Molecular Phylogenetics and Evolution 62 (2012) 573–578

Contents lists available at SciVerse ScienceDirect

Molecular Phylogenetics and Evolution

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

Short Communication

The phylogeny of the BEP clade in grasses revisited: Evidence from the whole-genome sequences of chloroplasts

Zhi-Qiang Wu, Song Ge *

State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China
Graduate University, Chinese Academy of Sciences, Beijing 100039, China

ARTICLE INFO

Article history:
Received 26 August 2011
Revised 22 October 2011
Accepted 24 October 2011
Available online 10 November 2011

Keywords:
Phylogeny
BEP clade
Grasses
Plastomes

ABSTRACT

Despite the considerable efforts to reconstruct the phylogeny of grasses, the relationships among the subfamilies Bambusoideae, Pooideae and Ehrhartoideae in the BEP clade remain unresolved. Here we completely sequenced three chloroplast genomes of representative species from Bambusoideae and Ehrhartoideae and obtained 19 additional chloroplast genome sequences of other grasses from GenBank. Using sequences of 76 chloroplast protein-coding genes from the 22 grass species, we fully resolved the phylogeny of the BEP clade. Our results strongly supported the (B,P)E hypothesis, i.e., Bambusoideae and Pooideae are more closely related than Ehrhartoideae. This result was not biased by systematic or sampling errors and was impervious to phylogenetic methods or model specification. The divergence time estimate suggests that the initial diversification of the BEP clade into three subfamilies happened within a short time period (~4 MY). The presence of these short internal branches may explain the inability of previous studies to achieve a confident resolution of the BEP clade. The combination of the sequences of the entire chloroplast genomes provided sufficient phylogenetic information to resolve the BEP phylogeny fully. These results provide a valuable evolutionary framework for comparative and functional genomic studies using the grass family as a model system.

1. Introduction

The grass family (Poaceae) is one of the most diverse angiosperm families and consists of approximately 700 genera and more than 10,000 species (Clayton and Renvoize, 1986; GPWG, 2001). The grasses not only include many economically important crops, such as rice (*Oryza sativa*), corn (*Zea mays*), wheat (*Triticum aestivum*) and sorghum (*Sorghum bicolor*), but also dominate various natural and agricultural landscapes of the world (Clayton and Renvoize, 1986). With the accomplished and ongoing genome sequencing of many cereal crops, grass species have become a model system for functional and comparative genomic research.

In the past decades, the phylogenetic relationships of Poaceae have been the focus of extensive studies. An evolutionary framework has been thoroughly established at the family level (Clark et al., 1995; GPWG, 2001; Duvall et al., 2007; Bouchenak-Khelladi et al., 2008; Vicentini et al., 2008; Kellogg, 2009). To date, two major clades within the family have been identified, with one comprising the three subfamilies Bambusoideae, Ehrhartoideae, and Pooideae (BEP clade) and the other consisting of the seven subfamilies Panicoideae, Arundinoideae, Chloridoideae, Centothecoideae, Micraioideae, Aristidoideae, and Danthonioideae (PACCMAD clade) (GPWG, 2001; Duvall et al., 2007; Bouchenak-Khelladi et al., 2008; Kellogg, 2009). However, the phylogenetic relationships within these clades have not been fully resolved (GPWG, 2001; Bouchenak-Khelladi et al., 2008; Kellogg, 2009). Using the plastid *ndhF* sequences, Clark et al. (1995) first recovered a tree with two major groups (the PACC and BOP clades) and found that bambusoid (i.e., Bambusoideae) and oryzoid (i.e., Ehrhartoideae) clustered together and were sister to pooid (i.e., Pooideae). The (B,E)P relationship, with Bambusoideae and Ehrhartoideae being more closely related than Pooideae, was confirmed by subsequent studies (GPWG, 2001; Vicentini et al., 2008; Sungkaew et al., 2009). Nevertheless, many other studies have revealed either a sister relationship of Bambusoideae and Pooideae, (B,P)E (Zhang, 2000; Bouchenak-Khelladi et al., 2008; Leseberg and Duvall, 2009; Peng et al., 2010) or a sister relationship of Ehrhartoideae and Pooideae, (E,P)B (Mason-Gamer et al., 1998; Mathews et al., 2000). The problem of the relationship among the BEP lineages is similar to the well-known "trichotomy problem" involving chimpanzee, gorilla, and human (Satta et al., 2000) in which all three possible relationships have been recovered in various studies.

Apart from the inconsistent phylogenies within the BEP clade, it is noteworthy that the different topologies were generated if different sampling strategies (gene and taxon) were employed

* Corresponding author at: State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Xiangshan, Beijing 100093, China. Fax: +86 10 62590843.
E-mail address: gesong@ibcas.ac.cn (S. Ge).

1055-7903/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2011.10.019

and if different methods of analysis were used. For instance, based on sequences of 43 orthologous loci selected from 10,608 putative full-length cDNA sequences of bamboos, Peng et al. (2010) obtained different phylogenetic relationships within the BEP clade using different phylogenetic methods. Ehrhartoideae was the basal lineage on the maximum likelihood and Bayesian trees, i.e., (B,P)E. In contrast, Pooideae was the basal lineage on the Neighbor-Joining tree, i.e., (B,E)P. Therefore, despite considerable efforts to reconstruct the phylogeny of the grasses, the BEP relationship remains unclear.

The cpDNA sequences are increasingly used for resolving the deep phylogeny of plants because of their low rates of nucleotide substitutions and structural changes (Soltis et al., 2004; Jansen et al., 2007; Moore et al., 2010). Concatenating sequences from many genes may overcome the problem of multiple substitutions that results in a loss of phylogenetic information between chloroplast lineages and can reduce sampling errors due to substitutional noise (Delsuc et al., 2005). Here, we sequenced the complete chloroplast genomes of three grass species and obtained 19 additional chloroplast genomes of grasses from GenBank. Using sequences of 76 plastid protein-coding genes (70,700 bp in total) that are derived from the complete plastid genomes of the 22 grass species, we were able to resolve the phylogenetic relationships of the BEP clade fully. Based on the divergence time estimates, we found that two rapid speciation events within the BEP clade occurred in a small time interval (roughly 4 million years). These events might be responsible for the inconsistent phylogenetic reconstructions that have been reported in previous studies.

2. Materials and methods

2.1. Complete plastid genome sequence of three grass species

Using the Fosmid library and PCR amplification methods, we completely sequenced three plastomes, including two species (*Leersia tisserantii* and *Rhynchoryza subulata*) from the subfamily Ehrhartoideae and one species (*Phyllostachys propinqua*) from Bambusoideae. Fresh leaves of the three species were collected from plants grown in the greenhouse or garden of the Institute of Botany of the Chinese Academy of Sciences in Beijing. The seeds of *L. tisserantii* and *R. subulata* were requested from the International Rice Research Institute (IRRI). The species' accession numbers were 101384 and 100913, respectively. Total cellular DNA was extracted using the cetyltrimethyl ammonium bromide (CTAB) method and purified with phenol extraction to remove proteins. For two Ehrhartoideae species, fosmid libraries were constructed following the method in Wang et al. (2009). Five overlapping clones containing chloroplast DNA fragments were identified through Southern hybridization with the following chloroplast gene probes: *psbC*, *rpoB*, *rps2*, *atpB*, *rbcL*, *psbB*, *rpoA*, *rrn16*, *rrn23*, *ndhB*, *ndhF*, *ccsA*. The positive clones were sheared again using a shotgun library with 1.5–3 k fragments. The shotgun libraries were sequenced and assembled at the Beijing Genomics Institute (BGI) in Beijing, China. For sequencing *P. propinqua*, a full set of primers have been designed based on two previously sequenced bamboo plastomes (Wu et al., 2009). PCR amplification and purification of the products were performed, as described in Tang et al. (2010). The purified products were directly sequenced on an ABI 3730 automated sequencer (Applied Biosystems, Foster City, CA, USA). The sequences were assembled with the ContigExpress program from the Vector NTI Suite 6.0 (Informax Inc., North Bethesda, MD).

We used DOGMA (Dual Organellar GenoMe Annotator, Wyman et al., 2004) for genome annotation. A draft annotation was subsequently inspected using the DOGMA tools for accurate assessment of the start and stop codons and the exon–intron boundaries of genes and adjusted manually. Both tRNAs and rRNAs were identified by BLASTN searches against the same database of chloroplast genomes.

2.2. Phylogenetic reconstructions

A total of 22 chloroplast genomes from Poaceae were used in our phylogenetic analyses, including three genomes sequenced in this study and 19 retrieved from NCBI. A detailed information on the 22 chloroplast genomes is provided in Table 1. Because 30 tRNA and four rRNA genes were notably conserved and provided almost no information, only 76 protein-coding genes from the chloroplast genomes were used in our phylogenetic analyses. The nucleotide sequences were translated into amino acids and aligned with ClustalX 1.81 (Thompson et al., 1997). The intron-containing genes were aligned using MUSCLE (Edgar, 2004) and were manually adjusted. The incongruence length difference (ILD) test, as implemented in PAUP 4.0b10 (Swofford, 2002), was conducted for testing the suitability of combining different fragments in the phylogenetic analyses.

The phylogenetic trees were reconstructed using maximum parsimony (MP) implemented with PAUP Version 4.0b10 (Swofford, 2002), maximum likelihood (ML) with PHYML Version 2.4.5 (Guindon and Gascuel, 2003), and Bayesian inference (BI) with MrBayes Version 3.1.2 (Ronquist and Huelsenbeck, 2003). Modeltest 3.7 (Posada and Buckley, 2004) was used to select the best nucleotide substitution models for each data set under the Akaike Information Criterion (AIC) (Shepherd et al., 2002). MP heuristic searches used tree bisection reconnection (TBR) branch swapping executed for 1000 replicates. ML analysis was performed with 1000 bootstrap replicates. For the estimation of Bayesian posterior probabilities (PP), two independent runs were simultaneously performed with four Metropolis-coupled Markov chains, with each run consisting of one cold chain and three incrementally heated chains and with all runs randomly started in the parameter space. Each run was conducted with 1×10^6 generations and sampled every 100 generations. When the log-likelihood scores were observed to stabilize, a consensus tree was calculated after discarding the first 25% of the trees as burn-in.

2.3. Analysis of systematic bias and congruence tests

It is widely recognized that systematic errors, such as compositional signals, rate signals, and heterotachous signals, might be reinforced as the sequence length or the number of sampled genes increases (Delsuc et al., 2005; Zou et al., 2008). Thus, we investigated the potential impact of systematic bias on our phylogenetic reconstruction. First, we assessed the compositional bias among species by performing tests of the stationarity of base composition in TREEPUZZLE 5.2 (Schmidt et al., 2002). To detect the possible bias from substitution saturation in the sequences, we plotted transitions and transversions against the pairwise sequence divergence, as implemented in DAMBE (Xia and Xie, 2001), with an asymptotic relationship indicating the presence of saturation. Finally, we used the HYPHY (Pond and Muse, 2005) for relative rate tests that examined whether the genes evolved at different rates for pairs of lineages using *Anomochloa marantoidea* as an outgroup. Sequences were also analyzed under the RY-coding strategy (A and G = R, C and T = Y), which discarded fast-evolving transitions and thus made phylogenetic reconstructions less susceptible to uneven evolutionary rates among lineages (Delsuc et al., 2005).

2.4. Divergence time estimation

Because rate constancy was rejected for the combined chloroplast data, Bayesian approaches with an uncorrelated relaxed clock

Table 1
Comparison of major features of 22 Poaceae chloroplast genomes.

Subfamily	Species	Total size		LSC region		IR region		SSC region		GenBank accession
		Length (bp)	GC (%)	Length (bp)	GC (%)	Length (bp)	GC (%)	Length (bp)	GC (%)	
Ehrhartoideae	*Oryza sativa* ssp. *indica*	134,496	39.00	80,553	37.11	20,798	44.35	12,347	33.32	NC_008155
	Oryza nivara	134,494	39.01	80,544	37.12	20,802	44.35	12,346	33.33	NC_005973
	Oryza sativa ssp. *japonica*	134,551	39.00	80,604	37.11	20,802	44.35	12,343	33.37	AY522330
	Oryza meridionalis	134,551	39.01	80,606	37.11	20,802	44.35	12,343	33.36	GU592208
	Oryza australiensis	134,549	38.98	80,614	37.07	20,796	44.36	12,343	33.25	GU592209
	Leersia tisserantii	136,550	38.88	81,865	37.01	21,329	44.05	12,027	33.23	JN415112[a]
	Rhynchoryza subulata	136,303	39.00	82,029	37.14	20,840	44.36	12,594	33.40	JN415114[a]
	Potamophila parviflora	134,551	39.07	80,604	37.19	20,800	44.32	12,347	33.58	GU592210
	Microlaena stipoides	134,551	39.22	80,613	37.28	20,793	44.18	12,343	33.77	GU592211
Bambusoideae	*Bambusa oldhamii*	139,350	38.92	82,889	37.01	21,790	44.22	12,881	33.27	NC_012927
	Dendrocalamus latiflorus	139,374	38.92	82,963	37.02	21,774	44.23	12,863	33.18	NC_013088
	Phyllostachys propinqua	139,704	38.88	83,227	36.96	21,800	44.23	12,877	33.14	JN415113[a]
Pooideae	*Triticum aestivum*	134,545	38.31	80,348	36.32	20,703	44.06	12,791	32.26	NC_002762
	Hordeum vulgare	136,462	38.32	80,597	36.31	21,582	43.83	12,701	32.33	NC_008590[b]
	Agrostis stolonifera	136,584	38.45	80,546	36.3	21,649	44.06	12,740	33.01	NC_008591[b]
	Lolium perenne	135,282	38.25	79,964	36.11	21,445	43.90	12,428	32.49	NC_009950
	Brachypodium distachyon	135,199	38.57	79,445	36.51	21,542	44.06	12,668	32.79	NC_011032
Panicoideae	*Sorghum bicolor*	140,754	38.49	82,685	36.34	22,783	43.91	12,503	32.96	NC_008602[b]
	Saccharum officinarum	141,182	38.44	83,048	36.29	22,795	43.90	12,544	32.86	NC_006084
	Coix lacryma-jobi	140,745	38.46	82,792	36.3	22,715	43.93	12,523	32.90	NC_013273
	Zea mays	140,384	38.46	82,352	36.28	22,748	43.95	12,536	32.85	NC_001666
Anomochlooideae	*Anomochloa marantoidea*	138,412	38.66	82,264	36.71	21,993	43.66	12,162	33.72	NC_014062
Average			38.74		36.75		44.12		33.11	

[a] Sequenced in this study.
[b] The annotations for *Agrostis stolonifera* (NC_008591), *Hordeum vulgare* (NC_008590), and *Sorghum bicolor* (NC_008602) incorrectly included two extra tRNAs (*trnM-cau* between *trnG-ucc* and *trnT-ggu*; *trnfM-cau* between *trnR-ucu* and *rps14*), and they have, therefore, been excluded from the calculation of the number of genes.

model were employed to estimate the divergence times of the four subfamilies in the Poaceae with a focus on the BEP clades. The BEAST program (Drummond and Rambaut, 2007) was used to perform these analyses. The model of sequence evolution was determined using Modeltest 3.7 (Posada and Buckley, 2004). Markov chains were run for 10,000,000 generations, while every 1000 generations were sampled with at least a 10% burn-in phase. We adopted two previously published divergence times for calibration. One divergence time was estimated by Vicentini et al. (2008), assuming that maize and rice diverged 52 million years ago (MYA), and the other time was 35 MYA for the divergence of the tribes Oryzeae and Ehrharteae (Tang et al., 2010). These two time points were used for calibrating the age of the stem nodes of the four subfamilies in the Poaceae. The chronograms shown were calculated using the median clade credibility tree plus 95% confidence intervals.

3. Results

We have sequenced the chloroplast genomes of two species in Ehrhartoideae (*L. tisserantii* and *R. subulata*) and one species in Bambusoideae (*P. propinqua*). The genome sizes were 136,550 bp, 136,303 bp and 139,704 bp, respectively (Table 1). The genomes include a pair of inverted repeats (IRs) with a length of 21,329 bp (*L. tisserantii*), 20,840 bp (*R. subulata*) and 21,800 bp (*P. propinqua*), separated by a small single-copy region of 12,027 bp (*L. tisserantii*), 12,594 bp (*R. subulata*) and 12,877 bp (*P. propinqua*), and a large single-copy region of 81,865 bp (*L. tisserantii*), 82,029 bp (*R. subulata*) and 83,227 bp (*P. propinqua*), respectively. The organization of the 22 chloroplast genomes largely reflects the highly conserved gene content (128 genes) and gene order of the grass chloroplasts (Michelangeli et al., 2003). The total length of the complete genomes ranges from 134,494 bp to 141,182 bp, with length variation less than 2 kb being observed within each subfamily. The overall average GC content of the grass species is 38.7% (±0.003), with the highest GC content being in the IR region (44.1%) and the lowest in the SSC region (33.1%) (Table 1).

We concatenated the sequences of the 76 protein-coding genes into a data set of 70,700 bp in length with 79.5% of exon sequences. Of these sequences, 11,772 bp (16.7%) and 5726 bp (8.1%) are variable and phylogenetically informative sites, respectively. Because an ILD test found no significant incongruence among 76 genes ($P = 0.38$), the concatenated data set was used for phylogenetic analysis. Three phylogenetic methods (ML, MP and BI) yielded an identical and fully resolved tree, with almost all internal nodes supporting 100% bootstrap values or 1.00 posterior probabilities (Figs. 1 and S1: Supplementary data). In the resulting phylogenies, Bambusoideae and Pooideae are more closely related than Ehrhartoideae. We further used different data sets to reconstruct the phylogeny. Regardless of whether the concatenated exons, 3rd codons or introns were used, trees with a topology identical to that in Fig. 1 were obtained (Table 2).

To investigate the potential impact of systematic bias on our phylogenetic reconstruction, we first tested the homogeneity of base composition across species and found no significant difference in base content ($P > 0.05$). Next, we plotted transitions and transversions against pairwise sequence divergence to investigate the possibility of a bias owing to substitution saturation. No saturation was found for all six types of partitions (Fig. S2: Supplementary data). Finally, we analyzed rate constancy among species using the relative rate test and found that the null hypothesis of rate constancy was rejected in many contrasts ($P < 0.01$). It is noteworthy that three Bambusoideae species evolved at a slower rate than species of other subfamilies (Fig. 1). To reduce the potential impact of rate heterogeneity on tree reconstruction, we re-coded the data set under the RY-coding strategy. The resulting phylogenetic tree was topologically identical to that shown in Fig. 1 (Table 2). Because the rate discrepancy, at least among two subfamilies, was found for 26 genes, we further reconstructed the phylogeny by excluding the 26 genes. The resulting topology, which was based on the remaining genes, was the same as that of Fig. 1 with

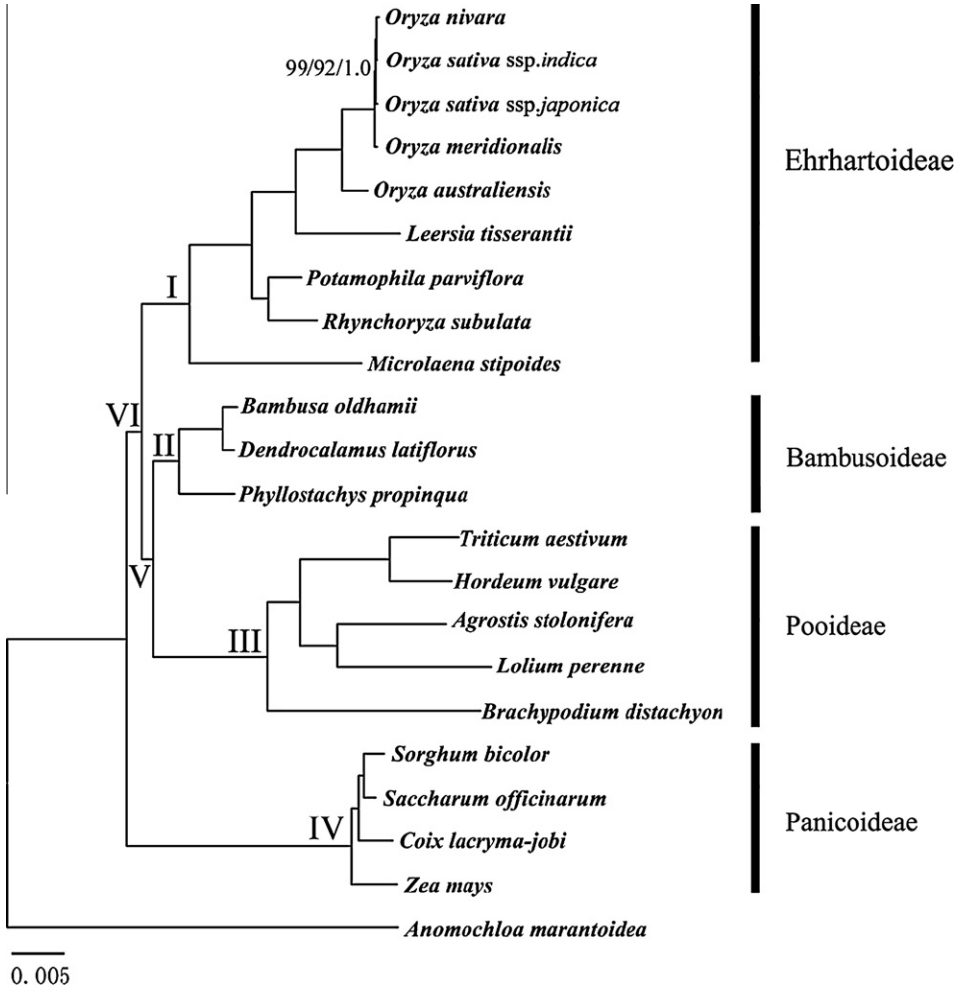

Fig. 1. ML tree inferred from the concatenated sequences of 76 protein-coding genes under the GTR + I + G model. The ML analysis generated a single tree with a ML value of $-\ln L = 199643.78$. Numbers near branches are bootstrap values of ML and MP and the posterior probability for BI, respectively. Branches without numbers indicate 100% bootstrap support and 1.0 posterior probability. Branch length is proportional to the number of substitutions, as measured by the scale bar. The same topology was obtained from MP and BI (Fig. S1, Supplementary data).

Table 2
Bootstrap support from 1000 replicates for the six internal branches of phylogenetic trees based on the concatenated sequences of different positions and using different methods.

Method	Bootstrap support (%)					
	I	II	III	IV	V	VI
Exon sites only (MP/ML)	100/100	100/100	100/100	100/100	99/99	100/99
3rd Codon sites (MP/ML)	100/100	100/100	100/100	100/100	99/98	100/95
Intron sites only (MP/ML)	100/100	100/100	100/100	100/100	93/94	100/97
RY-coding strategy for all codon sites (ML)	100	100	100	100	61	–
Exclude 26 rate non-constancy genes (MP/ML)	100/100	100/100	100/100	100/100	84/75	99/52

slightly lower statistical support for clades V and VI (Table 2). This result suggests that rate heterogeneity did not impact the phylogenetic inference.

We used a relaxed-clock approach to estimate the divergence time within the BEP clade. As shown in Fig. 2, the divergence between the subfamily Panicoideae and the BEP clade is estimated to be 51.9 (49.9–53.8) MYA. Within the BEP clade, Ehrhartoideae diverged from the clade consisting of Bambusoideae and Pooideae (node I in Fig. 2) at approximately 46.98 (40.80–51.60) MYA, whereas the Bambusoideae and Pooideae clades split (node II in Fig. 2) at ~42.80 (36.61–48.80) MYA. The divergence times for the three BEP subfamilies are estimated to be at the minimum of 35.05 (33.0–37.0) MYA (E), 32.25 (24.7–38.9) MYA (P), and 22.54 (9.1–35.6) MYA (B), respectively. Note that two short internal branches, corresponding to nodes I and II in Fig. 2, occur on the tree, implying that two rapid divergence events occurred within approximately 4 million years in or near the Middle Eocene.

4. Discussion

Despite extensive investigations of the phylogeny of the grass family, the phylogenetic relationships of the BEP clade remain unresolved. All three possible relationships among the three subfamilies have been suggested by a variety of previous studies. The inconsistent phylogenies might be attributed to a relatively small number of phylogenetic markers used in these studies.

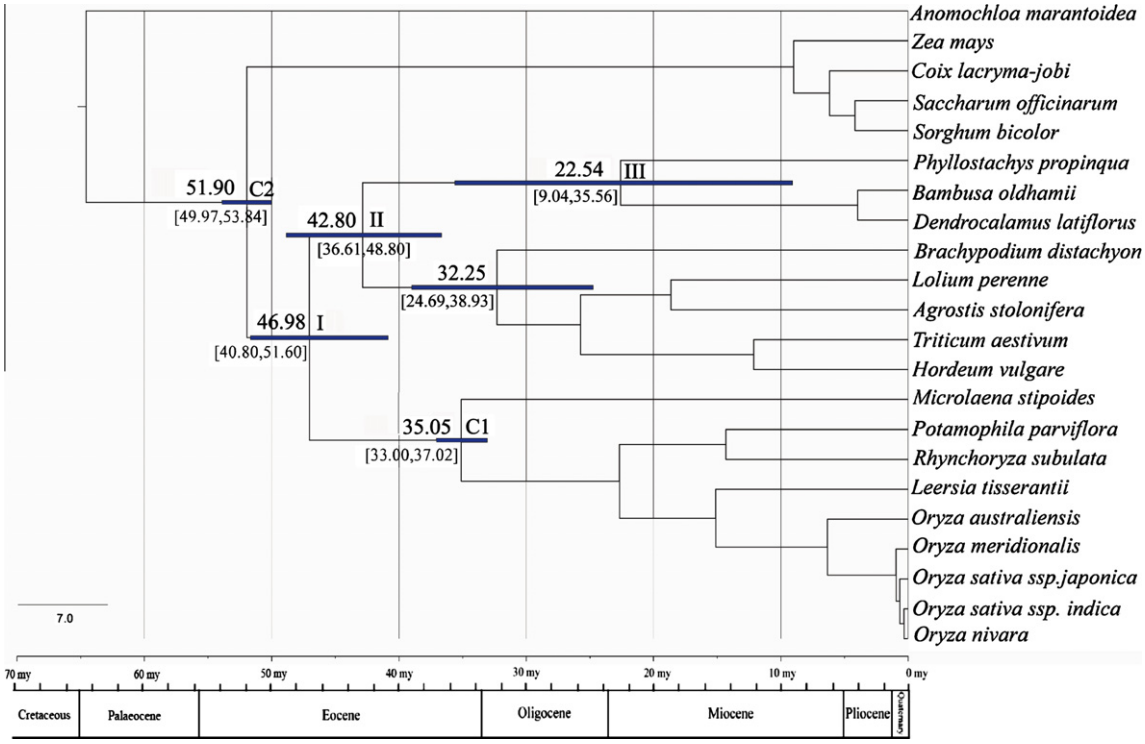

Fig. 2. Chronogram obtained for Poaceae under a Bayesian relaxed-clock approach, as implemented in the BEAST program (Drummond and Rambaut, 2007) and based on the combined 76 protein-coding genes. Gray boxes indicate 95% confidence intervals on nodal ages.

Because different genes/fragments might have evolved at different rates or possess different histories (Delsuc et al., 2005; Zou et al., 2008), a limited sample could be inadequate for recovering the true phylogeny. For instance, the (B,E)P and (B,P)E topologies were observed on the plastid *ndhF* (Clark et al., 1995) and *rpl16* intron (Zhang, 2000) trees, respectively, whereas a (E,P)B topology occurred on the nuclear *waxy* (Mason-Gamer et al., 1998) and phytochrome B (Mathews et al., 2000) trees, respectively. Two recent studies have used sequences of 61 chloroplast protein-coding genes (Leseberg and Duvall, 2009) and sequences of 43 nuclear genes (Peng et al., 2010) to reconstruct the grass phylogeny. In these two cases, a single species was sampled to represent Bambusoideae and/or Ehrhartoideae. It is possible that insufficient taxon sampling prevented a confident recovery of the evolutionary history of these subfamilies. (Soltis et al., 2004; Delsuc et al., 2005).

In the present study, we completely sequenced three Bambusoideae and Ehrhartoideae species and were able to include multiple species from each subfamily in our analyses. Using sequences of 76 chloroplast protein-coding genes, we fully resolved the phylogeny of the BEP clades and strongly supported the (B,P)E relationship. This result was not biased by systematic or sampling errors and was unaffected by different phylogenetic methods or model specification.

It is noteworthy that we identified two short internal branches on the tree (Fig. 1), implying that the BEP clade has experienced rapid divergence or radiation. Our estimation of divergence times indicates that the initial diversification of the BEP clade appears to have occurred at ~46.98 (40.80–51.60) MYA, thereby giving rise to three lineages (subfamilies) within a short time interval (~4 MY). Such short internal branches connecting the three subfamilies have been observed and were interpreted as rapid divergence or radiation (Clark et al., 1995; Peng et al., 2010). These events might be the main reason for the incongruent phylogenies of the BEP clade found by the previous studies. Although rapid divergence or radiation has confounded the resolution of relationships for many groups (Delsuc et al., 2005; Zou et al. 2008; Peng et al., 2010), both theoretical and empirical studies indicate that large-sequence datasets have the power to recover true and robust phylogenies (Satta et al., 2000; Delsuc et al., 2005; Jansen et al., 2007; Zou et al., 2008; Moore et al., 2010). Clearly, our combined sequences of the chloroplast genomes provide abundant information and greater power for fully resolving the BEP phylogeny. These findings can provide an important evolutionary framework for comparative and functional genomic studies using grasses as a model system.

Acknowledgments

We thank Dr. Wen-Li Chen for providing samples and taxonomic identification. We are grateful to Fu-Min Zhang, Hai-Fei Zhou, and Hai-Qin Sun for technical assistance and discussions and to Tao Sang for reading and improving the manuscript. We are also grateful to the International Rice Research Institute (Los Banos, Philippines) for providing seed samples. This work was supported by the National Natural Science Foundation of China (30990240).

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2011.10.019.

References

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

Clark, L.G., Zhang, W., Wendel, J.F., 1995. A phylogeny of the grass family (Poaceae) based on *ndhF* sequence data. Syst. Bot. 20, 436–460.

Clayton, W.D., Renvoize, S.A., 1986. Genera Graminum, Grasses of the World. Kew Bull Additional Series XIII.

Delsuc, F., Brinkmann, H., Philippe, H., 2005. Phylogenomics and the reconstruction of the tree of life. Nat. Rev. Genet. 6, 361–375.
Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7, 214.
Duvall, M.R., Davis, J.I., Clark, L.G., Noll, J.D., Goldman, D.H., Sánchez-Ken, J.G., 2007. Phylogeny of the grasses (Poaceae) revisited. Aliso 23, 237–247.
Edgar, R.C., 2004. MUSCLE, multiple sequence alignment with high accuracy and high throughput. Nucl. Acids Res. 32, 1792–1797.
GPWG (Grass Phylogeny Working Group), 2001. Phylogeny and subfamilial classification of the grasses (Poaceae). Ann. Mol. Bot. Gard. 88, 373–457.
Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52, 696–704.
Jansen, R.K., Cai, Z., Raubeson, L.A., Daniell, H., Depamphilis, C.W., Leebens-Mack, J., Muller, K.F., Guisinger-Bellian, M., Haberle, R.C., Hansen, A.K., Chumley, T.W., Lee, S.B., Peery, R., McNeal, J.R., Kuehl, J.V., Boore, J.L., 2007. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. Proc. Natl. Acad. Sci. USA 104, 19369–19374.
Kellogg, E., 2009. The evolutionary history of ehrhartoideae, Oryzeae, and *Oryza*. Rice 2, 1–14.
Leseberg, C.H., Duvall, M.R., 2009. The complete chloroplast genome of *Coix lachyrma-jobi* and a comparative molecular evolutionary analysis of plastomes in cereals. J. Mol. Evol. 69, 311–318.
Mason-Gamer, R.J., Weil, C.F., Kellogg, E.A., 1998. Granule-Bound starch synthase: structure, function, and phylogenetic utility. Mol. Biol. Evol. 15, 1658–1673.
Mathews, S., Tsai, R.C., Kellogg, E.A., 2000. Phylogenetic structure in the grass family (Poaceae), evidence from the nuclear gene Phytochrome B. Am. J. Bot. 87, 96–107.
Michelangeli, F.A., Davis, J.J., Stevenson, D.W., 2003. Phylogenetic relationships among Poaceae and related families as inferred from morphology, inversions in the plastid genome, and sequence data from the mitochondrial and plastid genomes. Am. J. Bot. 90, 93–106.
Moore, M.J., Soltis, P.S., Bell, C.D., Burleigh, J.G., Soltis, D.E., 2010. Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. Proc. Natl. Acad. Sci. USA 107, 4623–4628.
Peng, Z.H., Lu, T.T., Li, L.B., Liu, X.H., Gao, Z.M., Hu, T., Yang, X.W., Feng, Q., Guan, J.P., Weng, Q.J., Fan, D.L., Zhu, C.R., Lu, Y., Han, B., Jiang, Z.H., 2010. Genome-wide characterization of the biggest grass, bamboo, based on 10,608 putative full-length cDNA sequences. BMC Plant Biol. 10, 116.
Pond, S.L.K., Muse, S.V., 2005. HyPhy: hypothesis testing using phylogenies. Bioinformatics 21, 676–679.
Posada, D., Buckley, T.R., 2004. Model selection and model averaging in phylogenetics, advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. Syst. Biol. 53, 793–808.
Ronquist, F., Huelsenbeck, J.P., 2003. MrBAYES 3, Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.
Satta, Y., Klein, J., Takahata, N., 2000. DNA archives and our nearest relative: the trichotomy problem revisited. Mol. Phylogenet. Evol. 14, 259–275.
Schmidt, H.A., Strimmer, K., Vingron, M., von Haeseler, A., 2002. TREE-PUZZLE, maximum likelihood phylogenetic analysis using quartets and parallel computing. Oxford Univ. Press 18, 502–504.
Shepherd, P.R., Salvesen, G., Toker, A., Ladbury, J., Mayo, K.H., Baici, A., Board, P., Booth, P., Bourne, Y., Doyle, D., 2002. The molecular mechanism for the spectral shifts between vertebrate ultraviolet-and violet-sensitive cone visual pigments. Biochem. J. 367, 129–135.
Soltis, D.E., Albert, V.A., Savolainen, V., Hilu, K., Qiu, Y.L., Chase, M.W., Farris, J.S., Stefanović, S., Rice, D.W., Palmer, J.D., Soltis, P.S., 2004. Genome-scale data, angiosperm relationships, and "ending incongruence", a cautionary tale in phylogenetics. Trends Plant Sci. 9, 477–483.
Sungkaew, S., Stapleton, C.M.A., Salamin, N., Hodkinson, T.R., 2009. Non-monophyly of the woody bamboos (Bambusinae; Poaceae), a multi-gene region phylogenetic analysis of *Bambusoideae s. s.* J. Plant Res. 122, 95–108.
Swofford, D.L., 2002. PAUP*, Phylogenetic Analysis Using Parsimony (* and Other Methods). Sinauer Associates, Sunderland, Massachusetts.
Tang, L., Zou, X.H., Achoundong, G., Potgieter, C., Second, G., Zhang, D.Y., Ge, S., 2010. Phylogeny and biogeography of the rice tribe (Oryzeae): evidence from combined analysis of 20 chloroplast fragments. Mol. Phylogenet. Evol. 54, 266–277.
Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The ClustalX windows interface, flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucl. Acids Res. 24, 4876–4882.
Vicentini, A., Barber, J.C., Aliscioni, A.A., Giussani, L.M., Kellogg, E.A., 2008. The age of the grasses and clusters of origins of C4 photosynthesis. Global Change Biol. 14, 2693–2977.
Wang, B., Xue, J.Y., Li, L., Liu, L., Qiu, Y.L., 2009. The complete mitochondrial genome sequence of the liverwort *Pleurozia purpurea* reveals extremely conservative mitochondrial genome evolution in liverworts. Curr. Genet. 55, 601–609.
Wu, F.H., Kan, D.P., Lee, S.B., Daniell, H., Lee, Y.W., Lin, C.C., Lin, N.S., Lin, C.S., 2009. Complete nucleotide sequence of *Dendrocalamus latiflorus* and *Bambusa oldhamii* chloroplast genomes. Tree Physiol. 29, 847–856.
Wyman, S.K., Jansen, R.K., Boore, J.L., 2004. Automatic annotation of organellar genomes with DOGMA. Bioinformatics 20, 3252–3255.
Xia, X., Xie, Z., 2001. DAMBE, software package for data analysis in molecular biology and evolution. J. Hered. 92, 371–373.
Zhang, W., 2000. Phylogeny of the grass family (Poaceae) from rpl16 intron sequence data. Mol. Phylogenet. Evol. 15, 135–146.
Zou, X.H., Zhang, F.M., Zhang, J.G., Zang, L.L., Tang, L., Wang, J., Sang, T., Ge, S., 2008. Analysis of 142 genes resolves the rapid diversification of the rice genus. Genome Biol. 9, R49.