Molecular Plant
Research Article

CellPress
Partner Journal

A well-supported nuclear phylogeny of Poaceae and implications for the evolution of C_4 photosynthesis

Weichen Huang[1], Lin Zhang[2], J. Travis Columbus[3], Yi Hu[1], Yiyong Zhao[1,2], Lin Tang[1,4], Zhenhua Guo[5], Wenli Chen[6], Michael McKain[7], Madelaine Bartlett[8], Chien-Hsun Huang[1,2], De-Zhu Li[5], Song Ge[6] and Hong Ma[1,*]

[1]Department of Biology, 510 Mueller Laboratory, Huck Institutes of the Life Sciences, The Pennsylvania State University, University Park, State College, PA 16802, USA

[2]Ministry of Education Key Laboratory for Biodiversity Science and Ecological Engineering and State Key Laboratory of Genetic Engineering, Institute of Biodiversity Sciences and Institute of Plant Biology, School of Life Sciences, Fudan University, 2005 Songhu Road, Shanghai 200438, China

[3]Rancho Santa Ana Botanic Garden and Claremont Graduate University, 1500 North College Avenue, Claremont, CA 91711, USA

[4]College of Life Sciences, Zhejiang University, Hangzhou, Zhejiang 310058, China

[5]Plant Germplasm and Genomics Center, Germplasm Bank of Wild Species, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, Yunnan 650201 China

[6]State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China

[7]Department of Biological Sciences, University of Alabama, 411 Mary Harmon Bryant Hall, Tuscaloosa, AL 35487, USA

[8]Biology Department, University of Massachusetts Amherst, 611 North Pleasant Street, 221 Morrill 3, Amherst, MA 01003 USA

***Correspondence: Hong Ma (hxm16@psu.edu)**

https://doi.org/10.1016/j.molp.2022.01.015

ABSTRACT

Poaceae (the grasses) includes rice, maize, wheat, and other crops, and is the most economically important angiosperm family. Poaceae is also one of the largest plant families, consisting of over 11 000 species with a global distribution that contributes to diverse ecosystems. Poaceae species are classified into 12 subfamilies, with generally strong phylogenetic support for their monophyly. However, many relationships within subfamilies, among tribes and/or subtribes, remain uncertain. To better resolve the Poaceae phylogeny, we generated 342 transcriptomic and seven genomic datasets; these were combined with other genomic and transcriptomic datasets to provide sequences for 357 Poaceae species in 231 genera, representing 45 tribes and all 12 subfamilies. Over 1200 low-copy nuclear genes were retrieved from these datasets, with several subsets obtained using additional criteria, and used for coalescent analyses to reconstruct a Poaceae phylogeny. Our results strongly support the monophyly of 11 subfamilies; however, the subfamily Puelioideae was separated into two non-sister clades, one for each of the two previously defined tribes, supporting a hypothesis that places each tribe in a separate subfamily. Molecular clock analyses estimated the crown age of Poaceae to be ~101 million years old. Ancestral character reconstruction of C_3/C_4 photosynthesis supports the hypothesis of multiple independent origins of C_4 photosynthesis. These origins are further supported by phylogenetic analysis of the *ppc* gene family that encodes the phosphoenolpyruvate carboxylase, which suggests that members of three paralogous subclades (*ppc-aL1a*, *ppc-aL1b*, and *ppc-B2*) were recruited as functional C_4 *ppc* genes. This study provides valuable resources and a robust phylogenetic framework for evolutionary analyses of the grass family.

Keywords: Gramineae, transcriptome, nuclear phylogeny, molecular clock, C_4 photosynthesis, *ppc* gene evolution

Huang W., Zhang L., Columbus J.T., Hu Y., Zhao Y., Tang L., Guo Z., Chen W., McKain M., Bartlett M., Huang C.-H., Li D.-Z., Ge S., and Ma H. (2022). A well-supported nuclear phylogeny of Poaceae and implications for the evolution of C_4 photosynthesis. Mol. Plant. **15**, 755–777.

Published by the Molecular Plant Shanghai Editorial Office in association with Cell Press, an imprint of Elsevier Inc., on behalf of CSPB and CEMPS, CAS.

INTRODUCTION

The grass family, i.e., Poaceae (also called Gramineae), is widely distributed and is the fifth largest plant family, consisting of 12 subfamilies and over 11 000 species (Kellogg, 2015; Christenhusz and Byng, 2016; Soreng et al., 2017). Species from this family, such as rice, wheat, maize, millet, sorghum, and barley, have been domesticated by humans as major sources of staple foods, as important fodder and forage for farm animals, and as industrial materials, including sugarcane, bamboos, and reeds. Grass crop domestication and breeding efforts have been a major focus of agriculture since the dawn of civilization. Grasses are also essential components of many diverse ecosystems, such as grasslands, wetlands, and savannas. One characteristic of many grass species, including the crops maize, sorghum, millet, and sugarcane, is carbon fixation via the C_4 photosynthetic pathway, which involves a four-carbon intermediate, in addition to typical C_3 photosynthesis that uses a three-carbon intermediate (Sage, 2004; Christin et al., 2007a; Muhaidat et al., 2007; Schlüter and Weber, 2020). C_4 photosynthesis increases the local concentration of CO_2 near the carbon-fixing enzyme Rubisco, thereby improving the efficiency of photosynthesis and increasing the adaptability of C_4 plants, especially in hot and dry environments (Christin et al., 2007a; Edwards and Still, 2008). Grasses account for ~60% of all C_4 plants (~7500 species from multiple families; Sage, 2004), but other C_4 plants are found in Cyperaceae (a family in Poales, the same monocot order as grasses) and many eudicot families, such as Asteraceae, Brassicaceae, Euphorbiaceae, and families in the large order Caryophyllales (Sage, 2004; Christin et al., 2007a, 2007b; Muhaidat et al., 2007).

Common species of Poaceae have long been recognized and named by people around the world; the current grass classification is built on extensive analyses of the phenetic taxonomy described in Clayton and Renvoize, 1986, Tzvelev (1989), Watson and Dallwitz (1992), Clayton et al. (2006) and their subsequent works. More recently, molecular phylogenetic analyses have facilitated revision of the Poaceae classification, leading to the current division into 12 subfamilies and to phylogenies of large subfamilies such as Pooideae (~3900 species) and Panicoideae (~3300 species) (Grass Phylogeny Working Group et al., 2001; Simon, 2007; Grass Phylogeny Working Group II, 2012; Kellogg, 2015; Soreng et al., 2015; Soreng et al., 2017). Among the 12 subfamilies, Anomochlooideae, Pharoideae, and Puelioideae are small subfamilies that contain four, 12, and 11 species, respectively (Clark and Judziewicz, 1996; Clark et al., 2000; Kellogg, 2015; Soreng et al., 2017) and form a grade of successive sisters to the remainder of the family. The other nine subfamilies form two large sister clades: the BOP clade with Bambusoideae, Oryzoideae, and Pooideae and the PACMAD clade with Panicoideae, Aristidoideae, Chloridoideae, Micrairoideae, Arundinoideae, and Danthonioideae (Kellogg, 2000; Kellogg, 2001; Grass Phylogeny Working Group II, 2012; Saarela et al., 2015; Soreng et al., 2015; Soreng et al., 2017). All C_4 grasses are found in subfamilies of the PACMAD clade, whereas members of the BOP clade, which contains important crops such as rice and wheat, as well as bamboos, are all C_3 plants (Sage, 2004). A well-supported Poaceae phylogeny can facilitate evolutionary and comparative studies of topics such as the evolution of inflorescence structure (Vegetti and Anton, 1995; Perreta et al., 2009) and the origin of C_4 photosynthetic pathways (Vicentini et al., 2008; Christin and Besnard, 2009; Grass Phylogeny Working Group II, 2012).

At the same time, some phylogenetic uncertainties remain among Poaceae members, including among tribes and subtribes. The early-divergent subfamilies (Anomochlooideae, Pharoideae, and Puelioideae) are small, and their monophyly and relationships are supported by morphological characters and phylogeny using a few molecular markers (Clark and Judziewicz, 1996; Clark et al., 2000; Grass Phylogeny Working Group II, 2012). In the BOP clade, the tribe-level relationships in the largest grass subfamily Pooideae have not been fully resolved, and a well-resolved nuclear phylogeny of Bambusoideae remains elusive. The relationships among the PACMAD subfamilies are generally consistent, but some aspects still differ among studies (Prasad et al., 2011; Soreng et al., 2017; Saarela et al., 2018). Moreover, the phylogeny at the tribal and sub-tribal levels, especially within large subfamilies, is still unclear or incomplete. The subfamily Chloridoideae, for example, contains five tribes according to some studies, including the tribe Centropodieae with *Centropodia* and *Ellisochloa* (Peterson et al., 2011; Soreng et al., 2017). However, in other studies, Centropodieae is placed closer to other members of the PACMAD clade rather than to Chloridoideae and is therefore not regarded as a tribe in Chloridoideae (Fisher et al., 2016). Also, the largest tribe in Chloridoideae, Cynodonteae, consists of 21 subtribes (Soreng et al., 2017), but the relationships among them remain to be determined. The largest PACMAD subfamily, Panicoideae, contains 13 tribes and is diverse in morphology and other important traits, but the relationships among Panicoideae tribes are not consistent among previous studies (Bouchenak-Khelladi et al., 2008; Grass Phylogeny Working Group, 2012; Soreng et al., 2017; Saarela et al., 2018; Dunning et al., 2019; Welker et al., 2020). All known C_4 grasses are members of four subfamilies in the PACMAD clade, namely Aristidoideae, Chloridoideae, Micrairoideae, and Panicoideae, the last of which contains most C_4 grass species (e.g., Sinha and Kellogg, 1996; Christin et al., 2012). Previous studies have proposed multiple origins of C_4 photosynthesis in grasses (Sinha and Kellogg, 1996; Christin et al., 2007a, 2012; Edwards and Still, 2008). However, the relationships among some PACMAD subfamilies and among some lineages within Chloridoideae and Panicoideae remain uncertain. Examples include the position of Centropodieae (with both C_3 and C_4 taxa) relative to other Chloridoideae and the relationships among early-divergent tribes in Panicoideae, including the C_3 tribes Centotheceae, Chasmanthieae, and Thysanolaeneae and the C_4 tribe Tristachyideae (Grass Phylogeny Working Group II, 2012; Saarela et al., 2018). These phylogenetic relationships must be resolved in order to further understand the evolution of C_3/C_4 photosynthesis in Poaceae.

Previous Poaceae phylogenetic studies have relied largely on plastid and mitochondrial genes or a small number of nuclear genes (reviewed in Kellogg, 2015; Soreng et al., 2015; Soreng et al., 2017), although more recent studies have used over 100 nuclear genes, focusing on the subfamily Chloridoideae (Fisher et al., 2016), and 200 nuclear genes from over 140 species in the BOP and PACMAD clades (Dunning et al., 2019). Further analyses using a relatively large number of genes available from

the nuclear genome can potentially resolve many of the remaining questions in Poaceae phylogeny and are made feasible by rapidly advancing sequencing technologies that generate a large number of sequences in the form of transcriptomes or genomes at relatively low cost. Tens of thousands of nuclear genes can be obtained efficiently from transcriptome datasets, allowing the identification of candidate orthologous genes. Indeed, recent efforts using low-copy nuclear genes have proved successful in resolving previously difficult relationships in large families or among more divergent lineages (Wickett et al., 2014; Zeng et al., 2014; Huang et al., 2016; Xiang et al., 2017; Yang et al., 2018; Mandel et al., 2019; Zhang et al., 2020a). Furthermore, unlike organellar genes, nuclear genes are inherited biparentally and can provide more complete evidence for evolutionary history, including hybridization and other processes.

Here, we generated 342 transcriptome datasets and seven genome skimming datasets for a comprehensive phylogenetic analysis of the Poaceae family using nuclear genes. Combined with 35 public transcriptome/genome datasets, a total of 384 datasets from Poaceae and outgroup species were included. Our sampling covered all 12 subfamilies and 45 of 52 tribes in Poaceae. From the genome/transcriptome sequences of 10 representative Poaceae species, 1234 putative orthologous genes were identified as seeds for searching candidate orthologous sequences from all 384 datasets. Single-gene maximum-likelihood trees of 1234 genes were reconstructed. Organismal phylogenies were reconstructed using the coalescent method with single-gene trees from five subsets of genes selected based on different criteria and using coalescent and super-matrix methods with a 180-gene dataset. Our results confirmed the monophyly of 11 subfamilies (excluding Puelioideae) and the (O, (B, P)) topology of the BOP clade and provided maximal support for relationships among five of the PACMAD subfamilies (except for the placement of Micrairoideae). Using the phylogeny here, we performed ancestral state reconstruction using the maximum parsimony method implemented in Mesquite and molecular phylogenetic analyses of grass homologs of the *ppc* genes that encode the phosphoenolpyruvate carboxylases (PPECs) involved in C_4 photosynthesis. The results support multiple independent origins of C_4 photosynthesis in the PACMAD clade.

RESULTS AND DISCUSSION

Transcriptomic and genomic datasets and selection of nuclear genes

For nuclear phylogenetic analyses, we sequenced 342 transcriptomes and seven genomes (by genome skimming), with a median number of 68 153 unigene sequences and an average N50 value of 934 bp (see Supplemental Table 1 for more statistics on each data set). These and 35 public datasets represent 385 Poaceae samples (two Anomochlooideae, four Aristidoideae, five Arundinoideae, 51 Bambusoideae, 86 Chloridoideae, seven Danthonioideae, three Micrairoideae, 16 Oryzoideae, 79 Panicoideae, one Pharoideae, 111 Pooideae, and six Puelioideae for a total of 371 species, and 14 additional redundant samples) and 13 outgroups. The taxon sampling here includes 45 of the 52 tribes, and the remaining seven un-sampled tribes contain a total of ~40 species.

We used genomic/transcriptomic sequences of 10 Poaceae species from large subfamilies (see Supplemental Figure 1 for the 10 species) to identify 1234 conserved low-copy nuclear genes, and we searched for their homologous sequences in all other datasets here (see Supplemental Table 1 for the number of genes in each sample). Because the six species of Puelioideae (three in *Guaduella* and three in *Puelia*) had genome skimming datasets with relatively shallow sequencing depth, we maximized the gene coverage of Puelioideae by selecting genes that have homologs in at least one species in each of *Guaduella* and *Puelia*, resulting in 1150 genes. To reach relatively high taxon coverage, we selected genes with at least 90% coverage among the sampled taxa, yielding 895 genes. The coalescent method for phylogenetic reconstruction uses single-gene trees; thus, to ensure the quality of each gene tree, we favored longer genes with more phylogenetic information in order to produce gene trees with relatively high support values. We selected three additional sets of 775, 570, and 436 genes using progressively longer alignment length cutoffs (see Supplemental Figure 1 for a workflow). Furthermore, we examined the original set of 1234 genes and removed genes that might be more prone to long-branch attraction (see section "methods"; Supplemental Figures 1 and 2) to generate a set of 579 genes. The overlap of these genes with the previously identified 1150 genes, plus an additional coverage requirement of at least 370 taxa, resulted in a set of 180 genes. This smallest gene set was used for phylogenetic analysis by the maximum-likelihood method with a super-matrix approach because of the known systematic errors that occur when super-matrix datasets with large gene sets are used in phylogenetic reconstructions (Philippe et al., 2011).

A highly supported Poaceae phylogeny: early-divergent lineages

To construct a nuclear Poaceae phylogeny, we used the sets of 1150, 895, 775, 570, and 436 genes for coalescent analyses (Figures 1, 2, 3, and 4, Supplemental Figures 3 and 4). Our results were consistent, with maximum local posterior probability values on most branches (321/(364-1), 88.43%) (Supplemental Figure 3), and they agree with accepted classifications for most taxon groups from the subfamily to the genus level. Phylogenies were also constructed from the set of 180 genes using both the coalescent and super-matrix (maximum-likelihood) methods (Supplemental Figures 4 and 5).

The results from all analyses support the monophyly of 11 subfamilies but not of Puelioideae, which is divided into two highly supported paraphyletic branches, one each for *Guaduella* and *Puelia* (Figure 1 and Supplemental Figures 3–5). Previously, the monophyly of Puelioideae was supported using three genes from two species, *Puelia ciliata* and *Guaduella marantifolia* (Clark et al., 2000). Other Poaceae phylogenetic studies that included Puelioideae sampled a single species, *Puelia olyriformis*, and supported the placement of Puelioideae as the third divergent subfamily before the separation of the BOP and PACMAD clades (Grass Phylogeny Working Group II, 2012; Jones et al., 2014; Saarela et al., 2018). Although the monophyly of Puelioideae could not be rejected by approximately unbiased (AU) tests with the 180-gene

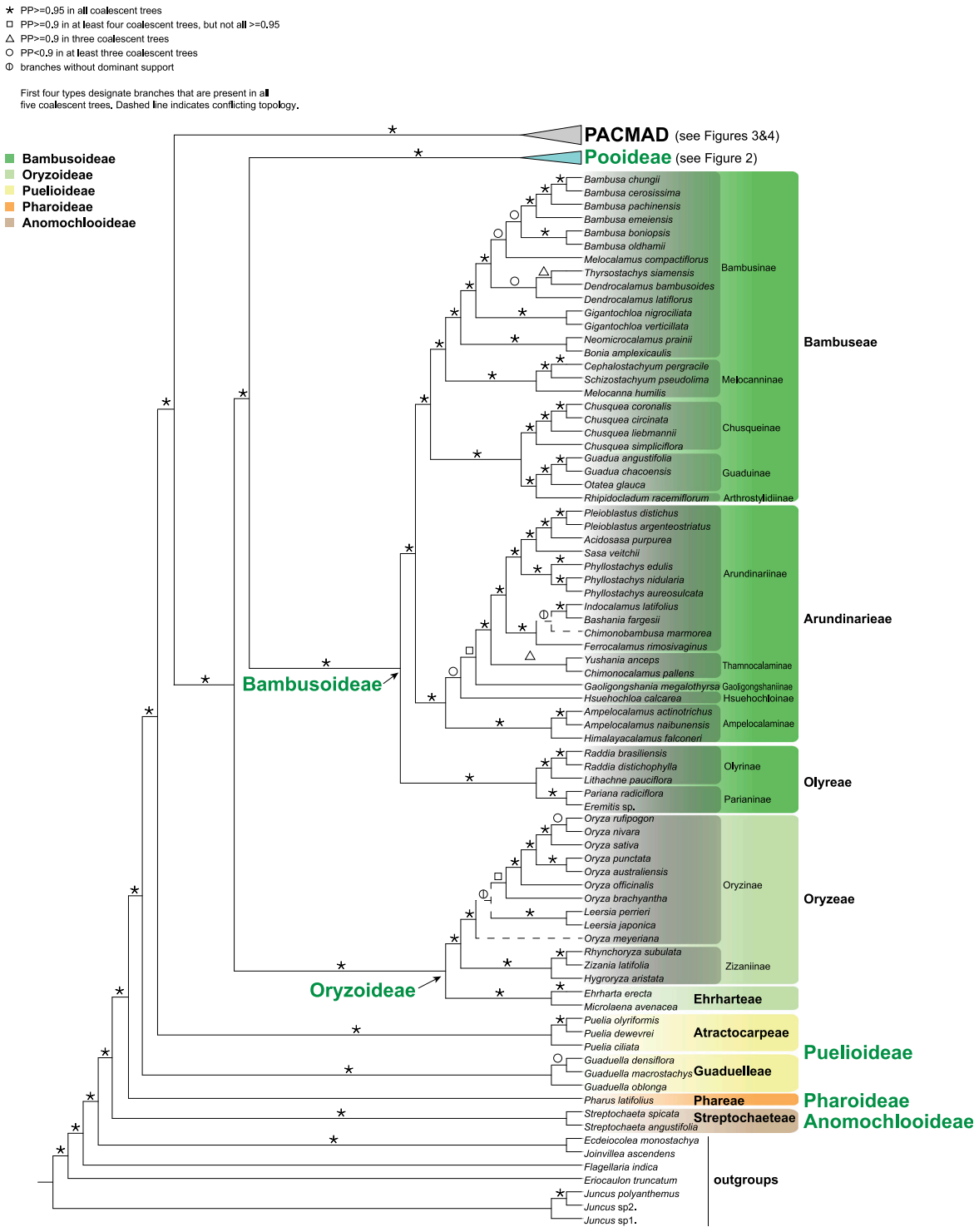

Figure 1. A summary for a portion of the Poaceae phylogeny (including Bambusoideae and Oryzoideae).

A portion of the Poaceae phylogeny is shown here for a summary of results from coalescent analyses using five different gene sets (with 1150, 895, 775, 570, and 436 genes); detailed phylogenetic relationships are shown for species in the three small, early-divergent subfamilies (Anomochlooideae, Pharoideae, Puelioideae) and Bambusoideae and Oryzoideae in the BOP clade. Symbols above the branches represent local PPs, and the corresponding values are indicated in the upper left corner. Pooideae and the PACMAD clade are represented by triangles, and the detailed phylogenetic relationships for these clades are shown in Figures 2, 3, and 4. Different colored backgrounds represent subfamilies, as explained in the upper left corner. Names of subfamilies are shown in green, and names of tribes are shown in black. Branches associated with alternative topology are shown in dashed lines. Detailed local posterior probability support values from the five coalescent analyses are shown in Supplemental Figure 3, and individual coalescent trees are shown in Supplemental Figure 4. The symbols and colors for backgrounds and names are the same in Figures 1, 2, 3, and 4.

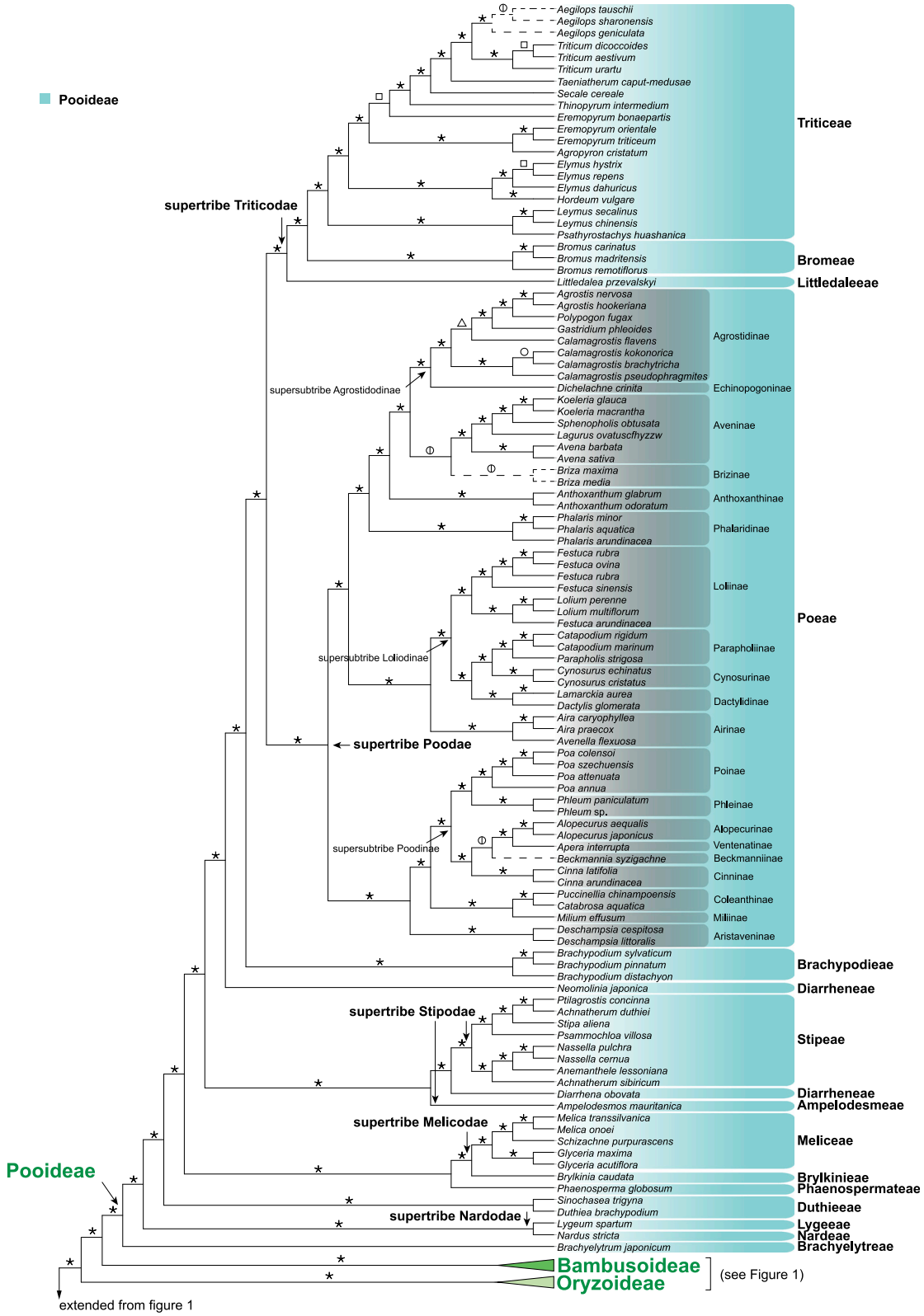

Figure 2. A summary for a portion of the Poaceae phylogeny (Pooideae).
The Pooideae portion of the summarized Poaceae phylogeny is shown. Supertribes are indicated with arrows pointing to the nodes, as are three supersubtribes in the tribe Poeae. See also legend for Figure 1.

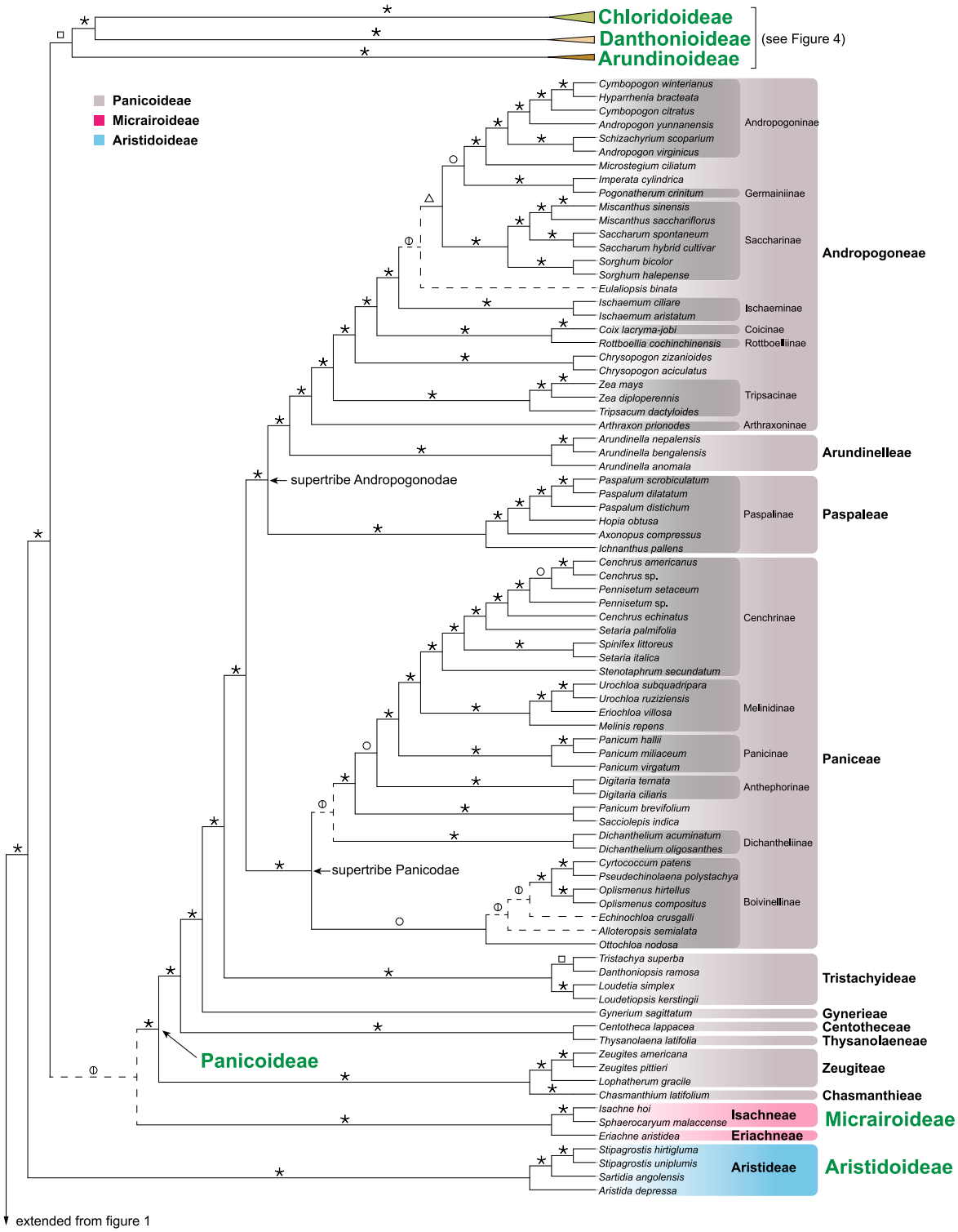

Figure 3. A summary for a portion of the Poaceae phylogeny (Aristidoideae, Micrairoideae, and Panicoideae).
A portion of the summarized Poaceae phylogeny is shown, with three subfamilies, Aristidoideae, Micrairoideae, and Panicoideae, all parts of the PACMAD clade. Supertribes are marked with arrows pointing to the nodes. Alternative topologies are shown in Supplemental Figure 10 for several taxa related to C_3/C_4 evolution. See also legend for Figure 1.

super-matrix dataset (Supplemental Table 4), the difference in monophyly of Puelioideae between this study and previous results could be due to the different histories of nuclear and plastid genes. In addition, Anomochlooideae is always sister to all other Poaceae, followed by Pharoideae and the two clades of Puelioideae (Figure 1).

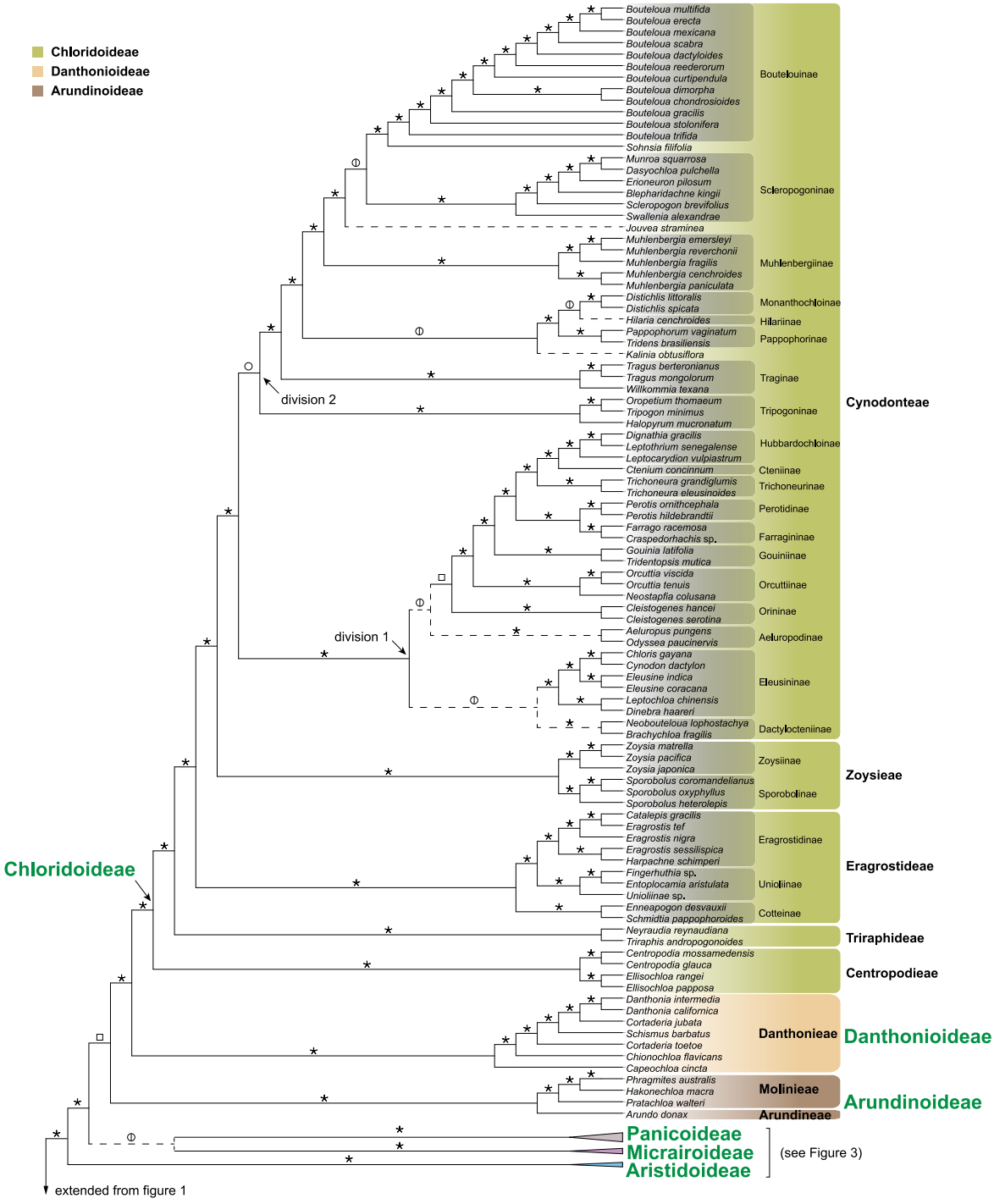

Figure 4. A summary for a portion of the Poaceae phylogeny (Arundinoideae, Danthonioideae, and Chloridoideae).
A portion of the summarized Poaceae phylogeny is shown, with three subfamilies, Arundinoideae, Danthonioideae, and Chloridoideae, all part of the PACMAD clade. Two large clades in the large tribe Cynodonteae are marked with "division 1" and "division 2". See also legend for Figure 1.

Phylogenetic relationships in the BOP clade

The BOP clade with Bambusoideae, Oryzoideae, and Pooideae was first identified by Clark et al. (1995) and is monophyletic in several studies, with alternative relationships among the three subfamilies; however, the topology (O, (B, P)) is supported by recent studies using plastid genes or whole plastomes (Grass Phylogeny Working Group II, 2012; Zhao et al., 2013; Jones et al., 2014; Saarela et al., 2018). The same (O, (B, P)) topology is supported maximally by our results (Figure 1 and Supplemental Figures 3–6).

In Oryzoideae, the two tribes here, Ehrharteae and Oryzeae, are monophyletic, as are two Oryzeae subtribes, Oryzinae and

Zizaniinae (Figure 1). In Oryzinae, a clade of seven *Oryza* species sampled here includes *Oryza sativa* (subspecies *japonica* of cultivated rice) and two closely related species *Oryza nivara* and *Oryza rufipogon*, with moderate support, consistent with the close but complex relationships among these three species (Zhu and Ge, 2005). The relationships of other *Oryza* species relative to *O. sativa* and *O. rufipogon* are different from previously reported relationships (Kellogg, 2009; Tang et al., 2010b). *Oryza meyeriana* and *Oryza granulata* have previously been considered the same species (Ge et al., 1999); *O. meyeriana* is placed at two different positions in different trees here (Figure 1, Supplemental Figures 3 and 7) and in previous studies (Aggarwal et al., 1999; Ge et al., 1999; Zou et al., 2008, 2013; Kumagai et al., 2010). In Zizaniinae, *Hygroryza* is sister to *Rhynchoryza* + *Zizania*, consistent with previous studies (Kellogg, 2009; Tang et al., 2010a, 2010b).

In Bambusoideae, three tribes, Olyreae, Arundinarieae, and Bambuseae, are each monophyletic with maximal support (Figure 1 and Supplemental Figures 3–6), with a topology of (O, (A, B)), where the sister relationship of the woody Arundinarieae and Bambuseae supports a single origin of woodiness in bamboos. Previously, a plastome phylogeny placed Arundinarieae as sister to the other bamboos (Wysocki et al., 2015), supporting two origins of woody bamboos or one loss of woodiness in Olyreae. An analysis of 38 bamboo species (Triplett et al., 2014) and a recent genome-based analysis (Guo et al., 2019) both strongly supported the herbaceous Olyreae being sister to the woody bamboos.

In Olyreae, two subtribes are monophyletic with maximum support in all trees. Members of Arundinarieae are temperate woody bamboos; they were previously placed in the single subtribe Arundinariinae but have recently been divided into five subtribes (Li et al., 2013; Zhang et al. 2018; Zhang et al. 2020b) (Figure 1 and Supplemental Figures 3–6). The phylogeny is consistent in all coalescent trees, except for the position of *Chimonobambusa marmorea* (in Arundinariinae) (Figure 1 and Supplemental Figure 3). The taxon groups (three subtribes and three genera) with two or more species are monophyletic. Among the five subtribes, Ampelocalaminae is placed as sister to the remaining four subtribes, with Hsuehochlinae and Gaoligongshaniinae consistently being successive sisters of the clade of Arundinariinae + Thamnocalamaminae (Figure 1 and Supplemental Figure 3).

In Bambuseae with tropical woody bamboos, five of 11 subtribes were sampled. Guaduinae and Arthrostylidiinae are sisters and form a neotropical clade together with Chusqueinae, consistent with previous studies (Wysocki et al., 2015). Melocanninae and Bambusinae form a paleotropical clade. In Bambusinae, 10 of the sampled taxa belong to the *Bambusa-Dendrocalamus-Gigantochloa* complex (Goh et al., 2013), where *Gigantochloa* is monophyletic and sister to a highly supported clade that includes the other two genera (Figure 1 and Supplemental Figure 3). Bambuseae and Arundinarieae have been reported to have allopolyploid ancestry, with Arundinarieae being tetraploids (subgenomes A and B) and Bambuseae including tetraploids (neotropical; subgenomes C and D) and hexaploids (paleotropical; subgenomes C, D, and E) (Triplett et al., 2014). A recent study of diploid and polyploid woody bamboo genomes presented an alternative hypothesis with ABCD subgenomes, in which subgenome C is shared by the three woody bamboo lineages (Guo et al., 2019). Such a polyploid history of woody bamboos suggests that their phylogenetic relationships are probably more complex than presented here (Guo et al., 2021) (see next section for more discussion).

Pooideae is the largest of the 12 Poaceae subfamilies and includes wheat, barley, and other crops, as well as the model grass *Brachypodium distachyon* (Vogel et al., 2010; Kellogg, 2015; Soreng et al., 2017). The analyses here, with 111 samples in 15 tribes, maximally support monophyly (Figure 2, Supplemental Figures 3, and 4) for seven of eight tribes with at least two species (in order from early to late divergent lineages): Duthieeae, Meliceae, Stipeae, Brachypodieae, Poeae, Bromeae, and Triticeae. However, Diarrheneae, with one species in each of two genera, *Diarrhena* and *Neomolinia*, is polyphyletic (see below). Ten of the 15 Pooideae tribes are grouped into five supertribes (Soreng et al., 2017), all of which are maximally supported as monophyletic. The phylogeny here provides well-resolved relationships among the tribes and supertribes (Figure 2). The separation of the monotypic Phaenospermateae from Duthieeae is consistent with recent reports (Schneider et al., 2011). The supertribe Stipodae with tribes Stipeae and Ampelodesmeae is not monophyletic, as *Diarrhena* of Diarrheneae is sister to Stipeae, and Ampelodesmeae is sister to (Stipeae + *Diarrhena*). The clade with Stipeae, *Diarrhena*, and Ampelodesmeae is sister to a large clade with five tribes and *Neomolinia* of the tribe Diarrheneae. Finally, within Triticodae, Littledaleae is sister to (Triticeae + Bromeae).

Phylogenetic relationships in the PACMAD clade

The PACMAD clade as a whole and its six subfamilies are maximally supported as monophyletic in a number of studies, although the relationships among the subfamilies are inconclusive and sensitive to phylogenetic methods (Grass Phylogeny Working Group II, 2012; Cotton et al., 2015; Soreng et al., 2017; Saarela et al., 2018) (see Supplemental Figure 6 for a comparison between two previous studies and the results here). Nevertheless, increasing evidence supports Aristidoideae as the sister to the other five subfamilies (Vicentini et al., 2008; Grass Phylogeny Working Group II, 2012; Kellogg, 2015; Soreng et al., 2015; Soreng et al., 2017). In the coalescent analyses of this study, Aristidoideae is consistently sister to the other five subfamilies with maximum support (Figures 3 and 4 and Supplemental Figures 3–5). The relationships among four of the remaining five subfamilies are consistent and highly supported as (((Chloridoideae, Danthonioideae), Arundinoideae), Panicoideae), but Micrairoideae is placed at one of two positions with varying support values (Supplemental Figure 7) (see next section for more discussion).

Our results also provide strong support for many relationships within the PACMAD subfamilies (Figures 3 and 4), and relationships among three genera of Aristidoideae are consistent with the GPWG phylogeny (Grass Phylogeny Working Group II, 2012). In Panicoideae (~3240 species), the largest subfamily in PACMAD, our sampling includes 10 out of 13 tribes, and the six tribes with two or more species are all

monophyletic (Figure 3). The tribes Chasmanthieae and Zeugiteae form a sister clade to the remaining Panicoideae, with Centotheceae and Thysanolaeneae in the next divergent clade, followed successively by Gynerieae and Tristachyideae. Previous studies (Sánchez-Ken et al., 2007; Grass Phylogeny Working Group II, 2012; Saarela et al., 2018) supported a branch with three tribes (Tristachyideae + (Thysanolaeneae + Centotheceae)) as sister to all other Panicoideae tribes and (Chasmanthieae + Zeugiteae) on the next divergent branch (see Supplemental Figure 6 for a comparison with previous studies).

The remaining four Panicoideae tribes belong to two maximally supported monophyletic supertribes, Panicodae and Andropogonodae. Panicodae contains only one tribe, Paniceae, with six out of seven subtribes sampled (Figure 3 and Supplemental Figure 3–5), and *Sacciolepis indica*, which was not previously assigned to a subtribe. The subtribes are monophyletic except for Panicinae (*Panicum*) and have consistent relationships (except for the placement of Dichantheliinae) (Figure 3 and Supplemental Figure 3), but the support for the monophyly of Boivinellinae is lower than that for the other subtribes, and other topologies are possible (Supplemental Figures 3 and 4). *Panicum brevifolium* is maximally supported as sister to *S. indica*, apart from other *Panicum* species. *Pennisetum* is nested in a clade with *Cenchrus* species, consistent with the recent treatment of *Pennisetum* as a synonym of *Cenchrus* (Chemisquy et al., 2010). Two *Setaria* species are not grouped together, with *Setaria palmifolia* next to the *Cenchrus*/*Pennisetum* clade and *Setaria italica* sister to *Spinifex littoreus*, consistent with previous studies showing that *Setaria* is not monophyletic and placing *S. palmifolia* and *S. italica* in separate lineages (Morrone et al., 2012).

The other supertribe in Panicoideae, Andropogonodae, has three previously defined tribes, Paspaleae, Arundinelleae, and Andropogoneae (Figure 3), and a recently described tribe, Jansenelleae (Bianconi et al., 2020), which includes two genera not sampled here. The three sampled tribes are maximally supported as monophyletic, with Paspaleae sister to a clade of Arundinelleae plus Andropogoneae, consistent with previous reports (Grass Phylogeny Working Group II, 2012; Saarela et al., 2018) (Supplemental Figure 6). In Paspaleae, *Ichnanthus*, *Axonopus*, and *Hopia* consistently form a grade in all trees here, outside a clade of three *Paspalum* species (Figures 3 and Supplemental Figures 3–5). In Andropogoneae, our sampling includes eight subtribes and four unplaced genera: *Chrysopogon*, *Eulaliopsis*, *Imperata*, and *Microstegium*. Our analyses support the monophyly of the subtribes Tripsacinae, Ischaeminae, and Andropogoninae but not of Saccharinae. In addition, the placement of Arthraxoninae and Tripsacinae as successive sisters to other Andropogoneae is consistent with a previous plastome study (Saarela et al., 2018) but not with the topology of another nuclear phylogeny (Estep et al., 2014). The next lineage to diverge has two *Chrysopogon* species, supporting a recently proposed designation of this genus as a new subtribe (Welker et al., 2020). The subtribes Rottboelliinae and Coicinae form a clade that is sister to the remaining Andropogoneae with four subtribes, which were not resolved previously (Mathews et al., 2002). The previously unplaced *Eulaliopsis* is either sister to a clade with the subtribes Saccharinae, Germainiinae, and Andropogoninae or placed elsewhere (Supplemental Figure 7), and *Microstegium* is maximally supported as sister to Andropogoninae.

Arundinoideae is represented here (Figure 4) by four genera/species from two tribes, Arundineae and Molinieae, the latter of which has two subtribes, Crinipinae and Molininae. The placement of *Pratochloa walteri* in Crinipinae is in agreement with a previous study (Ingram et al., 2011). In Danthonioideae (one tribe Danthonieae), *Danthonia* is monophyletic, but *Cortaderia* is not, in agreement with the reported paraphyly of *Cortaderia* (Grass Phylogeny Working Group, 2003).

Chloridoideae (~1600 species) is the second largest subfamily in PACMAD, and our study included 86 samples in 56 (out of 124) genera; all five tribes, Centropodieae, Triraphideae, Eragrostideae, Zoysieae, and Cynodonteae, are maximally supported as monophyletic (Figure 4). Centropodieae, with two monophyletic genera, is maximally supported as sister to other Chloridoideae. Triraphideae, Eragrostideae, and Zoysieae are monophyletic and comprise the next three successive sister lineages of the remaining Chloridoideae. Within Eragrostideae, all three subtribes are monophyletic, with Cotteinae sister to a clade of Unioliinae and Eragrostidinae, although *Eragrostis* is paraphyletic. The two Zoysieae subtribes, Zoysiinae and Sporobolinae, are both monophyletic.

The largest Chloridoideae tribe, Cynodonteae, is sister to Zoysieae, and our sampled species represented 19 subtribes and three genera that were not previously placed in a subtribe (Figure 4 and Supplemental Figures 3–5). These subtribes and genera form two large sister clades: division 1 and division 2 (Figure 4). In division 1, Dactylocteniinae and Eleusininae form the most basal lineage in two trees with larger numbers of genes, and Aeluropodinae forms the next lineage; however, in three trees with smaller numbers of genes, Dactylocteniinae, Aeluropodinae, and Eleusininae form a grade outside the remaining taxa of division 1 (Supplemental Figure 4). Next, Orininae and Orcuttiinae form a grade outside a maximally supported clade that contains six subtribes. In division 2, the subtribe Tripogoninae is monophyletic and sister to the other subtribes of this division. Different relationships of Pappophorinae with other subtribes were reported previously (Soreng et al., 2017) (Supplemental Figure 6). Among the three unplaced genera, *Kalinia* is sister to a clade of Pappophorinae, Hilariinae, and Monanthochloinae, whereas *Jouvea* is sister to a weakly supported clade containing Scleropogoninae + (the unplaced *Sohnsia* + Boutelouinae) in three of the trees (Supplemental Figure 7). The well-resolved relationships among *Bouteloua* species are generally consistent with previous studies (Columbus, 1999; Peterson et al., 2015).

Polyploidy in grasses and possible impact on the Poaceae phylogeny

Grasses have experienced multiple rounds of polyploidization, including one shared by all grasses (Tang et al., 2010b), those in the early history of the Bambusoideae subfamily (Guo et al., 2019, 2021), and more recent ones involving members of related genera or within a genus, such as those in the tribe Andropogeneae and other grasses (Mason-Gamer et al., 2010; Liu et al., 2011; Estep et al., 2014; Triplett et al., 2014).

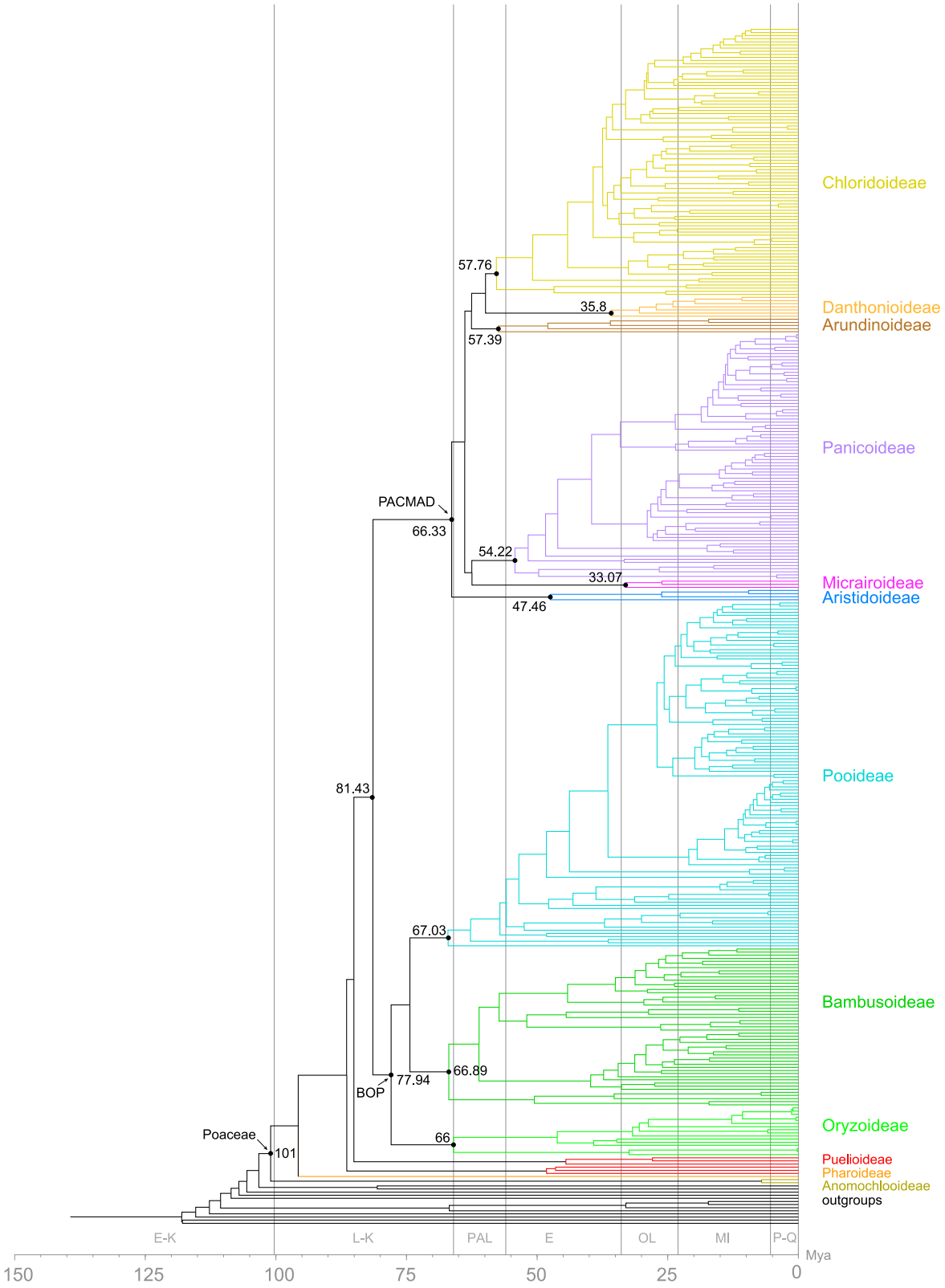

Figure 5. Divergence time estimation for Poaceae.

A Poaceae time tree was estimated from molecular clock analyses performed in treePL, using the topology of the 1150-gene coalescent tree as a backbone and the concatenated super-matrix from 180 genes for calculation of branch length. Information on fossils used for calibration is shown in

(legend continued on next page)

Therefore, the bifurcating topology of the Poaceae phylogeny shown here is likely to be a simplified and partial view of grass evolutionary history. Nevertheless, for the early polyploidy events shared by all grasses or many members of Bambusoideae, for which the parental lineages of the presumed polyploid hybrids have not been identified and are possibly extinct, the relationships among grasses or bamboos, respectively, can be generally reliable if individual orthologous groups (OGs) of marker genes do not include different paralogs from such polyploidy events. For relatively recent polyploidy events, we generally did not include more than two species affected by such an event; thus, the effect of recent polyploidy events on phylogeny awaits further analyses with greater sampling density. For Puelioideae, the proposed non-monophyly based on the nuclear phylogeny could also be affected by past polyploidy, although analyses of chromosome number suggest that *Puelia* species are diploid (Dujardin, 1978; Soderstrom, 1981), and no evidence for polyploidy of *Guaduella* has been reported. Also, as three species in each genus were included in the phylogeny here, polyploidy in a specific member of either genus is unlikely to have affected the overall phylogenetic placement of these genera.

We performed additional analyses to further examine the potential effect of gene duplications (such as those from polyploidy events) that occurred relatively early in the Poaceae history. First, we obtained the local posterior probabilities (PPs) and the quartet support values at each node of the three alternative topologies from coalescent analyses using the 1150 genes and the four smaller subsets of genes (the alternatives around a node are for a quadripartition and not the bipartition; see Supplemental Figure 4). We found that most nodes have a relatively high PP value and quartet support value for the first topology, as shown in Figures 1, 2, 3, and 4, including the monophyly of 11 subfamilies. We also found high PP support values for relationships among the subfamilies, although the quartet support was weak for some of the relationships among the PACMAD subfamilies. The generally high support for Poaceae relationships is consistent with the idea that the subgenomes of a recent polyploid often lose genes differentially, making it more likely that the single-copy genes sampled from multiple species are derived from the same ancestral copy and are thus orthologous.

For Bambusoideae, the PP values and the quartet support values were generally supportive of the species phylogeny, including the monophyly of two clades of tropical woody bamboos (Figure 1 and Supplemental Figures 3–5). The woody bamboos have been proposed to be tetraploids or hexaploids, with tropical woody bamboos and temperate woody bamboos belonging to the tribes Bambuseae and Arundinarieae, respectively (Kellogg, 2015; Guo et al., 2019). The tropical woody bamboos form two clades, a paleotropical clade (hexaploids, including subtribes Bambusinae and Melocanninae, Figure 1) and a neotropical clade (tetraploids, with the subtribes Arthrostylidiinae, Chusqueinae, and Guaduinae), and the temperate woody bamboos are also tetraploids. Guo et al. (2019) reported genomic evidence supporting the model that the paleotropical bamboos share the ABC subgenomes, the neotropical bamboos have the BC subgenomes, and the temperate woody bamboos carry the CD subgenomes, with their diploid progenitors thought to be long extinct. The Bambusoideae phylogeny here is consistent with the polyploid history proposed above, as the temperate, paleotropical, and neotropical woody bamboos form separate monophyletic groups.

The quartet analyses also revealed that there are alternative topologies at some nodes with considerable support, such as those for some relationships among the PACMAD subfamilies. As these subfamilies diverged within a relatively short period of time (see below, Figure 5), the differences in topologies among gene trees could be due to several possible factors, such as incomplete lineage sorting and insufficient phylogenetic resolving power, in addition to the possible use of paralogs from ancient polyploidy events. To test further for the effects of paralogs from past polyploidizations, we examined the gene trees for evidence of paralogy/non-orthology. To do this, we needed two conditions: (1) gene trees with sufficiently high support for key topologies, and (2) reliable knowledge of species relationships to assess orthology. For the first condition, we avoided genes in the larger gene sets that had either relatively low taxon coverage or short sequences (Supplemental Figure 1), and we focused on the set with 436 OGs because both low coverage and short sequences can lead to less reliable gene tree topologies. For the second condition, we chose to use monophyly of the five largest subfamilies (Bambusoideae, Chloridoideae, Oryzoideae, Panicoideae, and Pooideae) as support for gene orthology, as such monophyly is supported by analyses of both chloroplast genes and the five sets of nuclear genes here. The examination of the gene trees for the 436 OGs suggested that a relatively small number of sequences (1–10 sequences for 263 gene trees; 11–20 sequences for 87 gene trees; 21–43 for 51 gene trees) did not group with a majority of sequences from the same subfamily, suggesting that they are not orthologous to most sequences in the same OG. For the 401 OGs with at least one putative non-ortholog, gene trees were reconstructed after removal of the putative non-orthologous sequences, and a new coalescent tree was generated using the modified 436 gene set. We also generated two other coalescent trees using a 390-gene set with 90% species coverage and a 373-gene set with presence in both Puelioideae genera, as removal of the putative non-orthologs reduced the species coverage for some genes (see coalescent trees from filtered gene sets in Supplemental Figure 8). The phylogenetic relationships in these coalescent trees are generally consistent with those in Figures 1, 2, 3, and 4; for example, the monophyly of subfamilies, the paraphyly of Puelioideae, and most of the relationships among the subfamilies are the same. Micrairoideae was supported as sister to Panicoideae in four

Supplemental Table 6. Terminals of different subfamilies are marked in different colors, and subfamily names are indicated to the right. Terminal taxon names are omitted for brevity. Geological time scale is shown at the bottom, and periods are delimited with vertical gray lines. E-K, early Cretaceous; L-K, late Cretaceous; PAL, Paleocene; E, Eocene; OL, Oligocene; MI, Miocene; P-Q, Pliocene and Quaternary. Estimated divergence times of subfamilies are marked at corresponding nodes. Detailed divergence times with species names are shown in Supplemental Figure 9.

of the five earlier coalescent trees (Supplemental Figure 3), whereas it was sister to the clade of ((Chloridoideae, Danthonioideae), Arundinoideae) with support of PP = 0.86 from the original 436-gene set (Supplemental Figures 4 and 8). In the three coalescent trees obtained after removal of putative non-orthologs, Micrairoideae is sister to Panicoideae with PP values of 0.78–0.89 (Supplemental Figure 8), suggesting that its placement next to ((Chloridoideae, Danthonioideae), Arundinoideae) may have been influenced by non-orthologs. Therefore, although it is likely that the Poaceae history is more complex than that depicted in the phylogeny here, these phylogenetic relationships represent a major portion of the history, especially at the levels of subfamily and tribe.

Lower cretaceous origin of Poaceae

The newly reconstructed nuclear phylogeny of Poaceae provides an opportunity to estimate the origin and divergence times of major lineages using molecular clock analysis. Early studies suggested that there were a few pollen fossils related to Poaceae dated to no older than 70 million years ago (mya) (Muller, 1981). More recently, earlier fossils have been discovered (Shi et al., 2012; Wu et al., 2018), supporting older ages for Poaceae and its major clades, and the Poaceae crown age has been estimated as older than 100 million years (my) using these fossil calibrations (Schubert et al., 2019). We performed molecular clock analysis with treePL 1.0 (Smith and O'Meara, 2012) using the ML tree from a concatenated super-matrix of 180 nuclear genes. To include fossil calibrations outside the Poaceae family, we added six more outgroup species in addition to the seven species in the coalescent analyses. The upper and/or lower boundaries of these calibrations were set according to previous studies (see Supplemental Table 6).

Our analyses estimate the crown age of Poaceae to be ~101 my, in the early Cretaceous (Figure 5, and see confidence intervals [CIs] in Supplemental Figure 9 and the CI values for major nodes in Supplemental Table 7). Following successive divergences of the four early Poaceae lineages over a period of ~20 my, the crown age of (PACMAD + BOP) is estimated to be ~81 my in the late Cretaceous. Thus, the PACMAD and BOP clades probably diverged before the Cretaceous-Paleogene (K-Pg) boundary. Furthermore, the three BOP subfamilies also diverged from each other before the K-Pg boundary (between ~78 and 74 mya), whereas the six PACMAD subfamilies separated from each other shortly after the K-Pg boundary over a period of less than 7 my (~66.33–59.86 mya). Subsequently, most tribes in large subfamilies diverged over much longer periods. For example, among the tribes in Pooideae, Brachyelytreae (*Brachyelytrum*) diverged from other Pooideae at ~67 mya, whereas Bromeae (*Bromus*) separated from Triticeae (with *Leymus* and *Triticum*) much more recently at ~19 mya (Supplemental Figure 9); thus, the divergences among tribes of Pooideae spanned a period of ~48 my. Similarly, the tribes in Panicoideae diverged over a period of ~30 my (from ~54.22 to ~23.54 mya). Our analyses also provide divergence times for the subtribes and genera sampled here (Supplemental Figure 9); there is a general tendency for the divergence times in Oryzoideae and Bambusoideae to be older than those in Pooideae, Panicoideae, and Chloridoideae.

Ancestral character reconstruction supports multiple origins of C_4 photosynthesis in PACMAD grasses

The success of grasses is thought to be due in part to their ability to fix carbon via C_4 photosynthesis, which facilitates adaptation to habitats with stressful conditions, such as high temperature or light intensity, aridity, and salinity (Christin et al., 2007a; Edwards and Still, 2008). In hot and dry environments, plants tend to close stomata to retain water, reducing their access to CO_2. Many C_4 plants have evolved a specialized organization of leaf tissues called Kranz anatomy (Tregunna et al., 1970; Smith and Epstein, 1971) that can increase the local concentration of CO_2 near the carbon-fixing enzyme Rubisco by physically separating the light-dependent reactions from the Calvin cycle.

C_4 photosynthesis has been reported in several angiosperm families, including Amaranthaceae, Asteraceae, Brassicaceae, Cyperaceae, and Euphorbiaceae; however, Poaceae has the largest number (~4500 species, ~60% of all C_4 plants) of C_4 species (Sage, 2004). All C_4 grasses belong to the PACMAD clade, although they do not form a monophyletic group. To date, four subfamilies are reported to contain C_4 species, namely Aristidoideae, Micrairoideae, Panicoideae, and Chloridoideae. However, the existence of undiscovered C_4 species in Arundinoideae or Danthonioideae cannot be excluded, as the photosynthetic pathway is somewhat of a continuous, complex trait, and sometimes there are both C_3 and C_4 ecotypes/subspecies within a species, such as *Alloteropsis semialata* (Lundgren et al., 2016).

Among the Poaceae species sampled for this project, 150 have been described as C_4 species, including one species in Micrairoideae, three in Aristidoideae, 62 in Panicoideae, and 84 in Chloridoideae, according to Soreng et al. (2017). The C_3/C_4 photosynthetic ancestral states were reconstructed using the maximum parsimony method implemented in Mesquite (version 3.6) using the coalescent trees from five different gene sets (with 378 termini) (Figure 6 and Supplemental Figure 10). As C_4 species are only known in the four PACMAD subfamilies, our analyses support the hypotheses that the most recent common ancestor (MRCA) of Poaceae, the four nodes for the separation of the three earliest divergent subfamilies, the crown node of BOP + PACMAD, and the MRCAs of the BOP clade and the three BOP subfamilies were all C_3 (Figure 6 and Supplemental Figure 11).

The subfamily Aristidoideae is sister to the other PACMAD subfamilies, consistent with the relationship summarized recently (Soreng et al., 2017). The MRCA of PACMAD and that of the five subfamilies after the divergence of Aristidoideae are both proposed to be C_3 (Figure 6). However, the ancestral state of Aristidoideae is uncertain. All three genera in Aristidoideae are sampled here, with a phylogenetic topology in which *Aristida* is sister to (*Sartidia* + *Stipagrostis*), consistent with previous studies (Cerros-Tlatilpa and Columbus, 2009). The fact that most *Aristida* and *Stipagrostis* species are C_4 while *Sartidia* and *Aristida longifolia* are C_3 makes the ancestral state of Aristidoideae equivocal. If the Aristidoideae ancestor was C_3, then C_4 has originated at least twice, once in *Aristida* and once in *Stipagrostis*, whereas *Sartidia* has retained the ancestral state, consistent with a previous report (Cerros-Tlatilpa and

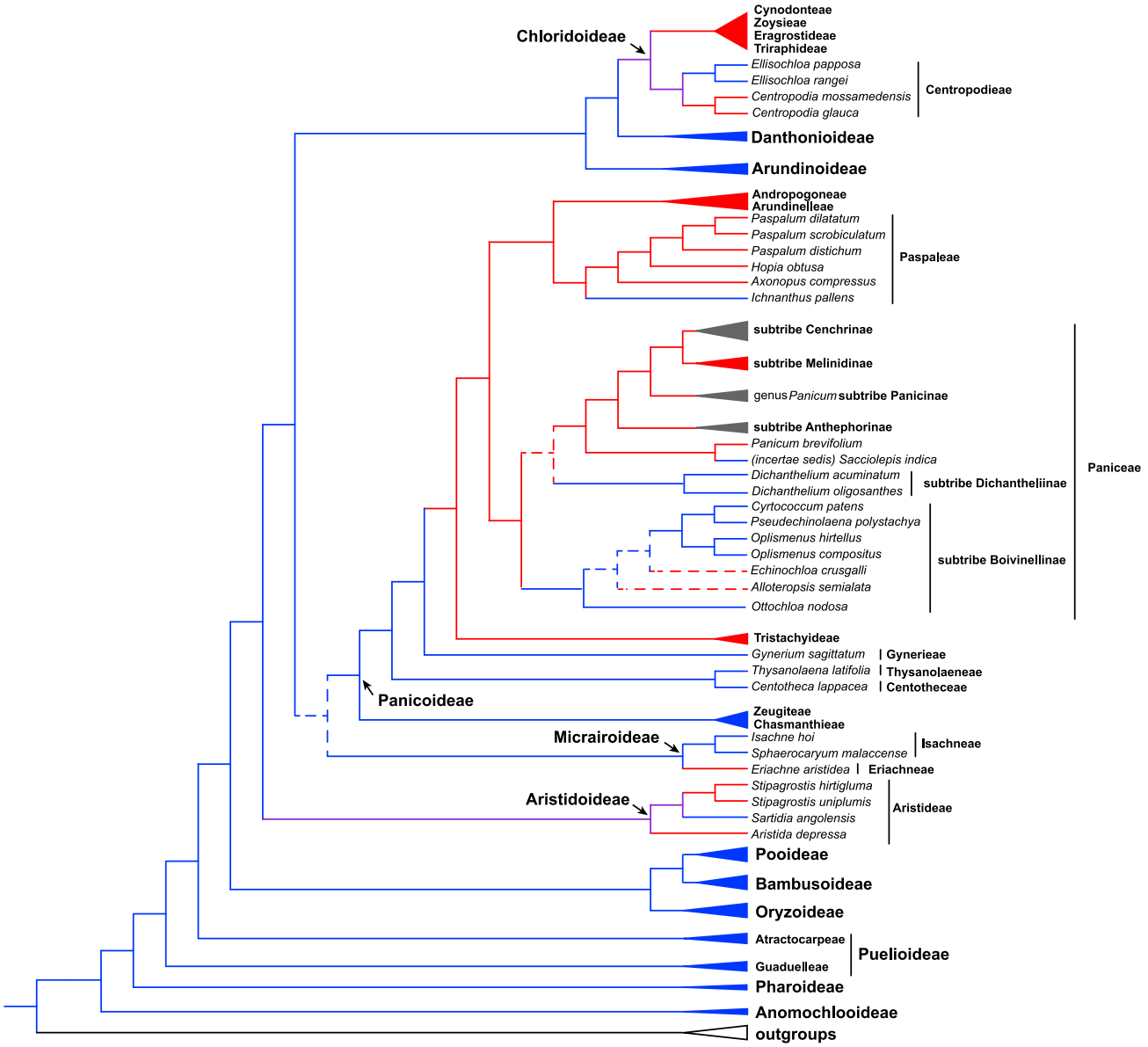

Figure 6. Ancestral state reconstruction of photosynthetic pathway type in Poaceae.
Ancestral states of photosynthetic pathway type (C_3/C_4) were estimated using information from extant taxa (species) by the maximum-likelihood method using the Mesquite program. Terminals and branches are marked with different colors to represent photosynthetic type. Red indicates C_4, blue indicates C_3, purple indicates an uncertain ancestral state, and gray indicates a mixed clade that includes both C_3 and C_4 species/genera, although some were not sampled. The mixed clades shown here are dominated by C_4 species, and our sampling included only the C_4 species. Branches associated with alternative topology are shown in dashed lines. The detailed ancestral character reconstruction is shown in Supplemental Figure 11.

Columbus, 2009). A less likely scenario is that the Aristidoideae ancestor was already C_4, and there were reversions to C_3 in *Sartidia* and *A. longifolia*.

The ancestral state of each of the other three subfamilies that contain C_4 species is proposed to be C_3, according to the reconstruction here. More specifically, Micrairoideae was estimated to be C_3 in all analyses from five different coalescent trees, regardless of the placement of this subfamily. Our sampling includes three Micrairoideae genera, and the C_4 genus *Eriachne* is sister to the clade of *Isachne* + *Sphaerocaryum* (both are C_3); alternatively, the ancestral state of Micrairoideae could also be C_4, with a reversion to C_3 in the MRCA of *Isachne* + *Sphaerocaryum*. Among the genera in Micrairoideae not sampled here, *Micraira* alone defines a tribe (Micraireae) and was placed as the first divergent lineage in the subfamily (Soreng et al., 2017). As *Micraira* is C_3, its placement as sister to the other Micrairoideae genera would support C_3 as the ancestral state of Micrairoideae.

Panicoideae is the largest subfamily in PACMAD, with 13 tribes and 10 tribes represented in our sampling, including five with C_3 species but no known C_4 species (Centotheceae, Chasmanthieae, Gynerieae, Thysanolaeneae, and Zeugiteae), three with only known C_4 species (Andropogoneae, Arundinelleae, and

Tristachyideae), and two with both C_3 and C_4 species (Paniceae and Paspaleae). The three small tribes not sampled here (Cyperochloeae [two species], Lecomtelleae [one species], and Steyermarkochloeae [two species]) contain only C_3 species. In the phylogeny here, the earliest divergent lineage of Panicoideae has two tribes, Chasmanthieae and Zeugiteae, and the sister relationship of Chasmanthieae and Zeugiteae is consistent with previous reports (Soreng et al., 2017; Saarela et al., 2018) (Supplemental Figure 6). The next two early separating branches are (Centotheceae + Thysanolaeneae) and Gynerieae. The three early-divergent lineages mentioned above are all C_3; thus, this relationship strongly supports C_3 as the ancestral state of Panicoideae. The MRCA of the other five tribes sampled here and the MRCA of each of these tribes (Tristachyideae, Paniceae, Paspaleae, Arundinelleae, and Andropogoneae) are supported as C_4 by the ancestral character reconstruction, even though Paniceae and Paspaleae also contain C_3 species (Figure 6). The three small C_3 tribes not sampled here were previously placed outside this large clade of five tribes (Soreng et al., 2017); therefore, their placements also support C_3 as ancestral for Panicoideae. Within Paniceae, the inferred ancestral state of subtribes Boivinellinae and Dichantheliinae varies between analyses, mainly because of uncertainties in the phylogeny (Supplemental Figures 3, 4, and 10). Nevertheless, our analyses support two probable C_4 to C_3 reversions for the two C_3 species *Ichnanthus pallens* and *S. indica* in tribes Paspaleae and Paniceae, respectively (Figure 6). In the phylogeny here, the mostly C_4 Paspaleae is sister to a combined clade of two C_4 tribes, Andropogoneae and Arundinelieae, supporting the ancestor of Paspaleae as C_4 and a reversion to C_3 in *I. pallens*. In Paniceae, *S. indica* was previously not assigned to a subtribe, but in our results, it is supported as sister to the C_4 species *P. brevifolium*. This relationship supports a reversion to C_3 in *S. indica*. Three subtribes in Paniceae (Anthephorinae, Cenchrinae, and Panicinae) contain both C_3 and C_4 species, but our sampling was incomplete; in addition, the tribe Paspaleae contains a few other C_3 genera that are not included here. Therefore, more sampling is needed to better understand the evolution of C_4 photosynthesis in these two tribes.

Among the remaining three subfamilies of the PACMAD clade, Arundinoideae and Danthonioideae are entirely C_3 and are placed as successive sisters of Chloridoideae, supporting the MRCA of the combined clade of these three subfamilies as C_3 (Figure 6). Within Chloridoideae, the tribe Centropodieae has two genera, *Centropodia* (C_4) and *Ellisochloa* (C_3), and is sister to all the other chloridoid grasses, making the ancestral states of Centropodieae and Chloridoideae uncertain in our analysis, even though the MRCA of the combined clade of the other four tribes is inferred to be C_4. If the MRCA of Chloridoideae was C_3, then there were two origins of C_4 photosynthesis, one for *Centropodia* and the other for the MRCA of the other four tribes that contain the majority of Chloridoideae. However, if the ancestral states of Chloridoideae and Centropodieae were both C_4, then there was one origin of C_4 for the subfamily and a reversion to C_3 for *Ellisochloa*.

The ancestral state reconstruction analyses here support separate origins of C_4 in each of the four subfamilies that contain C_4 species, possibly more than one origin in Aristidoideae and Chloridoideae, and multiple origins in Panicoideae, consistent with a previous report (Grass Phylogeny Working Group II, 2012) that proposed as many as 24 separate transitions from C_3 to C_4 photosynthesis. At the same time, there were possible reversions in Panicoideae (Paniceae and possibly others).

It should be noted that sampling limitations here, including the lack of some C_3 lineages, probably affected some of the ancestral state reconstruction results. In Panicoideae, especially in supertribe Panicodae and Andropogonodae, the sampling favored C_4 species and is likely to have increased the probability of inferring the ancestral nodes of these two supertribes as C_4. On the other hand, our sampling included early-divergent tribes of Panicoideae, such as Chasmanthieae and Zeugiteae, which are C_3, supporting the inferred ancestral state of Panicoideae as C_3. Previous studies on Panicoideae phylogeny, mostly using plastome genes, reconstructed different relationships among some tribes, especially basal tribes (e.g., the position of Tristachyideae, C_4). Therefore, although our sampling is indeed incomplete at the subtribe level, our well-supported phylogeny provides meaningful information. Additional studies with greater sampling are needed to investigate the previously proposed >20 transitions from C_3 to C_4 (Grass Phylogeny Working Group II, 2012) and to resolve relationships among some Paniceae subtribes. Moreover, C_4 photosynthesis is a complex trait that involves changes in both leaf anatomy and biochemical processes (Grass Phylogeny Working Group II, 2012; Washburn et al., 2015); thus, even closely related C_4 species may have experienced distinct evolutionary histories for C_4 photosynthesis, as noted previously (Sinha and Kellogg, 1996; Christin et al., 2010; Dunning et al., 2017; Moreno-Villena et al., 2018) and as supported by the evolutionary analyses of *ppc* homologs in the next section.

Phylogenetic analyses of the *ppc* gene family provide molecular evidence for independent origins of C_4 photosynthesis in grasses

C_4 photosynthetic processes depend on multiple genes that are responsible for biochemical pathways and leaf anatomy and are co-opted for the C_4 functionality (Moreno-Villena et al., 2018). Among these genes, the *ppc* gene that encodes PEPC responsible for the initial fixation of atmospheric CO_2 into organic compounds (Sage, 2004), has been studied in several plant families, including Poaceae, Asteraceae, and Fabaceae (Bläsing et al., 2000; Christin et al., 2007a; Christin and Besnard, 2009; Wang et al., 2016). The *ppc* gene belongs to a gene family that encodes several enzymes involved in photosynthesis and some stress-response processes. Previous studies of the *ppc* family indicated that *ppc* genes for C_4 photosynthesis encode proteins with shared sequence motifs (Bläsing et al., 2000; Christin et al., 2007a; Paulus et al., 2013) and that the C_4 *ppc* genes in Poaceae originated from non-C_4 paralogs in two different *ppc* clades (Christin and Besnard, 2009), sometimes involving possible lateral gene transfer (Christin et al., 2012). Previous phylogenetic analysis of *ppc* gene sequences from several Poaceae species and other Poales (*Eleocharis*, Cyperaceae), other monocots (*Aloe*, Asphodelaceae; *Hydrilla*, Hydrocharitaceae; *Vanilla*, Orchidaceae), and several eudicot families helped to define several clades of grass *ppc* genes (referred to here as subclades): *ppc-aL1a*, *ppc-aL1b*, *ppc-aL2*, *ppc-B1*, *ppc-B2*,

and *ppc-aR* (Christin and Besnard, 2009). However, the origins of these subclades were not clear.

Here, to gain additional insights into the evolution of C_4 photosynthesis in the PACMAD clade, we performed phylogenetic analyses of the Poaceae *ppc* gene family (see section "methods") using putative *ppc* genes from all grass subfamilies except Puelioideae (with only low-coverage genome sequences) and from nine of 15 non-grass Poales families (Figure 7A), representing all major Poales clades. Representatives of Musaceae (Zingiberales) and Asparagaceae (Asparagales) were included as outgroups. Our phylogenetic results support a model in which the grass *ppc* subclades originated first with a duplication shared by the MRCA of both Poales and Zingiberales (Figure 7A, indicated by one of the stars) after divergence from Asparagales. Subsequently, another duplication early in the history of Poales, probably after the separation of Typhaceae, generated the common ancestor of the *ppc-aL1a* and *ppc-aL1b* subclades and the ancestor of the *ppc-aL2* subclade, and a later duplication, probably at the MRCA of Poaceae, produced the *ppc-aL1a* and *ppc-aL1b* subclades (Figure 7A, Supplemental Figure 12). Although the origins of the *ppc-B1*, *ppc-B2*, and *ppc-aR* subclades are less clear, they seem to result from duplications of an ancestral gene shared by Poales and Zingiberales (Musaceae) after the divergence of most families in Poales (Figure 7A). However, the placements of a *ppc-B1*-like gene from Flagellariaceae, a family closely related to Poaceae, and genes from Anomochlooideae, the grass subfamily that is sister to all other Poaceae, indicate that further analysis is needed with genes from more representatives of families closely related to Poaceae in order to resolve the early histories of the *ppc-B1*, *ppc-B2*, and *ppc-aR* subclades.

Previous comparative analyses of PEPC amino acid sequences from Poaceae and other families for C_4 photosynthesis or non-C_4 functions revealed characteristic residues at multiple positions (Christin et al., 2007a; Christin and Besnard, 2009) (Supplemental Table 9). The available sequences from many Poaceae members provide an opportunity to further examine the conservation of these residues. Our comparison of over 500 PEPC sequences indicated that characteristic residues for either putative C_4 or non-C_4 enzymes were very similar to those reported previously (Supplemental Table 9; supplemental methods). It has also been reported that the *ppc* gene for C_4 photosynthesis is expressed at higher levels in some C_4 plants (Moreno-Villena et al., 2018). To investigate whether the putative C_4 *ppc* genes identified here were also more highly expressed, we examined their expression levels by mapping RNA sequencing (RNA-seq) reads to the mRNA sequences of different *ppc* genes. Our results suggested that, for some species, the putative C_4 *ppc* gene was probably expressed at a higher level than other *ppc* genes in the same species, such as *Centropodia glauca*, *Neyraudia reynaudiana*, *Eriachne aristidea*, *Loudetiopsis kerstingii*, *Echinochloa esculenta*, and *Hopia obtusa*. However, in several other species, the putative C_4 *ppc* genes appeared not to be the most highly expressed *ppc* genes (Supplemental Table 10; Supplemental Methods). It is possible that the transcriptomes of different species contain different amounts of photosynthetic organs/tissues, and more detailed information about *ppc* gene expression is needed to understand the expression patterns of C_4 and non-C_4 *ppc* genes.

The *ppc* gene phylogenetic analysis here also indicates that the putative *ppc* genes with the characteristic motif for C_4 photosynthesis belong to one of three subclades: *ppc-aL1a*, *ppc-aL1b*, and *ppc-B2* (Figure 7B and 7C), a result that strongly supports the hypothesis of multiple C_4 origins in Poaceae. Specifically, among the three genera in Aristidoideae, both *Aristida* and *Stipagrostis* contain numerous C_4 species, whereas *Sartidia* contains only C_3 species. Here, we identified *ppc* genes from members of each of these three genera (Figure 7 and Supplemental Figure 12), including non-C_4 *ppc* genes from *Aristida* (*Aristida adscensionis* and *Aristida rhiniochloa*; Figure 7B), *Sartidia angolensis*, and *Stipagrostis* (*Stipagrostis acutiflora* and *Stipagrostis plumosa*; Figure 7C) and C_4 *ppc* genes from *Aristida* and *Stipagrostis* species (Figure 7B and 7C). It was previously reported that the C_4 genes of *Aristida* (*A. adscensionis* and *A. rhiniochloa*) belonged to the *ppc-B2* subclade, whereas the C_4 genes of *Stipagrostis pennata* belonged to the *ppc-aL1a* clade (Christin and Besnard, 2009), indicating that *ppc* genes for C_4 photosynthesis probably originated at least twice in Aristidoideae. Our results suggest that the evolution of C_4 *ppc* genes in Aristidoideae may be more complex; in addition to confirming the previous findings (Figure 7B and 7C), our analyses showed that C_4 *ppc* genes from two *Stipagrostis* species not sampled previously (*Stipagrostis hirtigluma* and *Stipagrostis uniplumis*) and a third *Aristida* species (*Aristida depressa*) belonged to the *ppc-B2* subclade (Figure 7C).

Eriachne is the only C_4 lineage in Micraroideae, and it contains five identified *ppc* sequences predicted to function in C_4 photosynthesis. One of these is placed in *ppc-aL1a* and is related to C_3 *ppc* genes from several other PACMAD subfamilies, suggesting an independent and previously unknown origin of C_4 *ppc* (Figure 7B). Four other *ppc* sequences from *Eriachne* are placed in *ppc-B2*, and they are all closely related to sequences from two *Echinochloa* species. This relationship is further supported by a BLAST search showing that the sequences most similar to sequences of these four *Eriachne ppc-B2* sequences are from *Echinochloa*, a C_4 species in the tribe Paniceae of Panicoideae. Furthermore, the C_4 *ppc* genes from both *Echinochloa* and *Eriachne* are close to C_4 genes from other members of Panicoideae. Therefore, the possibility of lateral gene transfer between *Echinochloa* and *Eriachne* cannot be excluded and deserves further study.

Panicoideae and Chloridoideae are the two largest PACMAD subfamilies and contain the majority of C_4 species in Poaceae, although only a subset was included in the *ppc* gene family analysis here. In Panicoideae, all C_4 *ppc* genes identified here (from *Arundinella*, *Axonopus*, *Digitaria*, *Hopia*, *Loudetiopsis*, *Zea mays*, and *Echinochloa*) are in the *ppc-B2* subclade (Figure 7C). As mentioned in the previous section, our ancestral character analyses identified two possible C_4 to C_3 reversions (or retentions of the ancestral C_3 state) in the Panicoideae members *I. pallens* and *S. indica* (Figure 6). No C_4-type *ppc* sequences were found in the transcriptomes of these two species, further supporting their C_3 state. In Chloridoideae, the *ppc* gene family analysis showed that the C_4 *ppc* genes in several Chloridoideae genera (*Bouteloua*, *Centropodia*, *Dignathia*, *Enneapogon*, *Neyraudia*, *Muhlenbergia*, *Sohnsia*, and *Zoysia*) are in the *ppc-B2* subclade. On the other hand, C_4 *ppc* genes in three closely

Figure 7. Molecular phylogenetic analyses of the *ppc* gene family.

(A) A simplified presentation of the *ppc* gene family from a maximum-likelihood analysis using 516 nucleotide sequences of protein coding regions. The *ppc* gene clades (*ppc-aL2*, *ppc-aL1a*, *ppc-aL1b*, *ppc-aR*, and the combined *ppc-B1* + *ppc-B2*) are delimited by tree topology and reference sequences from *C. patens*. Gene clades are collapsed to the family level, Poaceae clades are shown in different colors, and clades of other families are in black. The nodes representing the common ancestors of Poales are marked by arrows. Bootstrap values are shown above branches. A detailed phylogeny is shown in Supplemental Figure 12.

(B) An illustration of subclades *ppc-aL1a* and *ppc-aL1b* in the *ppc* tree shown in **(A)** and Supplemental Figure 12. Monophyletic gene clades from the same subfamilies are collapsed. Names of sequences with a putative C_4 function (determined based on the presence of a conserved serine residue corresponding to Ser780 of the *Z. mays* C_4 PEPC) are marked in red, and non-C_4 sequences are marked in black. Bootstrap values are shown above branches.

(C) Summary of a maximum-likelihood gene tree based on 119 nucleotide coding sequences from Poaceae *ppc-B1* and *ppc-B2* only, plus four outgroup sequences. Colors of C_4/non-C_4 genes are the same as in **(B)**. Bootstrap values are shown above branches. Subfamily names are marked to the right of gene clades, as is the tribe name Tristachyideae (Panicoideae). A detailed phylogeny of the *ppc-B1* and *ppc-B2* clades is shown in Supplemental Figure 13.

related genera (*Blepharidachne*, *Dasyochloa*, and *Erioneuron*, all in the subtribe Scleropogoninae of the tribe Cynodonteae; Figure 4) are in the *ppc-aL1b* subclade (Figure 7B), supporting a different origin. The *ppc* gene phylogeny here is different from the species phylogeny, probably caused by the convergent evolution of C_4 *ppc* genes included here, as proposed by Christin et al. (2007a). In addition, lateral gene transfer was also hypothesized for *ppc* in *Alloteropsis* species in Panicoideae (Christin et al., 2012) and is a possible explanation for the *E. aristidea* and *Echinochloa* C_4 *ppc* genes.

Conclusions and implications

A well-resolved Poaceae nuclear phylogeny supporting monophyly of most subfamilies and tribes

We present a generally well-resolved Poaceae phylogeny that supports the monophyly of 11 out of 12 subfamilies and most of the tribes with two or more sampled taxa, mostly consistent with recent classifications (Kellogg, 2015; Soreng et al., 2015, 2017). The classification on the basis of plastid-gene phylogeny is generally stable and agrees with the nuclear phylogeny here. In addition, this nuclear phylogeny provides better resolution among subfamilies and also for many tribes and some subtribes. Specifically, the deep relationships in the PACMAD clade have long been difficult to resolve, and various topologies have been estimated using chloroplast and mitochondrial genes or a small number of nuclear genes (Grass Phylogeny Working Group II, 2012; Christin et al., 2014; Soreng et al., 2015; Soreng et al., 2017; Saarela et al., 2018). There have been conflicts between results from chloroplast and mitochondrial genes (Cotton et al., 2015) and between different sets of genes (Saarela et al., 2018). Although previous studies placed (Chloridoideae + Danthonioideae) as sister to (Arundinoideae + Micrairoideae) (Soreng et al., 2017; Saarela et al., 2018), either Aristidoideae or Panicoideae was the first divergent lineage among the PACMAD subfamilies (Supplemental Figure 6). The branches subtending the individual PACMAD subfamilies are usually short, suggesting rapid diversification among these subfamilies. Our analyses from both coalescent and super-matrix approaches estimated Aristidoideae as sister to the remaining PACMAD subfamilies and Arundinoideae as sister to (Danthonioideae + Chloridoideae) (Figures 3 and 4, and Supplemental Figures 3–6). Micrairoideae is supported in most coalescent analyses as sister to Panicoideae, whereas the signal for its placement as sister to (Arundinoideae + (Danthonioideae + Chloridoideae)) (Supplemental Figure 4) may be due to paralogous sequences, possibly generated by ancient polyploidization events.

Phylogenetic analysis of the ppc *family provides insights into evolution of C_4 photosynthesis*

Six subclades for grass *ppc* genes (*ppc-aL1a*, *ppc-aL1b*, *ppc-aL2*, *ppc-B1*, *ppc-B2*, and *ppc-aR*) were defined by previous molecular phylogenetic analysis using sequences from several Poaceae species, other Poales (*Eleocharis*, Cyperaceae), other monocots (*Aloe*, Asphodelaceae; *Hydrilla*, Hydrocharitaceae; *Vanilla*, Orchidaceae), and eudicots (Christin et al., 2007a; Christin and Besnard, 2009). However, the origins of these subclades have not been clear. Also, other distant *ppc* paralogs exist but are not closely related to genes known to function in photosynthesis (Moreno-Villena et al., 2018) and not analyzed here. The analysis here included a broad sampling of *ppc* homologs from Poaceae and 10 other Poales families, as well as Musaceae (Zingiberales) and Asparagaceae (Asparagales), providing a better understanding of the early histories of the grass *ppc* genes. The results indicate that the six grass *ppc* subclades belong to two ancient clades, *ppc-aL* and *ppc-aR/B* (each with three subclades), and both probably originated in the MRCA of Poales and Zingiberales (Figure 7A). The duplication of the ancestral *ppc-aL* gene in early Poales generated the *ppc-aL1* and *ppc-aL2* clades, and a subsequent duplication of *ppc-aL1* in early Poaceae produced the *ppc-aL1a* and *ppc-aL1b* subclades. However, the evolution of *ppc-aR/B* genes to *ppc-aR*, *ppc-B1*, and *ppc-B2* subclades is less clear, although one possible scenario is that a duplication in the MRCA of Poaceae generated the *ppc-aR* and *ppc-(B1+B2)* clades.

The putative C_4 *ppc* genes were identified based on a conserved serine residue (corresponding to residue 780 in the *Z. mays* PEPC, GRMZM2G083841) and belong mostly to the *ppc-(B1+B2)* clade, with a few in the *ppc-aL1a* and *ppc-aL1b* subclades (Figure 7 and Supplemental Figure 12). Previously, the *ppc-(B1+B2)* genes formed two subclades (Christin et al., 2007a; Christin and Besnard, 2009). Here, a phylogenetic analysis of the *ppc-(B1+B2)* genes also yielded two highly supported clades, *ppc-B1* and *ppc-B2* (Figure 7C and Supplemental Figure 13), containing known *ppc-B1* and *ppc-B2* genes, respectively (Christin et al., 2007a; Christin and Besnard, 2009). To avoid possible effects of natural selection on gene phylogeny, another analysis using the nucleotide residues at the third codon positions was performed. Although the detailed phylogenetic relationships among gene sequences are somewhat different, both *ppc-B1* and *ppc-B2* clades were recovered, and C_4 sequences were clustered in the *ppc-B2* clade (Supplemental Figure 14; supplemental methods). The *ppc-B1* clade contains genes from the early-divergent Poaceae subfamilies Anomochlooideae and Pharoideae and from both the BOP and PACMAD clades. The *ppc-B2* subclade includes most of the putative C_4 *ppc* genes, as well as non-C_4 *ppc-B2* homologs from several BOP and PACMAD subfamilies, but not from Oryzoideae and the early-divergent subfamilies. Therefore, *ppc-B1* and *ppc-B2* subclades probably resulted from a duplication in the MRCA of Poaceae, but *ppc-B2* genes were lost from (or not expressed in) members of several subfamilies sampled here, all containing C_3 plants.

In the *ppc-B2* clade, most putative C_4 genes are clustered into one large clade, with the exception of C_4 genes from the early-divergent Centropodieae of Chloridoideae (Figure 7C and Supplemental Figure 12), suggesting that the Centropodieae C_4 genes had a separate origin from the other C_4 genes in the *ppc-B2* clade. A putative origin of C_4 photosynthesis in *Centropodia* distinct from other Chloridoideae is also supported by ancestral character reconstruction (Figure 6). Most of the other C_4 *ppc-B2* genes form a clade with 86% BS support, suggesting that they may have a single origin; however, their relationships do not agree with the species relationships, as noted previously (Christin et al., 2007a). The relationships among the subfamilies in the PACMAD clade were difficult to resolve, even using multiple genes (Prasad et al., 2011; Soreng et al., 2017). Therefore, it is not surprising that the C_4 *ppc-B2* genes do not follow the species relationships. Nevertheless, C_4 *ppc-B2* genes of two Aristidoideae genera (*Aristida* and *Stipagrostis*) were placed close to genes from Panicoideae

(maize and sorghum) and Chloridoideae species (in Cynodonteae and three other tribes), respectively, suggesting that the *Aristida* and *Stipagrostis* C_4 genes may not have the same origin in the subfamily.

Previously, an *S. pennata* C_4 gene in the *ppc-aL1b* subclade (Christin and Besnard, 2009) (Figure 7B) supported a different origin from that of C_4 *ppc-B2* genes in *S. hirtigluma* and *S. uniplumis* (Figure 7C). We also identified C_4 *ppc-aL1b* genes from three other species (*Blepharidachne kingii*, *Dasyochloa pulchella*, and *Erioneuron pilosum*) from the same subtribe, Scleropogoninae, in the tribe Cynodonteae of Chloridoideae (Figure 7B and Supplemental Figure 12), suggesting a shared C_4 origin in the ancestor of the subtribe Scleropogoninae. The close relationship of these genes to that of *S. pennata* in Aristidoideae might be explained by horizontal transfer between respective members of Aristidoideae and Cyndonteae (Chloridoideae). Moreover, a putative C_4 gene was identified in the *ppc-aL1a* subclade from *E. aristidea* of Micraroideae (Figure 7B); this species also has C_4 *ppc-B2* gene(s) related to those from Panicoideae species, suggesting that the C_4 *ppc-B2* genes may have experienced horizontal transfer from a Panicoideae taxa to *E. aristidea*, as proposed previously among Paniceae members (Christin et al., 2012).

The phylogenetic analyses of *ppc* homologs expanded the coverage of subfamilies compared with previous studies to include all subfamilies, except Puelioideae, and identified more putative grass C_4 genes in the *ppc-B2* and *ppc-aL1b* subclades. In addition, the analyses here uncovered a new C_4 gene in the *ppc-aL1a* subclade. The results support at least three origins of C_4 genes in Chloridoideae (two in *ppc-B2* and one in *ppc-aL1b*), at least three origins in Aristidoideae (two in *ppc-B2* and one in *ppc-aL1b*), at least two origins in Micraroideae (one in *ppc-B2* and one in *ppc-aL1a*), and multiple origins in Panicoideae. These findings indicate not only that there were multiple origins of C_4 *ppc* but also that members of at least three *ppc* subclades were recruited. The clades containing most C_4 species in both Panicoideae (with large tribes Andropogoneae and Paniceae) and Chloridoideae (with the largest tribe Cynodonteae) (Figure 6) originated during the early to middle Eocene (Figure 5 and Supplemental Figure 9). As the Earth's temperature was relatively high during this period, the evolution of C_4 photosynthesis may have promoted adaptation to warm environments and contributed to the diversification of Andropogoneae/Paniceae and Cynodonteae in the two largest PACMAD subfamilies.

METHODS

Taxon sampling, RNA/DNA isolation, and high-throughput sequencing

Our taxon sampling aimed to represent Poaceae with all subfamilies and as many tribes as possible. For large tribes (for example, Andropogoneae in Panicoideae, Cynodonteae in Chloridoideae, and Poeae in Pooideae), we also tried to include members of many subtribes. We sampled a total of 357 Poaceae species, representing 45 of 52 tribes in Poaceae. In addition, we sampled 13 outgroup species, including one species each from Ecdeiocoleaceae (*Ecdeiocolea monostachya*) and Joinvilleaceae (*Joinvillea ascendens*), which form a sister clade to Poaceae. Also sampled were members of other Poales families, including Flagellariaceae, Restionaceae, Eriocaulaceae, Cyperaceae, Juncaceae, and Typhaceae, as well as members of three other orders close to Poales: Arecales (Arecaceae), Zingiberales (Zingiberaceae), and Commelinales (Commelinaceae). Information on taxa included in this study is listed in Supplemental Table 1.

Total RNA/DNA was isolated from samples of leaves, stems, inflorescences, or young fruits using the NucleoSpin RNA Plant RNA isolation kit (MACHEREY-NAGEL, REF 740949.50) or by a standard EDTA protocol. The RNA/DNA samples were used for library construction and sequencing by either the Penn State core facility or commercial sequencing companies. The Illumina platform was used to construct sequencing libraries and perform paired-end sequencing to obtain 150-bp reads. The procedure generally included the following steps: (1) total RNA was extracted from fresh or frozen plant tissues, then treated with DNase to remove DNA; (2) mRNAs were captured by purification using a column with oligo (dT); (3) mRNAs were used as templates to synthesize first-strand cDNA using random hexamer primers; (4) second-strand cDNAs were synthesized and purified, their 5′ ends repaired and 3′ ends adenylated, and finally ligated to adaptors; (5) cDNAs were amplified by PCR. Paired-end transcriptome sequencing (2 × 150 bp) was performed by GENERGY BIO using the Illumina HiSeq 3000 platform. Publicly available transcriptomes/genomes/SRA data were retrieved from NCBI databases (https://www.ncbi.nlm.nih.gov/) and EMBL-EBI (http://www.ebi.ac.uk/). See Supplemental Table 1 for the sources of samples.

Data availability

The sequence data generated for this study have been deposited at public databases, and accession numbers are listed in Supplemental Table 1.

Raw sequence processing and assembly

The procedures for transcriptomic and genomic sequence processing and assembly are illustrated in Supplemental Figure 1. For transcriptomic data, paired-end sequencing data sets were first trimmed with Trimmomatic (Trinity 2.2.0 plug-in) (Grabherr et al., 2011) using default settings. FastQC (0.11.8) (Andrews, 2010) quality checks were performed after trimming to confirm the removal of adapters and low-quality regions. Transcriptome assembly was performed using Trinity (V 2.2.0) (Grabherr et al., 2011) with default parameters on the Penn State ACI server. Deduplication of assembled contigs was performed with CD-HIT-EST (V 4.6.8) (Fu et al., 2012) using the parameter -c 0.98. Coding sequences were extracted from deduplicated contigs with TransDecoder (V 5.3.0) (https://github.com/TransDecoder/TransDecoder/wiki). For shotgun genome sequencing data, Trimmomatic was also implemented to remove sequencing adaptors and low-quality regions. KmerGenie (1.7048) (Chikhi and Medvedev, 2014) was used to optimize the k-mer value in the subsequent assembly process. Optimized K values were set for the assembly of genomic data sets by SOAPdenovo2 (2.04-r240) (Luo et al., 2012). Assembled genomic contig data sets were deduplicated with CD-HIT-EST (V 4.6.8). The genomic data generated by shotgun genome sequencing were relatively sparse, and some target coding sequences may have been partial (that is, missing some regions) and hence may not have been retained by TransDecoder. Therefore, to obtain more sequence for subsequent analyses, the assembled genomic contigs were not processed by TransDecoder to generate cds data sets. For public genomes/transcriptomes, non-redundant coding sequences were retrieved directly from the NCBI database. SRA data sets were processed in the same way as the transcriptome data sets generated in our own project. Statistics on the non-redundant coding sequence data sets and genomic contig data sets were calculated with statswrapper.sh (a BBMap tool, V 38.33) to check assembly quality and are provided in Supplemental Table 1.

Identification and retrieval of low-copy orthologous nuclear genes

We selected genome/transcriptome sequences of 10 Poaceae species (*B. distachyon*, *Eleusine coracana*, *Hordeum vulgare*, *Lygeum spartum*, *Oryza sativa*, *Phaenosperma globosum*, *Phyllostachys heterocycla*, *S. italica*, *Sorghum bicolor*, *Stipa aliena*) that represent the five largest

subfamilies but do not include the recent polyploids wheat and maize, with additional criteria related to data quality, to identify putative low-copy (one or two copies per species) nuclear genes across Poaceae. The putative orthologous genes were identified using OrthoMCL v1.4 (Li et al., 2003) with the parameters perl orthomcl.pl –mode 3 –blast_file 10sps.blastresult –gg_file 10sps.gg, where –mode 3 instructs OrthoMCL to perform the analysis using a user-provided BLAST output file (10sps.blastresult) and a genome gene relation file (10sps.gg). Other parameters were set to default settings. The HMM files of the 1234 identified OGs were used as the seeds for HaMStR (13.2.6) (Ebersberger et al., 2009) to search for and retrieve corresponding orthologous sequences from the assembled contig datasets from transcriptome and genome sequencing. Cutoff e-values for blast and hmm search were both set to 1e−20. Only one sequence was retained per dataset for each seed, and sometimes fragments matching non-overlapping parts of the seeds were combined to represent the whole sequence. The number of orthologous sequences retrieved for each genome contig dataset (sampled in this project) ranged from 252 to 1018, and the number of orthologous sequences retrieved for each cds dataset (all others except for the seven genome skimming datasets) ranged from 235 to 1234.

Sequence alignment and reconstruction of single-gene trees

Orthologous sequences retrieved by HaMStR were sorted by sequence ID (orthologous group ID), then reorganized and formatted into fasta files. Nucleotide sequences (cds) of each OG were aligned with MAFFT (v7.397) (Katoh et al., 2009) using the –auto option. Alignments were then trimmed with trimAl (1.4.1) (Capella-Gutiérrez et al., 2009) using the -automated1 option to remove poorly aligned regions and/or sequences. Single-gene ML trees based on the alignments of 1234 OGs were reconstructed using RAxML (8.2.1) (Stamatakis, 2014) with rapid bootstrapping of 100 replicates and the GTRCAT model.

Detection of sequences prone to long-branch attraction

To identify and remove genes that are prone to exhibiting long-branch attraction, TreSpEx (1.1) (Struck, 2014) was applied to all the 1234 single-gene alignments, together with the single-gene ML trees corresponding to orthologous genes, to analyze long-branch attraction (determined by heterogeneity or longest branches) and saturation (determined by the slope or R^2 of linear regression). The probability density function curves of these four indicators were plotted in R (Supplemental Figure 2). Genes that deviated from a normal distribution for each of the four indicators were removed. The numbers of genes removed based on heterogeneity or longest branches were 389 and 393, respectively, and 555 and 96 genes were removed based on the slope or R^2 of the linear regression. After deletion of the genes from these four sets, 571 genes out of the 1150-gene set were retained and were further filtered for super-matrix and molecular clock analysis.

Phylogenetic analyses using the astral coalescent method or a super-matrix dataset and AU test

For the coalescent analysis dataset with 378 samples, the number of samples with a positive detection for each gene ranged from 23 to 377. The numbers of genes retrieved by HaMStR searches of the six shotgun genomic contig datasets were relatively low (the lowest being 252), and these six genomic datasets represent the two genera in the basal subfamily Puelioideae (*Guaduella* and *Puelia*). We therefore filtered the set of genes to make sure that each gene was present in at least one species from each of these two genera. The remaining 1150 genes were further filtered by coverage and alignment length to generate smaller sets. We obtained six sets of genes for the coalescent analyses, containing 1150, 895, 775, 570, 436, and 180 genes. See Supplemental Figure 1 for the gene selection procedure.

Astral 5.6.3 (Sayyari and Mirarab, 2016) was used to infer multi-gene coalescent trees from different sets of single-gene trees. The local posterior probability value was chosen, as it was recommended by the authors of the program. It reflects the probability that a branch is present in a species tree, and it is not comparable to a multi-gene bootstrap value. The coalescent trees were edited with Dendroscope (V 3.6.2) (Huson and Scornavacca, 2012) and summarized using TreeGraph 2 (2.14.0) (Stöver and Müller, 2010). The 180-gene set was also used to generate a super-matrix dataset with a length of 184 993 and a total of 71 037 312 matrix cells. The percentage of missing sites was 10.977%, and the proportion of variable sites was 0.798. An AU test was performed using CONSEL v0.20 (Shimodaira and Hasegawa, 2001; see also http://cran.r-project.org/web/packages/scaleboot/index.html) on the 180-gene super-matrix with sequences from 384 species (Supplemental Figure 5).

Ancestral state reconstruction

The reconstruction of the ancestral state of photosynthetic pathway type was performed with Mesquite (3.6) (Massidon and Maddison, 2019). The state of each sampled species was coded as either 0 (C_3) or 1 (C_4) according to information summarized by Soreng et al. (2017), and the ancestral state was inferred by the maximum parsimony method using default parameters in the context of the topology from five coalescent trees (see Supplemental Table 1 for the state code of photosynthetic type).

Fossil calibrations and divergence time estimation

Thirteen fossil calibrations were used in our analyses, including phytolith data for Poaceae (Wu et al., 2018) (Supplemental Table 6), which provide informative calibration points and support older ages than those estimated using only the relatively scarce macrofossils (Christin et al., 2014; Kellogg, 2015). The phytoliths (silica bodies) from grasses are regarded as distinct from those of other families in Poales and can be assigned to subfamilies of Poaceae (Magallón et al., 2015). Taxonomic assignment and age of the fossils were designated according to the references cited in Supplemental Table 6. In our analyses, all the fossil calibrations were implemented as minimum constraints, except for the crown age of Commelinids, which was set to be no older than 118 my.

Given the large amount of sequence data from over 380 taxa, we used the PL method implemented in treePL (1.0) (Smith and O'Meara, 2012) to estimate the divergence time. The ML tree reconstructed by RAxML (8.2.1) from the smallest set of 180 genes with branch length information was used as the input tree to avoid systematic errors that can be associated with super-matrix datasets of hundreds of genes (Philippe et al., 2011). This tree was generated using the 180-gene set with the topology of the 1150-gene coalescent tree (also supported by most analyses) as a constraint. Parameter optimization and cross-validation were performed to select the best smoothing value, along with other parameters. A smoothing value of 0.1 was selected, which is low and indicates a large deviation from the strict molecular clock hypothesis (Huang et al., 2016). One hundred BS replicates with branch length information for the 180-gene ML tree were also generated by RAxML (8.2.1) to calculate the CIs of node ages (Supplemental Figure 9).

Analysis of the *ppc* gene family

The sampling for *ppc* gene family analysis was designed to represent all the subfamilies in the PACMAD clade and to cover most C_4 lineages. In addition, species from other subfamilies of Poaceae were included to cover major tribes, excluding the subfamily Puelioideae, for which only genome skimming data were available and from which *ppc* homologs could not be reliably retrieved by blast searches. Fifteen species from nine other families of Poales, as well as Musaceae (Zingiberales) and Asparagaceae (Asparagales), were also included as outgroups. A total of 107 samples were included for the *ppc* gene family analysis. Amino acid sequences representing the six *ppc* lineages from *Cyrtococcum patens* (Panicoideae) were used as queries to perform tblastn searches against the coding sequence datasets of the selected species. The six reference coding sequences from *C. patens* and some other species from public data sets were also included (see Supplemental Table 8). Duplicate copies with identical sequences from the same samples were removed,

and coding sequences significantly shorter than the others were also removed manually, but critical C_4-type *ppc* sequences that belonged to species in critical phylogenetic positions were retained. A total of 516 *ppc* sequences were retained. The *ppc* sequences were translated into amino acid sequences and aligned using ClustalO (1.2.4) (Sievers et al., 2011). An alignment of nucleotide sequences was then generated based on the corresponding amino acid alignment and was trimmed using trimal (1.4.1) (Capella-Gutiérrez et al., 2009) to remove poorly aligned regions. ML analysis was performed on the trimmed alignment using IQ-TREE (1.6.12) (Nguyen et al., 2015) to reconstruct gene family trees. C_4 *ppc* genes were distinguished by the presence of serine at position 780 (following the numbering of *Z. mays* C_4 *ppc*, GRMZM2G083841) in the corresponding amino acid sequences (Bläsing et al., 2000). Sequences with other amino acids at this residue in the alignment were not treated as C_4 *ppc* genes, as such sequences have not been shown experimentally to function as C_4 *ppc* genes.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at *Molecular Plant Online*.https://doi.org/10.1016/j.molp.2022.01.015

FUNDING

This work was supported by funds from Eberly College of Science and the Huck Institutes of the Life Sciences at the Pennsylvania State University and from grants from the National Natural Science Foundation of China (31770242 and 31970224) and funds from the Ministry of Education Key Laboratory of Biodiversity Science and Ecological Engineering and State Key Laboratory of Genetic Engineering at Fudan University.

AUTHOR CONTRIBUTIONS

H.M., J.T.C., S.G., and D.-Z.L. designed the project. Y.Z., L.Z., Y.H., J.T.C., H.M., Z.G., W.C., W.H., M.M., and M.B. sampled plant materials. Y.H., L.Z., Y.Z., and W.H. performed DNA and RNA isolation. W.H. and L.T. performed sequence data processing. W.H. and L.Z. performed seed gene sequence selection. W.H. performed phylogenetic, molecular clock, ancestral character reconstruction, and *ppc* gene family analyses with assistance and discussion from L.Z., Y.Z., and C.-H.H. W.H. and H.M. interpreted results. W.H. and H.M. wrote manuscript drafts. H.M., S.G., D.-Z.L., Z.G., W.C., and M.B. edited the manuscript. H.M. and C.-H.H. provided funds.

ACKNOWLEDGMENTS

We thank Lulu Xun, Qingwen Lin, Yang Yang, Yuxing Gu, Fei Ren, Weijie Li, Zhiming Zhong, Detuan Liu, Hailei Zheng, Shuai Liao, Binghua Chen, Liqiong Jiang, Guangda Tang, Xinxin Zhu, Holly Forbes, Clare Loughran, Peter Brownless, Suzanne Cubey, Lynn Clark, Andrew Wyatt, Jocelyn Frazelle, Pablo Bolaños-Villegas, and Sergio Castro-Pacheco for assistance in plant sampling. We thank Dr. Elizabeth A. Kellogg for encouragement and insightful comments and discussions on earlier drafts of the manuscript, and Drs. Ji Yang, Zhiping Song, Wenju Zhang, Ji Qi, Yuguo Wang, and Taikui Zhang for valuable discussions. We thank the Royal Botanic Gardens, Kew; Royal Botanic Garden, Edinburgh; UC Botanical Garden at Berkeley; Missouri Botanical Garden; New York Botanical Garden; and the Germplasm Bank of Wild Species, Kunming Institute of Botany for providing materials. No conflict of interest is declared.

Received: December 8, 2020
Revised: June 9, 2021
Accepted: January 24, 2022
Published: January 27, 2022

REFERENCES

Aggarwal, R.K., Brar, D.S., Nandi, S., Huang, N., and Khush, G.S. (1999). Phylogenetic relationships among *Oryza* species revealed by AFLP markers. Theor. Appl. Genet. **98**:1320–1328.

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Bianconi, M.E., Hackel, J., Vorontsova, M.S., Alberti, A., Arthan, W., Burke, S.V., Duvall, M.R., Kellogg, E.A., Lavergne, S., McKain, M.R., et al. (2020). Continued adaptation of C4 photosynthesis after an initial burst of changes in the Andropogoneae grasses. Syst. Biol. **69**:445–461.

Bläsing, O.E., Westhoff, P., and Svensson, P. (2000). Evolution of C4 phosphoenolpyruvate carboxylase in *Flaveria*, a conserved serine residue in the carboxyl-terminal part of the enzyme is a major determinant for C4-specific characteristics. J. Biol. Chem. **275**:27917–27923.

Bouchenak-Khelladi, Y., Salamin, N., Savolainen, V., Forest, F., van der Bank, M., Chase, M.W., and 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.

Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics **25**:1972–1973.

Cerros-Tlatilpa, R., and Columbus, J.T. (2009). C3 photosynthesis in *Aristida longifolia*: implication for photosynthetic diversification in Aristidoideae (Poaceae). Am. J. Bot. **96**:1379–1387.

Chemisquy, M.A., Giussani, L.M., Scataglini, M.A., Kellogg, E.A., and Morrone, O. (2010). Phylogenetic studies favour the unification of *Pennisetum*, *Cenchrus* and *Odontelytrum* (Poaceae): a combined nuclear, plastid and morphological analysis, and nomenclatural combinations in *Cenchrus*. Ann. Bot. **106**:107–130.

Chikhi, R., and Medvedev, P. (2014). Informed and automated k-mer size selection for genome assembly. Bioinformatics **30**:31–37.

Christenhusz, M.J.M., and Byng, J.W. (2016). The number of known plants species in the world and its annual increase. Phytotaxa **261**:201–217.

Christin, P.A., and Besnard, G. (2009). Two independent C4 origins in Aristidoideae (Poaceae) revealed by the recruitment of distinct phosphoenolpyruvate carboxylase genes. Am. J. Bot. **96**:2234–2239.

Christin, P.A., Salamin, N., Savolainen, V., Duvall, M.R., and Besnard, G. (2007a). C4 photosynthesis evolved in grasses via parallel adaptive genetic changes. Curr. Biol. **17**:1241–1247.

Christin, P.A., Salamin, N., Savolainen, V., and Besnard, G. (2007b). A phylogenetic study of the phosphoenolpyruvate carboxylase multigene family in Poaceae: understanding the molecular changes linked to C4 photosynthesis evolution. Kew Bull. **62**:455–462.

Christin, P.A., Freckleton, R.P., and Osborne, C.P. (2010). Can phylogenetics identify C4 origins and reversals? Trends Ecol. Evol. **25**:403–409.

Christin, P.A., Edwards, E.J., Besnard, G., Boxall, S.F., Gregory, R., Kellogg, E.A., Hartwell, J., and Osborne, C.P. (2012). Adaptive evolution of C4 photosynthesis through recurrent lateral gene transfer. Curr. Biol. **22**:445–449.

Christin, P.A., Spriggs, E., Osborne, C.P., Strömberg, C.A.E., Salamin, N., and Edwards, E.J. (2014). Molecular dating, evolutionary rates, and the age of the grasses. Syst. Biol. **63**:153–165.

Clark, L.G., and Judziewicz, E.J. (1996). The grass subfamilies Anomochlooideae and Pharoideae (Poaceae). Taxon **45**:641–645.

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

Clark, L.G., Kobayashi, M., Mathews, S., Spangler, R.E., and Kellogg, E.A. (2000). The Puelioideae, a new subfamily of Poaceae. Syst. Bot. **25**:181–187.

Clayton, W.D., and Renvoize, S.A. (1986). Genera Graminum. Grasses of the World (H.M. Stationery Office).

Clayton, W.D., Vorontsova, M.S., Harman, K.T., and Williamson, H. (2006). RBG Kew: GrassBase - The Online World Grass Flora. https://www.kew.org/data/grassbase/.

Columbus, T. (1999). An expanded circumscription of *Bouteloua* (Gramineae: Choridoideae): new combinations and names. Aliso **18**:61–65.

Cotton, J.L., Wysocki, W.P., Clark, L.G., Kelchner, S.A., Pires, J.C., Edger, P.P., Mayfield-Jones, D., and Duvall, M.R. (2015). Resolving deep relationships of PACMAD grasses: a phylogenomic approach. BMC Plant Biol. **15**:178.

Dujardin, M. (1978). Chromosome numbers of some tropical African grasses from western Zaïre. Can. J. Bot. **56**:2138–2152.

Dunning, L.T., Lundgren, M.R., Moreno-Villena, J.J., Namaganda, M., Edwards, E.J., Nosil, P., Osborne, C.P., and Christin, P.A. (2017). Introgression and repeated co-option facilitated the recurrent emergence of C_4 photosynthesis among close relatives. Evolution **71**:1541–1555.

Dunning, L.T., Olofsson, J.K., Parisod, C., Choudhury, R.R., Moreno-Villena, J.J., Yang, Y., Dionora, J., Paul Quick, W., Park, M., Bennetzen, J.L., et al. (2019). Lateral transfers of large DNA fragments spread functional genes among grasses. Proc. Natl. Acad. Sci. U S A **116**:4416–4425.

Ebersberger, I., Strauss, S., and Von Haeseler, A. (2009). HaMStR: profile hidden Markov model based search for orthologs in ESTs. BMC Evol. Biol. **9**:157.

Edwards, E.J., and Still, C.J. (2008). Climate, phylogeny and the ecological distribution of C4 grasses. Ecol. Lett. **11**:266–276.

Estep, M.C., McKain, M.R., Diaz, D.V., Zhong, J., Hodge, J.G., Hodkinson, T.R., Layton, D.J., Malcomber, S.T., Pasquet, R., and Kellogg, E.A. (2014). Allopolyploidy, diversification, and the Miocene grassland expansion. Proc. Natl. Acad. Sci. U S A **111**:15149–15154.

Fisher, A.E., Hasenstab, K.M., Bell, H.L., Blaine, E., Ingram, A.L., and Columbus, J.T. (2016). Evolutionary history of chloridoid grasses estimated from 122 nuclear loci. Mol. Phylogenet. Evol. **105**:1–14.

Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics **28**:3150–3152.

Ge, S., Sang, T., Lu, B.R., and Hong, D.Y. (1999). Phylogeny of rice genomes with emphasis on origins of allotetraploid species. Proc. Natl. Acad. Sci. U S A **96**:14400–14405.

Goh, W.L., Chandran, S., Franklin, D.C., Isagi, Y., Koshy, K.C., Sungkaew, S., Yang, H.Q., Xia, N.H., and Wong, K.M. (2013). Multi-gene region phylogenetic analyses suggest reticulate evolution and a clade of Australian origin among paleotropical woody bamboos (Poaceae: Bambusoideae: Bambuseae). Plant Syst. Evol. **299**:239–257.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-seq data without a reference genome. Nat. Biotechnol. **29**:644–652.

Grass Phylogeny Working Group (GPWG), Barker, N.P., Clark, L.G., Davis, J.I., Duvall, M.R., Guala, G.F., Hsiao, C., Kellogg, E.A., Linder, P., Mason-Garner, R.J., et al. (2001). Phylogeny and subfamilial classification of the grasses (Poaceae). Ann. Missouri Bot. Gard. **88**:373–457.

Grass Phylogeny Working Group. (2003). The paraphyly of *Cortaderia* (Danthonioideae; Poaceae): evidence from morphology and chloroplast and nuclear DNA sequence data. Ann. Missouri Bot. Gard. **90**:1–24.

Grass Phylogeny Working Group II (GPWG II). (2012). New grass phylogeny resolves deep evolutionary relationships and discovers C4 origins. New Phytol. **193**:304–312.

Guo, Z.H., Ma, P.F., Yang, G.Q., Hu, J.Y., Liu, Y.L., Xia, E.H., Zhong, M.C., Zhao, L., Sun, G.L., Xu, Y.X., et al. (2019). Genome sequences provide insights into the reticulate origin and unique traits of woody bamboos. Mol. Plant **12**:1353–1365.

Guo, C., Ma, P.F., Yang, G.Q., Ye, X.Y., Guo, Y., Liu, J.X., Liu, Y.L., Eaton, D.A.R., Guo, Z.H., and Li, D.Z. (2021). Parallel ddRAD and genome skimming analyses reveal a radiative and reticulate evolutionary history of the temperate bamboos. Syst. Biol. **70**:756–773.

Huang, C.H., Sun, R., Hu, Y., Zeng, L., Zhang, N., Cai, L., Zhang, Q., Koch, M.A., Al-Shehbaz, I., Edger, P.P., et al. (2016). Resolution of Brassicaceae phylogeny using nuclear genes uncovers nested radiations and supports convergent morphological evolution. Mol. Biol. Evol. **33**:394–412.

Huson, D.H., and Scornavacca, C. (2012). Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. Syst. Biol. **61**:1061–1067.

Ingram, A.L., Christin, P.A., and Osborne, C.P. (2011). Molecular phylogenies disprove a hypothesized C4 reversion in *Eragrostis walteri* (Poaceae). Ann. Bot. **107**:321–325.

Jones, S.S., Burke, S.V., and Duvall, M.R. (2014). Phylogenomics, molecular evolution, and estimated ages of lineages from the deep phylogeny of Poaceae. Plant Syst. Evol. **300**:1421–1436.

Katoh, K., Asimenos, G., and Toh, H. (2009). Multiple alignment of DNA sequences with MAFFT. Methods Mol. Biol. **537**:39–64.

Kellogg, E.A. (2000). The grasses: a case study in macroevolution. Annu. Rev. Ecol. Syst. **31**:217–238.

Kellogg, E.A. (2001). Evolutionary history of the grasses. Plant Physiol. **125**:1198–1205.

Kellogg, E.A. (2009). The evolutionary history of Ehrhartoideae, Oryzeae, and *Oryza*. Rice **2**:1–14.

Kellogg, E.A. (2015). Flowering plants. In Monocots: Poaceae. The Families and Genera of Vascular Plants, *13*, K. Kubitzki, ed. (Springer).

Kumagai, M., Wang, L., and Ueda, S. (2010). Genetic diversity and evolutionary relationships in genus *Oryza* revealed by using highly variable regions of chloroplast DNA. Gene **462**:44–51.

Li, L., Stoeckert, C.J., and Roos, D.S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. **13**:2178–2189.

Li, D.-Z., Yang, H.-M., and Zhang, Y.-X. (2013). Revision of *Arundinaria* and *Neomicrocalamus* (Poaceae: Bambusoideae) of the flora of China. Plant Divers. Resour. **35**:605–612.

Liu, Q., Triplett, J.K., Wen, J., and Peterson, P.M. (2011). Allotetraploid origin and divergence in *Eleusine* (Chloridoideae, Poaceae): evidence from low-copy nuclear gene phylogenies and a plastid gene chronogram. Ann. Bot. **108**:1287–1298.

Lundgren, M.R., Christin, P.A., Escobar, E.G., Ripley, B.S., Besnard, G., Long, C.M., Hattersley, P.W., Ellis, R.P., Leegood, R.C., and Osborne, C.P. (2016). Evolutionary implications of C3–C4 intermediates in the grass *Alloteropsis semialata*. Plant. Cell. Environ. **39**:1874–1885.

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., et al. (2012). SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Gigascience **1**:18.

Magallón, S., Gómez-Acevedo, S., Sánchez-Reyes, L.L., and Hernández-Hernández, T. (2015). A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. New Phytol. **207**:437–453.

Mandel, J.R., Dikow, R.B., Siniscalchi, C.M., Thapa, R., Watson, L.E., and Funk, V.A. (2019). A fully resolved backbone phylogeny reveals numerous dispersals and explosive diversifications throughout the history of Asteraceae. Proc. Natl. Acad. Sci. U S A **116**:14083–14088.

Mason-Gamer, R.J., Burns, M.M., and Naum, M. (2010). Reticulate evolutionary history of a complex group of grasses: phylogeny of *Elymus* StStHH allotetraploids based on three nuclear genes. PLoS One **5**:e10989.

Massidon, W.P., and Maddison, D.R. (2019). Mesquite: A Modular System for Evolutionary Analysis, Version 3.6. http://www.mesquiteproject.org.

Mathews, S., Spangler, R.E., Mason-Gamer, R.J., and Kellogg, E.A. (2002). Phylogeny of Andropogoneae inferred from phytochrome B, GBSSI, and NDHF. Int. J. Plant Sci. **163**:441–450.

Moreno-Villena, J.J., Dunning, L.T., Osborne, C.P., and Christin, P.A. (2018). Highly expressed genes are preferentially co-opted for C4 photosynthesis. Mol. Biol. Evol. **35**:94–106.

Morrone, O., Aagesen, L., Scataglini, M.A., Salariato, D.L., Denham, S.S., Chemisquy, M.A., Sede, S.M., Giussani, L.M., Kellogg, E.A., and Zuloaga, F.O. (2012). Phylogeny of the Paniceae (Poaceae: Panicoideae): integrating plastid DNA sequences and morphology into a new classification. Cladistics **28**:333–356.

Muhaidat, R., Sage, R.F., and Dengler, N.G. (2007). Diversity of Kranz anatomy and biochemistry in C4 eudicots. Am. J. Bot. **94**:362–381.

Muller, J. (1981). Fossil pollen records of extant angiosperms. Bot. Rev. **47**:1–142.

Nguyen, L.T., Schmidt, H.A., Von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. **32**:268–274.

Paulus, J.K., Schlieper, D., and Groth, G. (2013). Greater efficiency of photosynthetic carbon fixation due to single amino-acid substitution. Nat. Commun. **4**:1518.

Perreta, M.G., Ramos, J.C., and Vegetti, A.C. (2009). Development and structure of the grass inflorescence. Bot. Rev. **75**:377–396.

Peterson, P.M., Romaschenko, K., Barker, N.P., and Linder, H.P. (2011). Centropodieae and *Ellisochloa*, a new tribe and genus in Chloridoideae (Poaceae). Taxon **60**:1113–1122.

Peterson, P.M., Romaschenko, K., and Arrieta, Y.H. (2015). Phylogeny and subgeneric classification of *Bouteloua* with a new species, *B. herrera-arrietae* (Poaceae: Chloridoideae: Cynodonteae: Boutelouinae). J. Syst. Evol. **53**:351–366.

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

Prasad, V., Strömberg, C.A.E., Leaché, A.D., Samant, B., Patnaik, R., Tang, L., Mohabey, D.M., Ge, S., and Sahni, A. (2011). Late Cretaceous origin of the rice tribe provides evidence for early diversification in Poaceae. Nat. Commun. **2**:480.

Saarela, J.M., Wysocki, W.P., Barrett, C.F., Soreng, R.J., Davis, J.I., Clark, L.G., Kelchner, S.A., Pires, J.C., Edger, P.P., Mayfield, D.R., et al. (2015). Plastid phylogenomics of the cool-season grass subfamily: clarification of relationships among early-diverging tribes. AoB Plants **7**:plv046.

Saarela, J.M., Burke, S.V., Wysocki, W.P., Barrett, M.D., Clark, L.G., Craine, J.M., Peterson, P.M., Soreng, R.J., Vorontsova, M.S., and Duvall, M.R. (2018). A 250 plastome phylogeny of the grass family (Poaceae): topological support under different data partitions. PeerJ **6**:e4299.

Sage, R.F. (2004). The evolution of C_4 photosynthesis. New Phytol. **161**:341–370.

Sánchez-Ken, J.G., Clark, L.G., Kellogg, E.A., and Kay, E.E. (2007). Reinstatement and emendation of subfamily Micrairoideae (Poaceae). Syst. Bot. **32**:71–80.

Sayyari, E., and Mirarab, S. (2016). Fast coalescent-based computation of local branch support from quartet frequencies. Mol. Biol. Evol. **33**:1654–1668.

Schlüter, U., and Weber, A.P.M. (2020). Regulation and evolution of C4 photosynthesis. Annu. Rev. Plant Biol. **71**:183–215.

Schneider, J., Winterfeld, G., Hoffmann, M.H., and RöSer, M. (2011). Duthieeae, a new tribe of grasses (Poaceae) identified among the early diverging lineages of subfamily Pooideae: molecular phylogenetics, morphological delineation, cytogenetics and biogeography. Syst. Biodivers. **9**:27–44.

Schubert, M., Marcussen, T., Meseguer, A.S., and Fjellheim, S. (2019). The grass subfamily Pooideae: Cretaceous–Palaeocene origin and climate-driven Cenozoic diversification. Glob. Ecol. Biogeogr. **28**:1168–1182.

Shi, G., Grimaldi, D.A., Harlow, G.E., Wang, J., Wang, J., Yang, M., Lei, W., Li, Q., and Li, X. (2012). Age constraint on Burmese amber based on U-Pb dating of zircons. Cretac. Res. **37**:155–163.

Shimodaira, H., and Hasegawa, M. (2001). CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics **17**:1246–1247.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. **7**:539.

Simon, B.K. (2007). GrassWorld - interactive key and information system of world grasses. Kew Bull. **62**:475–484.

Sinha, N.R., and Kellogg, E.A. (1996). Parallelism and diversity in multiple origins of C4 photosynthesis in the grass family. Am. J. Bot. **83**:1458–1470.

Smith, B.N., and Epstein, S. (1971). Two categories of $^{13}C/^{12}C$ ratios for higher plants. Plant Physiol. **47**:380–384.

Smith, S.A., and O'Meara, B.C. (2012). TreePL: divergence time estimation using penalized likelihood for large phylogenies. Bioinformatics **28**:2689–2690.

Soderstrom, T.R. (1981). Some evolutionary trends in the Bambusoideae (Poaceae). Ann. Missouri Bot. Gard. **68**:15–47.

Soreng, R.J., Peterson, P.M., Romaschenko, K., Davidse, G., Zuloaga, F.O., Judziewicz, E.J., Filgueiras, T.S., Davis, J.I., and Morrone, O. (2015). A worldwide phylogenetic classification of the Poaceae (Gramineae). J. Syst. Evol. **53**:117–137.

Soreng, R.J., Peterson, P.M., Romaschenko, K., Davidse, G., Teisher, J.K., Clark, L.G., Barberá, P., Gillespie, L.J., and Zuloaga, F.O. (2017). A worldwide phylogenetic classification of the Poaceae (Gramineae) II: an update and a comparison of two 2015 classifications. J. Syst. Evol. **55**:259–290.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics **30**:1312–1313.

Stöver, B.C., and Müller, K.F. (2010). TreeGraph 2: combining and visualizing evidence from different phylogenetic analyses. BMC Bioinformatics **11**:7.

Struck, T.H. (2014). Trespex-detection of misleading signal in phylogenetic reconstructions based on tree information. Evol. Bioinformatics **10**:51–67.

Tang, L., Zou, X., Achoundong, G., Potgieter, C., Second, G., Zhang, D., and Ge, S. (2010a). Phylogeny and biogeography of the rice tribe (Oryzeae): evidence from combined analysis of 20 chloroplast fragments. Mol. Phylogenet. Evol. **54**:266–277.

Tang, H., Bowers, J.E., Wang, X., and Paterson, A.H. (2010b). Angiosperm genome comparisons reveal early polyploidy in the monocot lineage. Proc. Natl. Acad. Sci. U S A. **107**:472–477.

Tregunna, E.B., Smith, B.N., Berry, J.A., and Downton, W.J.S. (1970). Some methods for studying the photosynthetic taxonomy of the angiosperms. Can. J. Bot. **48**:1209–1214.

Triplett, J.K., Clark, L.G., Fisher, A.E., and Wen, J. (2014). Independent allopolyploidization events preceded speciation in the temperate and tropical woody bamboos. New Phytol. **204**:66–73.

Tzvelev, N.N. (1989). The system of grasses (Poaceae) and their evolution. Bot. Rev. **55**:141–204.

Vegetti, A., and Anton, A.M. (1995). Some evolution trends in the inflorescence of Poaceae. Flora **190**:225–228.

Vicentini, A., Barber, J.C., Aliscioni, S.S., Giussani, L.M., and Kellogg, E.A. (2008). The age of the grasses and clusters of origins of C4 photosynthesis. Glob. Chang. Biol. **14**:2963–2977.

Vogel, J.P., Garvin, D.F., Mockler, T.C., Schmutz, J., Rokhsar, D., Bevan, M.W., Barry, K., Lucas, S., Harmon-Smith, M., Lail, K., et al. (2010). Genome sequencing and analysis of the model grass *Brachypodium distachyon*. Nature **463**:763–768.

Wang, N., Zhong, X., Cong, Y., Wang, T., Yang, S., Li, Y., and Gai, J. (2016). Genome-wide analysis of phosphoenolpyruvate carboxylase gene family and their response to abiotic stresses in soybean. Sci. Rep. **6**:38448.

Washburn, J.D., Schnable, J.C., Davidse, G., and Pires, J.C. (2015). Phylogeny and photosynthesis of the grass tribe Paniceae. Am. J. Bot. **102**:1493–1505.

Watson, L., and Dallwitz, M.J. (1992). The Grass Genera of the World: Descriptions, Illustrations, Identification, and Information Retrieval; Including Synonyms, Morphology, Anatomy, Physiology, Phytochemistry, Cytology, Classification, Pathogens, World and Local Distribution, and References, Version: 18th December 2012. http://delta-intkey.com.

Welker, C.A.D., McKain, M.R., Estep, M.C., Pasquet, R.S., Chipabika, G., Pallangyo, B., and Kellogg, E.A. (2020). Phylogenomics enables biogeographic analysis and a new subtribal classification of Andropogoneae (Poaceae—Panicoideae). J. Syst. Evol. **58**:1003–1030.

Wickett, N.J., Mirarab, S., Nguyen, N., Warnow, T., Carpenter, E., Matasci, N., Ayyampalayam, S., Barker, M.S., Burleigh, J.G., Gitzendanner, M.A., et al. (2014). Phylotranscriptomic analysis of the origin and early diversification of land plants. Proc. Natl. Acad. Sci. U S A **111**:E4859–E4868.

Wu, Y., You, H.L., and Li, X.Q. (2018). Dinosaur-associated Poaceae epidermis and phytoliths from the early cretaceous of China. Natl. Sci. Rev. **5**:721–727.

Wysocki, W.P., Clark, L.G., Attigala, L., Ruiz-Sanchez, E., and Duvall, M.R. (2015). Evolution of the bamboos (Bambusoideae; Poaceae): a full plastome phylogenomic analysis. BMC Evol. Biol. **15**:50.

Xiang, Y., Huang, C.H., Hu, Y., Wen, J., Li, S., Yi, T., Chen, H., Xiang, J., and Ma, H. (2017). Evolution of Rosaceae fruit types based on nuclear phylogeny in the context of geological times and genome duplication. Mol. Biol. Evol. **34**:262–281.

Yang, Y., Moore, M.J., Brockington, S.F., Mikenas, J., Olivieri, J., Walker, J.F., and Smith, S.A. (2018). Improved transcriptome sampling pinpoints 26 ancient and more recent polyploidy events in Caryophyllales, including two allopolyploidy events. New Phytol. **217**:855–870.

Zeng, L., Zhang, Q., Sun, R., Kong, H., Zhang, N., and Ma, H. (2014). Resolution of deep angiosperm phylogeny using conserved nuclear genes and estimates of early divergence times. Nat. Commun. **5**:4956.

Zhang, Y.X., Ma, P.F., and Li, D.Z. (2018). A new genus of temperate woody bamboos (Poaceae, Bambusoideae, Arundinarieae) from a limestone montane area of China. PhytoKeys **109**:67–76.

Zhang, C., Zhang, T., Luebert, F., Xiang, Y., Huang, C.H., Hu, Y., Rees, M., Frohlich, M.W., Qi, J., Weigend, M., et al. (2020a). Asterid phylogenomics/phylotranscriptomics uncover morphological evolutionary histories and support phylogenetic placement for numerous whole-genome duplications. Mol. Biol. Evol. **37**:3188–3210.

Zhang, Y.X., Guo, C., and Li, D.Z. (2020b). A new subtribal classification of Arundinarieae (Poaceae, Bambusoideae) with the description of a new genus. Plant Divers. **42**:127–134.

Zhao, L., Zhang, N., Ma, P.F., Liu, Q., Li, D.Z., and Guo, Z.H. (2013). Phylogenomic analyses of nuclear genes reveal the evolutionary relationships within the BEP clade and the evidence of positive selection in Poaceae. PLoS One **8**:e64642.

Zhu, Q., and Ge, S. (2005). Phylogenetic relationships among A-genome species of the genus *Oryza* revealed by intron sequences of four nuclear genes. New Phytol. **167**:249–265.

Zou, X.H., Zhang, F.M., Zhang, J.G., Zang, L.L., Tang, L., Wang, J., Sang, T., and Ge, S. (2008). Analysis of 142 genes resolves the rapid diversification of the rice genus. Genome Biol. **9**:R49.

Zou, X.H., Yang, Z., Doyle, J.J., and Ge, S. (2013). Multilocus estimation of divergence times and ancestral effective population sizes of *Oryza* species and implications for the rapid diversification of the genus. New Phytol. **198**:1155–1164.