of large plastid data sets resolved the tribe into eleven major lineages, while relationships among and within the lineages are still unclear (Zeng *et al.* 2010; Yang *et al.* 2013), although chloroplast phylogenomic analyses provided some improvements in resolving relationships among

Correspondence: De-Zhu Li, Fax: +86 871 6521 7791;
E-mail: dzl@mail.kib.ac.cn

[1] These authors contributed equally to this work.

lineages (Zhang *et al.* 2011; Ma 2012). Similarly, analysis using the two nuclear genes GBSSI and LEAFY revealed several major lineages, yet the relationships both among and within the lineages were not resolved, and there was much incongruence between the topologies based on the nuclear genes and the chloroplast regions (Zhang *et al.* 2012c; Yang *et al.* 2013). Furthermore, the relationships inferred from the molecular data strongly conflict with morphological classification. In the tribe, morphological characters among the species are highly diverse, while the level of genetic variation is relatively low (Guo & Li 2004; Peng *et al.* 2008; Triplett & Clark 2010; Zeng *et al.* 2010; Zhang *et al.* 2011). Incomplete lineage sorting, hybridization and introgression, low variability in genes, and recent rapid radiation in the tribe have been inferred to be the causative agents behind such observed patterns. Overall, it is thought that the tribe has undergone a very complex evolutionary trajectory (Triplett & Clark 2010; Triplett *et al.* 2010; Zeng *et al.* 2010; Zhang *et al.* 2012c).

Given their uniparental heredity and haploidy, plastid genes have a limited utility for resolving the phylogeny of a tribe that has a complex evolutionary history. In such cases, nuclear genes offer a viable option to reveal evolutionary relationships, as they have biparental heredity and multiple independent loci (Small *et al.* 2004). Moreover, the nuclear genome usually evolves faster than the chloroplast genome in plants (Wolfe *et al.* 1987; Gaut 1998), which could provide more phylogenetic information. Despite these benefits, only three nuclear genes, GBSSI, ITS (Guo & Li 2004; Peng *et al.* 2008; Zhang *et al.* 2012c) and LEAFY (Yang *et al.* 2013), have been utilized in the temperate woody bamboos. Thus, to adequately resolve the phylogenetic relationships within the temperate woody bamboos, it is essential to develop more desirable nuclear gene markers.

With the development of sequencing technologies, more and more molecular data are available, especially whole-genome data. The advent of next-generation sequencing (NGS) has changed studies of molecular phylogenetics (McCormack *et al.* 2011; Egan *et al.* 2012). The greatest challenge when using nuclear genes to reconstruct phylogenies is identifying orthologous genes. By analysing whole-genome data sets using appropriate software, such as Inparanoid (O'Brien *et al.* 2005), OrthoMCL (Li *et al.* 2003) and HaMStR (Ebersberger *et al.* 2009), we now have the ability to distinguish orthologous genes from paralogous genes (Duarte *et al.* 2010; Zhang *et al.* 2012a). To resolve phylogenies at lower ranks, that is, at the generic or specific level, introns could be more informative than exons because intron regions, which are under lower selective pressure, generally display greater variation than exons (Sang 2002). However, to utilize introns for resolving phylogenies, we must consider the presence of intra-individual allele heterozygotes (IIAHs), a phenomenon referring to heterozygotic introns within the same individual, which may complicate phylogenetic analysis. Although often ignored in phylogenetic analyses, an increasing body of studies indicates that these IIAHs include rich evolutionary information (Sota & Vogler 2001; Sota & Sasabe 2006; Yu *et al.* 2011).

Whole-genome data sets for six species in Poaceae (*Oryza sativa*, *Brachypodium distachyon*, etc.) and two large transcriptome data sets in Bambusoideae [i.e. for *Phyllostachys edulis* (Arundinarieae) and *Dendrocalamus latiflorus* (Bambuseae)] have been recently made available (Peng *et al.* 2010; Zhang *et al.* 2012b). These data sets offer an important opportunity to reconstruct the phylogeny of the temperate woody bamboos using genome-wide markers. In this study, we identified a set of single copy orthologous genes using the genomic data sets from the grasses and the transcriptome data sets from the bamboos. Furthermore, we analysed the phylogenetic utility of these genes by identifying the relationships among five species from two genera of the temperate woody bamboos. Considering the polyploidy of bamboos, NGS, which has been suggested to resolve the sequencing problem of multiple gene copy in polyploidy (Griffin *et al.* 2011), was adopted to sequence large multiple genes for each species. Finally, we attempted to infer causes for the observed discordance among gene trees through a range of analyses.

Materials and methods

Identification of single copy orthologues

To identify desirable genes for phylogenetic analysis, two methods were implemented. For the first, we extracted coding sequences (CDs) from two model grasses, *Oryza sativa* from Ehrhartoideae (http://rice.plantbiology.msu.edu) and *Brachypodium distachyon* from Pooideae (http://www.brachybase.org), and downloaded the cDNA sequences of *Phyllostachys edulis* from NCBI (http://www.ncbi.nlm.nih.gov). We performed a self-BLASTN with the *P. edulis* database, which consists of five cDNA libraries, to ensure the database contained no identical sequences. The *P. edulis* database was used as the query to BLASTN search against *O. sativa* and *B. distachyon* databases, respectively, at *E*-values of 10^{-5}, and vice versa. We collected sequences with one hit, with identity ≥70%, and with an alignment length of at least 100 bp in each BLAST result. Finally, we retrieved sequences that were shared in all of the results, which were thus considered single copy genes. To ensure these genes were orthologues, we submitted them online to

OrthoMCL DB, using the default settings (http://www.orthomcl.org/cgi-bin/OrthoMclWeb.cgi). Sequences that clustered into OrthoMCL groups were identified as orthologues.

For the second method, we included another bamboo transcriptome database, *Dendrocalamus latiflorus*, cDNA sequences of which were extracted from NCBI. To search for putative orthologous genes, we then compared the *B. distachyon*, *D. latiflorus*, *O. sativa* and *P. edulis* (without removing identical sequences) databases against each other using ORTHOMCL v1.4, a graph-clustering algorithm for grouping proteins into orthologue groups based on sequence similarity, with default parameters (Li *et al.* 2003). We collected the clusters that included exactly one gene from each genome from the OrthoMCL results, and the genes from these clusters were identified as single copy orthologues. The workflow is shown in Fig. 1.

Primer development for PCR

To resolve the phylogenies at generic or specific level, we chose genes that had a moderate portion of introns. Since only bamboo transcriptome data were available, we referred to the *O. sativa* and *B. distachyon* genome annotations when developing primers for facilitating gene amplifications. To select genes with high phylogenetic resolution at lower rankers, we analysed the variability of the genes in three taxa from *Oryza*: *O. glaberrima*, *O. sativa* subsp. *indica* and *O. sativa* subsp. *japonica*. Finally, we selected the genes that were shared in common between the two methods and that demonstrated high variability and a putative amplified length of 1.5–2 kb. Based on the results, we used Primer premier 5 (Lalitha 2000) to design primers for the selected genes, with cDNA sequences from *P. edulis* as the reference. Moreover, we added four genes that were previously identified as single copies to the marker candidates, TPI (Xu & Hall 1993), LEAFY (Bomblies & Doebley 2005), GBSSI (Dian *et al.* 2003) and GPA1 (Seo *et al.* 1995), and one low-copy gene, PmFT (Hisamoto *et al.* 2008).

Taxa sampling and DNA extraction

Referring to molecular and morphological studies (Li *et al.* 2006; Zeng *et al.* 2010; Zhang *et al.* 2012c), we found

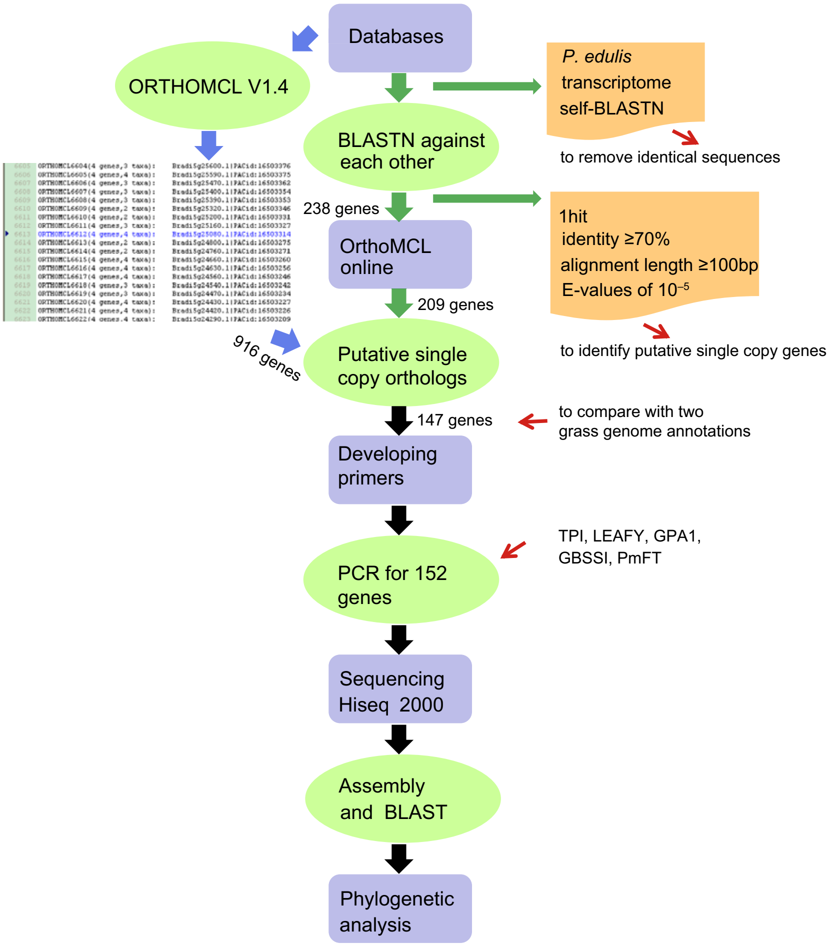

Fig. 1 Flow chart of identification and validation of putative orthologues. Green arrows indicate the first method used to identify putative orthologues. We performed bidirectional BLASTN between *Phyllostachys edulis* and the two model grass (*Oryza sativa* and *Brachypodium distachyon*), respectively, to select single copy genes which were then submitted to OrthoMCL DB to identify orthologues. Blue arrows direct the second method. We used local OrthoMCL to identify putative single copy orthologues. Dark arrows indicate steps carried out to test the phylogenetic utility of the marker candidates. We amplified these genes in five species sampled from Arundinarieae, sequenced successfully amplified genes by next-generation sequencing and then inferred the phylogenetic utility of these genes based on BI analysis.

that Clade V of Arundinarieae contains the most genera and is the most morphologically diverse. In this clade, *Phyllostachys* and *Fargesia* are the two greatest species-rich genera. The former has been treated as a well-defined monophyletic group by taxonomists, but its relationship to other genera was unclear. The circumscription of the latter genus is controversial, and its intergeneric relationship is even less well defined. To further investigate these two genera and Clade V of the temperate woody bamboos, we chose to analyse the phylogenetic utility of the identified genes in resolving the relationships between three species of *Phyllostachys* and two of *Fargesia* (Table 1). Total genomic DNA was extracted from silica-dried leaves using a modified CTAB procedure (Doyle & Doyle 1987).

PCR amplification and sequencing

For the five taxa, PCR amplifications were performed using all of the primers by standard methods. The PCRs contained 10 μL of 2× Taq PCR MasterMix (TIANGEN, Beijing, China) without dye, 1 μL of each primer (5 μM), 1 μL of template DNA (~50 ng/μL) and sterilized ddH_2O to a final volume of 20 μL. For some more difficult amplifications, we performed long PCRs that contained 0.25 μL of LA Taq (5 U/μL) (TaKaRa, Dalian, China), 2.5 μL of 10× LA PCR Buffer II (Mg^{2+} Plus), 3 μL of dNTP mixture (2.5 mM), 1 μL of each primer (5 μM), 1 μL of template (~50 ng/μL) and sterilized ddH_2O to a final volume of 25 μL. The PCR protocols for each marker are presented in Table S2 (Supporting information). The PCR products that did not produce more than three bands when visualized by 1.0% agarose gel electrophoresis with a 100-bp ladder (GeneRuler Plus DNA Ladder; Thermo Scientific, Carisbad, USA) were retained and quantified by a NanoDrop ND1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA). For each taxon, we then pooled the PCR products of every successfully amplified gene with the same weight, up to a final quantity of DNA of 20 μg. These pooled samples were prepared into short-insert 170-bp libraries according to the manufacturer's instructions (Illumina, San Diego, CA, USA). These samples were indexed before being mixed together and were then pair-end sequenced with a read length of 90 bp in an Illumina Genome Analyzer II at the BGI in Shenzhen, China.

Table 1 Information for taxa used in this study

Taxon	Voucher association	Locality
Fargesia Franchet		
Fargesia nitida (Mitford) Keng ex Yi	Zhang08017	Wolong, Sichuan, China
Fargesia spathacea Franchet	MPF10141	Wanyuan, Sichuan, China
Phyllostachys Siebold and Zuccarini		
Phyllostachys edulis (Carrière) Houzeau	MPF10163	Kunming Institute of Botany, China
Phyllostachys nigra var. *henonis* (Mitford) Stapf ex Rendle	MPF10172	Kunming Institute of Botany, China
Phyllostachys nidularia Munro	ZLN-2011069	Kuntong, Anji, Zhejiang, China

Voucher specimens are deposited in KUN (herbarium of the Kunming Institute of Botany, China).

De novo assembly and BLAST

Raw reads obtained from the sequencing run were first filtered to remove adapter contamination, and then reads containing more than 50% bases with a Phredscaled probability (Q) <20 were discarded. VELVET version 1.2.07 (Zerbino & Birney 2008; Zerbino *et al.* 2009) was used to assemble the clean reads into contigs with a *k*-mer value of 69 and a minimum contig length of 100 bp. To investigate the assembly, we used AMOS version 3.1.0 (Schatz *et al.* 2011) to convert the velvet_asm.afg file to the velvet_asm.ace file, which was then read by EAGLEVIEW version 2.2 (Huang & Marth 2008). This visualization of contigs with mapping reads allowed us to validate single-nucleotide polymorphisms (SNPs) in the contigs. The contigs were imported to GENEIOUS version 4.8.5 (Drummond *et al.* 2009) for assembly. Polymorphic sites (minor allele frequency >0.3) were coded according to the International Union of Pure and Applied Chemistry (IUPAC) nucleotide code. When sites from the Geneious assembly were identified as polymorphic (i.e. when sites overlapped by two heterogeneous contigs), the consensus sequence was divided into several sequences to separate those contigs. Finally, we performed BLASTN searches on the NCBI nucleotide database (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and the Rice Genome Annotation Project (http://rice.plantbiology.msu.edu/analyses_search_blast.shtml) with the resulting sequences and compared the outcomes with the previously identified orthologues to find target genes. If certain genes in some taxa were absent, we attempted to assemble them using corresponding gene sequences of the other taxa as references.

Phylogenetic analysis of individual genes

Genes shared in at least four species were used for phylogenetic analysis. We aligned every gene separately in Geneious using MUSCLE Alignment (Edgar 2004) and adjusted them manually where necessary. All characters in the data matrix were unordered and equally weighted, and gaps were treated as missing data without coding. In addition, for some downstream analysis, we created a data set containing genes common among all five taxa. For the data set, we retained one sequence for the genes with multiple sequences according to the following protocol: (i) if these sequences fell into a monophyletic clade in gene trees, we chose one randomly; (ii) if they did not group into one clade, we chose the sequence that was the most distant from the other genus based on a distance matrix (calculated in MESQUITE version 2.74 (Maddison & Maddison 2001) using uncorrected pairwise distance for each gene) and we discarded the sequence that was not grouped with the congeners. We referred to this data set as 'single-sequence data'. Unrooted gene trees were reconstructed by Bayesian inference (BI) method using matrices from each gene separately. For each gene in the single-sequence data, we used *Fargesia* as the outgroup. Prior to BI analysis, we applied JMODELTEST 2.1.2 (Guindon & Gascuel 2003; Posada 2008) to select optimized models for sequence evolutions under Akaike Information Criterion (Posada & Buckley 2004). Bayesian analysis was performed using the program MRBAYES version 3.1.2 (Ronquist & Huelsenbeck 2003), and the priors were set according to the best-fit model (details in Tables S3 and S4, Supporting information). The Markov chain Monte Carlo (MCMC) chains were run for 4 million generations, and trees were sampled every 1000 generations. In the final MCMC run, the average standard deviation of split frequencies dropped below 0.01. The first 25% of trees were discarded as burn-in, and the consensus tree and posterior probabilities were calculated from the remaining trees.

Species tree estimating in a coalescent process

We implemented the program BEST v2.3 (Liu & Pearl 2007) to infer species tree based on the single-sequence data. A priori, θ was set to an inverse gamma distribution [thetapr = invgamma (3, 0.003)], and the gene mutation rate was set to default [GeneMuPr = uniform (0.5, 1.5)]. Gene trees were generated from individual genes using the best-fit models (as described above). A modified MRBAYES was run for 100-million generations with four chains and two runs, sampling trees every 1000 generations. In the last run, the average standard deviation of split frequencies dropped below 0.01, and the first 25% of trees were discarded as burn-in.

Bayesian concordance of gene trees and phylogenetic network analysis

The discordance among gene trees was assessed and, to identify the concordant loci, we calculated Bayesian concordance factors (CFs) using BUCKY version 1.4.0 (Ané *et al.* 2007; Larget *et al.* 2010). We used the individual gene trees generated from single-sequence data in MrBayes for BUCKy analysis. The analysis was run for 10-million generations, with an initial burn-in of 1-million generations, four independent replicate runs, four MCMC chains, and a range of a priori α (α = 0.1, 1, 10, 20, 40, 60, 80, 100).

To investigate incomplete signals among the concordant genes resulting from the BUCKy analysis, we performed a 'phylogenetic network' analysis. The combined data from these genes were imported to SPLITSTREE4 (Huson & Bryant 2006) for neighbour-net analysis (Bryant & Moulton 2002, 2004) based on a model-corrected distance (using the best-fit model which we selected as described above).

Estimating hybridization in the presence of coalescence

To determine whether incomplete lineage sorting or introgression is responsible for the incongruence among gene trees, we used the HybTree method, which can identify hybridization with coalescence (Degnan & Salter 2005; Meng & Kubatko 2009; Gerard *et al.* 2011). The level of introgression is indicated by the value of γ. The consensus trees from single-sequence data in MrBayes were used by the method to detect hybridization events. The parameters were set to default, and 'iterate' was selected to have the three *Phyllostachys* species alternate as the potential hybrid species.

Results

Identification of putative single copy orthologues and development of markers

Through bidirectional BLASTN searches, 238 genes were initially identified (Table S1, Supporting information). After screening by OrthoMCL DB, we obtained 209 genes as putative single copy orthologues. From the local OrthoMCL method, we identified 916 genes as putative single copy orthologues that were shared among four genomes (TXT1). We compared the sequences resulting from these two methods, and, focusing on *Phyllostachys edulis*, identified the shared sequences as candidates for phylogenetic markers (Table S1, Supporting information). In total, 170 primer pairs for 152 genes were used as marker candidates; of those, we developed 165 primer pairs to amplify 147 putative single copy orthologues.

PCR amplification, sequencing, assembly and BLAST

Of the 152 genes, 112 were successfully amplified, with the partial failing of amplifications in some taxa. The amplified length ranged from 800 to 3000 bp, with most amplicons falling in the 1500–2000 bp range. Amplification failure may be due to unmatched primers that resulted from differences in the boundaries of introns and exons of genes between bamboos and rice. Illumina paired-end sequencing produced *c.* 400 Mb of clean data for each taxon. Using the Velvet assembler, we obtained 1209–2246 nodes. More information for the de novo assembly is found in Table 2. Of the 112 genes, 101 were obtained after sequencing, assembly and BLAST, with an average coverage of 100×. Of the 101 genes, 75 were shared in all five species, 11 were shared in four species, six were shared only in *Phyllostachys*, two were shared only in *Fargesia* and the remainder was shared in less than three species from the two genera. All the sequences are deposited in GenBank, and their Accession nos. are listed in Table S5 (Supporting information). The missing sequences of some genes in some taxa may be caused by low-quality PCR products, failure in sequencing or assembling for the complexity of gene structure.

Phylogenetic analysis using individual genes

Of the 101 genes, 86 were shared between at least four species and were therefore used for reconstructing phylogenies. When two divergent types of sequences presented in some gene and these sequences fell into two well-supported sister clades in the gene trees (e.g. Fig. 2), we divided the sequences of these genes into two data sets that we arbitrarily labelled ‘a’ and ‘b’. Besides, two

Table 2 Summary of NGS sequencing and de novo assembling

	fni	fsp	ped	pnh	pni
Number of nodes	2246	1256	1209	1419	1629
N50 (bp)	542	543	903	475	518
Aligned reads (%)	30.2	30.4	38.8	26.8	25.8
Total clean reads	4 743 056	2 917 036	5 063 556	4 017 918	4 432 408
SRA accession ByRun	SRR1175742	SRR1175743	SRR1175738	SRR1175710	SRR1175741

‘fni’, *Fargesia nitida*; ‘fsp’, *Fargesia spathacea*; ‘ped’, *Phyllostachys edulis*; ‘pnh’, *Phyllostachys nigra* var. *henonis*; ‘pni’, *Phyllostachys nidularia*; NGS, next-generation sequencing.

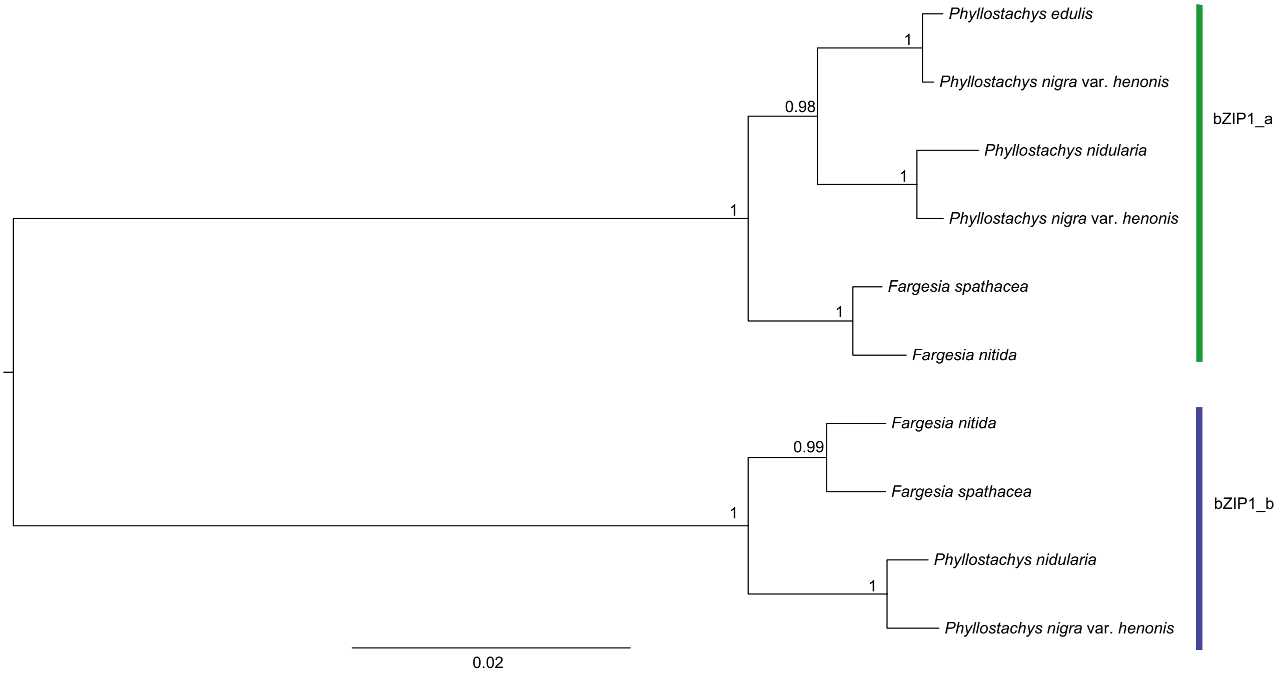

Fig. 2 The topology of a gene tree generated from a gene presenting two divergent types of sequences. Here, we presented the gene tree of bZIP1. Numbers on the branches represent posterior probabilities.

regions of some gene were sequenced which were labelled with '1' and '2'. Thus, we got 105 nucleotide sequence matrices rather than 86 (based on the 105 matrices, we obtained 82 matrices for single-sequence-data) for phylogenetic analysis. Topology of each gene tree generated by BI is listed in Table S6 (Supporting information). Among the gene trees, tree topological incongruence was primarily found in the relationships among the three species in *Phyllostachys*: 47% of gene trees supported the sister relationship of *P. edulis* and *P. nigra* var. *henonis* (A); 16% of gene trees supported the sister relationship of *P. nidularia* and *P. nigra* var. *henonis* (B); in one topology with 20% of gene trees supporting, *P. nigra* var. *henonis* contained sequences that clustered with the other two congeners (C); 5% of gene trees supported the sister relationship of *P. nidularia* and *P. edulis* (D); 5% of gene trees showed that *Phyllostachys* was nested with *Fargesia* (E); and 7% of gene trees were complex, for multiple types of sequences were obtained in some taxa (F). The proportions of the six main gene tree topologies are shown in Fig. 3. Based on the phylogenetic analysis, we found that, of the 86 genes, despite 12 supporting topologies E or F, 74 could be potential markers for resolving the phylogeny of the temperate woody bamboos.

BEST species tree analysis

In the species tree generated from single-sequence data, *Phyllostachys* and *Fargesia* were reciprocally monophyletic with posterior probabilities of 1.0 (Fig. 4). Within the *Phyllostachys* clade, *P. edulis* and *P. nigra* var. *henonis* were placed together as the sister group to *P. nidularia* with a posterior probability of 1.0, and the branches were very short.

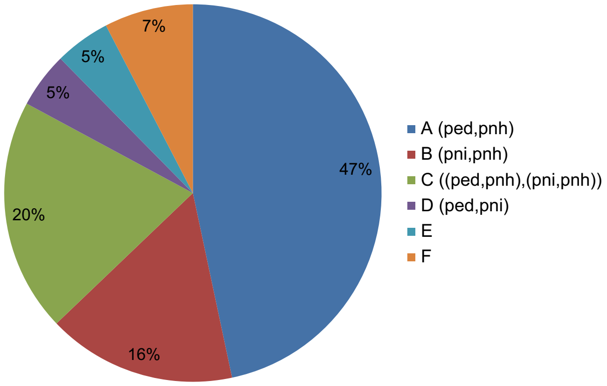

Fig. 3 Proportions of gene trees that support six topologies separately. 'ped', 'pnh' and 'pni' standard for *Phyllostachys edulis*; *P. nigra* var. *henonis*; and *P. nidularia*, respectively (the same below). E standards for the topology in which species from the two genera were nested. F standards for the complex topology which was caused by multiple types of sequences in some taxon.

Bayesian concordance analysis and phylogenetic network analysis

The results of the BUCKy analysis based on 82 genes with a range of a priori α sets showed no differences. All the tree topologies are listed in Table S7 (Supporting information) with $\alpha = 0.1$. The primary concordance tree (Fig. 5a) was similar to the estimated species tree by BEST. The CF for separating *Phyllostachys* from *Fargesia* was 78.995 (95% credible interval: 78, 79), which represents the number of genes supporting the branch. The clade where *P. edulis* grouped with *P. nigra* var. *henonis* received a CF of 49.010 (95% credible interval: 44, 53), while CF for the sister relationship of *P. nidularia* and *P. nigra* var. *henonis* was 27.097 (95% credible interval: 23, 32). The sister relationship of *P. nidularia* and *P. edulis* received the lowest CF of 3.899 (95% credible interval: 3, 6).

Through BUCKy analysis, we extracted the 49 concordant genes and investigated incomplete phylogenetic signals among these genes with network analysis. The neighbour-net graph (Fig. 5b) resembled the primary concordance tree. However, the network formed a box-like structure when representing the relationships of the three *Phyllostachys* species, suggesting some conflicting phylogenetic signals even among these genes.

Hybridization estimation

We had the three *Phyllostachys* species alternate as a potential hybrid, but we only detected a hybridization event in the case where *P. nigra* var. *henonis* was used as a hybrid, producing an estimated level of hybridization of 0.33. In the other two cases, γ was equal to one or zero, indicating no hybridization event. Estimated branch lengths (in coalescent units) along the hybrid species phylogeny were $t_1 = 0.000000011$, $t_2 = 1.2081$ and $t_3 = 2.1106$. Figure S1 (Supporting information) is a model tree illustrating to which branch t_i refers.

Discussion

Identification of single copy orthologues based on transcriptome and genome databases

When phylogenetic studies utilize nuclear genes, validation of orthologous genes is necessary (Fares *et al.* 2006). Although many methods have been used to identify orthologues in species with complete genomes (Li *et al.* 2003; O'Brien *et al.* 2005; Wall *et al.* 2008), few methods using incomplete genomes have been developed (Lee *et al.* 2002; Wu *et al.* 2006; Ebersberger *et al.* 2009). In this study, we used BLASTN combined with OrthoMCL DB and local OrthoMCL, respectively, to identify the single

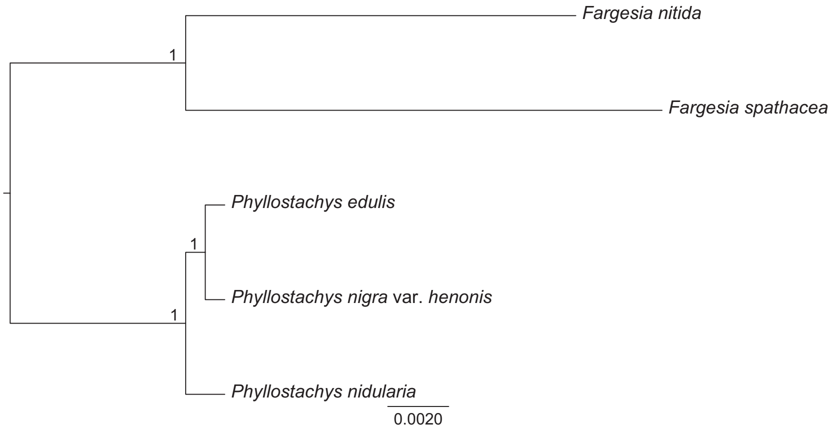

Fig. 4 Species tree estimated by BEST method (Liu & Pearl 2007). Numbers on the branches represent posterior probabilities.

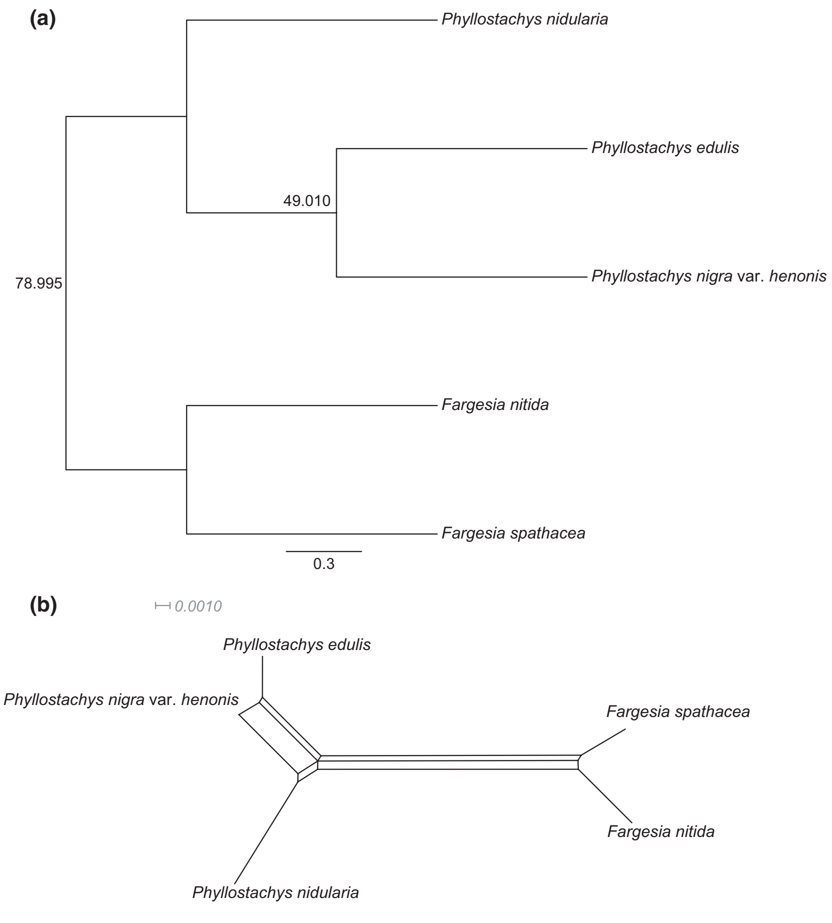

Fig. 5 The primary concordance tree from BUCKy analysis (Ané *et al.* 2007; Larget *et al.* 2010) (a) and the neighbour-net graph (Bryant & Moulton 2002, 2004) based on 49 concordant genes (b). (a) Numbers on branches are concordance factors and indicate the number of loci supporting a branch.

copy orthologues based on transcriptome and genome databases. We found that single copy orthologues obtained by the latter method included most of the genes identified by the former one; fewer genes were identified by the former one maybe due to the condition we set to select single copy genes. Furthermore, after comparing our results with the recently released draft genome of *Phyllostachys edulis* (Peng *et al.* 2013), we found that the orthologous genes identified in our study were mostly consistent with their gene-cluster results in which one gene was from *P. edulis* and one or two genes came from the other genomes. Although there were some discrepancies between the results from the transcriptome and genome data, most predicted single copy orthologues were consistent. Thus, it is an efficient approach to use transcriptome data and related species genome data as references for OrthoMCL to identify putative single copy orthologues in nonmodel organisms.

Discordance among gene trees and the potential causes

With the increased use of multiple genes for phylogenetic inference, incongruence among gene trees has been recognized as a prevalent problem in reconstructing phylogenies (Nichols 2001; Soltis *et al.* 2004; Qiu *et al.* 2006; Rokas & Carroll 2006). Incomplete lineage sorting, introgression and hybridization, horizontal gene transfer, and gene duplication and stochastic loss are speculated to account for the discordance (Doyle 1992; Maddison 1997). The discordance among gene trees detected in our results can be divided into three groups: one within *Phyllostachys*, as presented in topologies A, B, C and D; one in topology E where species from the two genera are nested with each other; and one in topology F, which was reconstructed based on genes having multiple divergent sequences in some species. Potential causes for the discordance found in topology E are unclear; however, it may result from incomplete lineage sorting, paralogy or introgression. Given that the temperate woody bamboos are polyploid (Soderstrom 1981) and that they may be undergoing diploidization (Peng *et al.* 2013), topology F could arise from the inconsistency of gene copy loss or asymmetrical rates of evolution between duplicated genes (Fares *et al.* 2006; Rong *et al.* 2010). In regard to the primary source of discordance, the sister relationship between *P. edulis* and *P. nigra* var. *henonis* and between *P. nidularia* and *P. nigra* var. *henonis* was the two major incongruent topologies. If incomplete lineage sorting led to the discordance, three major topologies, or one major and two minor ones (Jennings & Edwards 2005) rather than two major ones, should have been observed because lineage sorting of genes happens randomly. Hybridization usually leads to two major incongruent topologies, as the hybrid tends to group with the parental lineages. In the HybTree analysis, we obtained $\gamma = 0.33$, indicating some level of hybridization in *P. nigra* var. *henonis.* Furthermore, this analysis found that estimated branch length along the hybrid species phylogeny t_1 was much less than t_2 or t_3; in such a case, this rapid divergence time for the species suggests a possibility of incomplete lineage sorting. Morphologically, *P. nigra* var. *henonis* exhibits intermediate traits between *Phyllostachys* sect. *Phyllostachys* (including *P. edulis*) and *Phyllostachys* sect. *Heterocladae* (including *P. nidularia*), making its taxonomic placement controversial (Wang *et al.* 1980). While it seems that some introgression may exist in *Phyllostachys*, that is, *P. nigra* var. *henonis* is of a putative hybrid origin, the genus could have undergone incomplete lineage sorting. We speculate that the discordance among gene trees could be mainly caused by two processes, introgression and incomplete lineage sorting. More samples from *Phyllostachys* and other genera of Arundinarieae are needed to confirm our hypothesis and to infer the parents of the putative hybrid.

The effect of IIAHs in phylogenetic analysis

The variability of nuclear introns in an individual, or IIAHs, was first mentioned by Palumbi & Baker (1994) in their study of humpback whales. More recently, studies have suggested that IIAHs could provide phylogenetic signals and help to uncover underlying introgression (Sota & Sasabe 2006; Yu *et al.* 2011). However, working with IIAHs is currently challenging, especially given the utilization of multiple genes. We dealt with IIAHs in two ways: we retained IIAHs in the analysis of individual genes or we excluded them in the single-sequence data. In the analysis of individual genes, we found that most of the IIAHs fell into a monophyletic clade, except in tree topologies C and F. As described above, in topology C, the nonmonophyletic sequences from *P. nigra* var. *henonis* grouped with the other two congeners, indicating that some introgression may occur in *P. nigra* var. *henonis*. Interestingly, the data set that excluded the IIAHs produced very similar results. As BUCKy analysis revealed, the sister relationship of *P. edulis* and *P. nigra* var. *henonis* and of *P. nidularia* and *P. nigra* var. *henonis* received CF = 49.010 and CF = 27.097, respectively. Even in the concordant data, some conflict in resolving the relationship of the three species was detected. Moreover, the HybTree analysis indicated a level of hybridization of $\gamma = 0.33$, with *P. nigra* var. *henonis* as the putative hybrid. In our work, analysis based on the IIAHs did identify a likely introgression event, while the combined data set excluding them also indicated some level of hybridization. The combined data set that excluded IIAHs from multiple genes may compensate for the phylogenetic signals that IIAHs might provide.

Inferring phylogenies with multilocus data

Phylogenetic trees are used to infer the evolutionary history of species, and various methods exist for performing phylogenetic reconstructions. Concatenation and consensus are the two popular methods when using multiple genes (Gatesy *et al.* 2002; Sanderson *et al.* 2003; Delsuc *et al.* 2005). Although some studies have suggested that the concatenation approach is more accurate than consensus approach, the former ignores the different evolutionary patterns of genes and may generate misleading phylogenies (Kolaczkowski & Thornton 2004; Gadagkar *et al.* 2005; Mossel & Vigoda 2005; Kubatko & Degnan 2007; Degnan & Rosenberg 2009). Aside from combining multiple genes into a data set, another way to infer phylogenetic trees from multiple genes is by way of the 'democratic vote' in which the species tree is estimated using the most probable gene tree topology. When there are less than four taxa, the 'democratic vote' method may be able to accurately resolve phylogenetic relationships (Degnan *et al.* 2009). Alternatively, if there are more than five taxa, the most probable gene tree topology does not always represent the species tree, for anomalous gene trees (AGTs) may exist (Degnan & Rosenberg 2006, 2009). In this study, topology A, in which *P. edulis* was sister to *P. nigra* var. *henonis*, was the most common gene tree topology. Does this tree topology reveal the 'true' relationship among these species, since there were no AGTs found in this condition? Events such as introgression, horizontal gene transfer, and recombination may be involved in species evolution, so that the evolutionary history of the species may result in more of a reticulate, rather than tree-like, network. In our analysis of hybridization, we did observed some degree of hybridization, with *P. nigra* var. *henonis* as the putative hybrid, as well as the possibility of incomplete lineage sorting, given the much shorter divergence time. The latter process was also indicated in the estimated species tree by the very short branch lengths within the *Phyllostachys* clade. Inaccurate species relationships may result from using the BEST program, as it accounts for deep coalescence but not for other issues, as well as from using data generated from a model that is not included in BEST (Gerard *et al.* 2011). Thus, although our estimated species tree produced a high posterior probability for the sister relationship of *P. edulis* and *P. nigra* var. *henonis*, the estimated species relationships produced by BEST in our analysis may not represent the true evolutionary history. Thus far, we have only been able to use the genetic history of species to infer their evolutionary histories. Hence, phylogeny is often defined as the historical trajectory of the inheritance of genes from one generation to the next, and each genetic history is a part of the evolutionary history of the species (Maddison 1997). Considering that a species tree is composed of multiple genetic trees and that it is more than just the 'winner-take-all democracy', we should be cautious when inferring a species tree from multilocus data.

Acknowledgements

We are appreciative to Steven Callen of Saint Louis University for careful revision of the manuscript and suggestions. We thank Si-Yun Chen of the Kunming Institute of Botany, Chinese Academy of Sciences for help in data analysis. We are also indebted to Professor Hong Ma and his team at Fudan University for help in using the OrthoMCL program. This study was supported by grants from the National Natural Science Foundation of China (Grants Nos.: 31100148, 30990244 and U1136603) and the Yunnan Provincial Government through an innovation team program.

References

Ané C, Larget B, Baum DA, Smith SD, Rokas A (2007) Bayesian estimation of concordance among gene trees. *Molecular Biology and Evolution*, **24**, 412–426.

Bamboo Phylogeny Group (2012) An updated tribal and subtribal classification of the bamboos (Poaceae: Bambusoideae). In: *Proceedings of the 9th World Bamboo Congress* Vol. April 2012, (eds. Gielis J, Potters G), pp. 3–27. World Bamboo Organization, Antwerp, Belgium.

Bomblies K, Doebley JF (2005) Molecular evolution of FLORICAULA/LEAFY orthologs in the Andropogoneae (Poaceae). *Molecular Biology and Evolution*, **22**, 1082–1094.

Bouchenak-Khelladi Y, Salamin N, Savolainen V *et al.* (2008) Large multigene phylogenetic trees of the grasses (Poaceae): progress towards complete tribal and generic level sampling. *Molecular Phylogenetics and Evolution*, **47**, 488–505.

Bryant D, Moulton V (2002) NeighborNet: an agglomerative method for the construction of planar phylogenetic networks. In: *Algorithms in Bioinformatics* (eds Guigó R, Gusfield D), pp. 375–391. Springer, Berlin, Heidelberg.

Bryant D, Moulton V (2004) Neighbor-net: an agglomerative method for the construction of phylogenetic networks. *Molecular Biology and Evolution*, **21**, 255–265.

Degnan JH, Rosenberg NA (2006) Discordance of species trees with their most likely gene trees. *PLoS Genetics*, **2**, e68.

Degnan JH, Rosenberg NA (2009) Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology & Evolution*, **24**, 332–340.

Degnan JH, Salter LA (2005) Gene tree distributions under the coalescent process. *Evolution*, **59**, 24–37.

Degnan JH, DeGiorgio M, Bryant D, Rosenberg NA (2009) Properties of consensus methods for inferring species trees from gene trees. *Systematic Biology*, **58**, 35–54.

Delsuc F, Brinkmann H, Philippe H (2005) Phylogenomics and the reconstruction of the tree of life. *Nature Reviews Genetics*, **6**, 361–375.

Dian WM, Jiang HX, Chen QS, Liu FY, Wu P (2003) Cloning and characterization of the granule-bound starch synthase II gene in rice: gene expression is regulated by the nitrogen level, sugar and circadian rhythm. *Planta*, **218**, 261–268.

Doyle JJ (1992) Gene trees and species trees: molecular systematics as one-character taxonomy. *Systematic Botany*, **17**, 144–163.

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*, **19**, 11–15.

Drummond A, Ashton B, Cheung M *et al.* (2009) *Geneious v4. 8*. Biomatters Ltd., Auckland, New Zealand.

Duarte JM, Wall PK, Edger PP *et al.* (2010) Identification of shared single copy nuclear genes in *Arabidopsis*, *Populus*, *Vitis* and *Oryza* and their phylogenetic utility across various taxonomic levels. *BMC Evolutionary Biology*, **10**, 61.

Ebersberger I, Strauss S, Von Haeseler A (2009) HaMStR: profile hidden markov model based search for orthologs in ESTs. *BMC Evolutionary Biology*, **9**, 157.

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, **32**, 1792–1797.

Egan AN, Schlueter J, Spooner DM (2012) Applications of next-generation sequencing in plant biology. *American Journal of Botany*, **99**, 175–185.

Fares MA, Byrne KP, Wolfe KH (2006) Rate asymmetry after genome duplication causes substantial long-branch attraction artifacts in the phylogeny of Saccharomyces species. *Molecular Biology and Evolution*, **23**, 245–253.

Gadagkar SR, Rosenberg MS, Kumar S (2005) Inferring species phylogenies from multiple genes: concatenated sequence tree versus consensus gene tree. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, **304**, 64–74.

Gatesy J, Matthee C, DeSalle R, Hayashi C (2002) Resolution of a supertree/supermatrix paradox. *Systematic Biology*, **51**, 652–664.

Gaut BS (1998) Molecular clocks and nucleotide substitution rates in higher plants. *Evolutionary Biology*, **30**, 93–120.

Gerard D, Gibbs HL, Kubatko L (2011) Estimating hybridization in the presence of coalescence using phylogenetic intraspecific sampling. *BMC Evolutionary Biology*, **11**, 291.

Griffin PC, Robin C, Hoffmann AA (2011) A next-generation sequencing method for overcoming the multiple gene copy problem in polyploid phylogenetics, applied to *Poa* grasses. *BMC Biology*, **9**, 19.

Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*, **52**, 696–704.

Guo ZH, Li DZ (2004) Phylogenetics of the *Thamnocalamus* group and its allies (Gramineae: Bambusoideae): inference from the sequences of GBSSI gene and ITS spacer. *Molecular Phylogenetics and Evolution*, **30**, 1–12.

Hisamoto Y, Kashiwagi H, Kobayashi M (2008) Use of flowering gene FLOWERING LOCUS T (FT) homologs in the phylogenetic analysis of bambusoid and early diverging grasses. *Journal of Plant Research*, **121**, 451–461.

Huang W, Marth G (2008) EagleView: a genome assembly viewer for next-generation sequencing technologies. *Genome Research*, **18**, 1538–1543.

Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, **23**, 254–267.

Jennings WB, Edwards SV (2005) Speciational history of Australian Grass Finches (Poephila) inferred from thirty gene trees. *Evolution*, **59**, 2033–2047.

Kolaczkowski B, Thornton JW (2004) Performance of maximum parsimony and likelihood phylogenetics when evolution is heterogeneous. *Nature*, **431**, 980–984.

Kubatko LS, Degnan JH (2007) Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Systematic Biology*, **56**, 17–24.

Lalitha S (2000) Primer premier 5. *Biotech Software & Internet Report: The Computer Software Journal for Scient*, **1**, 270–272.

Larget BR, Kotha SK, Dewey CN, Ané C (2010) BUCKy: gene tree/species tree reconciliation with Bayesian concordance analysis. *Bioinformatics*, **26**, 2910–2911.

Lee Y, Sultana R, Pertea G *et al.* (2002) Cross-referencing eukaryotic genomes: TIGR orthologous gene alignments (TOGA). *Genome Research*, **12**, 493–502.

Li DZ (1999) Taxonomy and biogeography of the Bambuseae (Gramineae: Bambusoideae). In: *Bamboo-Conservation, Diversity, Ecogeography, Germplasm, Resource Utilization and Taxonomy*. Proceeding of training course cum workshop (eds Rao AN, Rao VR), pp. 14–23. Kunming and Xishuangbanna, Yunnan, China. IPGRI-APO, Serdagn, Malaysia.

Li L, Stoeckert CJ, Roos DS (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Research*, **13**, 2178–2189.

Li DZ, Wang ZP, Zhu ZD *et al.* (2006) Bambuseae (Poaceae). In: *Flora of China* (eds Wu ZY, Raven PH, Hong DY), vol. 22, pp. 7–180. Science Press and Missouri Botanical Garden Press, Beijing and St. Louis.

Liu L, Pearl DK (2007) Species trees from gene trees: reconstructing Bayesian posterior distributions of a species phylogeny using estimated gene tree distributions. *Systematic Biology*, **56**, 504–514.

Ma PF (2012) *Phylogenomics of Arudinarieae (Poaceae: Bambusoideae)*. PhD, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming.

Maddison WP (1997) Gene trees in species trees. *Systematic Biology*, **46**, 523–536.

Maddison WP, Maddison D (2001) Mesquite: a modular system for evolutionary analysis.

McCormack JE, Hird SM, Zellmer AJ, Carstens BC, Brumfield RT (2011) Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution*, **66**, 526–538.

Meng C, Kubatko LS (2009) Detecting hybrid speciation in the presence of incomplete lineage sorting using gene tree incongruence: a model. *Theoretical Population Biology*, **75**, 35–45.

Mossel E, Vigoda E (2005) Phylogenetic MCMC algorithms are misleading on mixtures of trees. *Science*, **309**, 2207–2209.

Nichols R (2001) Gene trees and species trees are not the same. *Trends in Ecology & Evolution*, **16**, 358–364.

O'Brien KP, Remm M, Sonnhammer EL (2005) Inparanoid: a comprehensive database of eukaryotic orthologs. *Nucleic Acids Research*, **33**, D476–D480.

Orhrberger D (ed.) (1999) *The Bamboos of the World: Annotated Nomenclature and Literature of the Species and the Higher and Lower Taxa*. Elsebier Science, Amsterdam.

Palumbi SR, Baker CS (1994) Contrasting population structure from nuclear intron sequences and mtDNA of humpback whales. *Molecular Biology and Evolution*, **11**, 426–435.

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

Peng ZH, Lu TT, Li LB *et al.* (2010) Genome-wide characterization of the biggest grass, bamboo, based on 10,608 putative full-length cDNA sequences. *BMC Plant Biology*, **10**, 116–129.

Peng ZH, Lu Y, Li LB *et al.* (2013) The draft genome of the fast-growing non-timber forest species moso bamboo (*Phyllostachys heterocycla*). *Nature Genetics*, **45**, 456–461.

Posada D (2008) jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution*, **25**, 1253–1256.

Posada D, Buckley TR (2004) Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. *Systematic Biology*, **53**, 793–808.

Qiu YL, Li L, Wang B *et al.* (2006) The deepest divergences in land plants inferred from phylogenomic evidence. *Proceedings of the National Academy of Sciences*, **103**, 15511–15516.

Rokas A, Carroll SB (2006) Bushes in the tree of life. *PLoS Biology*, **4**, e352.

Rong J, Feltus F, Liu L, Lin L, Paterson A (2010) Gene copy number evolution during tetraploid cotton radiation. *Heredity*, **105**, 463–472.

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**, 1572–1574.

Sanderson MJ, Driskell AC, Ree RH, Eulenstein O, Langley S (2003) Obtaining maximal concatenated phylogenetic data sets from large sequence databases. *Molecular Biology and Evolution*, **20**, 1036–1042.

Sang T (2002) Utility of low-copy nuclear gene sequences in plant phylogenetics. *Critical Reviews in Biochemistry and Molecular Biology*, **37**, 121–147.

Schatz MC, Phillippy AM, Sommer DD *et al.* (2011) Hawkeye and AMOS: visualizing and assessing the quality of genome assemblies. *Briefings in Bioinformatics*, **23**, 10.

Seo HS, Kim HY, Jeong JY, Lee SY, Cho MJ, Bahk JD (1995) Molecular cloning and characterization of RGA1 encoding a G protein α subunit from rice (*Oryza sativa* L. IR-36). *Plant Molecular Biology*, **27**, 1119–1131.

Small R, Cronn R, Wendel J (2004) Use of nuclear genes for phylogeny reconstruction in plants. *Australian Systematic Botany*, **17**, 145–170.

Soderstrom TR (1981) Some evolutionary trends in the Bambusoideae (Poaceae). *Annals of Missouri Botanical Garden*, **68**, 15–17.

Soltis DE, Albert VA, Savolainen V *et al.* (2004) Genome-scale data, angiosperm relationships, and 'ending incongruence': a cautionary tale in phylogenetics. *Trends in Plant Science*, **9**, 477–483.

Sota T, Sasabe M (2006) Utility of nuclear allele networks for the analysis of closely related species in the genus *Carabus*, subgenus *Ohomopterus*. *Systematic Biology*, **55**, 329–344.

Sota T, Vogler AP (2001) Incongruence of mitochondrial and nuclear gene trees in the carabid beetles *Ohomopterus*. *Systematic Biology*, **50**, 39–59.

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

Triplett JK, Clark LG (2010) Phylogeny of the temperate bamboos (Poaceae: Bambusoideae: Bambuseae) with an emphasis on *Arundinaria* and allies. *Systematic Botany*, **35**, 102–120.

Triplett JK, Oltrogge KA, Clark LG (2010) Phylogenetic relationships and natural hybridization among the north American woody bamboos (Poaceae: Bambusoideae: *Arundinaria*). *American Journal of Botany*, **97**, 471–492.

Wall PK, Leebens-Mack J, Muller KF, Field D, Altman NS, DePamphilis CW (2008) PlantTribes: a gene and gene family resource for comparative genomics in plants. *Nucleic Acids Research*, **36**, D970–D976.

Wang CP, Yu ZH, Ye GH, Chu CD, Chao CS (1980) A taxonomic study of *Phyllostachys* in China. *Acta Phytotaxonomica Sinica*, **18**, 168–193.

Wolfe KH, Li WH, Sharp PM (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences*, **84**, 9054–9058.

Wu F, Mueller LA, Crouzillat D, Pétiard V, Tanksley SD (2006) Combining bioinformatics and phylogenetics to identify large sets of single-copy orthologous genes (COSII) for comparative, evolutionary and systematic studies: a test case in the euasterid plant clade. *Genetics*, **174**, 1407–1420.

Xu Y, Hall TC (1993) Cytosolic triosephosphate isomerase is a single gene in rice. *Plant Physiology*, **101**, 683–687.

Yang HM, Zhang YX, Yang JB, Li DZ (2013) The monophyly of *Chimonocalamus* and conflicting gene trees in Arundinarieae (Poaceae: Bambusoideae) inferred from four plastid and two nuclear markers. *Molecular Phylogenetics and Evolution*, **68**, 340–356.

Yu L, Luan PT, Jin W *et al.* (2011) Phylogenetic utility of nuclear introns in interfamilial relationships of Caniformia (order Carnivora). *Systematic Biology*, **60**, 175–187.

Zeng CX, Zhang YX, Triplett JK, Yang JB, Li DZ (2010) Large multi-locus plastid phylogeny of the tribe Arundinarieae (Poaceae: Bambusoideae) reveals ten major lineages and low rate of molecular divergence. *Molecular Phylogenetics and Evolution*, **56**, 821–839.

Zerbino DR, Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research*, **18**, 821–829.

Zerbino DR, McEwen GK, Margulies EH, Birney E (2009) Pebble and rock band: heuristic resolution of repeats and scaffolding in the velvet short-read de novo assembler. *PLoS ONE*, **4**, e8407.

Zhang YJ, Ma PF, Li DZ (2011) High-throughput sequencing of six bamboo chloroplast genomes: phylogenetic implications for temperate woody bamboos (Poaceae: Bambusoideae). *PLoS ONE*, **6**, e20596.

Zhang N, Zeng LP, Shan HY, Ma H (2012a) Highly conserved low-copy nuclear genes as effective markers for phylogenetic analyses in angiosperms. *New Phytologist*, **195**, 923–937.

Zhang XM, Zhao L, Larson-Rabin Z, Li DZ, Guo ZH (2012b) De novo sequencing and characterization of the floral transcriptome of *Dendrocalamus latiflorus* (Poaceae: Bambusoideae). *PLoS ONE*, **7**, e42082.

Zhang YX, Zeng CX, Li DZ (2012c) Complex evolution in Arundinarieae (Poaceae: Bambusoideae): incongruence between plastid and nuclear GBSSI gene phylogenies. *Molecular Phylogenetics and Evolution*, **63**, 777–797.

D.L. conceived and designed the experiments; L.Z. and X.Z. performed the experiments; L.Z. and X.Z. analyzed data; L.Z. supported the data analysis. L.Z., X.Z., Y.Z., C.Z., P.M., Z.G. and D.L. wrote the paper.

Data Accessibility

GenBank Accession nos. (see Table S5, Supporting information) and sequence matrixes: doi:10.5061/dryad.kp5cn. Raw sequence data was submitted to Sequence Read Archive (SRA) and the Accession nos. are listed in Table 2.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Fig. S1 The mode of estimated hybrid species tree.

Table S1 Putative orthologous single copy genes identified as candidates for phylogenetic markers.

Table S2 A list of primers and protocols used for PCR amplification.

Table S3 Information of the best-fit model for each gene.

Table S4 Information of the best-fit model for each gene of the single-sequence-data set.

Table S5 Percentage of variable (V) and parsimony informative characters (I) of genes we used and their GenBank Accession nos..

Table S6 The topologies of gene trees generated in MrBaysen for each gene with IIAHs.

Table S7 Posterior distribution for all tree topologies of each gene in BUCKy analysis.

Appendix S1 Identification of putative single orthologues share in four taxa based on the local OrthoMCL.