MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2013) **13**, 938–945 doi: 10.1111/1755-0998.12136

Identification of SNP markers for inferring phylogeny in temperate bamboos (Poaceae: Bambusoideae) using RAD sequencing

X. Q. WANG,*†[1] L. ZHAO,*[1] D. A. R. EATON,‡ D. Z. LI§ and Z. H. GUO*

*Plant Germplasm and Genomics Center, Germplasm Bank of Wild Species, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, China, †College of Life Sciences, University of Chinese Academy of Sciences, Beijing 100049, China, ‡Committee on Evolutionary Biology, University of Chicago, Chicago IL 60637, USA, §Key Laboratory of Biodiversity and Biogeography, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, China

Abstract

Phylogenetic relationships among temperate species of bamboo are difficult to resolve, owing to both the challenge of detecting sufficiently variable markers and their polyploid history. Here, we use restriction site–associated DNA sequencing to identify candidate loci with fixed allelic differences segregating between and within two temperate species of bamboos: *Arundinaria faberi* and *Yushania brevipaniculata*. Approximately 27 million paired-end sequencing reads were generated across four samples. From pooled data, we assembled 67 685 and 70 668 *de novo* contigs from partial overlap among paired-end reads, with an average length of 240 and 241 bp for the two species, respectively, which were used to investigate functional classification of RAD tags in a BLASTX search. Analysed separately by population, we recovered 29 443 putatively orthologous RAD tags shared across the four sampled populations, containing 28 023 sequence variants, of which *c.* 13 000 are segregating between species, and *c.* 3000 segregating between populations within each species. Analyses based on these RAD tags yielded robust phylogenetic inferences, even with data set constructed from surprisingly few loci. This study illustrates the potential for reduced-representation genome data to resolve difficult phylogenetic relationships in temperate bamboos.

Keywords: phylogeny, polyploid, restriction site–associated DNA, SNP markers, temperate bamboos

Received 27 February 2013; revision received 8 June 2013; accepted 10 June 2013

Introduction

The temperate bamboos are a morphologically diverse grass lineage with *c.* 32 genera and 600 species distributed primarily in the north temperate zone or at high elevations in tropical regions of both the northern and southern hemispheres (Ohrnberger 1999; Das *et al.* 2008; Triplett & Clark 2010). Molecular phylogenetic studies indicate that temperate woody bamboos (Arundinarieae) are sister to a clade composing both tropical woody bamboos (Bambuseae) and herbaceous bamboos (Olyreae; Bouchenak-Khelladi *et al.* 2008; Sungkaew *et al.* 2009) and represent one-third of all bamboo species (Ohrnberger 1999). There are *c.* 430 species distributed in East Asia, of which 180 are endemic to south-western China (Suzuki 1978; Ohrnberger 1999; Li *et al.* 2006). The temperate woody bamboos are of great ecological and economic importance as they provide food and raw materials for construction and manufacturing (Li *et al.* 2006), and also serve as the main food source for numerous vertebrates and invertebrates, including specialists such as the giant panda (*Ailuropoda melanoleuca*) and the red panda (*Ailurus fulgens*; Yi 1985; McNeely 1999).

Previous phylogenetic studies strongly support the monophyly of temperate bamboos (Kelchner & Clark 1997; Zhang 2000; Bouchenak-Khelladi *et al.* 2008; Peng *et al.* 2008; Sungkaew *et al.* 2009), but taxonomic delineation and phylogenetic relationships at lower levels within the clade lack resolution (Triplett & Clark 2010; Zeng *et al.* 2010). The taxonomy of the temperate bamboos has traditionally relied on morphological features, but these morphological classifications are poorly supported by molecular data (Peng *et al.* 2008; Sungkaew *et al.* 2009; Triplett & Clark 2010; Zeng *et al.* 2010).

In rapidly diversified clades, generating variable genetic markers containing single-nucleotide polymorphisms (SNPs) is often difficult. However, recent technological

Correspondence: De-zhu Li, Fax: +86-871-65223503; E-mail: dzl@mail.kib.ac.cn Zhen-hua Guo, Fax: +86-871-65217791; E-mail: guozhenhua@mail.kib.ac.cn

[1]These authors contributed equally to this work.

advances in next-generation sequencing (NGS) now offer the possibility to generate large-scale sequence data from nonmodel organisms at a reasonable cost (Ekblom & Galindo 2011). For example, complete chloroplast genomes were recently applied to resolve phylogenetic relationships among six woody bamboo species (Zhang *et al.* 2011). Nuclear data offer even greater potential to resolve relationships among close relatives. However, the tetraploid nature of the temperate bamboos ($2n = 4\times = 48$) presents additional challenges for SNP marker development and genotyping (Ghorai & Sharma 1980; Kellogg & Watson 1993; Clark *et al.* 1995; Gielis *et al.* 1997; Peng *et al.* 2013).

Due to the large and complex genomes of bamboos, genome complexity reduction sequencing technology may offer a more efficient alternative to acquiring contigs for SNP discovery (Slate *et al.* 2009). One promising approach to reduced-representation genomics is restriction site–associated DNA (RAD) sequencing, which sequences short DNA fragments flanking restriction enzyme cut sites, allowing orthologous sequences to be targeted across multiple samples to identify and score thousands of genetic markers (Miller *et al.* 2007; Baird *et al.* 2008; Emerson *et al.* 2010; Hohenlohe *et al.* 2010). To date, the method has been successfully applied to SNP discovery in threespine stickleback (Baird *et al.* 2008; Hohenlohe *et al.* 2010), *Wyeomyia* mosquitoes (Emerson *et al.* 2010), rainbow and westslope cutthroat trout (Hohenlohe *et al.* 2011), eggplants (Barchi *et al.* 2011), *Cynara cardunculus* (Scaglione *et al.* 2012), *Sisymbrium austriacum* (Vandepitte *et al.* 2013) and sunflowers (Andrew *et al.* 2013). Thus, even without a reference genome, RAD sequencing can deliver huge numbers of SNPs for analysis, and in polyploids, deep sequencing coverage may provide the ability to distinguish among homoeologs.

Arundinaria Michaux and *Yushania* P. C. Keng are temperate genera with semelauctant inflorescences and three stamens (Li *et al.* 2006). *Arundinaria* has leptomorph rhizomes, while *Yushania* has pachymorph rhizomes (Li *et al.* 2006). *Arundinaria faberi* Rendle, a small subalpine bamboo usually found at elevations of 2300–3500 m, is distributed in southwest China (northeast Guizhou, southwest Sichuan and northern of Yunnan; Li *et al.* 2006). *Yushania brevipaniculata* (Handel-Mazzetti) T. P. Yi, a shrubby spreading bamboo occurring between 1800 and 3800 m, is commonly distributed in western Sichuan (Li *et al.* 2006). The two species share similar habitats and are often distributed sympatrically. Both species are important food for the giant panda (*A. melanoleuca*; Yi 1985). In previous studies, the phylogenetic relationships among sampled populations of *A. faberi* and *Y. brevipaniculata* could not be resolved using a small number of plastid and nuclear genes (Zeng *et al.* 2010; Zhang *et al.* 2012b). The major goal of this study is to identify SNP markers from two different populations of *A. faberi* and *Y. brevipaniculata*, and to evaluate the performance of RAD sequencing for phylogeny reconstruction in temperate bamboos.

Methods

Creation and sequencing of the RAD libraries

A total of 24 individuals representing four natural populations (two *A. faberi* populations and two *Y. brevipaniculata* populations) were sampled from Sichuan province, China (Table 1). Total genomic DNA was extracted from silica gel-dried leaf material using a modified CTAB procedure (Doyle & Doyle 1987). Then, DNA was pooled from six individuals in each population to form our four population samples, as described in Emerson *et al.* (2010). Each population sample was digested with *Eco*RI (5′-GAATTC-3′). RAD libraries were prepared and sequenced according to Miller *et al.* (2007) and Baird *et al.* (2008). Four paired-end libraries were constructed. All libraries were sequenced in a single lane of an Illumina HiSeq 2000 with read length of 91 bp.

De novo assembly

RAD sequences were preprocessed through three quality filtering steps. First, sequences containing sequencing errors in the cut site were removed. Second, any read containing sequencing errors in the population-specific barcode was discarded. Third, reads which had more than 50% of base calls with a low quality score ($Q \leq 5$) were eliminated. Reads were then demultiplexed, by sorting according to their barcodes into the four populations. Paired-end reads from both populations were pooled in each species, and *de novo* contigs assembled

Table 1 Location and sampling information for *Arundinaria faberi* and *Yushania brevipaniculata*

Taxon	Pop	Location	Latitude	Longitude	Altidude (m)
A. faberi	AEM	E'mei, Sichuan, China	N29°31′52″	E103°19′28″	2826
	AWL	Wolong, Sichuan, China	N31°02′36″	E103°09′50″	2696
Y. brevipaniculata	YEM	E'mei, Sichuan, China	N29°31′59″	E103°19′57″	2600
	YWL	Wolong, Sichuan, China	N31°01′34″	E103°11′13″	2476

using Velvet, version 1.0.02 (Zerbino & Birney 2008), run with a hash length of 27 bp, set the minimum contig size as 200 bp and with other parameters set to default.

Sequence annotation

A BLASTX search was performed against the NCBI non-redundant (Nr) protein database using BLAST, version 2.2.25 (Altschul *et al.* 1990) with an *E*-value cut-off of 1e-3 for the contigs *de novo* assembled from each species. Based on the results of the Nr protein database annotation, Blast2GO (Conesa *et al.* 2005) was employed to obtain the functional classification of the contigs by applying gene ontology (GO) terms (http://www.geneontology.org; Ashburner *et al.* 2000), which maps contigs to function according to the three principal GO categories: molecular function, cellular component and biological processes (Harris *et al.* 2004). The plot of GO classification was obtained by WEGO (http://wego.genomics.org.cn/cgi-bin/wego/index.pl; Ye *et al.* 2006). Then the GO annotations of the contigs were mapped to the plant-specific GO slim ontology (http://www.geneontology.org/GO.slims.shtml).

SNP discovery and phylogenetic analysis

To discover SNP markers for phylogenetic studies, we utilized the PYRAD software (Eaton & Ree 2013), applied only to the single end of the paired-end sequences. Reads were clustered using a 90% similarity threshold to create 'stacks', which were then aligned and used to call consensus sequences based on the counts of alleles at each site. These consensus sequences were then clustered using the same threshold across samples to cluster putative orthologs.

We only kept reads that had a minimum depth of coverage >9, which made the base calls more accurate. Also, putative loci with unusually high coverage (>1000) were excluded because these loci probably derive from paralogous regions with similar sequences or from highly repetitive regions (Wagner *et al.* 2012). The algorithm used to make consensus base calls is based on a diploid bi-allelic model, where the probability that a site is heterozygous (AB) vs. homozygous (AA or BB) is calculated as a binomial probability depending on an error rate (e) (Li *et al.* 2008). Because our samples are tetraploid, and represent pooled populations, this model is violated. Effectively, to the extent, stacks represent all four tetraploid alleles; a SNP segregating between homoeologs (AABB) will be called the same as a SNP segregating within copies of a homoeolog (AAAB), in that both will be marked with an ambiguity code in the consensus sequence, representing a polymorphism. We created data sets with consensus sequences called at two different error rates, 0.001 and 0.01, but because they yielded few differences, we present only results from the former. To evaluate how much data are necessary and sufficient to obtain phylogenetic resolution, we prepared seven data sets with different numbers of loci.

To evaluate the performance of PYRAD, we built and genotyped loci *de novo* from the single end of the paired-end sequences using STACKS (Catchen *et al.* 2011). In the pipeline, 10 reads were required as minimum depth of coverage to create a stack. We set the maximum distance between stacks within a locus as one. The minimum number of populations a locus must be present to be processed was set to four.

The data sets were analysed using maximum-parsimony, maximum-likelihood and Bayesian methods. Maximum-parsimony analyses were conducted using PAUP, version 4.0b10 (Swofford 2002) with 1000 replicates of tree bisection and reconnection branch swapping. Node support was estimated with 1000 bootstrap replicates (MPBS). Maximum-likelihood analyses were implemented in RAxML, version 7.2.8 (Stamatakis 2006) using the general time-reversible (GTR) model of nucleotide substitution with the gamma distributed rate heterogeneity. Nonparametric bootstrapping was implemented in the fast bootstrap algorithm of RAxML with 1000 replicates (MLBS). Bayesian inference was performed using MrBayes, version 3.1.2 (Ronquist & Huelsenbeck 2003). The best-fitting models were determined according to the akaike information criterion (Posada & Buckley 2004). The GTR+I+Γ model and GTR+Γ model were appropriate for the PYRAD data sets and the STACKS data sets, respectively. The Markov chain Monte Carlo algorithm was run for 2 000 000 generations with trees sampled every 100 generations for each data set.

Results

RAD tag generation and de novo assembly

A total of *c.* 108 million 82–85 bp (AEM: 82 bp, AWL: 83 bp, YEM: 85 bp, YWL: 84 bp) paired-end reads were obtained after barcode trimming, cleaning and quality checking. The Q20 (sequencing base calls with an error rate of <1%) of each sample was above 97%, meaning data quality was very high. The mean GC content of the sequence in four populations was *c.* 44%, similar to that found in *Dendrocalamus latiflorus* (49.48%; Zhang *et al.* 2012a), but lower than the GC content value in cDNA for *Phyllostachys edulis* (54.0%; Peng *et al.* 2010), although the AT-rich enzyme *Eco*RI (GAATTC) may bias our data set towards lower GC content. The mean sequencing depth of RAD tag was 27 in each sample. The data summaries are presented in Table 2. Assembly of reads using Velvet resulted in 67 685 and 70 668 contigs with mean sizes of

Table 2 Summary statistics of the RAD tags sequencing

	Arundinaria faberi		*Yushania brevipaniculata*	
	AEM	AWL	YEM	YWL
Number of reads (million)	27.15	27.08	27.09	27.23
Total length of reads (million bp)	2335	2342	2370	2369
GC Rate%	44.07	44.1	44.21	44.16
Depth	27.4	27.8	27.8	27.8
Number of contigs	67 685		70 668	
Average contig length (bp)	240		241	
N50 (bp)	246		249	
Contig length range (bp, min–max)	200–774		200–604	

240 and 241 bp for *A. faberi* and *Y. brevipaniculata*, respectively (Table 2).

Sequence annotation and GO categorization

In all, 21% of assembled contigs (14 353/67 685 in *A. faberi*, and 14 697/70 668 in *Y. brevipaniculata*) of each species were significantly matched to known genes in the public Nr databases (Fig. 1). The overall functional annotation is depicted in Table S1 (Supporting information). The top-hit lists of protein sequences for BLAST results were as follows: *Oryza sativa*, *Brachypodium distachyon*, *Sorghum bicolor*, *Hordeum vulgare*, *Zea mays* and *P. edulis* (Fig. S1, Supporting information). The *Y. brevipaniculata* yielded highly similar results.

Woody bamboos are characterized by having woody culms, complex branching systems of both culms and rhizomes, and infrequent flowering. We identified sets of contigs that have putative functions in shoot branching including those belonging to the GRAS family, F-box family and NAC family (Table S1, Supporting information). Also, contigs related to floral development such as zinc-finger protein family, WD repeat-containing protein family, basic helix–loop–helix family (bHLH), Myb family and basic leucine zipper (Bzip) transcription factors were identified in our data set. As is well known, bamboo is a major resource of nonwood fibre because the lignin content of bamboo is comparable to that of woody plants and higher than most herbaceous plants. Here, contigs belonging to cinnamyl alcohol dehydrogenase catalyses gene family, which is thought to be correlated with lignin biosynthesis, were identified.

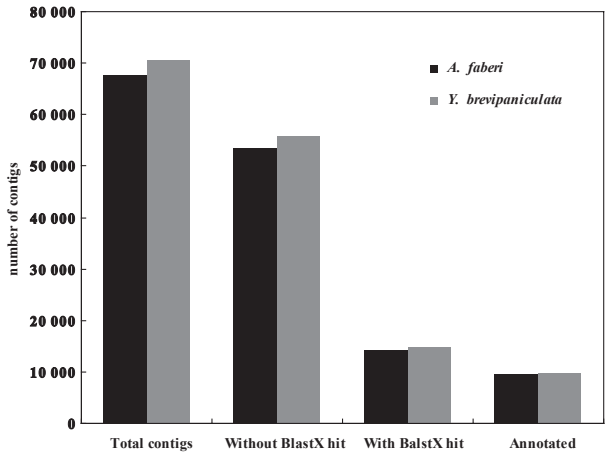

Fig. 1 The numbers of the annotated contigs.

There were 9525 and 9742 contigs assigned to one or more GO terms for *A. faberi* and *Y. brevipaniculata*, respectively (Figs 1 and 2). For biological processes, genes involved in cellular processes (GO: 0009987) or metabolic processes (GO: 0008152) are highly represented. For molecular functions, binding (GO: 0005488) is the most represented GO term. Regarding cellular components, the most highly represented categories are cell (GO: 0005623), cell part (GO: 0044464) and organelle (GO: 0043226).

A further functional classification of contigs was performed using a set of plant-specific GO slims. The most highly represented groups were 'plastid', 'mitochondrion' and 'cytoplasmic membrane-bounded vesicle', followed closely by 'ATP binding', 'DNA binding', 'membrane', 'RNA binding', 'DNA recombination' and 'RNA-directed DNA polymerase activity' (Table S2, Supporting information). Contigs involved in shoot development, root development, flower development, pollination, response to stress and lignin biosynthetic process were also represented.

SNP identification and phylogenetic analyses

After filtering and clustering, we recovered *c.* 29 443 putative orthologs shared across the four populations, for a total length of 2 129 079 bp. This includes 28 023 variable sites, of which 13 650 (Fig. 3) are fixed between species; the remainder being polymorphic. The SNPs occur at a rate of 0.95 SNP/tag. For *A. faberi*, we identified 3055 fixed differences between population E'mei and Wolong, excluding polymorphic sites, while in *Y. brevipaniculata* 3095 SNPs were fixed between the two populations (Fig. 3).

Seven data sets, ranging from 6 (0.05%) loci to 29 443 (100%) loci, were prepared for phylogenetic analyses. The results of the MP, ML and Bayesian analyses were largely congruent. Table 3 summarizes these results.

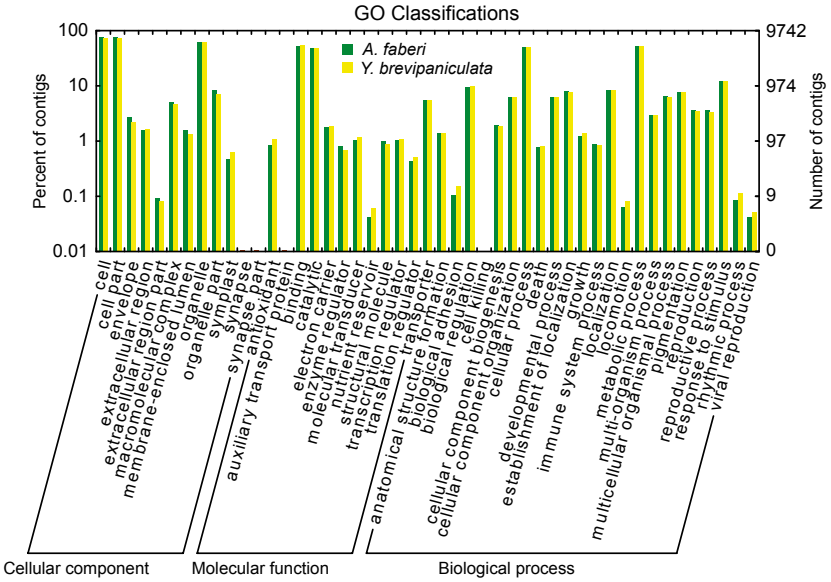

Fig. 2 Gene ontology classification of assembled contigs.

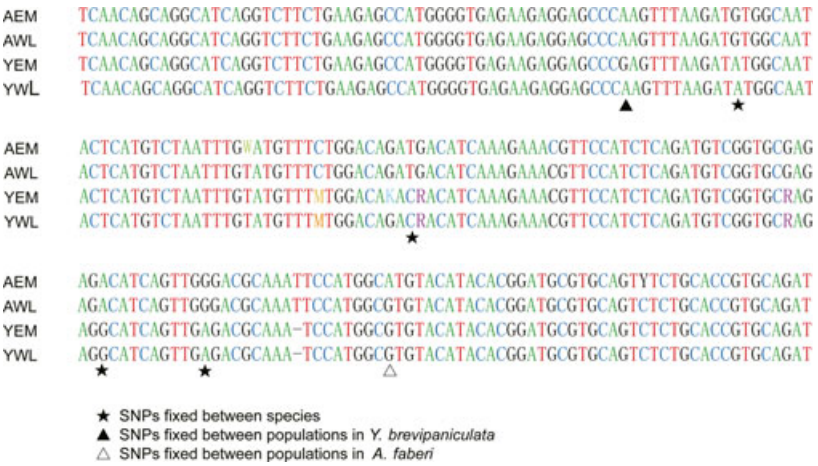

Fig. 3 Schematic representation of categories of SNPs fixed between species and populations.

Phylogenetic analysis using the largest data set revealed robust support for the relationship between the two species (100% MPBS, 100% MLBS, 1.00 PP), as well as the monophyly of each species. For the data set in which six loci were included, the phylogenetic analysis shows clear species-specific clade, but the node received decreased support (95% MPBS, 100% MLBS, 0.89 PP).

Maximum-parsimony (MP) analysis of the 29 443-locus combined data set recovered a single most-parsimonious tree with a tree length of 27 328, a consistency index (CI) of 0.9989 and a retention index (RI) of 0.9979. The ML analysis found a single optimal tree identical to the MP consensus tree. The topology from the Bayesian analysis was consistent with the MP and ML analyses (Fig. 4).

Analyses using STACKS revealed similar results. STACKS produces 11 348 loci for which all four populations have sequence data. A phylogeny based on 6638 parsimony informative characters was highly supported, recovering the two species as distinct (100% MPBS, 100% MLBS, 1.00 PP). These results show that pyRAD performed as well as STACKS, a commonly used software for RAD sequencing, and therefore, pyRAD is appropriate for inferring phylogeny from RAD sequence data.

Discussion

In this study, high-throughput sequencing of RAD tags enabled us to resolve fine-scale genetic divergence among *A. faberi* and *Y. brevipaniculata*. In previous studies, Zeng *et al.* (2010) and Zhang *et al.* (2012b) employed plastid and nuclear genes to infer phylogenetic relationships in Arundinarieae, and including *A. faberi* (*Bashania fangiana*) and *Y. brevipaniculata*. The results showed that the

Table 3 Statistics for different data sets used in phylogenetic analyses

Number of loci	Total characters	PIC	MP tree length	CI	RI	MPBS	MLBS	PP
29 443	2 129 079	13 962	27 328	0.9989	0.9979	100	100	1.00
2944	212 952	1348	2708	0.9996	0.9993	100	100	1.00
589	42 629	247	535	1	1	100	100	1.00
294	21 299	119	259	1	1	100	100	1.00
59	4271	25	52	1	1	100	100	1.00
29	2094	17	37	1	1	100	100	1.00
6	436	3	8	1	1	95	100	0.89

PIC, parsimony informative characters; MP, maximum parsimony; CI, consistency index; RI, retention index; MPBS, maximum-parsimony bootstrap analyses; MLBS, maximum-likelihood bootstrap analyses; PP, posterior probabilities.

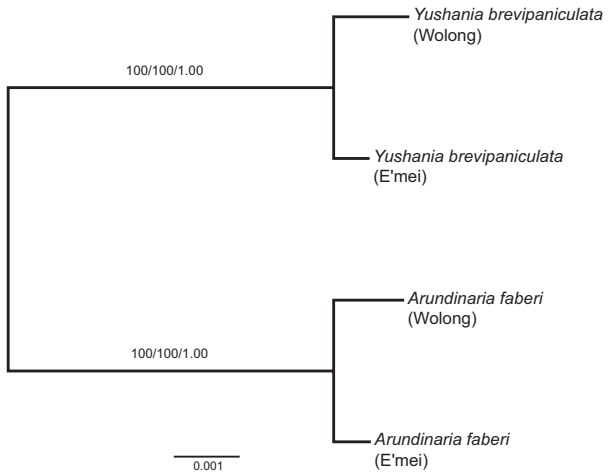

Fig. 4 Phylogram of the 50% majority-rule consensus tree from Bayesian inference based on 29 443 loci. Support values are shown for nodes as maximum-parsimony bootstrap/maximum-likelihood bootstrap/Bayesian inference posterior probability. Branch lengths were calculated through Bayesian analysis, and scale bar denotes substitutions per site.

two species were in the *Phyllostachys* clade, but the phylogenetic relationship between the two species was not resolved. Temperate woody bamboos are notorious for being a taxonomically difficult group with a low rate of molecular evolution (Gaut *et al*. 1997). Previous studies suggested that temperate woody bamboos may have diverged rapidly early in their evolutionary history (Triplett & Clark 2010; Zeng *et al*. 2010; Zhang *et al*. 2012b). Recently, phylogenetic analyses based on whole chloroplast genomes have resolved major relationships within the subfamily Bambusoideae, but the diversification among three clades (*Arundinaria*, *Shibataea* and *Phyllostachys* clades) of temperate woody bamboos remained unresolved even with complete chloroplast genomes sequences (Zhang *et al*. 2011). In previous chloroplast phylogenomic studies, it was suggested that even the entire chloroplast genome may be insufficient to fully resolve the rapidly radiating lineages (Parks *et al*. 2009; Moore *et al*. 2010). In the current study, by successively reducing the number of loci, we presented that 29 loci with a total length of 2094 bp would be sufficient to resolve the phylogeny. The relationships within the two species were better resolved using the larger data set, suggesting that insufficient parsimony information characters were the major cause for poor resolution in temperate bamboos. Here, we demonstrate the power and efficiency of RAD sequencing to resolve differences among closely related species in temperate bamboos.

Rubin *et al*. (2012) first applied RAD sequence data to reconstruct phylogenies, showing that clade age was an important factor in how many RAD tags were recovered across species, and that it was probably to perform best in young clades. The temperate bamboos are a relatively young group in the Bambusoideae, originating in the Miocene, *c*. 23 Ma (Bouchenak-Khelladi *et al*. 2009, 2010; Hodkinson *et al*. 2010). Similarly, Wagner *et al*. (2012) used NGS data from RAD markers to infer phylogenetic relationships among 16 species of cichlid fishes from Lake Victoria, recovering a strongly supported tree even in this rapidly radiated clade.

Little is known about the genetic diversity or population structure of bamboo germplasm (Barkley *et al*. 2005). With their low rate of molecular evolution, the predominant obstacle for population genetic studies of Bambusoideae is identifying sufficiently variable molecular markers. In our study, 3055 (0.14%) and 3095 (0.15%) SNPs were identified between population E'mei and Wolong for *A. faberi* and *Y. brevipaniculata*, respectively, showing that RAD sequencing has great potential for resolving relationships among the Bambusoideae.

Previous studies have identified and characterized several genes related to floral development (Tian *et al*. 2005, 2006; Lin *et al*. 2009), rhizome bud development (Wang *et al*. 2010), leaf senescence (Chen *et al*. 2011), plant tolerance to stress (Liu *et al*. 2012) and lignin biosynthesis (Zhou *et al*. 2012) in bamboos. In our study,

a large number of contigs were annotated by searching the orthologous genes of highest sequence similarity in public database (NCBI and GO). Some are relevant to important traits or developmental process. The results can be a valuable resource and are worthy of further investigation.

The current study represents an initial attempt at resolving complex evolutionary relationships in the temperate bamboos, using the largest collection of genome-wide SNPs yet used in bamboo phylogenetics. Extending this approach to a broader taxonomic sampling will help to elucidate the evolutionary history of temperate bamboos.

Acknowledgements

We thank Xiaoyan Wang, Xuyao Zhao, Xianzhi Zhang, Lina Zhang for joining in field work to collect samples, and also thank Dr. Chunxia Zeng, Yuxiao Zhang, Pengfei Ma and Xuemei Zhang for various help and support. This project was supported by the Knowledge Innovation Project of the Chinese Academy of Sciences (KSCX2-YW-N-067); the National Natural Science Foundation of China (30990244); NSFC-Yunnan province joint foundation (U1136603); Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry and the Young Academic and Technical Leader Raising Foundation of Yunnan Province (No.2008PY065, awarded to Zhen-Hua Guo); and the Yunnan Provincial Government through an innovation team programme.

Conflict of interest

The authors have no conflict of interest to declare and note that the sponsors of the issue had no role in the study design, data collection and analysis, decision to publish or preparation of the manuscript.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *Journal of Molecular Biology*, **215**, 403–410.

Andrew RL, Kane NC, Baute GJ, Grassa CJ, Rieseberg LH (2013) Recent nonhybrid origin of sunflower ecotypes in a novel habitat. *Molecular Ecology*, **22**, 799–813.

Ashburner M, Ball CA, Blake JA *et al.* (2000) Gene Ontology: tool for the unification of biology. *Nature Genetics*, **25**, 25–29.

Baird NA, Etter PD, Atwood TS *et al.* (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE*, **3**, e3376.

Barchi L, Lanteri S, Portis E *et al.* (2011) Identification of SNP and SSR markers in eggplant using RAD tag sequencing. *BMC Genomics*, **12**, 304.

Barkley NA, Newman ML, Wang ML, Hotchkiss MW, Pederson GA (2005) Assessment of the genetic diversity and phylogenetic relationships of a temperate bamboo collection by using transferred EST-SSR markers. *Genome*, **48**, 731–737.

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.

Bouchenak-Khelladi Y, Verboom GA, Hodkinson TR *et al.* (2009) The origins and diversification of C_4 grasses and savanna-adapted ungulates. *Global Change Biology*, **15**, 2397–2417.

Bouchenak-Khelladi Y, Verboom GA, Savolainen V, Hodkinson TR (2010) Biogeography of the grasses (Poaceae): a phylogenetic approach to reveal evolutionary history in geographical space and geological time. *Botanical Journal of the Linnean Society*, **162**, 543–557.

Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011) Stacks: building and genotyping loci de novo from short-read sequences. *G3-Genes Genomes Genetics*, **1**, 171–182.

Chen YX, Qiu K, Kuai BK, Ding YL (2011) Identification of an NAP-like transcription factor BeNAC1 regulating leaf senescence in bamboo (*Bambusa emeiensis* 'Viridiflavus'). *Physiologia Plantarum*, **142**, 361–371.

Clark LG, Zhang WP, Wendel JF (1995) A phylogeny of the grass family (Poaceae) based on *ndhf*- sequence data. *Systematic Botany*, **20**, 436–460.

Conesa A, Gotz S, Garcia-Gomez JM *et al.* (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, **21**, 3674–3676.

Das M, Bhattacharya S, Singh P, Filgueiras TS, Pal A (2008) Bamboo taxonomy and diversity in the era of molecular markers. In: *Botanical Research: Incorporating Advances in Plant Pathology* (eds Kader JC, Delseny M), vol. 47, pp. 225–268. Elsevier Academic Press Inc, San Diego.

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

Eaton DAR, Ree RH (2013) Inferring phylogeny and introgression using RADseq data: an example from flowering plants (*Pedicularis*: Orobanchaceae). *Systematic Biology*, doi:10.1093/sysbio/syt032.

Ekblom R, Galindo J (2011) Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity*, **107**, 1–15.

Emerson KJ, Merz CR, Catchen JM *et al.* (2010) Resolving postglacial phylogeography using high-throughput sequencing. *Proceedings of the National Academy of Sciences, USA*, **107**, 16196–16200.

Gaut BS, Clark LG, Wendel JF, Muse SV (1997) Comparisons of the molecular evolutionary process at *rbcL* and *ndhF* in the grass family (Poaceae). *Molecular Biology and Evolution*, **14**, 769–777.

Ghorai A, Sharma A (1980) Cyto-taxonomy of Indian Bambuseae. 2. Dendrocalameae and Melocanneae. *Acta Botanica Indica*, **8**, 134–138.

Gielis J, Valente P, Bridts C, Verbelen JP (1997) *Estimation of DNA Content of Bamboos Using Flow Cytometry and Confocal Laser Scanning Microscopy*. Academic Press Ltd, London.

Harris MA, Clark J, Ireland A *et al.* (2004) The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Research*, **32**, D258–D261.

Hodkinson TR, Chonghaile GN, Sungkaew S *et al.* (2010) Phylogenetic analyses of plastid and nuclear DNA sequences indicate a rapid late Miocene radiation of the temperate bamboo tribe Arundinarieae (Poaceae, Bambusoideae). *Plant Ecology & Diversity*, **3**, 109–120.

Hohenlohe PA, Bassham S, Etter PD *et al.* (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics*, **6**, e1000862.

Hohenlohe PA, Amish SJ, Catchen JM, Allendorf FW, Luikart G (2011) Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. *Molecular Ecology Resources*, **11**, 117–122.

Kelchner SA, Clark LG (1997) Molecular evolution and phylogenetic utility of the chloroplast *rpl16* intron in *Chusquea* and the Bambusoideae (Poaceae). *Molecular Phylogenetics and Evolution*, **8**, 385–397.

Kellogg EA, Watson L (1993) Phylogenetic studies of a large data set.1. Bambusoideae, and Ropogonodae, and Pooideae (Gramineae). *Botanical Review*, **59**, 273–343.

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. 57–96. Science Press and Missouri Botanical Garden Press, Beijing and St. Louis.

Li H, Ruan J, Durbin R (2008) Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Research*, **18**, 1851–1858.

Lin EP, Peng HZ, Jin QY *et al.* (2009) Identification and characterization of two Bamboo (*Phyllostachys praecox*) AP1/SQUA-like MADS-box genes during floral transition. *Planta*, **231**, 109–120.

Liu L, Cao XL, Bai R *et al.* (2012) Isolation and characterization of the cold-induced *Phyllostachys edulis* AP2/ERF family transcription factor, peDREB1. *Plant Molecular Biology Reporter*, **30**, 679–689.

McNeely JA (1999) Biodiversity and bamboo genetic resources in Asia: in situ, community-based and ex situ approaches to conservation. *Chinese Biodiversity*, **7**, 38–51.

Miller MR, Dunham JP, Amores A, Cresko WA, Johnson EA (2007) Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Research*, **17**, 240–248.

Moore MJ, Soltis PS, Bell CD, Burleigh JG, Soltis DE (2010) Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. *Proceedings of the National Academy of Sciences, USA*, **107**, 4623–4628.

Ohrnberger D (1999) *The Bamboos of the World: Annotated Nomenclature and Literature of the Species and the Higher and Lower Taxa*. Elsevier Science Publishers B.V., Amsterdam.

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

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.

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, 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.

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

Rubin BER, Ree RH, Moreau CS (2012) Inferring phylogenies from RAD sequence data. *PLoS ONE*, **7**, e33394.

Scaglione D, Acquadro A, Portis E *et al.* (2012) RAD tag sequencing as a source of SNP markers in *Cynara cardunculus* L. *BMC Genomics*, **13**, 3.

Slate J, Gratten J, Beraldi D *et al.* (2009) Gene mapping in the wild with SNPs: guidelines and future directions. *Genetica*, **136**, 97–107.

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

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.

Suzuki S (1978) *Index to Japanese Bambusaceae*. Gakken Co. Ltd, Tokyo.

Swofford DL (2002) *Paup * 4.0: Phylogenetic Analysis Using Parsimony (* and Other Methods). Version 4.0.b10*. Sinauer, Sunderland, Massachusetts.

Tian B, Chen YY, Yan YX, Li DZ (2005) Isolation and ectopic expression of a bamboo MADS-box gene. *Chinese Science Bulletin*, **50**, 217–224.

Tian B, Chen YY, Li DZ, Yan YX (2006) Cloning and characterization of a bamboo LEAFY HULL STERILE1 homologous gene. *DNA Sequence*, **17**, 143–151.

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.

Vandepitte K, Honnay O, Mergeay J *et al.* (2013) SNP discovery using Paired-End RAD-tag sequencing on pooled genomic DNA of *Sisymbrium austriacum* (Brassicaceae). *Molecular Ecology Resources*, **13**, 269–275.

Wagner CE, Keller I, Wittwer S *et al.* (2012) Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Molecular Ecology*, **22**, 787–798.

Wang KH, Peng HZ, Lin EP *et al.* (2010) Identification of genes related to the development of bamboo rhizome bud. *Journal of Experimental Botany*, **61**, 551–561.

Ye J, Fang L, Zheng HK *et al.* (2006) WEGO: a web tool for plotting GO annotations. *Nucleic Acids Research*, **34**, W293–W297.

Yi TP (1985) Classification and distribution of the food bamboos of the giant panda (I). *Journal of Bamboo Research*, **4**, 11–27.

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.

Zhang WP (2000) Phylogeny of the grass family (Poaceae) from *rpl16* intron sequence data. *Molecular Phylogenetics and Evolution*, **15**, 135–146.

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

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

Zhou MJ, Hu SL, Cao Y *et al.* (2012) Cloning and bioinformation analysis of C3H gene in *Neosinocalamus affinis*. *Bulletin of Botanical Research*, **32**, 38–46.

Z.H.G. and D.Z.L. designed the experiments. X.Q.W. performed the experiments. X.Q.W. and L.Z. analyzed data. D.A.R.E. provided technical support for analysis. X.Q.W. and L.Z. wrote the paper. Z.H.G., D.Z.L. and D.A.R.E. helped with writing and editing.

Data Accessibility

Contig assemblies in fasta format (contigs_arundinaria.fa & contigs_yushania.fa), SNPs type and shared loci between four populations (shared loci), matrix used to construct phylogenetic trees (4 bamboss.phy), and the raw data are available on Dryad, doi:10.2061/dryad.lmj31.

Supporting Information

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

Fig. S1 Characteristics of sequence homology of contigs BLASTED against GenBank databases for *A. faberi*.

Table S1 Top BLAST hits from public databases. Lists of the top results from BLASTING *A. faberi* and *Y. brevipaniculata* contigs against public databases (*E*-value cut-off of 10^{-3}).

Table S2 This GO terms for contigs. Lists of the GO terms for contigs of *A. faberi* and *Y. brevipaniculata*.