Rapid report

New grass phylogeny resolves deep evolutionary relationships and discovers C_4 origins

Author for correspondence:
Erika J. Edwards
Tel: +1 401 863 2081
Email: erika_edwards@brown.edu

Received: *21 September 2011*
Accepted: *11 October 2011*

Grass Phylogeny Working Group II*

New Phytologist (2012) **193**: 304–312
doi: 10.1111/j.1469-8137.2011.03972.x

Key words: C_4 photosynthesis, character evolution, phylogeny, Poaceae, state-dependent diversification.

Summary

- Grasses rank among the world's most ecologically and economically important plants. Repeated evolution of the C_4 syndrome has made photosynthesis highly efficient in many grasses, inspiring intensive efforts to engineer the pathway into C_3 crops. However, comparative biology has been of limited use to this endeavor because of uncertainty in the number and phylogenetic placement of C_4 origins.
- We built the most comprehensive and robust molecular phylogeny for grasses to date, expanding sampling efforts of a previous working group from 62 to 531 taxa, emphasizing the C_4-rich PACMAD (Panicoideae, Arundinoideae, Chloridoideae, Micrairoideae, Aristidoideae and Danthonioideae) clade. Our final matrix comprises *c.* 5700 bp and is > 93% complete.
- For the first time, we present strong support for relationships among all the major grass lineages. Several new C_4 lineages are identified, and previously inferred origins confirmed. C_3/C_4 evolutionary transitions have been highly asymmetrical, with 22–24 inferred origins of the C_4 pathway and only one potential reversal.
- Our backbone tree clarifies major outstanding systematic questions and highlights C_3 and C_4 sister taxa for comparative studies. Two lineages have emerged as hotbeds of C_4 evolution. Future work in these lineages will be instrumental in understanding the evolution of this complex trait.

---

*Sandra Aliscioni[1], Hester L. Bell[2], Guillaume Besnard[3,4], Pascal-Antoine Christin[5], J. Travis Columbus[2], Melvin R. Duvall[6], Erika J. Edwards[5], Liliana Giussani[7], Kristen Hasenstab-Lehman[2], Khidir W. Hilu[8], Trevor R. Hodkinson[9], Amanda L. Ingram[10], Elizabeth A. Kellogg[11], Saeideh Mashayekhi[2], Osvaldo Morrone[7], Colin P. Osborne[12], Nicolas Salamin[13,14], Hanno Schaefer[15], Elizabeth Spriggs[5], Stephen A. Smith[5,16] and Fernando Zuloaga[7]
[1]Cátedra de Botánica Agrícola, Facultad de Agronomía, Universidad de Buenos Aires, Av. San Martín 4453, C1417DSE, Buenos Aires, Argentina; [2]Rancho Santa Ana Botanic Garden and Claremont Graduate University, 1500 North College Avenue, Claremont, CA 91711-3157, USA; [3]CNRS, UPS, ENFA, Laboratoire Evolution & Diversité Biologique, UMR 5174, 31062 Toulouse 4, France; [4]Imperial College London, Silwood Park Campus, Buckhurst Road, Ascot SL5 7PY, UK; [5]Department of Ecology and Evolutionary Biology, Brown University, Box G-W, Providence, RI 02912, USA; [6]Department of Biological Sciences, Northern Illinois University, 1425 W Lincoln Hwy, DeKalb, IL 60115-2861, USA; [7]Instituto de Botánica Darwinion, Labardén 200, Casilla de Correo 22, B1642HYD, San Isidro, Buenos Aires, Argentina; [8]Department of Biological Sciences, Virginia Tech, Blacksburg, VA 24061, USA; [9]Department of Botany, School of Natural Sciences, University of Dublin, Trinity College, Dublin D2, Ireland; [10]Department of Biology, Wabash College, PO Box 352, Crawfordsville, IN 47933, USA; [11]Department of Biology, University of Missouri-St Louis, St Louis, MO 63121, USA; [12]Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK; [13]Department of Ecology and Evolution, University of Lausanne, 1015 Lausanne, Switzerland; [14]Swiss Institute of Bioinformatics, Quartier Sorge, 1015 Lausanne, Switzerland; [15]Department of Organismic and Evolutionary Biology, Harvard University, 22 Divinity Avenue, Cambridge, MA 02138, USA;[16]Heidelberg Institute for Theoretical Studies, Heidelberg, Germany.

Introduction

The grass family (Poaceae) includes > 11 000 recognized species with a cosmopolitan distribution and occupies an enormous range of habitats (Clayton & Renvoize, 1986; Osborne *et al.*, 2011). Grasses also include the three most important crops in the world (wheat (*Triticum aestivum*), maize (*Zea mays*) and rice (*Oryza sativa*)) and several productive species with great biofuel potential (Byrt *et al.*, 2011). Many grass lineages have evolved C_4 photosynthesis, a complex and coordinated set of anatomical and biochemical modifications that act to concentrate CO_2 at the site of fixation by Rubisco during the Calvin cycle (Sage, 2004; Edwards *et al.,* 2010). The direct effect of the C_4 pathway is to reduce photorespiration and saturate photosynthesis with CO_2, which has allowed C_4 grasses to colonize open and drier habitats in tropical and subtropical regions (Osborne & Freckleton, 2009; Edwards & Smith, 2010). Extant C_4 grass diversity is upwards of 4500 species, and C_4 grasses dominate many important ecosystems and contribute 20–25% of terrestrial primary productivity (Still *et al.*, 2003).

Despite the enormous economic and ecological importance of grasses, the evolutionary history of the group is still only partially understood. Phylogenies have accumulated over the past 20 yr, but most studies focused on specific groups below the subfamily level. The few family-wide phylogenetic studies (e.g. Clark *et al.*, 1995; GPWG, 2001; Duvall *et al.*, 2007) identified three species-poor lineages that are successively sister to all other grasses (Anomochlooideae, Pharoideae and Puelioideae) and placed the bulk of grass diversity in two main clades, known by their acronyms as BEP (Bambusoideae, Ehrhartoideae (formerly Oryzoideae) and Pooideae) and PACMAD (Panicoideae, Arundinoideae, Chloridoideae, Micrairoideae, Aristidoideae and Danthonioideae). More recently, the use of morphological traits (Bouchenak-Khelladi *et al.*, 2008) as well as supermatrix approaches (Edwards & Smith, 2010) has allowed extensive taxonomic coverage. These strategies, however, have not resolved relationships among the subfamilies in either the BEP or the PACMAD clade, mainly because data gathering approaches were not optimal and led to large amounts of missing data. A concerted effort was thus needed to produce a molecular phylogenetic study of the family that combined dense taxon sampling with a large and sufficiently complete molecular data set.

All C_4 grasses belong to the PACMAD group, but their polyphyly has been long recognized (Kellogg, 2000). Variations in the genetic basis and anatomical and biochemical details of the C_4 pathway among phylogenetic groups strongly support the hypothesis of multiple C_4 origins from C_3 ancestors (Sinha & Kellogg, 1996; Christin *et al.*, 2010). However, the exact number of C_4 lineages has been constantly increasing with the addition of more taxa, ranging from the early estimates of four origins (Kellogg, 2000) to 17–20 in more recent studies (Christin *et al.*, 2008; Edwards & Smith, 2010). Many genera of tropical grasses have only recently been analysed, preventing a precise evaluation of the number of C_4 groups and their relationships to C_3 grasses.

Here, we built a nearly complete data matrix of three chloroplast markers commonly used in grass phylogenetics to obtain a densely sampled and well-supported phylogeny for the grass family. Our first aim was to obtain a solid phylogenetic framework to study evolution in grasses. This new backbone tree will also provide the starting point for future work towards a complete, species-level phylogeny for the grasses. Our second aim was to improve the identification of photosynthetic transitions by drastically increasing taxon sampling in clades containing multiple C_3 and C_4 taxa. This phylogenetic information will be crucial for comparative and multidisciplinary studies addressing C_4 ecology, evolution, and genetics.

Materials and Methods

Strategies for taxon sampling

We selected three genetic markers from the chloroplast genome: the coding genes *rbcL* (ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit) and *ndhF* (NADH dehydrogenase subunit F), and the region encompassing the *matK* (maturase K) coding gene and the *trnK* (tRNA-Lys) introns (*trnK/matK*). These markers have been widely used in grass phylogenetics (e.g. Hilu *et al.*, 1999; Hilu & Alice, 2001; Giussani *et al.*, 2001; Christin *et al.*, 2008) but not in concert. Our strategy was aimed at filling in the gaps to achieve a dense and relatively balanced sampling of species across the major grass lineages, particularly in the PACMAD clade. We screened GenBank for these markers and supplemented the available data by sequencing genomic DNA (gDNA) for selected taxa available from previous studies (Hilu *et al.*, 1999; Hilu & Alice, 2001; Aliscioni *et al.*, 2003; Christin *et al.*, 2008; Vicentini *et al.*, 2008; Taylor *et al.*, 2011a; Morrone *et al.*, 2011) or isolated from herbarium specimens.

The master data set includes 545 accessions representing 531 species and 311 genera, representing nearly two-thirds of currently recognized PACMAD genera. We focused sampling efforts in groups that were suspected to contain photosynthetic transitions, especially the Panicoideae, which encompasses the majority of putative C_4 origins (Sinha & Kellogg, 1996; Giussani *et al.*, 2001; Christin *et al.*, 2009; Edwards & Smith, 2010). In this subfamily we included as many genera as possible. As most genera contain only one photosynthetic pathway, and assuming most genera to be monophyletic, this should make the count of photosynthetic transitions more accurate.

Genomic regions and DNA sequencing

For newly generated sequences, the three markers were PCR-amplified in 600–800-bp overlapping fragments with available primers (Taylor *et al.*, 2011a). However, much of the genomic DNA extracted from herbarium specimens was of poor quality, and amplification of long fragments failed. We therefore developed a battery of primers to amplify the different markers in overlapping segments as short as 250 bp (Supporting Information Table S1).

PCRs were carried out in a total volume of 25 µl, including *c.* 40–100 ng of gDNA template, 5 µl of 5× GoTaq reaction buffer, 0.1 mM dNTPs, 0.1 µM of each primer, 1 mM of $MgCl_2$, and 0.5 unit of *Taq* polymerase (GoTaq DNA Polymerase, Promega, Madison, WI, USA). The PCR mixtures were incubated in a thermocycler for 3 min at 94°C followed by 36 cycles consisting of 1 min at 94°C, 30 s at 48°C and 1 min at 72°C. This was followed by 10 min at 72°C. Successful amplifications were cleaned with an Exo-SAP-IT treatment (Affymetrix, Santa Clara, CA, USA) and sequenced using the Big Dye 3.1 Terminator Cycle Sequencing chemistry (Applied Biosystems, Foster City, CA, USA). All sequences have been deposited in GenBank (Table S2).

Sequence analyses

Sequences were initially aligned in ClustalW (Thompson *et al.*, 1994) and adjusted manually to account for gaps, particularly in *trnK* introns, *matK* and *ndhF*, following the criteria of Kelchner (2000). Homology assessment was difficult in some regions of the *trnK* alignment, so those regions, comprising 403 aligned bp total, were excluded from the analyses.

Phylogenetic trees were obtained from the three markers simultaneously through Bayesian inference as implemented in MrBayes 3.1 (Ronquist & Huelsenbeck, 2003). The substitution model was set to a GTR + G + I, determined as the best-fit model through hierarchical likelihood ratio tests. To avoid overparameterization and to reduce computational time, the data set was not partitioned among genes. Two different analyses, each of four parallel chains, were run for 11 717 000 generations, sampling a tree each 1000 generations after a burn-in period of 3 000 000. The convergence of the MCMC (Markov Chain Monte Carlo) run and the adequacy of the burn-in length were confirmed using the program tracer (Rambaut & Drummond, 2007). A majority rule consensus tree was computed on the 17 434 sampled trees. Phylogenetic trees were also inferred under a maximum likelihood criterion using the software RAxML (Stamatakis, 2006), under the GTRCAT substitution model. Support values for the branches were obtained from 1000 standard bootstrap pseudoreplicates.

Reconstruction of photosynthetic transitions

We typed all species in our tree as C_3 or C_4 according to different sources, summarized in GrassPortal (Osborne *et al.*, 2011). *Steinchisma hians* is a C_3-C_4 intermediate and was included in the C_4 category. We implemented various approaches to infer transitions between C_3 and C_4 photosynthesis, including stochastic mapping (Minin & Suchard, 2008) and ancestral state estimation using likelihood and a Markov model of discrete trait evolution (Pagel, 1999). All analyses were run using our Bayesian consensus tree (Supporting Information Fig. S1). To evaluate the influence of phylogenetic uncertainty on our analyses, we performed additional likelihood reconstruction and stochastic mapping analyses on an additional 1025 topologies sampled randomly from our post burn-in Bayesian posterior tree distribution. We estimated the number and placement of photosynthetic transitions on each tree, summarized the reconstructions across all trees, and identified two small but key regions where phylogenetic rearrangements affected our inferences of C_4 evolution.

Finally, to determine the potential influence of differential diversification rates in C_3 vs C_4 lineages on our estimates of transition rates (e.g. Maddison, 2006), we also implemented a maximum likelihood approach that simultaneously estimates diversification rates and transition rates for binary characters (Maddison *et al.*, 2007). We used only the PACMAD portion of our phylogeny for these analyses because it is far better sampled and because it contains all of the C_4 taxa and suspected transitions between character states. An appropriate λ value was estimated using the cross validation procedure in r8s (Sanderson, 2003), and used to smooth our Bayesian consensus tree with a root age set at 1. We then multiplied the branch lengths by 100 to make the computational steps more feasible. We distributed as many PACMAD species as possible among the tips of our tree based on genus richness estimates from either the Grass Genera of the World or the Tropicos taxonomic database (Watson & Dallwitz, 1992; Tropicos, 2011). Where more than one member of a genus was present, we considered only a single representative, and where a genus was polyphyletic, several genera were combined and a composite richness value was assigned for the entire clade. The final phylogenetic data set contained 209 representatives, with *c.* 70% of all PACMAD species assigned as related to a particular included taxon. We used the adjustments provided by Fitzjohn *et al.* (2009) to incorporate this information as unresolved clades at the tips of our tree and ran analyses using the 'diversitree' package in R.

Results and Discussion

Phylogeny of the grasses

Consistent with most previous studies, our analyses recover a grade of three lineages – Anomochlooideae, Pharoideae, and Puelioideae – subtending the BEP and PACMAD clades (Fig. 1). Both the BEP and PACMAD clades are strongly supported in our analyses, as are each of the constituent subfamilies. The branching order of the six PACMAD subfamilies is resolved with strong support for the sister taxon relationship of Arundinoideae plus Micrairoideae (AruM clade), and the *Centropodia*-Chloridoideae and Danthonioideae (CD clade). There is less bootstrap support for the sister taxon status of these two clades, and their relationship to Aristidoideae and Panicoideae, though Bayesian support is strong (Figs 1, S1 and S2). In general, Bayesian and RAxML inferred topologies and support were quite congruent (Figs S1, S2).

The position of Aristidoideae in our analyses is consistent with a number of studies (Clark *et al.*, 1995; Mathews & Sharrock, 1996; Soreng & Davis, 1998; Hilu *et al.*, 1999; Hsiao *et al.*, 1999; GPWG, 2001; Duvall *et al.*, 2007; Sánchez-Ken *et al.*, 2007; Christin *et al.*, 2008; Vicentini *et al.*, 2008). The CD clade has been recovered in several studies (Barker *et al.*, 1995; Soreng & Davis, 1998; Hilu *et al.*, 1999; Duvall *et al.*, 2007; Sánchez-Ken *et al.*, 2007; Christin *et al.*, 2008; Bouchenak-Khelladi *et al.*, 2008; Edwards & Smith, 2010; Peterson *et al.*, 2011), whereas the AruM clade appears in fewer (Duvall *et al.*, 2007; Sánchez-Ken *et al.*, 2007; Christin *et al.*, 2008; Edwards

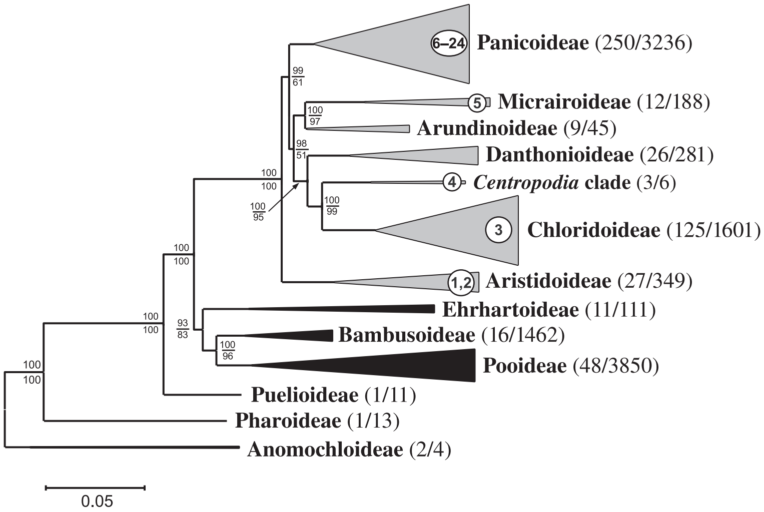

Fig. 1 Relationships among the subfamilies of Poaceae, inferred from Bayesian analysis of three chloroplast markers. The BEP (Bambusoideae, Ehrhartoideae (formerly Oryzoideae) and Pooideae) clade is in black, and PACMAD (Panicoideae, Arundinoideae, Chloridoideae, Micrairoideae, Aristidoideae and Danthonioideae) is in grey. Numbers alongside subfamilial names represent the proportion of species we sampled relative to the total species richness of each clade. All named lineages received 100% maximum likelihood bootstrap support (BS) and Bayesian posterior probabilities (PPs) of 1.0, with the exception of Arundinoideae; nodes receiving lower support are noted, with PP values above the line and BS below the line. Locations of C_4 origins are indicated by numbers, which correspond to Table 1.

& Smith, 2010; Peterson *et al.*, 2011; Teerawatananon *et al.*, 2011). The CD + AruM clade was recovered elsewhere only by Duvall *et al.* (2007) and Peterson *et al.* (2011), and, in fact, relationships among PACMAD subfamilies in their analyses mirror ours. We used SH tests (Shimodaira & Hasegawa, 1999) to determine whether any other previously suggested topologies could be rejected by the data. We were unable to reject other possible relationships, a somewhat surprising result given the high bootstrap and posterior probability values for several clades.

Although not a particular focus of this study, we also resolved the BEP clade, finding strong support for the sister relationship of Bambusoideae and Pooideae. However, our analysis here did not include *Streptogyna*, whose uncertain position in the BEP clade has often confused relationships. Our final data matrix, Bayesian consensus tree, and maximum likelihood tree are available for download on TreeBASE (TreeBASE accession #11973; http://purl.org/phylo/treebase/phylows/study/TB2:S11973).

Evolution of C_4 photosynthesis in grasses

Our analyses strongly rejected a symmetric model of photosynthetic transitions in the grasses. Using molecular or smoothed ultrametric branch lengths, the optimal likelihood model implied that reverse transitions from C_4 to C_3 photosynthesis are exceedingly unlikely (molecular: $Q_{C3,C4} = 3.45$, $Q_{C4,C3} = 8e^{-4}$, $\log_e L = -88.426$; smoothed: $Q_{C3,C4} = 0.42$, $Q_{C4,C3} = 9e^{-5}$, $\log_e L = -91.56$, where Q is the instantaneous rate of transition between character states, and L is the likelihood; Table S3). This asymmetry is also recovered when accounting for possible differences in diversification rates between C_3 and C_4 lineages. A six-parameter model, which allowed for unequal transition rates between character states, was strongly preferred to one enforcing transition rates to be equal using the Akaike Information Criterion (AIC): (ΔAIC = 38.5, $P = 1.9e^{-10}$). Under this model, we inferred an instantaneous transition rate of C_3/C_4 that was 50-fold higher than C_4/C_3 ($Q_{C3,C4} = 5.09e^{-3}$, $Q_{C4,C3} = 1.0e^{-4}$; Table S3). These results are consistent with other work emphasizing the prevalence of C_4 origins over losses, and are further supported by variation in the anatomy, biochemistry and genetic determinism of the C_4 pathway used by the different phylogenetic lineages (Christin & Besnard, 2009; Christin *et al.*, 2010; Roalson, 2011). With the recent discovery of C_3 *Eragrostis walteri* as a member of C_3 Arundinoideae (Ingram *et al.*, 2011), *Alloteropsis semialata* subsp. *eckloniana* stands alone as the sole remaining candidate for a loss of C_4 photosynthesis in grasses (Ibrahim *et al.*, 2009). While lengthy discussion of *Alloteropsis* is beyond the scope of this paper, it is becoming increasingly plausible that *Alloteropsis* includes multiple parallel transitions from C_3 to C_4 (Christin *et al.*, 2010).

Depending on the topology, ancestral character reconstructions inferred between 22 and 24 origins of the C_4 pathway (Fig. 2, Table 1). Although many of these origins have been identified in previous analyses (summarized in Christin *et al.*, 2009), their stability in light of our expanded taxon sampling increases our confidence that we have now correctly placed most of the C_4 grass lineages. These include *Eriachne* (+ *Pheidochloa*; see Morrone *et al.*, 2011), *Aristida* (excluding *Aristida longifolia*), *Stipagrostis*, Chloridoideae, *Centropodia*, Tristachyideae, Andropogoneae, *Paraneurachne muelleri*, *Steinchisma* (C_3-C_4 intermediates), the large 'MPC' clade (Melinidinae + Panicinae + Cenchrinae), and the *Paspalum* and *Axonopus* groups of Paspalineae (two origins based on the position of C_3 *Streptostachys asperifolia*). *Digitaria* stands as an additional strongly supported independent C_4 lineage in our data set, although the nuclear marker *phyB* (phytochrome B) has placed *Digitaria* as sister to the MPC clade (Vicentini *et al.*, 2008), and phosphoenolpyruvate carboxylase (*ppc*) places it within the MPC clade (Christin *et al.*, 2007).

While our intensive sampling captures nearly all of the purported closely related C_3 and C_4 lineages in grasses (Table 2), we still lack a clear picture of photosynthetic evolution in two specific areas of the tree. The uncertainties in these two groups are responsible for all of the variation in our sensitivity analyses, with the number of inferred origins in each group ranging from two to five. The first is the Boivinellinae (sensu Morrone *et al.*, 2011), a well-supported lineage in Paniceae that includes *Alloteropsis* and a second C_4 group, *Echinochloa*. Because they

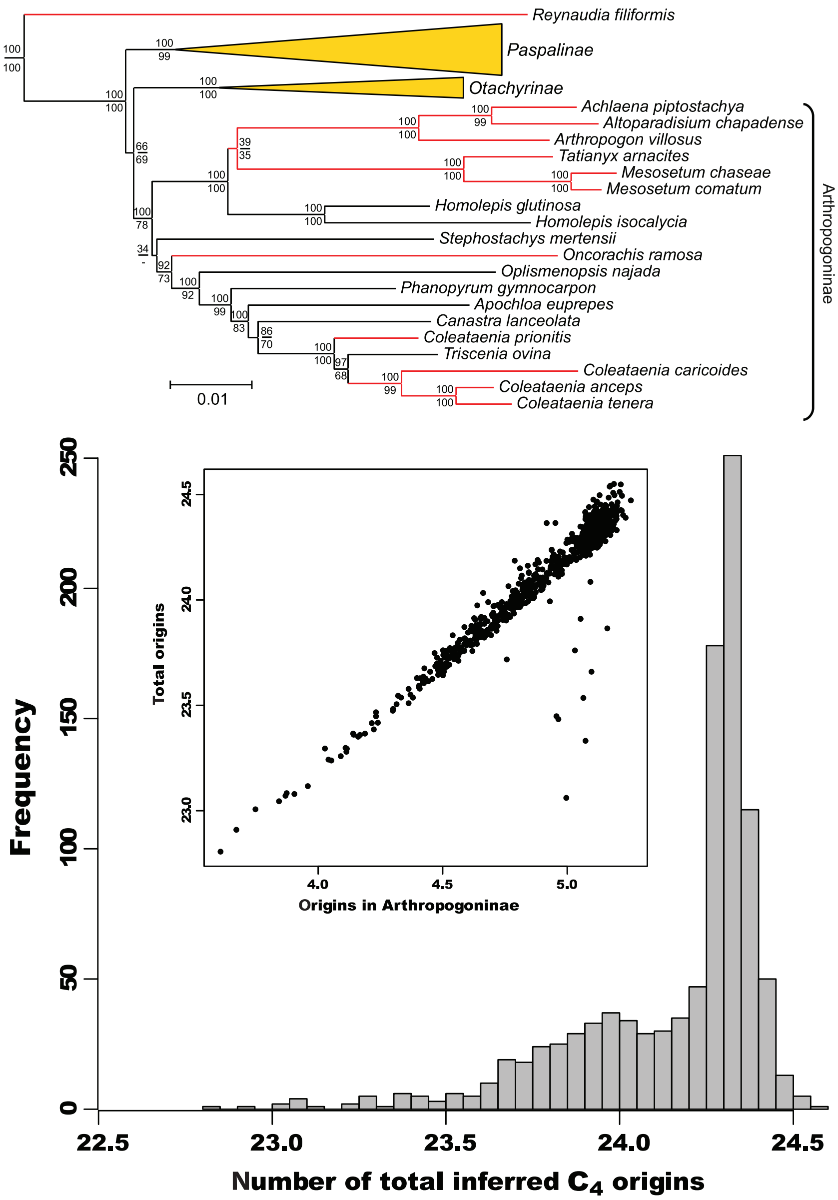

Fig. 2 Inference of C_4 evolution in grasses. Histogram represents variation in the number of inferred C_4 origins across all of PACMAD (Panicoideae, Arundinoideae, Chloridoideae, Micrairoideae, Aristidoideae and Danthonioideae) using stochastic mapping and 1025 alternative phylogenies sampled from a Bayesian posterior distribution of trees. The scatterplot inset indicates that most of the uncertainty in reconstructing C_4 evolution results from poor phylogenetic resolution of one small, mixed C_3/C_4 clade, the Arthropogonineae, shown above. C_3 only, black; C_4 only, red; mixed C_3 and C_4, yellow.

have never been placed as sisters in any phylogenetic analysis known to us, we feel confident that these represent distinct C_4 lineages. However, the identity of their closest C_3 relatives remains unclear. *Alloteropsis* is often united with *Entolasia*, as is the case here, though with limited statistical support. This clade also presents additional uncertainty regarding C_4 evolution within *Alloteropsis*, as discussed above. The nuclear gene *ppc* places *Echinochloa* as sister to the MPC clade (Christin *et al.*, 2007), suggesting a possibly complex origin for this genus.

A second interesting and problematic area of the tree is the Arthropogoninae clade. This lineage is especially well sampled here, with 14 of 16 genera (19 of 50 species) included. The two missing genera are reported as C_4 and have previously been tightly associated with C_4 lineages that are included in our analyses: *Keratochlaena* has been aligned with *Mesosetum*, and *Cyphonanthus* has been reported as sister to *Oncorachis* (Morrone *et al.*, 2011). In spite of this good taxon sampling, we have little confidence in our topology here. The current placement of *Triscenia*, a monotypic C_3 taxon from Cuba, breaks C_4 *Coleataenia* into two lineages, but *Triscenia* falls outside of *Coleataenia* in many trees from our posterior distribution. Furthermore, the relationships between the *Mesosetum* and *Altoparadisium* C_4 clades and the C_3 *Homolepis* lack support, blurring statistical inferences in this area of the phylogeny.

Both the *Alloteropsis* lineage and Arthropogoninae illustrate the complex nature of accounting C_4 origins, and how arriving at a single number may be misleading in the end. Even a perfectly resolved phylogeny will not overcome the difficulty of modelling past photosynthetic transitions. The predominance of C_3 to C_4 transitions and the extreme rarity of back transitions are strongly supported by different lines of evidence (this study; Christin *et al.*, 2010), but the extreme clustering of C_4 groups in certain

Table 1 C_4 lineages identified in this study, with recommendations for comparative studies of closely related C_3/C_4 species pairs

	C_4 lineage	Clear C_3 sister for comparative work?
1	***Aristida***	***Aristida longifolia***
2	***Stipagrostis***	***Sartidia***
3	**Chloridoideae**	**No, and not likely**
4	**Centropodia**	***Ellisochloa rangei***
5	**Eriachne**	**Isachne**
6	**Tristachyideae**	***Centotheceae/Thysanolaeneae***
7	**Andropogoneae**	**No, and not likely**
8	***Reynaudia***	**No, and not likely**
9	*Axonopus*	*Streptostachys asperifolia*
10	*Paspalum*	No, and not likely
11	***Anthaenantia***	**Otachyriinae p.p.**
12	***Steinchisma***	***Steinchisma laxa***
13	*Arthropogon*	Not yet, but likely
14	*Mesosetum*	*Homolepis*
15	*Oncorachis*	Not yet, but likely
16	*Coleataenia 1*	Not yet, but likely
17	*Coleataenia 2*	*Triscenia*
18	*Digitaria*	No, and not likely
19	*Echinochloa*	Not yet, but likely
20	***Paraneurachne***	***Neurachne***
21	**MPC**	***Homopholis***
22–24	*Alloteropsis*	*Alloteropsis ecklonian**

Bold indicates high confidence in that particular origin (both that it is an independent origin and that it is correctly placed; regular text indicates uncertainty in either or both). Numbers refer to Fig. 1. *Alternatively, this may represent a reversal to C_3 photosynthesis. See text for details. MPC, Melinidinae + Panicinae + Cenchrinae.

areas of the phylogenetic tree also questions the true 'independence' of many C_4 origins. Certain precursor traits probably evolved early in these lineages which increased the accessibility of the C_4 phenotype, and certain elements of the C_4 phenotype in these 'independent' lineages were probably inherited from their common ancestor. This pattern of extended, parallel evolution of the C_4 pathway has been demonstrated in several eudicot groups (e.g. McKown *et al.*, 2005; Christin *et al.*, 2011). Grass lineages highlighted in the present study that comprise closely related C_4 lineages separated by C_3 taxa, such as Arthropogoninae or Boivinellinae, represent ideal systems in which to investigate these hypotheses. We anticipate that the well-resolved phylogeny produced in this study will stimulate new comparative research aimed towards an integrative understanding of the processes that led to repeated evolution of C_4 photosynthesis in the grasses.

Conclusions

Combining the large amount of data generated during 20 yr of grass phylogenetics with a formidable and targeted new sequencing effort, we produced a family-wide phylogeny for grasses with a large amount of supporting DNA sequence data. The vast majority of grass species can now be assigned to clades, and the relationships among these groups received the strongest support obtained to date (Figs S1, S2). This new phylogenetic framework should facilitate comparative work on this important group of plants (e.g. Ghannoum *et al.*, 2005; Cousins *et al.*, 2008; Taylor *et al.*, 2011a,b). In particular, our dense sampling of C_3/C_4 transitions should be especially beneficial to the C_4 research community. We also structured our sampling within C_4 and C_3 lineages so as to produce a tree with a clade representation that is roughly proportional to extant grass diversity. Our phylogeny should thus be useful for research on various issues, such as morphological

Table 2 Unsampled genera in Paspaleae and Paniceae, their phylogenetic placement in Morrone *et al.* (2011), and their potential to represent an additional C_3/C_4 transition

Genus	No. of species	C_3/C_4	Subtribe	Morrone placement	Would inclusion result in a new transition?
Cyphonanthus	1	C_4	Arthropogoninae	Sister to *Steptostachys ramosa*	No
Keratochlaena	1	C_4	Arthropogoninae	Sister to *Mesosetum chaseae*	No
Ocellochloa	12	C_3	Paspalinae	Sister to *Echinolaena*	No
Renvoizea	10	C_3	Paspalinae	In polytomy at the base of Paspaleae	Possibly
Spheneria	1	C_4	Paspalinae	Sister to *Thrasyopsis*	No
Lecomtella	1	C_3	Paspalinae	Sister to *Gerritea* (morphological data only)	No
Baptorhachis	1	C_4	Paspalinae	Sister to *Ophiochloa* (morphological data only)	No
Acostia	1	C_4	Paspalinae	Sister to *Ophiochloa*, *Axonopus*, and *Centrochloa*	No
Thrasya	20	C_4	Paspalinae	In *Paspalum*	No
Tarigidia	1	C_4	Anthephorinae	Sister to *Chaetopoa* (morphological data only)	No
Trachys	1	C_4	Anthephorinae	Sister to *Chaetopoa* (morphological data only)	No
Odontelytrum	1	C_4	Cenchrinae	Included in *Cenchrus*	No
Zygochloa	1	C_4	Cenchrinae	Sister to *Spinifex* and *Pseudochaetochloa*	No
Paratheria	2	C_4	Cenchrinae	Sister to *Panicum antidotale*	No
Holcolemma	4	C_3	Cenchrinae	Sister to *Ixophorus unisextus* (morphological data only)	Possibly
Streptolophus	1	C_4	Cenchrinae	Sister to *Paratheria* (morphological data only)	No
Tricholaena	12	C_4	Melinidinae	Sister to *Leucophrys* with both sister to *Melinis repens*	No
Moorochloa	3	C_4	Melinidinae	Sister to *Melinis repens* + *Leucophrys* + *Tricholaena*	No
Leucophrys	1	C_4	Melinidinae	Sister to *Tricholaena*	No
Chaetium	3	C_4	Melinidinae	Sister to *Eriochloa*	No
Megathyrsus	3	C_4	Melinidinae	Sister to *Urochloa mutica*	No
Arthragrostis	3	C_4	Panicineae	Base of Panicinae (morphological data only)	No

and ecological diversification, variation in speciation/extinction rates, genomic evolution, biological invasions, and domestication of the world's most important crops.

Despite these very significant improvements, our phylogeny still covers only 5% of all recognized grass species. While we aimed to include as many genera as possible, some of these may not be monophyletic. Previous analyses have revealed many cases of nonmonophyly within grasses, including highly polyphyletic genera (e.g. *Panicum*, Aliscioni *et al.*, 2003; *Setaria,* Kellogg *et al.*, 2009; *Calamagrostis,* Saarela *et al.*, 2010), and members of the same genus even placed in different subfamilies (e.g. *Eragrostis* and *Merxmuellera*, Barker *et al.*, 1999; Ingram *et al.*, 2011). While recent and ongoing taxonomic revisions are improving matters greatly (Zuloaga *et al.*, 2006, 2007a,b, 2010, 2011; Morrone *et al.*, 2007, 2008; Sede *et al.*, 2008, 2009; Peterson *et al.*, 2011), the exact relationships among the numerous grasses will remain only approximated until most species are sequenced.

Phylogenetic studies across the entire Tree of Life over the past decades have left us with improved understanding of how the major groups of organisms are related to one another. Arguably the greatest remaining challenge is one of 'filling in the tips'; we see grasses as now currently poised to be a model lineage for experimenting with finding the best approach to this difficult problem. The number of grass species investigated is continuously increasing thanks to numerous taxonomically motivated sequencing studies of specific, smaller groups (e.g. Schneider *et al.*, 2009; Sungkaew *et al.*, 2009; Pirie *et al.*, 2010; Peterson *et al.*, 2010; Salariato *et al.*, 2010; Tang *et al.*, 2010). These studies in part utilize fast-evolving noncoding markers that are frequently difficult to align between distant grasses, but the backbone phylogeny developed in this study could be used to combine these independently produced phylogenies. A supermatrix approach would allow simultaneous analysis of a high number of grasses (e.g. Salamin *et al.*, 2002; Edwards & Smith, 2010), but studies using this approach typically only include loci that are widely sampled and can be aligned across the entire group, thus leaving out large amounts of available phylogenetic information. An alternative would be to use the present family-wide phylogeny as a backbone reference on which to graft more detailed phylogenies of specific groups. Setting topological and temporal constraints based on more deeply sampled phylogenies such as the one presented here would depict an evolutionary scenario congruent with phylogenetic and paleobotanical knowledge accumulated at larger taxonomic scales.

Acknowledgements

This paper is dedicated to the memory of Dr Osvaldo Morrone, who devoted his career to understanding the evolution of grasses. His expertise, particularly in subfamily Panicoideae, made this paper and many others possible. This study was conducted by members of the NESCent working group 'Grass Phylogeny Working Group II: Inferring the Complex Evolutionary History of C_4 Photosynthesis in Grasses', led by E.J.E, N.S, and S.A.S. The sequencing effort was funded by the grants Marie Curie IOF 252568 to P.A.C, NSF DEB-0920147 to J.T.C, NSF IOS-0843231 to E.J.E, NSF DEB-0921203 to A.I., Swiss NSF 3100A0_122433 to N.S, and a grant from the Plant Molecular Biology Center, Northern Illinois University, to M.R.D. We thank Neil Snow (Bishop Museum) for providing plant material. Carrie A. Kiel (RSABG), Colin P. Grennan, Sean V. Burke, and Sam S. Jones (Northern Illinois University) helped with the sequencing process. Several DNAs were kindly provided to G.B. by the Royal Botanic Gardens, Kew, UK.

References

Aliscioni SS, Giussani LM, Zuloaga FO, Kellogg EA. 2003. A molecular phylogeny of *Panicum* (Poaceae: Paniceae): tests of monophyly and phylogenetic placement within the Panicoideae. *American Journal of Botany* **90**: 796–821.

Barker NP, Linder HP, Harley EH. 1995. Polyphyly of Arundinoideae (Poaceae): evidence from *rbcL* sequence data. *Systematic Botany* **20**: 423–435.

Barker NP, Linder HP, Harley EH. 1999. Sequences of the grass-specific insert in the chloroplast *rpo*C2 gene elucidate generic relationships of the Arundinoideae (Poaceae). *Systematic Botany* **23**: 327–350.

Bouchenak-Khelladi Y, Salamin N, Savolainen V, Forest F, van der Bank M, Chase MW, Hodkinson TR. 2008. Large multi-gene phylogenetic trees of the grasses (Poaceae): progress towards complete tribal and generic level sampling. *Molecular Phylogenetics and Evolution* **47**: 488–505.

Byrt CS, Grof CPL, Furbank RT. 2011. C_4 plants as biofuel feedstocks: optimising biomass production and feedstock quality from a lignocellulosic perspective. *Journal of Integrative Plant Biology* **53**: 120–135.

Christin PA, Besnard G. 2009. Two independent C_4 origins in Aristidoideae (Poaceae) revealed by the recruitment of distinct phosphoenolpyruvate carboxylase genes. *American Journal of Botany* **96**: 2234–2239.

Christin PA, Besnard G, Samaritani E, Duvall MR, Hodkinson TR, Savolainen V, Salamin N. 2008. Oligocene CO_2 decline promoted C_4 photosynthesis in grasses. *Current Biology* **18**: 37–43.

Christin PA, Freckleton RP, Osborne CP. 2010. Can phylogenetics identify C_4 origins and reversals? *Trends in Ecology & Evolution* **25**: 403–409.

Christin PA, Sage TL, Edwards EJ, Ogburn RM, Khoshravesh R, Sage RF. 2011. Complex evolutionary transitions and the significance of C_3–C_4 intermediate forms of photosynthesis in Molluginaceae. *Evolution* **65**: 643–660.

Christin PA, Salamin N, Savolainen V, Duvall MR, Besnard G. 2007. C_4 photosynthesis evolved in grasses via parallel adaptive genetic changes. *Current Biology* **87**: 1241–1247.

Christin PA, Salamin N, Vicentini A, Kellogg EA, Besnard G. 2009. Integrating phylogeny into studies of C_4 variation in the grasses. *Plant Physiology* **149**: 82–87.

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

Clayton WD, Renvoize SA. 1986. Genera *Graminum.* Grasses of the world. *Kew Bulletin, Additional Series* **13**: 1–389.

Cousins AB, Badger MR, von Caemmerer S. 2008. C_4 photosynthetic isotope exchange in NAD-ME and NADP-ME type grasses. *Journal of Experimental Botany* **59**: 1695–1703.

Duvall MR, Davis JI, Clark LG, Noll JD, Goldman DH, Sánchez-Ken JG. 2007. Phylogeny of the grasses (Poaceae) revisited. *Aliso* **23**: 237–247.

Edwards EJ, Osborne CP, Stromberg CAE, Smith SA, C_4 Grass consortium. 2010. The origins of C_4 grasslands: integrating evolutionary and ecosystem science. *Science* **328**: 587–591.

Edwards EJ, Smith SA. 2010. Phylogenetic analyses reveal the shady history of C_4 grasses. *Proceedings of the National Academy of Sciences, USA* **107**: 2532–2537.

Fitzjohn RG, Maddison WP, Otto SP. 2009. Estimating trait-dependent speciation and extinction rates form incompletely resolved phylogenies. *Systematic Biology* **58**: 595–611.

Ghannoum O, Evans JR, Chow WS, Andrews TJ, Conroy JP, von Caemmerer S. 2005. Faster Rubisco is the key to superior nitrogen use efficiency in NADP-ME relative to NAD-ME C_4 grasses. *Plant Physiology* **137**: 638–650.

Giussani LM, Cota-Sanchez JH, Zuloaga FO, Kellogg EA. 2001. A molecular phylogeny of the grass subfamily Panicoideae (Poaceae) shows multiple origins of C_4 photosynthesis. *American Journal of Botany* **88**: 1993–2012.

GPWG (Grass Phylogeny Working Group). 2001. Phylogeny and subfamilial classification of the grasses (Poaceae). *Annals of the Missouri Botanical Garden* **88**: 373–457.

Hilu KW, Alice LA. 2001. A phylogeny of Chloridoideae (Poaceae) based on *matK* Sequences. *Systematic Botany* **26**: 386–405.

Hilu KW, Alice LA, Liang H. 1999. Phylogeny of Poaceae inferred from *matK* sequences. *Annals of the Missouri Botanical Garden* **86**: 835–851.

Hsiao C, Jacobs SWL, Chatterton NJ, Asay KH. 1999. A molecular phylogeny of the grass family (Poaceae) based on the sequences of nuclear ribosomal DNA (ITS). *Australian Systematic Botany* **11**: 667–688.

Ibrahim DG, Burke T, Ripley BS, Osborne CP. 2009. A molecular phylogeny of the genus *Alloteropsis* (Panicoideae, Poaceae) suggests an evolutionary reversion from C_4 to C_3 photosynthesis. *Annals of Botany* **103**: 127–136.

Ingram AL, Christin PA, Osborne CP. 2011. Molecular phylogenies disprove a hypothesized C_4 reversion in *Eragrostis walteri* (Poaceae). *Annals of Botany* **107**: 321–325.

Kelchner SA. 2000. The evolution of noncoding chloroplast DNA and its application in plant systematics. *Annals of the Missouri Botanical Garden* **87**: 482–498.

Kellogg EA. 2000. The grasses: a case study in macroevolution. *Annual Review of Ecology and Systematics* **31**: 217–238.

Kellogg EA, Aliscioni S, Morrone O, Pensiero J, Zuloaga F. 2009. A phylogeny of *Setaria* (Poaceae, Panicoideae, Paniceae) and related genera based on the chloroplast gene *ndh*F. *International Journal of Plant Sciences* **170**: 117–131.

Maddison WP. 2006. Confounding asymmetries in evolutionary diversification and character change. *Evolution* **60**: 1743–1746.

Maddison WP, Midford PE, Otto SP. 2007. Estimating a binary character's effect on speciation and extinction. *Systematic Biology* **56**: 701–710.

Mathews S, Sharrock RA. 1996. The phytochrome gene family in grasses (Poaceae): a phylogeny and evidence that grasses have a subset of the loci found in dicot angiosperms. *Molecular Biology and Evolution* **13**: 1141–1150.

McKown AD, Moncalvo JM, Dengler NG. 2005. Phylogeny of *Flaveria* (Asteraceae) and inference of C_4 photosynthesis evolution. *American Journal of Botany* **92**: 1911–1928.

Minin VN, Suchard MA. 2008. Fast, accurate and simulation-free stochastic mapping. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **363**: 3985–3995.

Morrone O, Aagesen L, Scataglini MA, Salariato DL, Denham SS, Chemisquy MA, Sede SM, Giussani LM, Kellogg EA, Zuloaga FO. 2011. Phylogeny of the Paniceae (Poaceae: Panicoideae): integrating plastid DNA sequences and morphology into a new classification. *Cladistics*, in press.

Morrone O, Denham SS, Aliscioni SS, Zuloaga FO. 2008. *Parodiophyllochloa*, a new genus segregated from *Panicum* (Paniceae, Poaceae) based on morphological and molecular data. *Systematic Botany* **33**: 66–76.

Morrone O, Scataglini MA, Zuloaga O. 2007. *Cyphonanthus*, a new genus segregated from *Panicum* (Poaceae: Panicoideae: Paniceae) based on morphological, anatomical and molecular data. *Taxon* **56**: 521–532.

Osborne CP, Freckleton RP. 2009. Ecological selection pressures for C_4 photosynthesis in the grasses. *Proceedings of the Royal Society Series B* **276**: 1753–1760.

Osborne CP, Visser V, Chapman S, Barker A, Freckleton RP, Salamin N, Simpson D, Uren V. 2011. GrassPortal: an online ecological and evolutionary data facility. [WWW document]. URL http://www.grassportal.org [accessed on 1 June 2011].

Pagel M. 1999. The maximum likelihood approach to reconstructing ancestral character states of discrete characters on phylogenies. *Systematic Biology* **48**: 612–622.

Peterson PM, Romaschenko K, Barker NP, Linder HP. 2011. Centropodieae and *Ellisochloa*, a new tribe and genus in Chloridoideae (Poaceae). *Taxon* **60**: 1113–1122.

Peterson PM, Romaschenko K, Johnson G. 2010. A phylogeny and classification of the Muhlenbergiinae (Poaceae: Chloridoideae: Cynodonteae) based on plastid and nuclear DNA sequences. *American Journal of Botany* **97**: 1532–1554.

Pirie MD, Lloyd KM, Lee WG, Linder HP. 2010. Diversification of *Chionochloa* (Poaceae) and biogeography of the New Zealand Southern Alps. *Journal of Biogeography* **37**: 379–392.

Rambaut A, Drummond AJ. 2007. Tracer v1.4. [WWW document]. URL http://beadt.bio.ed.ac.uk/Tracer. [accessed on 1 August 2011].

Roalson EH. 2011. Origins and transitions in photosynthetic pathway types in monocots: a review and reanalysis. In: Raghavendra AS, Sage RF, eds. *C_4 photosynthesis and related CO_2 concentrating mechanisms*. Dordrecht, the Netherlands: Springer, 319–338.

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

Saarela JM, Liu Q, Peterson PM, Soreng RJ, Paszko B. 2010. Phylogenetics of the grass 'Aveneae-Type Plastid DNA Clade' (Poaceae: Pooideae, Poeae) based on plastid and nuclear ribosomal DNA sequence data. In: Seberg O, Petersen G, Barfod AS, Davis J, eds. *Diversity, phylogeny, and evolution in the monocotyledons*. Aarhus, Denmark: Aarhus University Press, 557–586.

Sage RF. 2004. The evolution of C_4 photosynthesis. *New Phytologist* **161**: 341–370.

Salamin N, Hodkinson TR, Savolainen V. 2002. Building supertrees: an empirical assessment using the grass family (Poaceae). *Systematic Biology* **51**: 136–150.

Salariato DL, Zuloaga FO, Giussani LM, Morrone O. 2010. Molecular phylogeny of the subtribe Melinidinae (Poaceae: Panicoideae: Paniceae) and evolutionary trends in the homogenization of inflorescences. *Molecular Phylogenetics and Evolution* **56**: 355–369.

Sánchez-Ken JG, Clark LG, Kellogg EA, Kay EE. 2007. Reinstatement and emendation of subfamily Micrairoideae (Poaceae). *Systematic Botany* **32**: 71–80.

Sanderson MJ. 2003. R8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* **19**: 301–302.

Schneider J, Doring E, Hilu KW, Roser M. 2009. Phylogenetic structure of the grass subfamily Pooideae based on comparison of plastid *matK* gene-3' *trnK* exon and nuclear ITS sequences. *Taxon* **58**: 405–424.

Sede SM, Morrone O, Giussani LM, Zuloaga FO. 2008. Phylogenetic studies in the Paniceae (Poaceae): a realignment of section Lorea of *Panicum*. *Systematic Botany* **33**: 284–300.

Sede SM, Zuloaga FO, Morrone O. 2009. Phylogenetic studies in the Paniceae (Poaceae-Panicoideae): *Ocellochloa*, a new genus from the New World. *Systematic Botany* **34**: 684–692.

Shimodaira H, Hasegawa M. 1999. Comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution* **16**: 1114–1116.

Sinha NR, Kellogg EA. 1996. Parallelism and diversity in multiple origins of C_4 photosynthesis in the grass family. *American Journal of Botany* **83**: 1458–1470.

Soreng RJ, Davis JI. 1998. Phylogenetics and character evolution in the grass family (Poaceae): simultaneous analysis of morphological and chloroplast DNA restriction site character sets. *Botanical Review* **64**: 1–85.

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

Still CJ, Berry JA, Collatz GJ, DeFries RS. 2003. Global distribution of C_3 and C_4 vegetation: carbon cycle implications. *Global Biogeochemical Cycles* **17**: 1006.

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

Tang L, Zou XH, Achoundong G, Potgieter C, Second G, Zhang DY, Ge S. 2010. Phylogeny and biogeography of the rice tribe (Oryzeae): evidence from combined analysis of 20 chloroplast fragments. *Molecular Phylogenetics and Evolution* **54**: 266–277.

Taylor SH, Franks PI, Hulme SP, Spriggs E, Christin PA, Edwards EJ, Woodward FI, Osborne CP. 2011a. Photosynthetic pathway and ecological adaptation explain stomatal trait diversity amongst grasses. *New Phytologist*. doi: 10.1111/j.1469-8137.2011.03935.x

Taylor SH, Ripley BS, Woodward FI, Osborne CP. 2011b. Drought limitation of photosynthesis differs between C$_3$ and C$_4$ grass species in a comparative experiment. *Plant, Cell & Environment* **34**: 65–75.

Teerawatananon A, Jacobs SWL, Hodkinson TR. 2011. Phylogenetics of *Panicoideae* (Poaceae) based on chloroplast and nuclear DNA sequences. *Telopea* **13**: 115–142.

Thompson JD, Higgins DG, Gibson TJ. 1994. ClustalW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and matrix choice. *Nucleic Acids Research* **22**: 4673–4680.

Tropicos. 2011. Tropicos.org. Missouri Botanical Garden. [WWW document]. URL http://www.tropicos.org. [accessed on 15 August 2010].

Vicentini A, Barber JC, Aliscioni SS, Giussani LM. 2008. The age of the grasses and clusters of origins of C$_4$ photosynthesis. *Global Change Biology* **12**: 2963–2977.

Watson L, Dallwitz MJ. 1992 onwards. 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: 23 April 2010. [WWW document]. URL: http://delta-intkey.com/grass.

Zuloaga FO, Giussani LM, Morrone O. 2006. On the taxonomic position of *Panicum aristellum* (Poaceae: Panicoideae: Paniceae). *Systematic Botany* **31**: 497–505.

Zuloaga FO, Giussani LM, Morrone O. 2007b. *Hopia*, a new monotypic genus segregated from *Panicum* (Poaceae). *Taxon* **56**: 145–156.

Zuloaga FO, Morrone O, Davidse G, Pennington SJ. 2007a. Classification and biogeography of Panicoideae (Poaceae) in the New World. In: Columbus JT, Friar EA, Hamilton CW, Porter JM, Prince LM, Simpson MG, eds. *Monocots: comparative biology and evolution – Poales*. California, CA, USA: Rancho Santa Ana Botanic Garden, 503–529.

Zuloaga FO, Morrone O, Scataglini MA. 2011. Monograph of *Trichanthecium*, a new genus segregated from *Panicum* (Poaceae, Paniceae) based on morphological and molecular data. *Systematic Botany Monographs*, in press.

Zuloaga FO, Scataglini MA, Morrone O. 2010. A phylogenetic evaluation of *Panicum* sects. Agrostoidea, Megista, Prionita and Tenera (Panicoideae, Poaceae): two new genera, *Stephostachys* and *Sorengia*. *Taxon* **59**: 1535–1546.

Supporting Information

Additional supporting information may be found in the online version of this article.

Fig. S1 Bayesian consensus tree of 545 accessions, with statistical support for all nodes.

Fig. S2 Bayesian consensus tree of 545 accessions with major clades collapsed.

Table S1 Primers used for sequence amplification in this study

Table S2 List of taxa included in the phylogenetic analysis

Table S3 Estimated parameters from character evolution analyses