ABSTRACT

RESOLVING DEEP RELATIONSHIPS OF PACMAD GRASSES:
A PHYLOGENOMIC APPROACH

Joseph Cotton, M.S.
Department of Biological Sciences
Northern Illinois University, 2014
Melvin R. Duvall, Director

The phylogenetically recognized PACMAD (Panicoideae, Aristidoideae, Chloridoideae, Micrairoideae, Arundinoideae, Danthonioideae) clade of grasses has been the subject of numerous phylogenetic studies that have made an attempt at determining subfamilial relationships of the clade. The purpose of this thesis was to examine chloroplast genome sequences for 18 PACMAD species and analyze them phylogenomically. These analyses were conducted to provide resolution of deep subfamilial relationships within the clade. Divergence estimates were assessed to determine potential factors that led to the rapid radiation of this lineage and its dominance of open habitats. This was accomplished via next-generation sequencing methods to provide complete plastome sequence for 12 species. Sanger sequencing was performed on one species, *Hakonechloa macra*, to provide a reference. Phylogenomic analyses and divergence estimates were conducted on these plastomes in conjunction with six other previously banked plastomes.

The results presented here support Panicoideae as the earliest diverging PACMAD lineage. The initial diversification of PACMAD subfamilies was estimated to occur 32.4 mya. Phylogenomic analyses of complete plastome sequences provide strong support for deep relationships of PACMAD grasses. The divergence estimate of 32.4 mya at the crown node of the PACMAD clade coincides with the Eocene-Oligocene Transition (EOT). Throughout the

Eocene, prior to the EOT, was a period of global cooling and drying, which led to forest fragmentation and the expansion of open habitats now dominated by these grasses.

NORTHERN ILLINOIS UNIVERSITY
DE KALB, ILLINOIS

DECEMBER 2014

RESOLVING DEEP RELATIONSHIPS OF PACMAD GRASSES;

A PHYLOGENOMIC APPROACH

BY

JOSEPH L. COTTON

A THESIS SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE

MASTER OF SCIENCE

DEPARTMENT OF BIOLOGICAL SCIENCES

Thesis Director:
Melvin R. Duvall

ACKNOWLEDGEMENTS

First and foremost, I would like to thank Dr. Mel Duvall for his guidance and support. I would also like to acknowledge Bill Wysocki for his help with plastome assembly, annotation and analysis, as well as Sean Burke. Finally, I would like to thank the Plant Molecular Biology Center and the Department of Biological Sciences at Northern Illinois University for financial support.

TABLE OF CONTENTS

Page

Chapter

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACRE	Anchored Conserved Region Extension
BEAST	Bayesian Evolutionary Analysis Sampling Trees
BEAUTi	Bayesian Evolutionary Analysis Utility
BEP	Bambusoideae Ehrhartoideae Pooideae
bp	Base pair
bs	Bootstrap Support Value
CIPRES	Cyber Infrastructure for Phylogenetic Research
CMAD	Chloridoideae Micrairoideae Arundinoideae Danthonioideae
DC	Danthonioideae Chloridoideae
EOT	Eocene-Oligocene Transition
GPWGI (II)	Grass Phylogeny Working Group I (II)
GTR+G+I	General Time Reversal plus Gamma distribution plus proportion of Invariant sites
IR	Inverted Repeat
LSC	Long Single Copy
MA	Micrairoideae Arundinoideae
MAFFT	Multiple Alignment using Fast Fourier Transform
MCMC	Markov Chain Monte Carlo
ML	Maximum Likelihood

MP	Maximum Parsimony
NGS	Next-Generation Sequencing
PACC	Panicoideae Arundinoideae Chloridoideae Centothecoideae
PACCAD	Panicoideae Arundinoideae Chloridoideae Centothecoideae Arundinoideae Danthonioideae
PACCMAD	Panicoideae Arundinoideae Chloridoideae Centothecoideae Micrairoideae Arundinoideae Danthonioideae
PACMAD	Panicoideae Arundinoideae Chloridoideae Micrairoideae Arundinoideae Danthonioideae
PAUP*	Phylogenetic Analysis Using Parsimony *(and other methods)
PETM	Paleocene-Eocene Thermal Maximum
SSC	Short Single Copy
XSEDE	Extreme Science and Engineering Discovery Environment

PREFACE

Poaceae, the grass family, covers roughly 52.5 million km^2 (40.5%) of the total terrestrial area of the world (White, Murray, and Rohweder, 2000). As a plant family, grasses are the largest source of food for the world. Many species are readily cultivated such as *Zea mays* (corn), *Triticum aestivum* (wheat), and *Oryza sativa* (rice), yet overall global production of these grains has been declining as a by-product of climate change (Von Braun, 2007). Understanding the phylogenetic relationships of these grasses and the pressures leading to their diversification will allow for future predictions of their global distribution (Bouchenak-Khelladi et al., 2010).

Poaceae can be classified into three recognized lineages: the basal grade, BEP clade, and PACMAD clade, as currently established by the Grass Phylogeny Working Group II (GPWGII, 2012). The earliest diverging lineage is a deep grade comprised of three subfamilies (Anomochlooideae, Pharoideae, and Puelioideae) that are found in forest understories (Watson and Dallwitz, 1992). The BEP clade (Bambusoideae, Ehrhartoideae, Pooideae) consists strictly of C_3 photosynthetic species found in both open and shady habitats of cool and tropical regions. A lineage with a more recent divergence, the PACMAD clade (Panicoideae, Aristidoideae, Chloridoideae, Arundinoideae, Danthonioideae), consists of both C_3 and C_4 species (Edwards et al., 2010) that are typically found in open habitats and warmer climates, but includes some shade-tolerant species (Edwards and Still, 2008). There is strong support for the subfamilial relationships of the basal grade and BEP subfamilies, but support is still low for deep PACMAD relationships. The PACMAD clade has been the subject of numerous phylogenetic studies (many

of which predate the PACMAD acronym established by Duvall et al., 2007) that have made an attempt at determining subfamilial relationships of the clade (Clark et al., 1995; GPWGI, 2001; Duvall et al., 2007; GPWGII, 2012). These analyses have been conducted using a variety of data matrices consisting of several chloroplast markers that are known to provide strong support for plant relationships. In most cases a large set of taxa was analyzed, and although these data do provide strong support for some subfamilial relationships within the clade, support for the early diverging lineages is weak. These deep relationships include the relative positions of Panicoideae and Aristidoideae. This is because the PACMAD clade has rapidly radiated, as indicated by the relatively short, deep branch lengths of the clade (GPWGII, 2012), so that relatively few phylogenetically informative characters are available for study. A different approach is necessary to draw accurate conclusions regarding these relationships.

The development of next-generation sequencing (NGS) technology has allowed for assembly of complete chloroplast sequences in a relatively short period of time. This new method offers an approach to teasing out the relationships of Poaceae and, more specifically, of PACMAD grasses. It provides the ability to apply a greater amount of molecular data to phylogenetic analyses, as well as other analyses, such as divergence estimates.

Previous divergence time estimates of Poaceae have been conducted on the basis of relatively few nucleotide sequence sites (Christin et al., 2013). By limiting the number of taxa to critical representatives, phylogenomic-scale sequences can provide a statistically significant set of data for divergence estimation.

With both an accurate phylogenomic topology and greater confidence in divergence estimates of the clade, we can begin to answer why this rapid radiation occurred in the

PACMAD grasses. The divergence estimates can be compared to a geologic timeline to offer insight into evolutionary pressures that may have led to this rapid radiation.

CHAPTER 1 - INTRODUCTION

Poaceae have been the subject of numerous phylogenetic studies due to their economic and ecological importance and dominance in major terrestrial biomes (e.g., Clark et al., 1995; GPWGI, 2001; Duvall et al., 2007; GPWGII, 2012). The current phylogenetic classification of Poaceae includes a deep grade of three lineages: Anomochlooideae, Pharoideae, and Puelioideae, as well as the crown grasses represented by the BEP (Bambusoideae, Ehrhartoideae, Pooideae) and PACMAD (Panicoideae, Aristidoideae, Chloridoideae, Micrairoideae, Arundinoideae, Danthonioideae) clades (GPWGII, 2012). The PACMAD clade is of particular interest in this study because phylogenetic uncertainty has resulted in repeated reclassifications of subfamilies within the clade.

The sister group to the BEP clade has been variously defined as the PACC, PACCAD, PACCMAD, or PACMAD clade with different constituent subfamilies. A study conducted by Clark et al. (1995) utilized the chloroplast gene sequence, *ndhF*, which supported a monophyletic PACC (Panicoideae, Arundinoideae, Chloridoideae, Centothecoideae) clade, as well as indicated the polyphyletic nature of Arundinoideae. Subsequent work by the Grass Phylogeny Working Group I (GPWGI, 2001) addressed weak support within and among the grass subfamilies by making use of informative characters in seven molecular datasets along with morphological characteristics. For comparative purposes we will refer to their results for three chloroplast sequences (*ndhF*, *rbcL*, *trnK/matK*), not their eight-character data matrix analysis, as they did not differ in subfamilial arrangement or provide further resolution. The GPWGI (2001) study also

increased taxon sampling over those of previous phylogenetic studies to include representatives of 62 genera, 30 of which fell within a group described under the newly established PACCAD (Panicoideae, Arundinoideae, Chloridoideae, Centothecoideae, Aristidoideae, Danthonioideae) clade. Three taxa that nested within the PACCAD clade (*Eriachne*, *Gynerium*, and *Micraira*) were classified as *incertae sedis*, or of uncertain placement. Arundinoideae were also found to lack unifying morphological or molecular synapomorphies to establish it as monophyletic. The genera classified as *incertae sedis* were subjected to further analysis by Sanchez-Ken et al. (2007) along with other representatives from *Eriachne* and *Micraira* through the use of 69 structural characters as well as *ndhF* and *rpl16* chloroplast sequences. Their reinstatement of Micrairoideae as a distinct subfamily changed the PACCAD acronym to PACCMAD.

With increased taxon sampling across Panicoideae and Centothecoideae in a subsequent study, Sanchez-Ken and Clark (2010) concluded that Centothecoideae were paraphyletic with Panicoideae "…and the name should not have phylogenetic implications" (p. 1738). Their study defined the constituent subfamilies of the PACMAD clade, an acronym first used by Duvall et al. (2007), and established a backbone phylogenetic hypothesis against which deeper phylogenetic relationships could be explored. The Grass Phylogeny Working Group II (2012) constructed the most detailed grass phylogeny to date. One of their major goals was to determine the origins of C_4 photosynthesis across the family. They analyzed 452 PACMAD species, encompassing two thirds of the genera of the clade using the same chloroplast markers from the previous GPWG study (*rbcL*, *ndhF*, *trnK*/*matK*). Through multiple phylogenetic analyses, and an increase in taxonomic sampling, they were able provide strong support for Aristidoideae as the sister subfamily to the rest of the clade. However, the relationship between Panicoideae and the

CMAD (Chloridoideae, Micrairoideae, Arundinoideae, Danthonioideae) clade was only weakly supported (bootstrap [bs] value: 61%, posterior probability [pp]: 99), as well as the relationship between the MA (Micrairoideae, Arundinoideae) and DC (Danthonioideae, Chloridoideae) clades (bs value: 51%, pp 98). Furthermore, the arundinoids were weakly monophyletic. This may be due in part to the rapid radiation of the PACMAD clade.

Deep divergence time estimates of PACMAD grasses have been relatively few. This is partially because of the low availability of confidently dated grass fossils for use as calibration points at specific nodes (Bouchenak-Khelladi et al., 2010; Christin et al., 2013). The fossils used for calibration include pollen, phytoliths, and spikelets (Vicentini et al., 2008; Bouchenak-Khelladi et al., 2010; Prasad et al., 2011). Another contributing factor is the lack of a well-supported topology at the subfamily level, especially for deep relationships within the PACMAD clade due to the use of minimal molecular sequence to generate a well-supported phylogeny and accurate molecular clock. Previous divergence estimates of the PACMAD clade are highly variable: stem Aristidoideae (28.8 to 61.1 mya), crown PACMAD (38 to 61.1 mya), and stem Panicoideae (26 to 42.1 mya) (Christin et al., 2008, Vicentini et al., 2008, Bouchenak-Khelladi et al., 2010, Prasad et al., 2011). These four studies utilized a relatively small number of molecular markers in their phylogenetic analyses, and their lack of informative characters caused the topologies to vary between them. The analyses presented here utilize more phylogenetically informative characters and well-supported phylogenomic relationships to provide a greater accuracy of divergence estimates through the use of complete plastomes.

Phylogenomic studies of complete chloroplast genomes from Poaceae have provided strong support for the relationships within and among other subfamilies (Burke et al., 2012; Wu

and Ge, 2012; Jones et al., 2014). This study addresses the weak support in previous research for specific nodes in the PACMAD clade by utilizing complete plastomes. Mitochondrial sequence data were also analyzed with the goal of increasing character sampling among representative taxa (Clifton et al., 2004; Guo and Ge, 2005).

A reference plastome from one arundinoid species (*Hakonechloa macra*) was sequenced using Sanger technology and 12 other complete plastomes were determined by next-generation sequencing (NGS) methods for PACMAD taxa. These data were analyzed phylogenomically and divergence dates estimated to seek potential selective causes for the PACMAD radiation.

CHAPTER 2 - MATERIALS AND METHODS

Taxon Sampling

Taxa were sampled based on subfamilial membership to obtain representation for all major groups of interest. Outgroup sampling with respect to taxon selection and number posed a challenge due to the potential sensitivity of the positions of Aristidoideae and Panicoideae as the earliest diverging PACMAD lineages. Initially, *Oryza rufipogon* (GenBank accession NC_022668) was selected as the outgroup, but support for Aristidoideae as sister to the CMAD clade was low (bs value 76%). Another BEP representative, *Bambusa oldhamii* (NC_012927), was chosen as the outgroup and the topology remained consistent, but weak support was still seen at the stem Aristidoideae node (bs value 80%). Up to 17 representative BEP taxa, as well as basal-grade taxa (*Anomochloa marantoidea*, NC_014062; *Pharus latifolius* NC_021372; *Puelia olyriformis,* NC_023449), were combined in ten subsets to determine the effect of outgroup selection on ingroup topology and support. Ingroup topology proved to be stable across these analyses, although support for short deep branches was variable. Ultimately, the single outgroup species *Rhynchoryza subulata* (NC_016718; Ehrhartoideae) was found to produce results congruent to those of the largest outgroup species sets analyzed here.

Taxa from Panicoideae include representatives of five major tribes, *Panicum virgatum* (Paniceae, NC_015990), *Zea mays* (Andropogoneae, NC_001666), and *Coleataenia prionitis* (Paspaleae). *Centotheca prionitis* (Centotheceae) and *Thysanolaena maxima* (Thysanolaeneae) were also included from the previously recognized Centothecoideae (Clark et al., 1995; GPWG I,

2001; Sanchez-Ken et al., 2007), which GPWGII (2012) classified as Panicoideae. Three representative Arundineae, *Hakonechloa macra*, *Monachather paradoxa*, and *Phragmites australis,* that represent three major clades of Arundinoideae as retrieved by the Grass Phylogeny Working Group II (2012) were included. One member of an arundinoid genus classified as *incertae sedis*, *Elytrophorus spicatus*, was also included in an attempt to resolve this arundinoid relationship (http://www.tropicos.org/projectwebportal.aspx?pagename= ClassificationNWG&projectid=10). Three taxa from the most recently reinstated subfamily Micrairoideae were also analyzed to provide support for the Micrairoid/Arundinoid clade. Taxa were chosen to represent three of the four tribes of micrairoids: *Micraira spiciformis* (Micraireae), *Eriachne stipacea* (Eriachneae), and *Isachne distichophylla* (Isachneae). Four genera within Danthonieae were also included to represent the danthonioid lineage as well as a representative species of Aristidoideae, *Aristida purpurea* (Aristideae). The only published plastome for a chloridoid species, *Neyraudia reynaudiana*, was also included in our analyses (Wysocki et al., 2014).

Mitochondrial sequence data for each taxon included in the plastome analysis were retrieved from the Illumina read files. If mitochondrial sequences could not be obtained from read files, they were retrieved from the NCBI database. Because of this, our mitochondrial sampling was limited to species that were sequenced using NGS and those represented in Genbank. For this portion of the study, the taxon set was limited to 15 species with *B. oldhamii* as the outgroup taxon.

DNA Extraction and Sequencing

Leaf tissue samples were obtained for each species of interest (sources listed in Table 1) and dried with silica gel desiccant. Liquid nitrogen was used to lyse cells and homogenize plant tissue to maximize DNA yields. Extractions were performed using the DNeasy Plant Mini Kit (Qiagen, Valencia, CA, USA) following the manufacturer's protocol.

Table 1

List of Species Sequenced Using NGS Methods

Species	**Subfamily**	**Voucher**	**GenBank #**
Aristida purpurea	Aristidoideae	J. Saarela 605 (CAN)	KJ920224
Centotheca lappacea	Panic/Centoth	G. Sanchez-Ken s.n. (ISC)	KJ920225
Coleataenia prionitis	Panicoideae	O. Morrone 6195 (SI)	KJ920228
Thysanolaena maxima	Panic/Centoth	L. Clark & Gallaher 1717 (ISC)	KJ920236
Elytrophorus spicatus	Arundinoideae	J. Saarela 1685 (CAN)	KJ920230
Monachather paradoxus	Arundinoideae	J. Saarela 1629 (CAN)	KJ920235
Eriachne stipacea	Micrairoideae	S. Jacobs 8773 (NSW)	KJ920231
Isachne distichophylla	Micrairoideae	C. Morden 1227 (HAW)	KJ920233
Micraira spiciformis	Micrairoideae	S. Jacobs 8850 (NSW)	KJ920234
Chaetobromus dregeanus	Danthonioideae	N. Barker 978 (PRE)	KJ920226
Chionochloa macra	Danthonioideae	CHR 475278 (BOL)	KJ920227
Danthonia californica	Danthonioideae	J. Saarela 860 (CAN)	KJ920229

The plastome of one species, *Hakonechloa macra*, was amplified and Sanger sequenced using grass-specific primers (Leseberg and Duvall, 2009) and the general methods of Dhingra and Folta (2005). Only one copy of the IR (inverted repeat) was sequenced as well as all four IR boundaries. The alternative methods of Morris and Duvall (2010) were followed when amplifications failed, including the design and use of primers tailored to *H. macra* (Table 2). Amplifications were prepared for sequencing using the Wizard SV PCR Clean-up System

Table 2

Species-specific Primers Designed for *Hakonechloa macra*

Primer Name	Sequence
1FHma	CATTACCCACTTGTCCGACTGTTGC
2RHma	CGATACTGGAACTCAGAGCATAGGAGG
12RHma	GCATGATATTGGGGAATCTCCTTGC
31FHma	GCTGCGTGTATAAGAGCCGAAATGGG
31RHma	GGCAACGAGCATCCAAAACCAAAAG
31FHma2	GCCCGAAATGGAGTGGGTCCTTCC
31RHma2	GGTTTTGGATGCTCGTTGCCGCTAG
39FHma	CCTTGGGGTAAAGAGTTTACACTGC
99RHma	CCTGTAGTAGGGATCTGGTTCACTGC
101FHma	GCGACCCCACAGGCTTGTACTTTCG
101RHma	CTTCGTCTACCCACTTCTCTTTCAGG
116FHma	GAACCTGGAGGAGTTAGGTATGTAGG
116RHma	GGAGTCGGCAAATGGACTGAGAATG
148FHma	GGCGTTTTTTTAGTGAGTCCTGG
148RHma	CTCTTTGAGAGACCCCGCTTATTGC
148FHma2	GAGACCGACCCACTTCCTATCTAG
169FHma	GCCGTTATTGTCGCAAGCATACGAC
169RHma	GCAGAATTTCTCTAGTAGGAGCGTC
169FHma2	CTCTATGGTTTGGGGCTTTAGCAGG
169RHma2	CACTCGGATAAACGCAGCAACACGG
184FHma	GCAACAGTCGGACAAGTGGGTAATG

(Promega, Madison, WI, USA). Sanger sequencing was performed at Macrogen, Seoul, South Korea. A complete plastome was manually assembled from the overlapping sequences and adjacent segments using Geneious Pro version 7.0.1 (Biomatters, Auckland, New Zealand). The other 12 plastomes, not acquired through the NCBI nucleotide database, were sequenced using NGS. Starting quantities of total genomic DNA from three DNA extracts (*Danthonia californica, Elytrophorus spicatus, Monachather paradoxus*) were determined with a Nanodrop 1000 (ThermoFisher Scientific, Wilmington, DE, USA) to be 1.5 µg each. DNAs were diluted to 2 ng/µl and sheared into 300 base pair (bp) fragments using a Bioruptor® sonicator (Diagenode, Denville, NJ, USA) in two 12 min periods, with inversion of tubes between them. DNA preparations were then purified and concentrated using the MinElute Extraction Kit (Qiagen Inc., Valencia, CA, USA). Single read libraries were prepared with the TruSeq low throughput protocol (gel method) following manufacturer instructions (Illumina, San Diego, CA, USA). Single-end sequencing was conducted on a HiSeq 2000 instrument (Illumina, San Diego, CA, USA) at Iowa State University, Ames, USA. Reads produced this way were 99 bp in length.

An improved NGS method was used for the remaining nine species. Total genomic DNA extracts for the remaining taxa were diluted to 2.5 ng/µl in 20µl water. The Nextera Illumina library preparation kit (Illumina, San Diego, CA, USA) was used to prepare libraries for sequencing and the DNA Clean and Concentrator kit (Zymo Research, Irvine, CA, USA) was used for library sample purification. Single-end sequencing was performed with the HiSeq 2000 instrument at the Iowa State University Sequencing Facility as above.

The reads were first quality filtered using DynamicTrim v2.1 from the SolexaQA software package (Cox et al., 2010) with default settings, and then sequences less than 25 bp in

length (default setting) were removed with LengthSort v2.1 in the same package. The quality of the reads was then assessed using FastQC v0.10.1(www.bioinformatics.babraham.ac.uk/projects/fastqc/). The complete quality-trimmed set of reads was used for assembly.

Plastome Assembly and Annotation

Plastome assembly was performed with entirely *de novo* methods. The Velvet software package (Zerbino & Birney, 2008) was run iteratively following methods from Wysocki et al. (2014). Contigs were scaffolded using the anchored conserved region extension (ACRE) method (Wysocki et al., 2014). Any remaining gaps in the plastomes were resolved using contigs or reads by locating overlapping regions until the circular map was complete. Assembled plastomes were annotated in Geneious Pro by aligning them to a closely related and previously annotated reference plastome and transferring the annotations from the reference to the assembled plastome when the annotation shared a similarity of 70% or above. Each coding sequence was then examined and necessary adjustments were made to preserve intron boundaries, reading frames, and stop codons or to identify pseudogenes. The IR boundaries were located using the methods of Burke et al. (2012).

Plastome Analysis

Fully assembled plastomes were aligned in Geneious Pro using the MAFFT plugin (Katoh et al., 2005). Characters for which gaps were introduced in one or more sequences by the alignments were removed along with one of the IR sequences. This length of the alignment was 96,493 bp after IR and gap removal. A maximum-likelihood (ML) analysis was performed using RAxML-HPC2 on XSEDE (Stamatakis, 2006) and was accessed using the CIPRES science gateway (Miller et al., 2010). The tree with the highest likelihood was obtained and a rapid

bootstrap support analysis was performed with 1000 pseudoreplicates using default parameters and the ehrhartoid grass *Rhynchoryza subulata* as the specified outgroup. The same methods were used for other combinations of outgroup taxa. A consensus bootstrap tree was produced with the Consense function of the Phylip software package (Felsenstein, 2005) on CIPRES (Fig. 1). Tree files were visualized and edited using FigTree v1.4.0. Several other likelihood analyses were conducted using the separate outgroup taxa, *Puelia olyriformis*, *Bambusa oldhamii*, and *Oryza rufipogon*, but these analyses provided low bootstrap support (bs) values for the stem Aristidoideae.

A Bayesian inference (BI) analysis was performed using MrBayes 3.2.2 on XSEDE (Ronquist et al., 2012), which was accessed using the CIPRES science gateway. The Markov chain Monte Carlo (MCMC) analysis was run twice at 10,000,000 generations each. The analysis was run until the average standard deviation of split frequencies approximated zero.

Branch and bound maximum parsimony (MP) analyses were performed using PAUP* v4.0b10 (Swofford, 2003). An MP bootstrap analysis with 1000 pseudoreplicates, each with 10 random addition sequence replicates, was also performed.

A microstructural mutation analysis was conducted with a plastome alignment consisting of three outgroup species (*Rhynchoryza subulata, Oryza rufipogon,* and *Bambusa oldhamii*) and the 18 ingroup taxa with Geneious Pro using the MAFFT plugin. The length of the alignment was 130,475 bp after one IR was removed. The alignment was examined and any large inversions and indels involving the deeply diverging PACMAD taxa were noted.

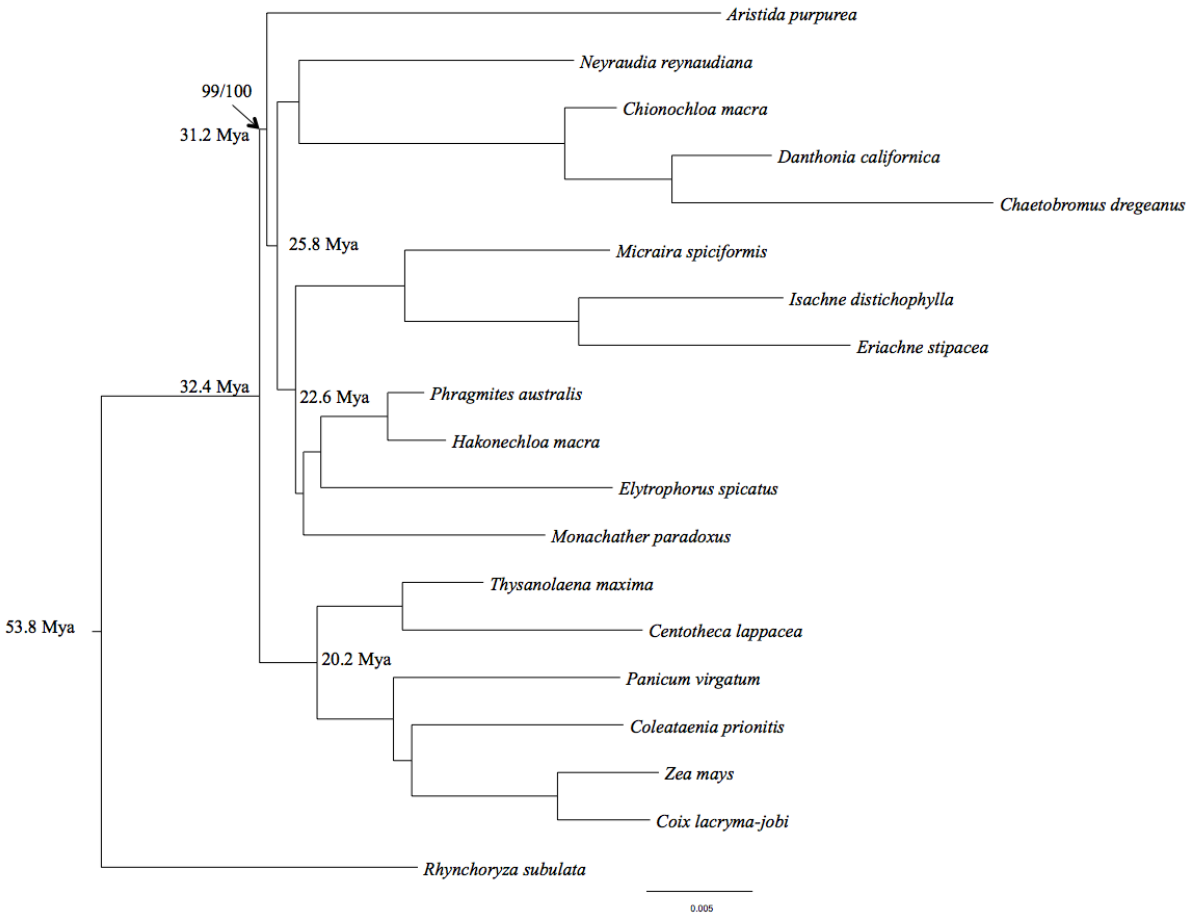

Figure 1: Maximum likelihood phylogram produced from a complete plastome analysis of 19 species of PACMAD grasses with removal of gaps and one copy of IR sequence. Estimated divergence dates are included. Branch lengths are proportional to the substitution rate along the branch. *Rhynchoryza subulata* was selected as the outgroup. Each node was fully supported with a bs value = 100 and a pp = 100 except where noted (bsv/pp). Divergence estimates for given nodes are listed.

Mitochondrial Analysis

Mitochondrial sequences were assembled from the same NGS read files and analyzed in a likelihood framework using the same utilities as the plastome analysis. The *matR* gene sequence was chosen based on prior use in phylogenetic analyses (Clifton et al., 2004). Six relatively variable intron sequences (Guo and Ge, 2005) were also chosen: *nad1* intron two, *nad4* intron 1, and *nad7* introns 1, 2, 3, and 4. Each intron sequence had less than 99% pairwise identity when aligned (see Table 2). Mitochondrial sequences from *Zea mays* and *Bambusa oldhamii* were obtained directly from GenBank. A reference mapping analysis was performed for each species that was included in the NGS libraries by mapping each read file to each coding sequence of interest from the full mitochondrial sequence of *Zea mays*. The consensus sequences from each reference mapping were examined for missing nucleotide sites and for frameshift mutations to detect erroneous assemblies. The corresponding *matR* and intron sequence data (Table 3) were assembled and aligned for each species. These sequences were analyzed using identical protocols and settings as the chloroplast data to generate a maximum likelihood tree (Fig. 2).

Divergence Estimates

Divergence dates were estimated using BEAST v2.1.2 (Bouckaert et al., 2014) and parameters were set in BEAUTi. Preliminary divergence dates with different combinations of calibrations and different prior distributions were estimated. Note that there are few fossils useful for calibrations of PACMAD grasses and that dates are sometimes contradictory. Two divergence analyses were selected and presented here to show the maximum range of estimated dates. Both estimates were each given two different seed values and chain lengths of 10 million

Table 3

Mitochondrial Sequence Analysis

Taxa	Sequences	Pairwise % Identitiy
Aristida purpurea	mat-r	99.3
Bambusa oldhamii	cox2 intron	91.1
Centotheca lappacea	nad1 intron 2	97.4
Chaetobromus dregeanus	nad4 intron 1	98.1
Chionochloa macra	nad7 intron 1	95.5
Coleataenia prionitis	nad7 intron 2	98.3
Danthonia californica	nad7 intron 3	98.9
Elytrophorous spicatus	nad7 intron 4	98.0
Eriachne stipacea		
Isachne distichophylla		
Micraira spiciformis		
Monachather paradoxus		
Zea mays		
Neyraudia reynaudiana		
Thysanolaena maxima		

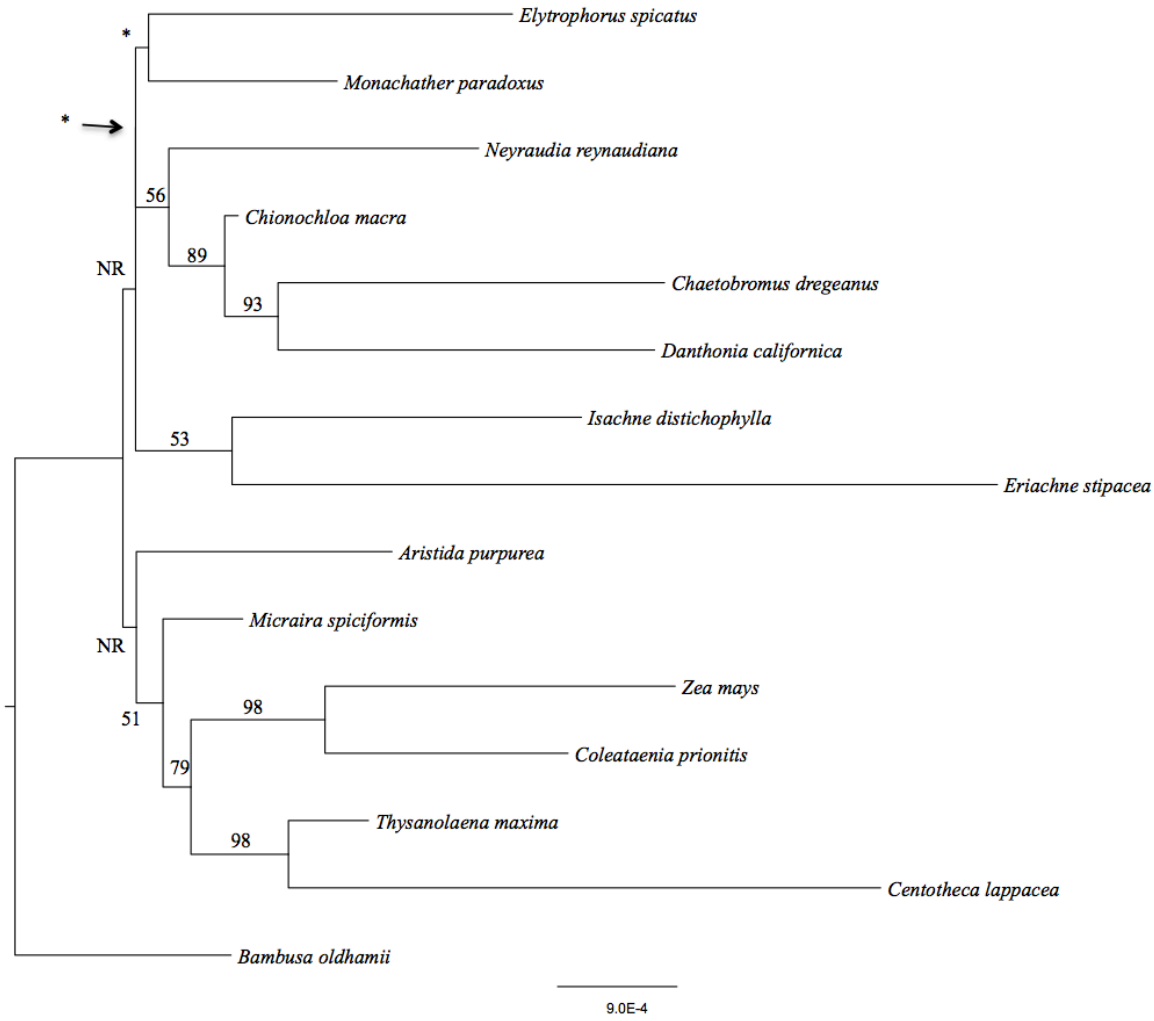

Figure 2: Maximum likelihood phylogram produced from analysis of 15 assembled and aligned mitochondrial *matR* and intron sequences. Branch lengths are proportional to the substitution rate along the branch. *Bambusa oldhamii* was selected as the outgroup. Bs values >50 and <100 are noted. Nodes labeled with * denote bs values <50. Each node marked NR was unresolved due to topological conflicts with our ML plastome analysis (Fig. 1).

for a total of 20 million generations each. Priors were set to constrain relationships that preserved the topology generated by the maximum likelihood plastome analysis. The substitution model utilized was GTR+G+I under an uncorrelated relaxed log-normal clock. A total of three fossils were used to calibrate the analyses. For each of these a log-normal distribution was selected. Parametersolver v3.0 (Cook et al. 2013) was used to calculate the mean and variance of these log-normal prior distributions. Fossil estimates for calibration of each run differed slightly. A total of three fossils were used to calibrate the analyses. The first run was assessed using a lower bound at the stem Chloridoideae node of 14 mya calibrated with a fossil identified as a member of Chloridoideae, due to the shape of the stomatal subsidiary cells and silica bodies (Bouchenak-Khelladi et al., 2010). There has been some debate, however, on the taxonomic identity of this fossil upon further microstructural evaluation (Bell and Columbus, 2008). An upper bound of 19 mya was placed at the stem Chloridoideae node using a chloridoid phytolith fossil (Strömberg, 2005). The crown Panicoideae were also constrained for the first run with a lower bound of seven mya using a *Setaria* fossil and an upper bound of eight mya with a fossil assigned to *Dichanthelium* sp. (Vicentini et al., 2008). The second BEAST run was conducted using the same calibrations as the first run, but with the addition of a lower bound on the crown BEP/PACMAD node of 65 mya and an upper bound of 67 mya using Oryzeae phytolith fossils (Prasad et al., 2011). The tree and log files produced from the BEAST analysis were combined with Logcombiner v2.1.2 and convergence assessed using Tracer v1.6 (Rambaut et al., 2014). FigTree v1.4.2 (Rambaut et al., 2014) was used to view and edit the combined tree file generated for each run of BEAST with a burn-in of 20%. Divergence estimates for nodes relevant to the deep relationships of PACMAD grasses are given (Table 4).

Table 4

Divergence Estimates for Each BEAST Analysis and Their Respective Calibrations and Age of Each Given Node

Beast Analysis	Calibrations		Crown PACMAD	Aristidoideae	MADC	Crown Arun/Micr	Chloridoideae	Crown Danthonioideae	Crown Panicoideae	centothecoid
1	(Chloridoideae: 14-19 mya)	Age	32.44	31.19	25.77	22.58	13.48	7.48	20.24	6.92
	(Crown Panicoideae: 7-8 mya)	95% HPD Lower	11.89	12.75	10.33	9.80	7.63	1.80	7.93	0.50
		95% HPD Upper	50.55	46.57	38.26	34.33	19.58	15.60	36.82	27.12
2	(Chloridoideae: 14-19 mya)	Age	32.74	20.46	19.08	17.31	14.43	12.99	23.61	16.52
	(Crown Panicoideae: 7-8 mya)	95% HPD Lower	16.99	10.55	9.51	7.26	5.74	2.86	8.21	2.81
	(Crown BEP/PACMAD: 65-67 mya)	95% HPD Upper	45.20	25.61	22.61	19.79	18.97	13.29	36.75	18.82

CHAPTER 3 - RESULTS

Outgroup Selection, Plastome

Although the PACMAD topology remained consistent across likelihood analyses conducted with different outgroups, selection greatly influenced support values for the position of Aristidoideae. Considering only single-taxon outgroups, the choice of *Puelia olyriformis* generated a bs value of 67% for this node. The use of the somewhat more closely related ehrhartoid species, *Oryza rufipogon*, increased the bs value to 76%. When *Bambusa oldhamii* was the outgroup this node had a bs value of 80%, but the greatest support was obtained when *Rhynchoryza subulata* was the outgroup with a bs value of 99%. *Bambusa oldhamii* was selected as the outgroup for the mitochondrial analysis since mitochondrial data were not available for *R. subulata*.

Plastome Characterization

The 13 new plastomes were largely conserved in gene content and organization. The short single copy (SSC) regions had ranges of 11,771 to 14,756 bp in length, long single copy regions (LSC) from 78,798 to 82,525 bp, and inverted repeats (IR) from 20,103 to 22,730 bp (see Table 3). All genes were conserved across the 19 species analyzed with the exception of *rpl14* and *rpl16* in *Hakonechloa macra*. A unique deletion of ~1900 bp in *H. macra* was found. The deletion includes all of *rpl16* as well as the first 70 bp of *rpl14* and the noncoding sequence between them. A 10 bp inversion in the SSC was present in most members of Panicoideae (with the exception of *Thysanolaena maxima*) and the outgroup taxon *R. subulata*. The sequence was

noncoding and found between the *petG* and *trnW*-CCA tRNA genes. A deletion totaling 46 bp was also found in each member of Panicoideae as well as *R. subulata* in the noncoding region between *ndhF* and *rpl2*. One last mutation of interest was a 1470 bp deletion found roughly 1000 bp upstream of *trnL*-CAA tRNA and 1383 bp downstream of *trnL*-CAU. The deletion was present in *O. rufipogon* and *R. subulata* as well as each member of Arundinoideae, Chloridoideae, Danthonioideae and Micrairoideae.

Phylogenomic Analyses

The ML analysis produced a tree with −lnL = 274737.67. The tree had mean terminal branch lengths (0.009) more than 2.5 times greater than the mean of the internal branch lengths (0.0035). The topology generated from the maximum likelihood (ML) analysis differed from that of the best topology of GPWGII (2012) with respect to the positions of Aristidoideae and Panicoideae. Here, Panicoideae (six species) are the most deeply diverging subfamily within the PACMAD clade. The next subfamily to diverge is Aristidoideae, which is united with the CMAD clade with a bs value of 99%. Chloridoideae is supported as sister to Danthonioideae with a bs value of 100%. The sister relationship between Micrairoideae and Arundinoideae is also supported with a bs value of 100%. The chloridoid/danthonioid clade is maximally supported as sister to the arundinoid/micrairoid clade.

The BI analysis produced a tree with -lnL = 280353.08. The BI topology was identical to the ML topology. The position of Aristidoideae and all other nodes in the topology were supported posterior probability (pp) values of 1.0.

The same data matrix under parsimony analysis had 6,457 parsimony informative sites. The maximum parsimony (MP) analysis produced a single tree of length 23,013 steps. The MP

analysis had an ensemble consistency index of 0.6072 and retention index of 0.6400. The divergence order for the MP analysis varied slightly from the ML and BI analyses with Aristidoideae diverging first, followed by the Panicoideae, consistent with GPWGII (2012). Bootstrap support values at each node were 100 with the exception of the crown Arundinoideae (bs value = 90%).

Mitochondrial Analysis

The mitochondrial analysis produced a tree with -lnL = 17041.15. The topology differed greatly from that of the plastome analysis. Panicoideae remained monophyletic, but with relatively weak support (bs value: 79%). The sister relationships of *T. maxima* and *Centotheca lappacea* as well as *Z. mays* and *Coleataenia prionitis* were each supported with bs values of 98%. A bs value of 89% constrained Danthonioideae as monophyletic. The sister relationship of Danthonioideae and Chloridoideae was retained, but with little support (bs value: 56%). Arundinoideae, as sampled here, were characterized as monophyletic and placed sister to the Chloridoideae/Danthonioideae clade with a bs value < 50%, unlike the relationships in GPWGII (2012). *Aristida purpurea* was paraphyletic with Micraroideae, and had bs <50% for many nodes. The mitochondrial topology was substantially incongruent to those from the plastome analyses and was not used in further analyses.

Divergence Date Estimation

A divergence date for the stem node of the PACMAD clade was estimated to be 53.81 mya using two fossil calibration points (see Table 4). A second divergence estimate was conducted with the addition of a calibration point at the stem PACMAD node. The estimated divergence date for the stem node of the PACMAD clade with these calibrations was 65.48 mya.

The divergence date of greatest interest, at the crown PACMAD node, was 32.44 mya and 32.74 mya for calibration sets one and two respectively. The Aristidoideae divergence of 31.19 mya for the first set and 20.46 mya for the second set was the most variable between the two. The addition of the calibration point in set two caused the stem and crown arundinoid divergences to decrease, while the divergence dates of the crown micrairoid, danthonioid, panicoid, and stem chloridoid lineages increased (see Table 3). Also of note is the crown panicoid divergence date of 20.34 mya for set one and 23.61 mya for set two.

CHAPTER 4 - DISCUSSION

The rapid radiation of PACMAD grasses makes it difficult to tease out their evolutionary relationships. The phylogenomic analysis of Poaceae presented here offers a clear advantage to understanding the deep divergences within the PACMAD grasses. Analysis of full plastome sequences and estimation of divergence times allows us to speculate on the cause of this accelerated diversification.

Phylogenomic Analyses

Phylogenetic analyses of rapid radiations in plant lineages such as the PACMAD clade tend to be challenging because the outgroups are on relatively long branches that equally attract the long terminal branches of the ingroup. The ingroup taxa are connected to each other only by short internal branches with relatively little phylogenetic information for robust resolution of deep relationships (Rothfels et al., 2012). Note that our MP analysis likely produced a different topology than our ML/BI topology due to long branch attraction to which parsimony is somewhat more susceptible (Swofford et al., 2003).

Each ML analysis generated an identical topology, but several outgroup taxa produced minimal support for the stem Aristidoideae. Outgroup taxa selected for their relatively close relationships to the PACMAD grasses, *Oryza rufipogon*, *Bambusa oldhamii*, and *Rhynchoryza subulata*, presented stronger support for this node than the more distantly related *Puelia olyriformis. B. oldhamii* exhibited a shorter terminal branch length than *R. subulata* and *O. rufipogon* (Wu and Ge, 2012), which was correlated with stronger support at the stem

Aristidoideae (80%) that fell between those values when *R. subulata* (99%) and *O. rufipogon* (76%) were selected as outgroups. One possible explanation is that the plastome of *B. oldhamii* exhibits greater levels of homoplasy with the PACMAD clade than *R. subulata*.

Past studies on the taxonomy of Poaceae (GPWGII, 2012) have provided weak support for the sister relationship of Panicoideae to the CMAD clade (bs value: 61%) as well as between the micrairoid/arundinoid and danthonioid/chloridoid clades (bs values: 51%). Although the taxon sampling of GPWGII (2012) included approximately 70% of PACMAD species, only three genetic markers of 600-800 bp each were analyzed. The phylogenomic method of analysis here allows for an increase in the number of molecular markers by several orders of magnitude to provide more informative sites and raise support values for the phylogeny. In our most well-supported likelihood topology (see Fig. 1) Panicoideae are sister to the other PACMAD grasses, and Aristidoideae are sister to the CMAD clade. All subfamilies sampled with two or more species were recovered as monophyletic. Although sample size does not allow us to phylogenomically investigate Centothecoideae as a subfamily (Clark et al., 1995; GPWGI, 2001; Sanchez-Ken et al., 2007), the two taxa formerly characterized as centothecoids (*Thysanolaena maxima* and *Centotheca lappacea*) were supported as monophyletic and sister to Panicoideae. Complete plastome sequence analysis is thus able to provide strong support for the phylogenetic relationships and suggests that further sampling of complete plastomes from PACMAD taxa would be useful to address relationships at lower taxonomic levels.

Evidence of three infrequent mutation events relevant to our phylogenomic analyses were also discovered (Fig. 3). The first was an inversion that occurred in the LSC region between *petG* and *trnW*-CCA tRNA, which was found in two of three outgroup species (*O. rufipogon and R.*

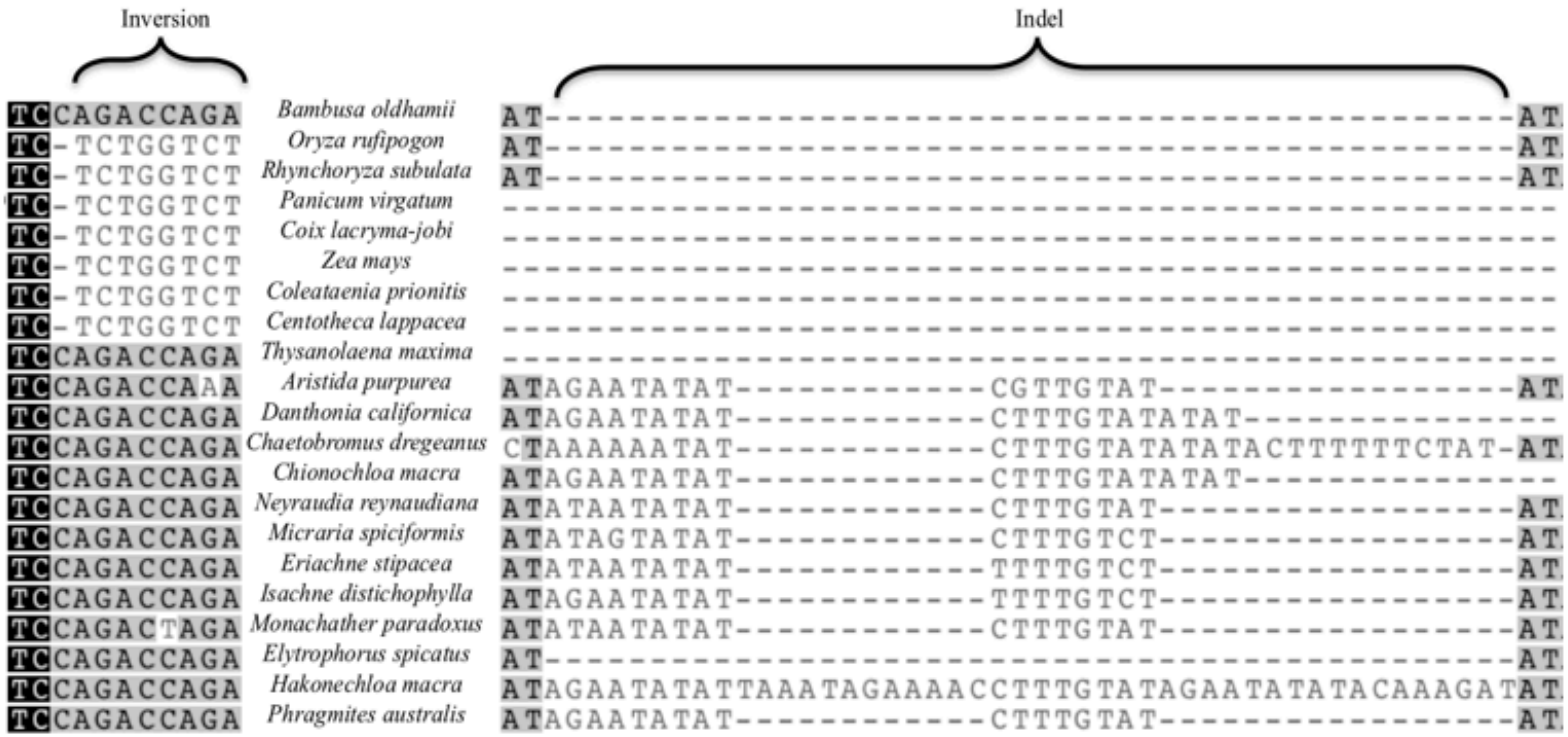

Figure 3: Inversion and Indel mutation events across 18 PACMAD species and three outgroup taxa.

subulata) and also present in most members of Panicoideae, with the exception of *T. maxima*. The inversion was not found in *A. purpurea*, or any other ingroup taxa. The second event was a 46 bp indel in the SSC region of noncoding sequence between *ndhF* and *rpl2.* The deletion was found in each outgroup species as well as each member of Panicoideae and an arundinoid (*E. spicatus*). These rare mutation events are generally consistent with our phylogenomic analysis but show some homoplasy for individual taxa. The third event was a 1470 bp deletion present in each member of the Micrairoideae, Arundinoideae, Danthonioideae and Chloridoideae as well as two outgroup taxa, *O. rufipogon*, and *R. subulata*, which provides support for the CMAD clade but is equivocal with regard to the deep branches of the PACMAD ingroup.

Divergence Estimates

Previous studies have set out to determine divergence dates for PACMAD grasses using many taxa, but with a relatively minimal number of molecular markers. The increase in molecular data in this study allows for a more accurate assessment of divergence dates at deep nodes due to the presence of greater phylogenetically informative sites. The age of the BEP/PACMAD clade was recently determined to be 54.9 mya (Christin et al., 2013), which is similar to the estimate here of 53.8 mya for the BEP/PACMAD crown node. This estimate suggests that the BEP and PACMAD clades diverged at the approximate time of the Paleocene-Eocene thermal maximum (PETM) (55-56 mya) and the transition to the Eocene era. The Eocene is characterized as a period of cooling and drying which led to forest fragmentation and created new habiats for open habitat and marginal forest species (Bellosi and Krause, 2013). The divergence of the PACMAD subfamilies has been formerly estimated to fall between 38 mya (Christin et al., 2008) and 45 mya (Bouchenak-Khelladii et al., 2010) but was resolved to an age

of 32.4 mya in this analysis. The rapid radiation of the PACMAD clade occurs along the Eocene-Oligocene transition (EOT). Throughout the Eocene there was a global cooling trend following Antarctic glaciation events (Mudelsee et al., 2014) as well as declining atmospheric CO_2, promoting C_4 photosynthesis (Christin et al., 2008). These climatic changes influenced habitat diversification across the globe, which allowed these wind pollinators to rapidly diversify and colonize open areas and forest margins following the EOT. Aristidoideae are often open-habitat grasses (GPWGI, 2001). If, as suggested by GPWGII (2012) the Aristidoideae are sister to the PCMAD grasses, then the exploitation of open habitats long preceded the radiation of the PACMAD clade for which there is no explanatory hypothesis. Panicoideae, however, are comprised of genera that colonize shady habitats (*Centotheca*), forest margins (*Coix*), and open habitats (*Thysanolaena*) (GPWGI, 2001), which would be expected of the deepest diverging subfamily following if climatic changes began to select open-habitat species. Species composition of the earliest diverging PACMAD lineage would be expected to fill both shade-tolerant and open-habitat niches, which is seen in Panicoideae and is supported by the phylogenomic analyses.

Conclusions

The PACMAD grasses have been repeatedly reclassified through several varying analyses of chloroplast data subsets with minimal molecular information. Deep intersubfamilial relationships were here retrieved with greater support than previous studies through the use of a phylogenomic approach. Our results provide strong support for Panicoideae as sister to the ACMAD clade allowing for further exploration of terminal relationships. It also offers an

approach for teasing out phylogenetic relationships of weakly supported, rapidly radiating lineages.

Divergence estimates for the PACMAD clade provides insight into the putative role of climate changes leading to habitat diversification, which possibly triggered the rapid radiation of these grasses. The date of 32.4 mya for the initial radiation of the lineage is consistent with the EOT. Prior, there were glaciation events and an overall global cooling trend throughout the Eocene, which led to environmental diversification via forest fragmentation and expansion of open habitats. These changes may have allowed grasses to rapidly speciate, hybridize, and ultimately dominate newly developed habitats.

REFERENCES

Bell, H. L., & Columbus, J. T. (2008). Proposal for an expanded Distichlis (Poaceae, Chloridoideae): support from molecular, morphological, and anatomical characters. *Systematic Botany*, *33*(3), 536-551.

Bellosi, E. S., & Krause, J. M. (2013). Onset of the Middle Eocene global cooling and expansion of open-vegetation habitats in central Patagonia. *Andean Geology*, *41*(1), 29-48.

Bouchenak‐Khelladi, Y., Verboom, G. A., Savolainen, V., & Hodkinson, T. R. (2010). Biogeography of the grasses (Poaceae): a phylogenetic approach to reveal evolutionary history in geographical space and geological time. *Botanical Journal of the Linnean Society*, *162*(4), 543-557.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C. H., Xie, D., ... & Drummond, A. J. (2014). BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS computational biology*, *10*(4), e1003537.

Burke, S. V., Grennan, C. P., & Duvall, M. R. (2012). Plastome sequences of two New World bamboos—Arundinaria gigantea and Cryptochloa strictiflora (Poaceae)— extend phylogenomic understanding of Bambusoideae. *American journal of botany*, *99*(12), 1951-1961.

Christin, P. A., Besnard, G., Samaritani, E., Duvall, M. R., Hodkinson, T. R., Savolainen, V., & Salamin, N. (2008). Oligocene CO_2 Decline Promoted C_4 Photosynthesis in Grasses. *Current Biology*, *18*(1), 37-43.

Christin, P. A., Spriggs, E., Osborne, C. P., Strömberg, C. A., Salamin, N., & Edwards, E. J. (2013). Molecular Dating, Evolutionary Rates, and the Age of the Grasses. *Systematic biology*, syt072.

Clark, L. G., Zhang, W., & Wendel, J. F. (1995). A phylogeny of the grass family (Poaceae) based on ndhF sequence data. *Systematic Botany*, 436-460.

Clifton, S. W., Minx, P., Fauron, C. M. R., Gibson, M., Allen, J. O., Sun, H., ... & Newton, K. J. (2004). Sequence and comparative analysis of the maize NB mitochondrial genome. *Plant physiology*, *136*(3), 3486-3503.

Cook J., Wathen K. J., and Nguyen H. (2013) Parametersolver v3.0 Available from https://biostatistics.mdanderson.org/SoftwareDownload/SingleSoftware.aspx?Software_Id=6

Cox, M. P., Peterson, D. A., & Biggs, P. J. (2010). SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. *BMC bioinformatics*, *11*(1), 485.

Dhingra, A., & Folta, K. M. (2005). ASAP: amplification, sequencing & annotation of plastomes. *BMC genomics*, *6*(1), 176.

Duvall, M. R., Davis, J. I., Clark, L. G., Noll, J. D., Goldman, D. H., & Sánchez-Ken, J. G. (2007). Phylogeny of the grasses (Poaceae) revisited. *Aliso: A Journal of Systematic and Evolutionary Botany*, *23*(1), 237-247.

Edwards, E. J., Osborne, C. P., Strömberg, C. A., & Smith, S. A. (2010). The origins of C4 grasslands: integrating evolutionary and ecosystem science. *science*, *328*(5978), 587-591.

Edwards, E. J., & Still, C. J. (2008). Climate, phylogeny and the ecological distribution of C4 grasses. *Ecology Letters*, *11*(3), 266-276.

Felsenstein, J. PHYLIP (2005). (Phylogeny Inference Package) version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle. *Efficiently Finding the Most Parsimonious Phylogenetic Tree*, *47*.

Grass Phylogeny Working Group I, Barker, N. P., Clark, L. G., Davis, J. I., Duvall, M. R., Guala, G. F., ... & Linder, H. P. (2001). Phylogeny and subfamilial classification of the grasses (Poaceae). *Annals of the Missouri Botanical Garden*, 373-457.

Grass Phylogeny Working Group II, (2012). New grass phylogeny resolves deep evolutionary relationships and discovers C4 origins. *New Phytologist*, *193*, 304- 312.

Guo, Y. L., & Ge, S. (2005). Molecular phylogeny of Oryzeae (Poaceae) based on DNA sequences from chloroplast, mitochondrial, and nuclear genomes. *American Journal of Botany*, *92*(9), 1548-1558.

Jones, S. S., Burke, S. V., & Duvall, M. R. (2014). Phylogenomics, molecular evolution, and estimated ages of lineages from the deep phylogeny of Poaceae. *Plant Systematics and Evolution*, *300*(6), 1421-1436.

Katoh K, Kuma KI, Toh H, Miyata T (2005). MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res*, 33(2): 511-518. doi: 10.1093/nar/gki198.

Leseberg, C. H., & Duvall, M. R. (2009). The complete chloroplast genome of Coix lacryma-jobi and a comparative molecular evolutionary analysis of plastomes in cereals. *Journal of molecular evolution*, *69*(4), 311-318.

Miller, M. A., Pfeiffer, W., & Schwartz, T. (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In *Gateway Computing Environments Workshop (GCE), 2010* (pp. 1-8). IEEE.

Morris, L. M., & Duvall, M. R. (2010). The chloroplast genome of Anomochloa marantoidea (Anomochlooideae; Poaceae) comprises a mixture of grass-like and unique features. *American journal of botany*, *97*(4), 620-627.

Mudelsee, M., Bickert, T., Lear, C. H., & Lohmann, G. (2014). Cenozoic climate changes: A review based on time series analysis of marine benthic δ18O records. *Reviews of Geophysics*.

Prasad, V., Strömberg, C. A. E., Leaché, A. D., Samant, B., Patnaik, R., Tang, L., ... & Sahni, A. (2011). Late Cretaceous origin of the rice tribe provides evidence for early diversification in Poaceae. *Nature Communications*, *2*, 480.

Rambaut A., et al. (2014) FigTree v1.4.2, Available from http://tree.bio.ed.ac.uk/software/figtree/

Rambaut A., Suchard M.A., Xie D. & Drummond A.J. (2014) Tracer v1.6, Available from http://beast.bio.ed.ac.uk/Tracer

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., ... & Huelsenbeck, J. P. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space.*Systematic biology*, *61*(3), 539-542.

Rothfels, C. J., Larsson, A., Kuo, L. Y., Korall, P., Chiou, W. L., & Pryer, K. M. (2012). Overcoming deep roots, fast rates, and short internodes to resolve the ancient rapid radiation of eupolypod II ferns. *Systematic biology*, *61*(3), 490-509.

Sánchez-Ken, J. G., & Clark, L. G. (2010). Phylogeny and a new tribal classification of the Panicoideae sl (Poaceae) based on plastid and nuclear sequence data and structural data. *American Journal of Botany*, *97*(10), 1732-1748.

Sánchez-Ken, J. G., Clark, L. G., Kellogg, E. A., & Kay, E. E. (2007). Reinstatement and emendation of subfamily Micrairoideae (Poaceae). *Systematic Botany*, *32*(1), 71- 80.

Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, *22*(21), 2688-2690.

Strömberg, C. A. (2005). Decoupled taxonomic radiation and ecological expansion of open-habitat grasses in the Cenozoic of North America. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(34), 11980-11984.

Swofford, D. L. (2003). {PAUP*. Phylogenetic analysis using parsimony (* and other methods). Version 4.}.

Vicentini, A., Barber, J. C., Aliscioni, S. S., Giussani, L. M., & Kellogg, E. A. (2008). The age of the grasses and clusters of origins of C4 photosynthesis. *Global Change Biology*, *14*(12), 2963-2977.

Von Braun, J. (2007). *The world food situation: new driving forces and required actions*. Intl Food Policy Res Inst.

Watson, L., & Dallwitz, M. J. (1992). *The grass genera of the world*. CAB international.

White, R. P., Murray, S., & Rohweder, M. (2000). *Grassland ecosystems* (p. 81). Washington, DC, USA: World Resources Institute.

Wu, Z. Q., & Ge, S. (2012). The phylogeny of the BEP clade in grasses revisited: Evidence from the whole-genome sequences of chloroplasts. *Molecular phylogenetics and evolution*, *62*(1), 573-578.

Wysocki, W. P., Clark, L. G., Kelchner, S. A., Burke, S. V., Pires, J. C., Edger, P. P., ... & Duvall, M. R. (2014). A multi-step comparison of short-read full plastome sequence assembly methods in grasses. *Taxon*, *63*(4), 899-910.

Zerbino, D. R., & Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome research*, *18*(5), 821-829.