PROJECT DESCRIPTION

PREVIOUS SUBMISSIONS

Our project, to study ecological diversification of grasses in a phylogenetic framework, was developed in three earlier proposals (MCB0744009, DEB0918605, and DEB1021271). Last year "…all reviewers agree that the team of PIs have an outstanding expertise and strong publication record." Most reviewers "…found this proposal very interesting and the use of anatomical and morphological goals was well received..." The panel indicated that the work would "…certainly… advance our understanding of the evolution of grass morphology." However, the program director and the panel were clear about an essential improvement: a "…major weakness…is the absence of nuclear genes that would benefit the evolutionary history of this plant group" (DEB1021271, Panel summary; S. Scheiner, pers. comm.).

Here we keep our strong team, although some roles have shifted—PIs: L. Clark (Iowa State University), M. Duvall (Northern Illinois University), and S. Kelchner (Idaho State University); Consulting/collaborative partners, E. Edwards (Brown University), E. Kellogg (University of Missouri, St. Louis), Mike Muszynski (Iowa State University), J. Saarela (Canadian Museum of Nature), and F. Zuloaga (Instituto de Botánica Darwinion). We maintain high sampling within crown grasses and emphasize ecological diversification by detailed studies of vegetative adaptations. For phylogenetics, we add ten nuclear loci to be sequenced and analyzed with the plastome data.

INTRODUCTION AND OBJECTIVES

Introduction. The grasses (Poaceae) dominate ecosystems throughout the world. The prairies of North America and the vast grasslands of Africa, the annual grasslands of the Mediterranean and California, the steppes of Russia, the pampas of Argentina, and the great bamboo forests of South America and Asia all attest to the power of this one family of plants. But each of these ecosystems has its own set of species. The North American and African grasslands are largely covered with members of the subfamilies Panicoideae and Chloridoideae, with their high efficiency form of photosynthesis (the C_4 pathway); the Mediterranean, Californian, Russian, and Argentine grasslands are populated by members of subfamily Pooideae, with their tolerance of and indeed need for cool winters; and the bamboo forests are made up of members of subfamily Bambusoideae, with their striking and uniquely derived plant architecture. In all cases the origins of these phenotypes and their responses to their environments present an evolutionary puzzle the context for which will be provided by the proposed research.

Poaceae are especially appropriate for a taxonomically focused exploration using plastomes. Grasses radiated rapidly (Bouchenak-Khelladi et al. 2008; GPWG I 2000, 2001; Mathews et al. 2000; Prasad et al. 2005) into a species-rich lineage—among monocot families Poaceae is second in size only to Orchidaceae. The family is now considered to comprise 12 subfamilies (Fig. 1), each of which is demonstrably monophyletic (GPWG I 2001; Sánchez-Ken et al. 2007; Duvall et al. 2007; Bouchenak-Khelladi et al. 2008; Sánchez-Ken & Clark 2010). Following a grade of early-diverging, species-poor lineages (subfamilies Anomochlooideae, Pharoideae and Puelioideae) are two major lineages of grasses that define a fundamental bifurcation, the BEP and the PACMAD clades. Three of the subfamilies—Bambusoideae, Ehrhartoideae and Pooideae—form the BEP clade and six belong to the PACMAD clade—Panicoideae, Arundinoideae, Chloridoideae, Micrairoideae, Aristidoideae, and Danthonioideae. Together these nine subfamilies include over 99% of the ca. 10,000 species in the family.

Complex structural and physiological adaptations, both vegetative and reproductive, in subgroups of grasses have arisen in diverse habitats sustaining the ongoing radiation. The phylogenetic signal of multi-gene data sets such as those studied by the first Grass Phylogeny Working Group (GPWG I), however, has proved insufficient to resolve branching order within the BEP and PACMAD clades, likely due to rapid diversification, selection, or other processes. Relationships between grass taxa and the origins of specific adaptations have thus remained obscure, despite considerable work by many investigators (e.g., Bouchenak-Khelladi et al. 2008; GPWG I 2001; Sánchez-Ken & Clark 2007; Sánchez-Ken et al. 2007). We propose here to rectify this problem by using data from whole plastomes and informative

nuclear loci, which provide over ten-fold more informative characters than in the most comprehensive previous multi-gene analyses (Leseberg & Duvall 2009).

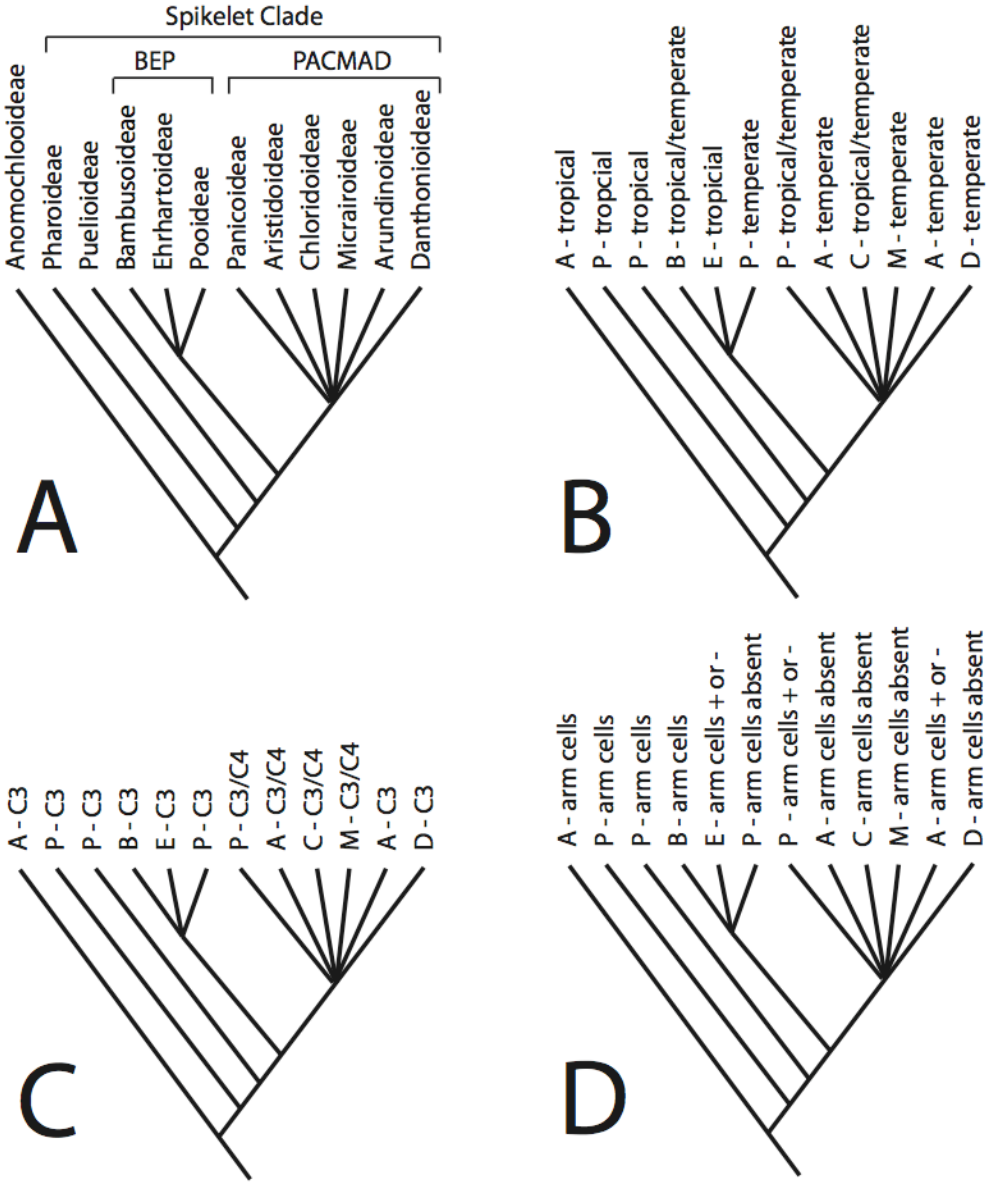

Fig. 1. A. Phylogeny of the grasses. All resolved nodes and all subfamilies are strongly supported. B-D. Examples of major ecological, physiological and morphological characteristics whose evolution depends on resolution of the BEP and PACMAD clades; others described in the text. B. Temperate vs. tropical habitat. Note that the distribution of Pooideae in temperate zones may be derived and synapomorphic, but could conceivably be optimized as ancestral depending on the resolution of the BEP clade. C. C_3 vs. C_4 photosynthesis. C_4 is likely to be independently derived in each of the subfamilies in which it occurs, but optimization of the character and dating of its origin depends on the resolution of the PACMAD subfamilies. D. Arm cells are mesophyll cells in the leaf with invaginated cell walls in cross section; these are plesiomorphic for the family but characterize the Bambusoideae and also appear elsewhere. They are likely to be plesiomorphic in that subfamily, but some character optimizations suggest that they could be independently derived.

A few recent analyses apparently have confirmed a sister relationship between the Bambusoideae and Pooideae (e.g., Bouchenak-Khelladi et al. 2008; Triplett & Clark 2010), but other analyses are unable to reject alternate topologies (e.g., Bamboo Phylogeny Group, in prep.). The uncertain position of *Streptogyna* (not shown in Fig. 1 but part of the BEP polytomy) also complicates phylogenetic inference in this clade. Some analyses infer a sister relationship between the Panicoideae and a clade containing the other five subfamilies of the PACMAD clade (the ligule of hairs clade, GPWG I 2001; Bouchenak-Khelladi et al. 2008), but resolution among the remaining five subfamilies is lacking or poorly supported and alternate topologies have been recovered in other analyses (Christin et al. 2008; Vicentini et al. 2008). We therefore represent the constituent subfamilies of each of the BEP and PACMAD clades as polytomies (Fig. 1) as this more accurately reflects the current state of knowledge.

Our ability to understand the evolution and rise to dominance of the grasses is hamstrung without resolution within these major clades. As described in more detail below, questions such as the origin of C_4 photosynthesis and the accompanying leaf anatomical changes in the PACMAD clade, the origin of cold (vernalization) requirements in the Pooideae, and the origin of the striking architecture, leaf, and floral structure of the bamboos all require that the two major polytomies be resolved. Our preliminary data suggests that the problem can be solved by addition of more taxa and more base pairs of DNA sequence from the plastome.

Objectives. The objectives of this proposal are to: 1) sequence ca. 100 plastomes of grasses and analyze them (with 57 other available plastomes) phylogenetically to break polytomies at the base of the BEP and PACMAD clades and obtain the most fully resolved phylogeny possible; 2) explore the tempo and mode of evolution in grass plastomes; 3) sequence these ca. 160 taxa for 10 nuclear loci to more fully explore the evolutionary history of the family; and 4) use the results to test the following major hypotheses.

A) The synapomorphic structure of grass leaves is a response to the moist shady habitats in which the family evolved, rather than the dry open habitats into which they diversified.

B) The developmental program of the distinctive fusoid cells shares elements with developmental programs deployed elsewhere in the leaf.

C) The 18 previously identified lineages of C_4 grasses largely represent independent origins with few if any reversals; leaf anatomical features associated with C_4 show different patterns among the major C_4 lineages.

D) Microstructural mutations, such as indels and inversions, are as informative as the substitution mutations in coding regions of the plastome.

E) Comparative analyses of plastome and nuclear loci will clarify complex evolutionary processes in the grass crown group; incongruence is expected at shallower levels.

We will thus address unresolved issues of the timing of evolutionary divergences and better understand the evolution and development of a subset of the factors responsible for the ecological diversification of the grasses to dissect major shifts in grass ecological dominance and the morphological changes that correlate with the environment.

SIGNIFICANCE

Outcomes/products of this project will include 1) a fully resolved phylogeny of the grasses with respect to the major lineages of the BEP and PACMAD clades; 2) a much improved understanding of plastome evolution with a focus on microstructural changes in non-coding regions; 3) grass-specific nuclear and chloroplast molecular tools; 4) Web resources for Poaceae in both English and Spanish organized around the phylogeny; 5) elucidation of the evolution and development of vegetative characters (especially the leaf) and physiology (C_3/C_4 photosynthesis); and 6) a deeper understanding of how these features promoted the ecological diversification of the grasses.

This project offers an unprecedented opportunity to obtain full resolution of major branching events in the grasses, a species-rich plant lineage of paramount ecological and economic importance. We focus on the evolution and development of vegetative features because of the existing body of work on grass leaf structure and C_4 photosynthesis and the likelihood that this suite of features played a critical

role in the ecological diversification of the grasses. With the phylogeny, we can examine this relationship and test a variety of hypotheses about morphological evolution and ecological diversification. The large amount of data from the complete plastomes and 10 nuclear loci we will analyze will also permit us to refine our understanding of the timing of evolutionary divergences within the grasses. Beyond our proposed work, a robust phylogeny for the grasses will facilitate exploration of a wide variety of evolutionary and ecological questions, including those related to responses to global climate change.

The availability of this number of plastomes from a single, diverse lineage also provides the potential to greatly extend our understanding of molecular evolution and the phylogenetic utility of mutational events within the plastome. Analysis of non-coding regions will provide the first well-sampled examination and modeling of microstructural changes in plastome DNA, which has been a limiting step in the use of introns and intergenic spacers in phylogenetic analyses. The results from this study will also complement on-going projects generating whole plastomes.

Much of the available information on grass structure, diversity, classification, and evolution is out-of-date, geographically restricted, or highly technical. The robust phylogeny we propose to generate provides a way to organize this information on grasses in an image-rich and user-friendly manner on the proposed Web site. The focus will be on major clades, subfamilies, and tribes. Content will be provided in both English and Spanish, and will be complementary to existing on-line resources.

The project is feasible within the timeframe given existing and in-progress grass plastomes, advances in high-throughput sequencing, the large number of samples in hand or available from international collaborators and other sources, and its collaborative approach, with expertise in grass systematics and evolution, molecular phylogenetics, modeling, and development and gene expression.

PROPOSED WORK

Sampling. Complete plastome and selected nuclear loci (see below) will be sequenced from 100 species of BEP and PACMAD grasses (Table 1). These species were selected from all subfamilies to represent ecological diversification and leaf specializations in Poaceae and permit testing of the hypotheses outlined in the introduction. Other plastomes will be included in the analysis [17 are completed and 40 more are underway (Table 1)] so that a total of ca. 158 grass plastomes can be analyzed. These include representatives of the small, early-diverging Anomochlooideae (2 spp.; Givnish et al. in press; Morris & Duvall 2010), Pharoideae, and Puelioideae (Duvall et al. 2010). Draft plastomes of six outgroups, which emphasize conserved coding regions are also completed for Ecdeiocoleaceae (2 spp.), Joinvilleaceae (2 spp.), Flagellariaceae and Restionaceae (Givnish et al. in press).

Specifically, we have included representatives of the 18 C_4 lineages and their C_3 sister groups that were previously identified (Christin et al. 2008). For the most species-rich subfamilies of Poaceae (Bambusoideae, Chloridoideae, Panicoideae, Pooideae), we selected additional taxa to represent as many tribes or major clades (e.g., x = 9 and x = 10 Paniceae) as possible. In a few cases, two or more congenerics (*Bambusa*) or conspecifics (*Alloteropsis*) are represented to test the resolving power of the plastome or to address specific evolutionary transitions. For example, *Steinchisma hians* is a C_3/C_4 intermediate whereas *S. laxum* is C_3 (see Duvall et al. 2003). Of the species we propose to sequence, frozen or silica-dried material is either on hand in the labs of the PIs, or can be obtained from members of the GPWG I or colleagues involved in this project (please see letters from J. Saarela and F. Zuloaga). Leaf samples are also available at the "DNA bank" of the Missouri Botanical Garden, St. Louis, or other gardens and herbaria. In all cases, specimen vouchers or voucher information will be obtained.

Sequencing. Illumina and 454 sequencing methods offer different combinations of read length, coverage, and cost per base. The relatively longer reads (up to 400 bases with a potential increase to 1000 bases in the near future) of the economical 454 Genome Sequencing Method will facilitate the complex task of complete plastome assembly. This method is available to us at the University of Iowa DNA facility under an instrument-sharing agreement with Iowa State University. For this project, DNAs will be obtained from either silica-dried or fresh tissues. The method of pooling long-range plastome and shorter nuclear amplicons will be used (Cronn et al. 2008; Goremykin et al. 2005). These fragments will be amplified with conserved grass-specific primers (Dhingra & Folta 2005; Leseberg & Duvall 2009).

Libraries will be constructed (Jansen et al. 2005) by Duvall's lab group and submitted through Clark for 454 sequencing. Sequence fragments will be assembled (Geneious Pro vers. 3.8; Drummond et al. 2007) and annotated with DOGMA software Wyman et al. 2004) and submitted to GenBank.

Table 1. BEP and PACMAD taxa in this study. The number of taxa in subgroups is given in parentheses. **Boldface** = plastomes completed/underway; * = plastomes in the Monocot Tree-of-Life project. C_4 taxa are highlighted in grey; C_3/C_4 intermediates or those potentially so, are boxed. GenBank accessions are given for banked plastomes.

POACEAE 154 + Early-diverging lineages (4) + Outgroups (6) = 164
BEP CLADE: BAMBUSOIDEAE (16)—***Acidosasa purpurea, Arundinaria gigantea, Bambusa oldhamii*** (FJ970915), **Bambusa stenostachya, **Chimonocalamus pallens**, *Chusquea circinata, **Cryptochloa strictiflora, Dendrocalamus latiflorus** (NC013088), **Fargesia nitida, Ferrocalamus rimosivaginus**, Guadua angustifolia, Melocanna baccifera, *Olyra latifolia, **Phyllostachys nigra var. henonis**, Thamnocalamus tessellatus, Yushania niitakayamensis* EHRHARTOIDEAE (7)—**Ehrharta calycina, E. dura, Humbertochloa bambusoides, Leersia oryzoides, **Oryza nivara** (NC_005973), **O. sativa** var. **japonica** (NC_001320), Phyllorachis sagittata* POOIDEAE (17)—***Agrostis stolonifera*** (NC_008591), **Ampelodesmos mauritanica*, *Anisopogon avenaceus*, **Avena sativa, *Brachyelytrum aristosum, **Brachypodium distachyon** (EU325680), *Bromus inermis*, *Diarrhena obovata, **Festuca arundinacea** (FJ466687), **Hordeum vulgare** (NC_008590), *Lolium multiflorum*, ***L. perenne*** (AM777385), *Lygeum spartum, *Melica uniflora, Phaenosperma globosum, *Piptatherum hymenoides, **Triticum aestivum** (NC_002762)*
PACMAD CLADE: PANICOIDEAE (70)—*Acroceras tonkinense*, *Alloteropsis semialata* (2), *Anthaenantiopsis rojasiana, Arthropogon lanceolatus, A. villosus, Arundinella hirta, Axonopus anceps, Brachiaria villosa, Cenchrus ciliaris, Centotheca lappacea, **Chasmanthium latifolium**, C. laxum, **Coix lacryma-jobi*** (FJ261955)*, Cyperochloa hirsuta, Cyphonanthus discrepans, Danthoniopsis dinteri, Dichanthelium dichotomum, Digitaria insularis, D. sanguinalis, Echinochloa colona, Echinolaena inflexa, *Gynerium sagittatum, Homolepis glutinosa, H. isocalycia, Hymenachne donacifolia, Ichnanthus pallens, Lasiacis sorghoides, Leptocoryphium lanatum, Loudetia simplex, Melinis minutiflora, Mesosetum chaseae, **Microcalamus convallarioides**, Neurachne alopecuroides, N. munroi, N. tenuifolia, Ophiochloa hydrolithica, Oplismenus hirtellus, Orthoclada laxa, Otachyrium versicolor, Panicum anceps, P. euprepes, P. bulbosum, P. laxum, P. miliaceum, P. ovuliferum, P. prionitis, *P. virgatum, Parodiophyllochloa cordovensis, Paspalum conjugatum, P. fimbriatum, P. geminatum, Pennisetum glaucum, Plagiantha tenella, Pseudechinolaena polystachya, **Saccharum officinarum*** (NC_006084)*, Sacciolepis indica, *Setaria italica, **Sorghum bicolor*** (NC_008602)*, Steinchisma hians, S. laxum, Streptostachys asperifolia, S. rasoma, Tatianyx arnacites, *Thysanolaena latifolia, Tristachya leucothrix, Urochloa maxima, **Zea mays*** (NC_001666), *Zeugites mexicana, Zuloagaea bulbosa* ARUNDINOIDEAE (6)—*Amphipogon strictus*, *Arundo donax, Dregeochloa calvinensis, D. pumila, Molinia caerulea, Phragmites australis* CHLORIDOIDEAE (20)—**Bouteloua curtipendula, Centropodia glauca, Chloris verticillata, Cottea pappophoroides, *Eleusine coracana, Eragrostis capensis, E. minor, **E. tef, Hilaria cenchroides**, Merxmuellera rangei, Muhlenbergia schreberi, **Neyraudia reynaudiana**, Orcuttia californica, *Spartina pectinata, *Sporobolus indicus, Triodia desertorum, Triraphis ramosissima, Uniola paniculata,*Zoysia matrella, **Z. macrantha**.* MICRAIROIDEAE (6)—**Eriachne festucacea, E. triseta, Isachne arundinacea, I. mauritania, Micraira spiciforma, *M. subulifolia* ARISTIDOIDEAE (5)—**Aristida adscensionis, A. longifolia, Sartidia jucunda, *Stipagrostis plumose, S. zeyheri* DANTHONIOIDEAE (6)—*Austrodanthonia laevis*, *Cortaderia richardii*, *Danthonia californica, Karroochloa purpurea, Merxmuellera macowanii, *M. lupulina*
INCERTAE SEDIS (1)—**Streptogyna americana*

Nuclear Loci. In response to the reviewers of our 2010 proposal, previously characterized nuclear loci were evaluated to add to our phylogenetic studies. In Poaceae nuclear sequences suggest complex evolution accelerated by processes such as polyploid hybridization, transposon activity, lineage sorting, and selection. Most grasses are believed to be polyploids or paleopolyploids with extensive duplication of loci (GPWG 2001; Levy & Feldman 2002). Widespread interspecific hybridizations in Poaceae, often accompanied by polyploidization, have been reported (Fan et al. 2009; Mason-Gamer 2004; Triplett et al. 2010; and many others). Transposon insertions have increased overall genome sizes (Kellogg 2000; SanMiguel et al. 1998) and lengthened specific loci (Mason-Gamer et al. 2010). Incomplete lineage sorting is a special concern for recently diverged taxa, as in *Hordeum* (Jakob & Blattner 2006). Disruptive positive selection of specialized isoforms has also been documented (Monson 2003), which can interfere with phylogenetic inference.

Ten nuclear loci were selected for this project on the basis of low-copy number, feasibility of sequencing, and demonstrated phylogenetic utility in Poaceae (Table 2). Although PEPC has been used recently (Mason-Gamer et al. 2010), it was not included because of the confounding effects of disruptive selection (Christin et al. 2007). These ten loci will be sequenced from the same taxa as in Table 1. Other loci, such as those that encode endo-1,4-β glucanase, poly-A binding protein, etc., which are under development as phylogenetic tools in various Poaceae, will be considered for future studies.

Table 2. Nuclear loci to be sequenced.

Loci	Protein	Sources
Acc1	acetyl-CoA carboxylase	Fan et al. 2009
Adh2	alcohol dehydrogenase	Gaut et al. 1999; Ge et al. 2004
BAM	β-amylase	Mason-Gamer 2005; Mason-Gamer et al. 2010
PHYA, *PHYB*, *PHYC*	phytochromes A, B, and C	Mathews et al. 2000; Mathews & Sharrock 1996
GBSSI (*waxy*)	granule-bound starch synthase	Mason-Gamer et al. 1998
zfl1, zfl2	FLORICAULA/LEAFY	Bomblies & Doebley 2005
tb1	teosinte branched-1	Lukens & Doebley 2001

Phylogenetic analyses and estimation of divergences. A grass plastome is expected to represent a linkage group of loci that share the same evolutionary history. In this project at least 81 conserved coding sequences will be analyzed (following Jansen et al. 2007; Givnish et al. in press). Incongruence of signal among plastome regions can occur, however, and is commonly due to insufficient data variability, model inadequacy, or analytical bias (Kelchner 2009). True historical incongruence among chloroplast data sets is less common, but may occur from heteroplasmy, recombination, or paralogy (Wolfe & Randle 2004). Because grass chloroplasts are usually maternally inherited, there is also a possibility that plastome signal will not reflect true organismal history, though the potential is much less pronounced at higher taxonomic levels (e.g., among major lineages or subfamilies as in this study). A multilocus analysis of nuclear data, performed in conjunction with a plastome phylogeny estimation, can provide insight into how nuclear and chloroplast histories might differ, possibly because of specific processes known to affect evolution in grasses such as polyploid hybridization and incomplete lineage sorting.

Given the rapid developments in phylogenetic methodology, and complexity of potential causes for incongruence within and among data sets in this study, we propose an analytical framework that seeks to answer several questions of interest about data, model, signal, and estimate quality. The tool set can easily be altered if improvements are made to software or theory prior to analysis. For example, methods to infer species trees from multilocus molecular data that may be affected by multiple processes such as incomplete lineage sorting and introgression are being developed (BEST: Liu & Pearl, 2007; Liu et al., 2008; STEM-hy: Kubatko 2009). Accuracy and precision might prove critically important in some parts of the tree: if a rapid radiation event generated the PACMAD polytomy, for example, any estimated

resolution of the node will necessarily rely on high quality characters, appropriate model selection, and careful assessment of signal, including an exploration of potentially weak features of the estimate.

Data quality will be explored by visual inspection of DNA sequence reads, criterion-based optimization of homology statements (see ***Molecular evolutionary analyses***), and network-type data exploration within and among data partitions using SPLITSTREE (Huson & Bryant 2006). Base composition bias and site saturation will also be investigated. Character sets will be defined for problematic data to assist in character addition-removal experiments during phylogenetic inference.

Model adequacy for individual and combined data sets will be assessed with an adapted version of JMODELTEST (Posada 2008) that improves control over the candidate pool of models. Results of dynamic likelihood ratio tests will be compared to a list of models with Akaike Information Criterion differences ≤ 2 (Posada & Buckley 2004). Goodness-of-fit testing will also be performed using parametric bootstrapping (Sullivan & Joyce 2005).

Signal quality can be investigated by testing for bias and error in the analytical process. There is no guarantee that evolutionary history will be a hierarchical process in each data set, or that a hierarchical signal best represents phylogeny in grasses. Hence, we will experiment with taxon and data removal, outgroup rooting positions, and robustness of each topology to reasonable changes in the model of character evolution. Neighbor-net analyses will help identify data partitions with high character conflict and uncertainty that could still produce strong bootstrap values or posterior probabilities on a bifurcating tree topology.

Quality of the estimate can be assessed by consensus network analyses, clade support measures (nonparametric bootstrap, posterior probabilities), and alternative hypothesis testing (Shimodaira & Hasegawa 1999; Shimodaira 2002). Robustness of the phylogeny estimation can be tested in part by varying the models of character evolution used; hence, our study will include independently derived estimates from maximum parsimony (PAUP*, Swofford 2002), maximum likelihood (GARLI, Zwickl 2006), and Bayesian inference (MRBAYES, Ronquist & Huelsenbeck 2003). Indel characters will be analyzed separately to test their phylogenetic performance, and combined with nucleotide data if signal is not strongly incongruent and the analytical framework can accept such data (e.g., parsimony, Bayesian inference).

Divergence times for all phylogenetic nodes will be estimated to relate them to each other and to documented historical events such as known climate shifts. A Bayesian framework will be used to estimate divergence dates; analyses will employ an uncorrelated lognormal model of a relaxed clock analysis (BEAST, Drumond & Rambaut 2007). Conservatively interpreted fossil descriptions will be used to set a minimum of four calibration points including the stem node of the BEP-PACMAD clade (see Christin et al. 2008, supplemental procedures). Preliminary "tuning" runs will allow us to optimize the final analyses; we expect posterior distributions of parameters will require at least two runs of four independent Markov chain Monte Carlo analyses for 60 million generations (150 taxa), depending on convergence estimates. Convergence of the chains will be assessed using TRACER (Rambaut & Drummond 2007).

Morphological and physiological characters can be optimized on the phylogeny using maximum likelihood, as implemented in MESQUITE (Maddison & Maddison 1999-2009) and LASRDISC (Jackson 2004). Improved methods may be developed by the time of analysis and will be considered.

Molecular evolutionary analyses. One goal of this project is to identify the taxonomic distribution and frequency of indels and inversions (together constituting microstructural mutations). Such changes are easy to identify if they are rare and if indels themselves do not vary in length, but more complex changes (e.g., nested or overlapping indels) are difficult to interpret, particularly because alignment software performs poorly when there is significant length variation among sequences (Morrison 2006; Wong et al. 2008). One solution is criterion-based manual alignment (Morrison 2009b) a process that attempts to identify sequence patterns and secondary structures (e.g., potential hairpins) that are associated with microstructural changes in DNA sequences (Kelchner 2000; Graham et al. 2000). Using such sequence-based information can improve primary homology estimates of indel characters

(Kelchner & Clark 1997; Quandt & Stech 2004), especially when stepwise gaps are recognized as multiple independent length mutation events (Morrison 2009a). Although the approach will likely be incorporated into future alignment algorithms (Morrison 2009a), the proposed study currently requires manual confirmations for alignments of all coding and non-coding regions using the method outlined in Kelchner (2000) and updated most recently by Morrison (2009a).

Length mutations and inversions can be categorized by reference to sequence pattern and inferred mutational mechanism (Kelchner 2000). Microstructural changes will be sorted and counted by region (with single-copy and inverted repeat regions of the plastome treated separately). Genome location, proximity to secondary structure, and distances between mutations will be recorded. Ambiguous regions will be partitioned and treated separately during genomic phylogeny estimation.

A survey of four non-coding regions across 40 species of bamboos (Bamboo Phylogeny Group, data not shown) indicated that microstructural changes occur in this lineage at a frequency of roughly 40 per kilobase in introns and 60 per kilobase in intergenic spacers. Similar levels were seen in complete plastomes of *Coix* and *Anomochloa* (Leseberg & Duvall 2009; Morris & Duvall 2010). By extrapolation, we anticipate that the proposed study should recover a dataset of several thousand microstructural changes. The indel dataset would derive from a well-sampled lineage with growing phylogenetic resolution and a sample size large enough for statistical tests of alternative hypotheses of indel evolution. It would also provide enough observations to predict the evolution of microstructural changes and test likelihood models for phylogenetic analysis of indel characters.

With a data set of conserved microstructural changes among the plastome sequences, we can then explore (1) size classes and different approaches to categorizing indel and inversion types, (2) frequency of each type and their correlation with primary and secondary structure of DNA and RNA sequences, (3) the distribution of such changes in introns, spacers, and coding regions, (4) the treatment of microstructural changes as discrete characters in phylogenetic analysis. The phylogeny based on substitutions in plastome coding regions will provide a comparison against which can be tested the phylogenetic performance of indels, the adequacy of current phylogenetic models for microstructural data, and the potential for improvement of indel and inversion use in phylogenetic inference. It may be that simple markov models for discrete character data (e.g., Lewis 2001) will be sufficient once repeatable categorization of indels is established. Alternatively, our observations might point the way to developing a more complex rate matrix approach to phylogenetic models for microstructural changes.

Evolution and development of vegetative features. We will use our most robust phylogeny to investigate the developmental biology and potential adaptive significance of over 20 characters of grass vegetative structure and anatomy. Among these are life form, life cycle, vegetative growth and reproduction (rhizomes, tillering, aerial branching), and leaf structure and physiology (C_3/C_4). Space does not permit detailed description of all the characters we plan to investigate, so we focus here on representative ones. Some or most of these characters are likely to be homoplasious, although because of the lack of resolution within the BEP and PACMAD clades, it is hard to tell exactly where the losses occur and if the characters reverse. This project will permit assessment of the pattern of evolution of the characters. Testable hypotheses are highlighted in boldface.

At the same time we plan a set of developmental studies, investigating histology and expression of relevant genes. We anticipate that we will be able to explore questions of homology and identify additional developmental events that provide morphological characters for subclades. In the process, we expect to develop better hypotheses of the function of various structures to test a number of these (see below).

In addition to scoring the vegetative characters, members of the Clark lab will investigate the anatomy and development of grass leaves in some detail. Clark plans to integrate some of this work into her plant systematics and plant anatomy courses with undergraduate students collecting portions of the primary data. For anatomy, we will sample as many species as possible for which we have whole plastomes; we anticipate that most will be available (and a few have been studied in detail already—e.g., *Streptogyna*, Soderstrom et al. 1987). We will sample a subset of these for developmental studies,

targeting a minimum of two species per major clade. Vegetative shoot apices will be dehydrated in an ethanol series, embedded in paraffin, sectioned and stained. Because leaves of different ages surround the young shoot meristem, it is easy to obtain leaves of multiple stages in a single specimen. Sections will be done both crosswise and lengthwise, relative to the long axis of the leaf. Leaf clearings will also be prepared for the same taxa. Our goal is to develop a three-dimensional picture of the structure of leaf cells as they develop and at functional maturity. Our observations will allow us to make the character descriptions more precise, and in many cases more quantitative, in that they will focus on detailed aspects of cell structure. This in turn will permit greater precision of mapping chars on the phylogeny.

Mesophyll cell wall lobes (arm cells): If the leaf of rice is cut crosswise (transverse section), the cell walls are clearly invaginated. Such cells constitute a synapomorphy for the grasses, and are present in the early-diverging lineages (Fig. 1D). They also occur in all bamboos (Fig. 3A, *C. depauperata*), in some members of subfamily Ehrhartoideae (e.g., rice and its relatives) and in *Phragmites* (Arundinoideae), but appear to be absent in many clades such as the Pooideae and Danthonioideae (Fig. 1D). Because of the lack of resolution at the base of the BEP clade, it is unclear whether this represents a plesiomorphy or a reversal following loss. Preliminary work (Chonan 1970; Sánchez-Ken & Clark, unpubl. data; see **Preliminary Results**) also indicates that there is significant variation in these cells in longitudinal view. We propose to examine additional taxa and re-evaluate and test hypotheses regarding the evolution and adaptive significance of arm cells in light of the phylogeny. For example, we will test the hypothesis that **arm cells occur in plants of wet or shady sites** and the complementary hypothesis that **longitudinally elongated lobed cells occur in plants of open habitats**. We will examine the ecological distribution of all types of arm cells that occur in grasses.

To determine whether the pattern of microtubules or actin prefigures the formation of lobes, we will do immunolocalizations of tubulin and actin using commercial antibodies on leaves throughout development. Immunolocalization experiments are rapid and straightforward (e.g., Sinha & Kellogg 1996). It is convenient to mount sections from several species on the same microscope slide so that interspecific comparisons can be done as part of a single experiment. We will use a secondary antibody with a fluorescent tag so that the cells can be imaged on a confocal microscope.

Fusoid cells: Fusoid cells are mysterious, often large, rectangular to cigar-shaped and curiously empty cells in leaf cross sections (Fig. 3B). Like arm cells, they are synapomorphic for the grasses; the plants in which they occur are shade-loving and prefer (require?) moist habitats. Unlike arm cells, fusoid cells occur only in the early-diverging lineages, *Streptogyna*, and the Bambusoideae. A correlation has been demonstrated between growth in shade and presence of fusoid cells (March & Clark, in press) in bamboos. Superficially similar cells have been reported in Panicoideae s.l., but their position (part of the bundle sheath) indicates that they are non-homologous. The function and development of fusoid cells are unknown. They have been suggested to be filled with water, with CO_2, or simply with air, but it is difficult to study their contents. It is not even clear whether they are alive or dead. They could be dead cells created by programmed cell death of multiple mesophyll cells, or could be a single enlarged cell that has died (and possibly also collapsed), or perhaps it is alive and has simply lost its organelles.

We propose to investigate fusoid cells in section using a series of histochemical and immunochemical stains to identify particular cellular compartments. First, to determine whether the cells are alive at all, we will use one or more vital dyes such as Evans blue or Phenosafranin (excluded from living cells) or fluorescein diacetate or rhodamine (which turn living cells yellow green or red, respectively). We will assess the cells at different stages of development to determine when they die (if they do). We will stain cells with DAPI, a fluorescent stain for DNA, to determine whether fusoids ever had nuclei at any stage of their development. Likewise, because chloroplasts autofluoresce strongly, looking at a developmental series under fluorescent light will indicate if and when chloroplasts break down. MitoTracker will be used to follow the fate of mitochondria. We wish to determine whether a fusoid cell has an intact plasma membrane and tonoplast. While a variety of stains will distinguish between phospholipids and other lipids, the best specificity can be obtained by membrane specific markers. Accordingly, we will do immunolocalizations using commercial antibodies to an H+ATPase (Agrisera, Sweden), an enzyme that is localized to the plasma membrane. Likewise Vaculolar ATPase is

a marker for the tonoplast and will test whether there is a vacuolar membrane. UGPase and sucrose phosphate synthase are markers of cytoplasm.

These tests will allow us to answer several basic questions about fusoid cells and their development and to test the hypothesis, for example, that **fusoid cells borrow a developmental program that is already active somewhere else in the leaf**. We will further examine the differential development of fusoid cells in sun and shade leaves, and test the hypothesis that **fusoid cells function in light-scattering in shady environments**. We will also then be able to repeat the most informative assays on a wide range of species to determine the similarities and differences among the fusoids of the early-diverging grasses, the bamboos, and even the presumably non-homologous "fusoids" in the centothecoid grasses (Panicoideae).

Cell density: In some species cells are closely packed in both cross and longitudinal sections (Fig. 3C-D, *Sporobolus*) whereas in others they are closely packed in cross section (*Chusquea*, Fig. 3A) but show extensive air spaces in longitudinal section. In other species, cells are loosely packed when viewed in cross section but they may be tightly packed in longitudinal section (*Brachyelytrum*, not shown). In a few taxa, the upper (adaxial) part of the mesophyll is tightly packed, forming a palisade-like layer, whereas the lower (abaxial) part is much looser. We will test the hypothesis that **C_4 taxa have minimal air spaces (i.e., tightly packed mesophyll in all dimensions) relative to C_3 taxa**.

We also hypothesize that the **differences in cell density in the leaf reflect heterochronic changes in cell division versus expansion**. The palisade layer continues to divide as the leaf is expanding, whereas the lower spongy layer is less dense because cell division ceases while the leaf is expanding, so the cells pull apart. Similarly, in leaves in which the cells appear dense in cross section, cell division in the lateral plane keeps up with tissue expansion; if the cells are loosely packed in longitudinal section it suggests that growth in the proximo-distal direction outpaces cell division.

Bundle sheath plastids: The presence and orientation of plastids in the bundle sheath is correlated with C_4 (high efficiency) photosynthesis. In general, plastids in C_4 grasses proliferate in the bundle sheath, whereas bundle sheaths of C_3 grasses have only a few poorly developed plastids or proplastids. As with the characters described above, this is usually considered to be a binary character by phylogeneticists. However, the number and position of plastids in the bundle sheath varies among C_4 plants. For example, the leaves of *Eriachne* (C_4) can be identified easily in cross section by the extreme density of the plastids, whereas in other C_4 plants the plastids may be aligned along the wall away from the vein (centrifugal) or towards the vein (centripetal). One hypothesis we will test is that **all C_3 grasses have poorly developed plastids in their bundle sheaths**.

Pseudopetioles: As in other monocots, the base of the leaf sheaths the stem and the upper (distal) part forms the blade. Between the sheath and blade, a thin stalk develops; because it is clearly not homologous with a dicot petiole, it is known as a pseudopetiole. Like cells with invaginated walls, and fusoid cells, presence of a pseudopetiole is a synapomorphy for the grasses, but is then repeatedly lost (and possibly regained) in evolution.

Developmentally, the blade forms first and the sheath is then intercalated afterward. But how the pseudopetiole forms is unknown. Likewise its adaptive significance is unclear, although it tends to occur in species of shady environments and could be involved in orientation of the leaf; a pulvinus is often present at least at the base of the pseudopetiole in many taxa. We will examine its structure and development, and test the hypothesis that **the pseudopetiole is correlated with shady environments**.

Leaf intercalary meristems: Grasses are well known for the presence of intercalary meristems in their leaves and stems. These are meristematic regions, normally at the base of the internode or the base of the blade, where cell divisions occur more or less throughout the life of the leaf so that the cells never fully differentiate. Leaf intercalary meristems are assumed to be present in most members of the BEP + PACMAD clades, and absent in taxa with pseudopetioles (e.g., the early-diverging lineages and the Bambusoideae). We intend to investigate this with greater precision by sectioning leaf blades at different stages of development to track cell division in the base of the blade. By investigating several leaves per plant, we can develop a quantitative estimate of the percentage of dividing cells. One

hypothesis we will test is that **taxa with pseudopetioles lack a leaf interacalary meristem at any stage of development**.

Ecological diversification. Methods and data are now available so that morphological characteristics can be correlated directly with environmental variables. We will use the available data in the Worldclim database (temperature and precipitation as summarized by bioclim), coupled with locality data for grass specimens in GBIF. A new resource, the Grass Portal is under development, which will make available large numbers of georeferenced grass specimens with bioclim data already attached. Duvall has been in contact with the developer of the portal (Dr. Colin Osborne, Sheffield University) and will be a beta tester for the resource as it comes on-line in early 2011; this should speed up much of this part of the project (see also the letter from E. Edwards).

These data will be assembled to determine the mean and variance for each species and also for each clade for each of the assessments of temperature and precipitation and their interactions. We will test for correlations between variables using Principal Components Analysis (PCA). For variables that are strongly correlated, we will use only one. For comparison, we will also create new variables based on the first two or three Principal Components.

We will correlate the environmental data with the character data from each of the morphological and physiological characters described above, testing whether clades with or without a given character (e.g., fusoid cells, leaf intercalary meristems) also have significantly different values of environmental variables (e.g., temperature in the wettest month of the year, or annual precipitation). Methods will generally follow those presented by Edwards & Smith (2010) in which they correlated environment with shifts in photosynthetic pathway. In brief, for each morphological character of interest we will identify phylogenetically independent contrasts, using Phylocom's AOT module (Webb et al. 2008) and LASRdisc (Jackson 2004). We will then reconstruct the optimal mean value for each climate variable and test for significant differences across the phylogeny. We will compare results using a Brownian motion model versus the Ornstein-Uhlenbeck model described by Butler & King (2004). The former model assumes that the process being investigated is neutral with respect to selection, whereas the latter incorporates the possibility of selection and drift. We expect that for at least some of the characters we are investigating, a model incorporating selection will be a better fit for the data than the neutral model.

Website development. The Grass Phylogeny Web site will be developed to provide 1) an easily accessible repository of information on grass phylogeny and 2) image-rich content about grass structure, classification and evolution for a broad range of English and Spanish-speaking users. We will post an up-to-date summary phylogeny of grasses based on published sources along with a review of grass phylogenetics; this content will also be submitted to the Tree of Life Web site. The summary phylogeny will show families, subfamilies and tribes as the terminal taxa. Grass-specific molecular tools including primer sequences, regions affected by large deletions or inversions and genetic maps of completed plastomes will also be posted as part of the phylogenetics component. As results emerge from this study, we will continue to update the phylogeny. Content on grass structure, classification and evolution will be organized around the summary phylogeny. We envision that a user will be able to click on any given branch of the summary phylogeny and obtain images (line drawings, diagrams, and photos) and explanatory text in user-friendly language about the morphology, anatomy, and diversity of that clade. Users will thus be able to track changes in character states along these branches. Each grass subfamily and tribe (where relevant) will have a source page including images, text and literature citations. Although the Web site will not provide identification tools *per se*, it will complement existing keys and floristic treatments and it will be useful as a teaching tool for formal courses on grasses or for anyone who wishes to learn more about grasses. The target audience is non-specialists, which differs from the Grass Portal, but we also plan to link to that site.

Clark currently serves as the director of the Biological/Pre-Medical Illustration (BPMI) program at Iowa State University (see the BPMI Web site for additional details). BPMI is an interdisciplinary, pre-professional undergraduate major in which students take introductory and advanced biology courses while

receiving intensive training in traditional illustration and the incorporation of computer technology in the creation of imagery, modeling and animation. BPMI majors are also required to complete an internship leading to work for publication or presentation. Clark has worked with a number of these students in the past and will recruit BPMI majors to work on image creation for the Grass Phylogeny Web site for internships and as hourly workers. Current BPMI majors and local graduates provide a pool for recruitment of one or more skilled Web developers.

SEQUENCING AND INITIAL PHYLOGENETIC ANALYSES OF NUCLEAR AND PLASTOME PARTITIONS

Plastomes. Duvall's group has sequenced plastomes or draft plastomes from seven species: *Anomochloa marantoidea* (Morris & Duvall, 2010; GQ329703), *Chasmanthium latifolium* (HQ363060-HQ363120), *Coix lacryma-jobi* (Leseberg and Duvall, FJ261955), *Microcalamus convallarioides* (Duvall et al., 2010), *Pharus latifolius* (Jones & Duvall, unpublished), *Puelia olyriformis* (Givnish et al., in press; HQ603991-HQ604067), and *Joinvillea plicata* (Leseberg and Duvall, 2009; FJ486219-FJ486269). Also underway are plastome sequences for a temperate woody bamboo and four chloridoid species (Table 1).

Nuclear Loci. Of the ten nuclear loci evaluated for this project (Table 2), the previously published sampling of single copy sequences of *GBSSI* and *PHYB* available in GenBank were best correlated with those of the plastome sequences from grasses. Partial GBSSI sequences included four conserved exons at the 3′ terminus. Partial PHYB sequences included about 1.2 kb of exon I.

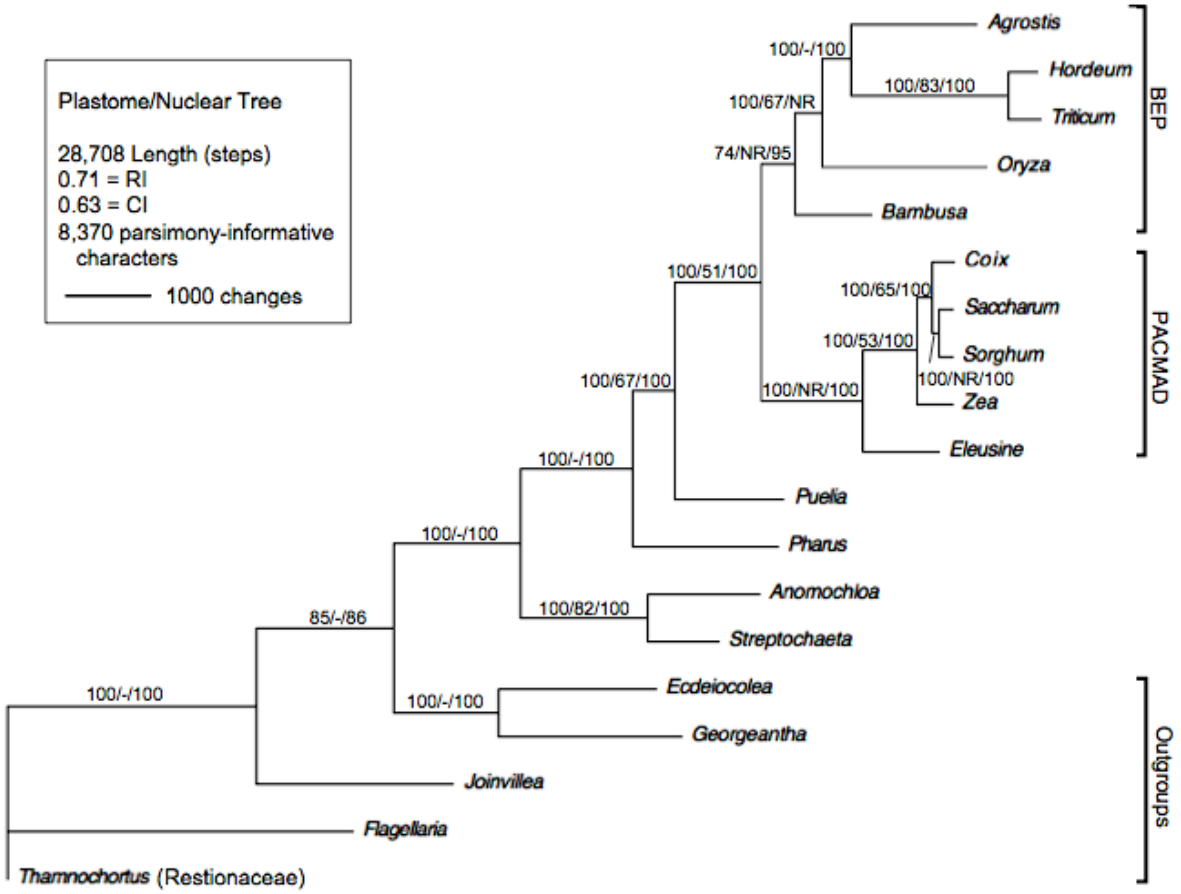

Fig. 2. Preliminary parsimony analysis of 81 plastome and two nuclear loci from 19 graminoids. Branch lengths are proportional to the inferred substitutions along the branches. Bootstrap values along the branches are ordered by partitioned analyses: combined analysis/nuclear partition/plastome partition. NR = "not resolved," referring to the positions of *Bambusa* and *Coix*. Nodes excluded from the nuclear analysis because of missing data are indicated with "-." Major clades and outgroups are indicated.

***Phylogenetics*.** A matrix of 81 plastome and two nuclear loci from 19 species was aligned and analyzed (Fig. 2). Species of Poaceae were selected for the availability of data from all partitions. The plastome and two nuclear loci were initially analyzed as three separate partitions. The two nuclear gene trees were topologically congruent, except for the weakly supported position of *Bambusa*. Note that the published *PHYB* sequence for *Bambusa* is only 30% as long as the other *PHYB* sequences. Because of substantial congruence the two nuclear loci were then concatenated and reanalyzed.

Analyses of the nuclear and plastome partitions were also topologically congruent, again except for the position of *Bambusa,* which was not resolved in either analysis, and *Coix*, which was supported at two adjacent nodes in the panicoid subtree. An exploratory total evidence analysis was then performed, in which the position of *Bambusa* is resolved, though only moderately supported (Fig. 2), but conflicts with that in Bouchenak-Khelladi et al. (2008) and Triplett & Clark (2010). The backbone of the grass subtree is well supported. Rapid divergence and/or extinctions, indicated by long branches, are generally indicated throughout the tree, and especially in the outgroups and early-diverging Poaceae.

In our preliminary analyses, sampling is most dense in the panicoid subtree, including *Sorghum*, *Saccharum*, *Zea* and *Coix*. Internal branches in this portion of the tree are relatively short (Fig. 2) and the addition of more panicoids would potentially exhaust the phylogenetic information in the selected loci. However, microstructural characters in noncoding regions can be used to increase signal by nearly 40% (Leseberg & Duvall 2009). Indel and inversion characters had lower homoplasy than point mutations in the conserved loci and succeed in resolving the position of *Coix* (Leseberg & Duvall 2009). Microstructural characters are valuable sources of phylogenetic information that can be encoded and analyzed to provide the information needed for robust comparisons at lower taxonomic levels. Characterization of this class of mutations will allow for the development of mathematical models to eventually permit analysis of these types of mutations by maximum likelihood.

Leaf anatomy. In phylogenetic studies for the last two decades, invaginated cell walls have usually been coded simply as present or absent as seen in cross section (e.g., GPWG I 2001). However, recent work and careful examination of the literature shows that this is an over-simplification (e.g., Chonan 1970). In cross section, chlorenchyma cells may be ± rectangular (no lobing) to circular, irregular, scalloped all around (rosette or plicate), or asymmetrically invaginated (shallowly or deeply) on the adaxial or abaxial sides or both. Invaginated cell walls may be present but visible only in longitudinal section (Sánchez-Ken & Clark, unpubl. data; Fig. 3D). The character is thus more complex and more subtle than originally thought and will clearly repay more detailed investigation.

The extent of the development of air spaces also varies considerably (Fig. 3D). For example, *Sporobolus* (C_4) in cross section appears to lack air spaces but small air spaces are visible longitudinally. A similar situation appears to exist in bamboos in cross section (*Chusquea*, C_3; Fig. 3A), but longitudinal sections reveal the presence of plates of arm cells separated by spaces, much as seen in pine leaves.

Preliminary observations in bamboos (Bambusoideae) indicate that the development of fusoid

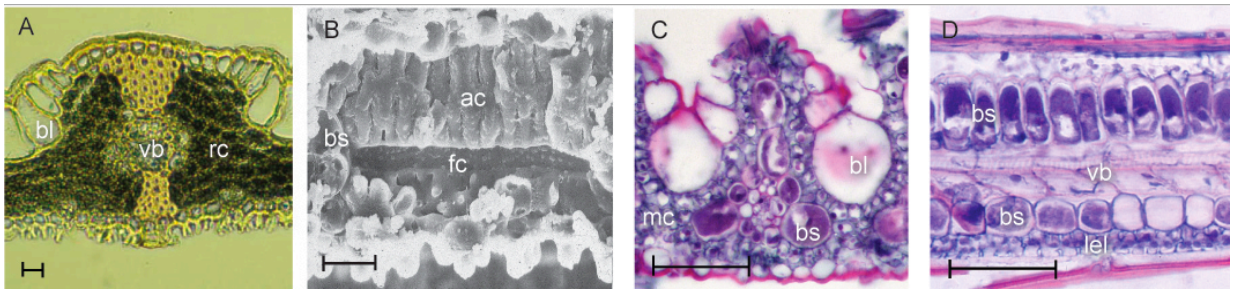

Fig. 3. Grass leaf anatomy. A-B. *Chusquea* (Bambusoideae), bar = 20 µm. A. Hand cross section. B. SEM of transverse cryofracture. C-D. *Sporobolus* (Chloridoideae), bar = 50 µm. C. Paraffin cross section. D. Paraffin longitudinal section. ac—arm cell; bl—bulliform cell; bs—bundle sheath; fc—fusoid cell; lel—longitudinally elongated lobed cell; mc—mesophyll cell; rc—rosette cell; vb—vascular bundle.

cells is strongly correlated with light intensity (within the same species, fusoid cells are present in shade leaves and absent in sun leaves) (March & Clark, in press). Freeze-fracture studies have not identified anything that looks like ice crystals in the cells (Fig. 3B), but these need to be repeated. An extensive literature on grass leaf anatomy and its relationship to photosynthetic pathway exists (e.g., Metcalfe 1960; Brown 1975; Watson & Dallwitz 1992; Sinha & Kellogg 1996), and many of the taxa (at least at the generic level) that we propose to sample have been examined for leaf anatomy. However, the emphasis has been on cross sections, often documented only by camera lucida drawings or low resolution photos (e.g., Metcalfe 1960; Tateoka 1963). We have both cross and longitudinal paraffin sections for ca. 30 taxa, of which virtually all are sampled from genera on our list, and of which approximately half coincide as to species.

COLLABORATIVE RESPONSIBILITIES

Duvall will be responsible for overall coordination of the project, including communication with international collaborators, data distribution, data alignment and analyses, and preparation of manuscripts. DNA extractions, amplifications, and library preparations will also take place in Duvall's lab. Clark will be responsible for supervising the sequencing of libraries at the DNA Facility, University of Iowa and for leaf anatomical and developmental work, in consultation with Kellogg and Muszynski (see letters in supplemental materials). She will be primarily responsible for Web site development in both English and Spanish. Note that Clark is bilingual and has previously taught agrostology/bamboo courses in Spanish in Costa Rica and Venezuela. Duvall, Kelchner, and Muszynski will also contribute web content. Kelchner will be responsible for characterization and analysis of microstructural changes and phylogenetic testing. Other collaborators will contribute samples or will be involved in data analysis and manuscript preparation. The two international collaborators are J. Saarela, Canadian Museum of Nature (Pooideae); and F. Zuloaga, Instituto de Botánica Darwinion, Argentina (Panicoideae) (see letters in supplemental materials). E. Edwards (Brown University) will assist with the ecological analysis (please see accompanying letter). Data relevant to this project have been assembled (Edwards & Smith 2010).

TIMETABLE

Most of the samples are already in hand or easily obtainable, so DNA extraction and amplification will begin immediately. Once the libraries are prepared, they will be shipped to Clark for sequencing. Once plates are submitted, turnaround time is usually very fast, on the order of hours to a few days, so data files will be sent to Duvall as soon as they are received. ***Year 1***: Complete DNA extractions, begin amplifications and sequencing of plastome and nuclear loci; initiate work on content for the Web site. ***Year 2***: Continue sequencing and initiate annotations; conduct preliminary analyses; initiate leaf anatomical studies; continue development of images and Web site content. Present preliminary results. ***Year 3***: Finish the sequencing and annotation; perform analyses and begin modeling the data; finish leaf anatomical and developmental work; continue Web site development; present results to date at the 2013 Monocots V/Grasses VI meeting at the New York Botanical Garden; prepare and submit manuscripts. ***Year 4***: Complete analyses; prepare a revised subfamilial classification if warranted; finish the Web site, including a phylogeny; prepare and submit manuscripts. The PIs or one or more students or postdoctoral scholars will attend at least one professional meeting per year to present results.

BROADER IMPACTS

The work proposed here will contribute to an improved understanding of grass evolution and ecological diversification through the generation of a robust phylogeny. It will also demonstrate the phylogenetic utility of whole plastome sequences and mutational events within a species-rich clade and contribute to a better understanding of plastome molecular evolution. Analysis of non-coding regions will provide the first well-sampled examination and modeling of microstructural changes in plastome DNA, which has been a limiting step in the use of introns and intergenic spacers in phylogenetic analyses. The Web-accessible phylogeny and grass-specific molecular tools (primer sequences, regions affected by large deletions or inversions, genetic maps of completed plastomes) will be useful to cereal scientists,

molecular geneticists, paleobotanists/ecologists, and biologists interested in grass evolution. The image-rich Web content about grass structure, classification and evolution will serve as a teaching tool. It will promote international collaborations with Latin Americans, English and Portuguese speakers, including ecologists and conservation biologists working with grasses or grass-dominated habitats, agrostologists, and any non-specialist needing a readily understandable introduction to grass structure and diversity.

The technical work will involve the training of two graduate and two undergraduate students in the Duvall lab; at least one graduate student and two or more undergraduates (focusing on BPMI majors, who are overwhelmingly female) in the Clark lab; and one postdoctoral researcher each in the Clark and Kelchner labs. The PIs regularly publish with their students and have trained members of under-represented groups (see below). If this proposal is funded, REU supplements will be sought to facilitate additional participation by students in the project. Leaf anatomical studies will be incorporated into labs in plant biology and plant anatomy courses taught by Clark and Duvall as appropriate.

Completed sequences will be deposited in GenBank. In addition to content posted on the Grass Phylogeny Web and Tree of Life Web sites, papers on molecular evolutionary and phylogenetic analyses will be submitted to peer-reviewed scientific journals and presented at scientific conferences, including the international Monocots V/Grasses VI conference scheduled for August 2013 at the New York Botanical Garden. Student participation in both presentations and publications will be a high priority.

RESULTS FROM PRIOR NSF SUPPORT

DEB-0515712 to PI Clark ($387,356) and **DEB-0515828 to PI Kelchner** ($264,893), Collaborative Research: Phylogeny and Generic Classification of the Woody Bamboos (Poaceae: Bambusoideae: Bambuseae), 2005-2009 (with 2 1-year extensions to 2011).

The major objectives of this project are to 1) generate a robust phylogeny for the woody bamboos using multiple plastid sequence data sets, AFLPs, structural characters, and a rigorous phylogenetic approach; 2) examine selected aspects of morphological evolution with the bamboos; 3) construct predictive, more stable subtribal and generic classifications for the woody bamboos based on the phylogeny; and 4) establish an umbrella Web site for bamboo biodiversity to host an interactive identification key to genera, provide links to relevant information, and make images and descriptive information on bamboos more readily available. The Bamboo Phylogeny Group (BPG; now 20 researchers in 10 countries) was formed to facilitate sampling and provide complementary expertise to achieve the objectives. Major findings to date include evidence of paraphyly of the woody bamboos, recovery of 10 major lineages within the temperate bamboos, documentation of extensive reticulation among temperate bamboos, some changes in generic concepts, and the need for recognition of new genera. The analytical approach also demonstrated stability of the chloroplast phylogeny estimation and support values, and robustness to taxon sampling, systematic error, and bias. The Bamboo Biodiversity Web site (Clark & Gardner, 2005) includes sample interactive keys and illustrated character lists for the morphological analysis, while the Bamboo Phylogeny Database (Kelchner) is annotating and archiving sequences from the project for public use and future database development.

Clark has supported 4 graduate students (1 M.S., 3 Ph.D.); Kelchner has supported 5 graduate students (1 Special, 3 M.S., 1 Ph.D.). A total of 3 high school students, 14 undergraduates, and 2 high school teachers have received training in various aspects of DNA sequencing, phylogenetic analysis, construction of interactive keys, leaf anatomy, and botanical illustration during the course of this project. Of these 14, 12 are female (including one Asian-American and one Native American). Clark also has hosted visits from 5 of the BPG collaborators (4 female) and 3 international Ph.D. students (Brazil, Venezuela, Thailand) for training in DNA sequencing and bamboo systematics. An REU supplement to Kelchner provided support and data for 2 female students, one of whom developed the project into an M.S. (completed); results include 3 conference presentations and 3 forthcoming papers. Thus far, 24 publications (published or in press) have been supported wholly or in part by this project, with another 7 in the final stages of preparation. 37 talks (both contributed at meetings and invited workshops and seminars) and 6 posters based on this project have been presented since December, 2005. Publications and manuscripts related to the project are indicated with an asterisk in the Clark or Kelchner Biosketches.