1 -Revised Aug 8, 2016-

2
3
4 Title: Evolutionary history of chloridoid grasses estimated from 122 nuclear loci

5
6
7 Authors:

8
9 Amanda E. Fisher ^{a,*}, Kristen M. Hasenstab ^b, Hester L. Bell ^b, Ellen Blaine ^c,
10 Amanda L. Ingram ^d, J. Travis Columbus ^b

11
12
13 ^a Department of Biological Sciences, California State University Long Beach, Long
14 Beach, California 90840, USA. Amanda.Fisher@csulb.edu

15
16 ^b Rancho Santa Ana Botanic Garden and Claremont Graduate University, 1500 North
17 College Avenue, Claremont, California 91711, USA.

18
19 ^c Department of Computer Science, Stanford University, 353 Serra Mall, Stanford,
20 California 94305, USA.

21
22 ^d Department of Biology, Wabash College, Crawfordsville, Indiana 47933, USA.

23
24 *Corresponding Author Address: Department of Biological Sciences, California State
25 University Long Beach, Long Beach, California, 90840, USA.

26 *Email address:* Amanda.Fisher@csulb.edu.

27 *Phone number:* 562-985-4814 (office)

ABSTRACT

Chloridoideae (chloridoid grasses) are a subfamily of ca. 1700 species with high diversity in arid habitats. Until now, their evolutionary relationships have primarily been studied with DNA sequences from the chloroplast, a maternally inherited organelle. Next-generation sequencing is able to efficiently recover large numbers of nuclear loci that can then be used to estimate the species phylogeny based upon biparentally inherited data. We sought to test our chloroplast-based hypotheses of relationships among chloridoid species with 122 nuclear loci generated through targeted-enrichment next-generation sequencing, sometimes referred to as hyb-seq. We targeted putative single-copy housekeeping genes, as well as genes that have been implicated in traits characteristic of, or particularly labile in, chloridoids: e.g., drought and salt tolerance. We recovered ca. 70% of the targeted loci (122 of 177 loci) in all 47 species sequenced using hyb-seq. We then analyzed the nuclear loci with Bayesian and coalescent methods and the resulting phylogeny resolves relationships between the four chloridoid tribes. Several novel findings with this data were: the sister lineage to Chloridoideae is unresolved; *Centropodia* + *Ellisochloa* are excluded from Chloridoideae in phylogenetic estimates using a coalescent model; *Sporobolus subtilis* is more closely related to *Eragrostis* than to other species of *Sporobolus*; and *Tragus* is more closely related to *Chloris* and relatives than to a lineage of mainly New World species. Relationships in Cynodonteae in the nuclear phylogeny are quite different from chloroplast estimates, but were not robust to changes in the method of phylogenetic analysis. We tested the data signal with several partition schemes, a concatenation analysis, and tests of alternative hypotheses to assess our confidence in this new, nuclear estimate of evolutionary relationships. Our work provides markers and a framework for additional phylogenetic studies that sample more densely within chloridoid tribes. These results represent progress towards a robust classification of this important subfamily of grasses, as well as proof-of-concept for hyb-seq next-generation sequencing as a method to generate sequences for phylogenetic analyses in grasses and other plant families.

Keywords: Chloridoideae, grass phylogeny, hyb-seq targeted sequencing, phylogenomics, Poaceae.

1. Introduction

Grasses (Poaceae) are one of the most important plant families, as dominant constituents of grasslands and savannas, forage for wild and domestic animals, building materials (bamboos), and the major source of carbohydrates for most humans (rice, corn, wheat; GPWG I, 2001; Spriggs et al., 2014). Subfamily Chloridoideae is in the PACMAD clade of grasses (Panicoideae, Arundinoideae, Chloridoideae, Micrairoideae, Aristidoideae, and Danthonioideae), which is sister to the BOP clade (Bambusoideae, Oryzoideae, and Pooideae; GPWG II, 2011; GPWG I, 2001; Soreng et al., 2015). There are 1400-1700+ species and 131-140 genera of chloridoids (Kellogg 2015; Peterson et al., 2010a; Soreng et al., 2015), and the subfamily is monophyletic in phylogenetic analyses (Bouchenak-Khelladi et al., 2010; Christin et al., 2014, 2008; Clark et al., 1995; Duvall et al., 2007, 2016; GPWG II, 2011; GPWG I, 2001; Hilu and Alice, 2001; Hilu et al., 1999; Peterson et al., 2010a).

Chloridoids have a cosmopolitan distribution with centers of diversity in subtropical and tropical deserts. Many of these grasses are specialized to thrive in arid, resource-poor habitats (Clayton and Renvoize, 1986) and all core chloridoid species sampled to date have C₄ Kranz leaf anatomy [since the re-classification of the C₃ species *Eragrostis walteri* to Arundinoideae (Ingram et al., 2011)]. Several recent papers (Kellogg, 2015; P. Peterson et al., 2011; Soreng et al., 2015) treat Centropodieae as a tribe in Chloridoideae, resulting in the inclusion of a C₃ lineage, *Ellisochloa*, within the subfamily. Chloridoid species are highly variable in their morphology, but most have "chloridoid" bicellular microhairs (short, club-shaped hairs where the basal cell is longer than the apical cell). The growth habit of chloridoid grasses ranges from diminutive caespitose annuals (e.g., several *Muhlenbergia* species) to perennials that are tall and reed-like (e.g., *Neyraudia*) or spreading and mat-forming (e.g., *Cynodon*). Inflorescence architecture is highly variable and spikelets may possess 1-100+ florets (Liu et al., 2005). Various sexual

systems have been documented in chloridoids, including monoecy and dioecy (Kinney et al., 2008, 2003).

Prior to insights from molecular phylogenies, grass classifications over the past 70 years have assigned chloridoid species to four to nine tribes based on inflorescence and spikelet morphology (reviewed in GPWG I, 2001). Some recent classifications of the subfamily based on chloroplast and nrITS data recognize five tribes: Centropodieae, Triraphideae, Eragrostideae, Zoysieae, and Cynodonteae (Kellogg, 2015; Peterson et al., 2010a; Soreng et al., 2015). However, the Grass Phylogeny Working Group II (2011) did not accept Centropodieae within Chloridoideae, a decision we support until additional molecular and anatomical evidence is made available. Molecular phylogenetic studies that sampled across chloridoids are largely consistent in that Eragrostideae are sister to Zoysieae + Cynodonteae (Columbus et al., 2007; Duvall et al., 2016; Hilu and Alice, 2001; Mathews et al., 2000; Peterson et al., 2010a; Roodt-Wilding and Spies, 2006). Studies focusing on species-level relationships have found that many genera were not monophyletic (e.g., Bell and Columbus, 2008; Columbus et al., 2000; Ingram and Doyle, 2003; Peterson et al., 2012, 2010b) and that hybridization and polyploidization have played a role in the evolution of some species (Ainouche et al., 2003; Bell and Columbus, 2008; Liu et al., 2011; Siqueiros-Delgado et al., 2013). Relationships among genera in tribe Cynodonteae have been particularly difficult to resolve with chloroplast and nuclear ribosomal ITS datasets (Columbus et al., 2007; Duvall et al., 2016; Peterson et al., 2010a).

Previous studies of chloridoid phylogeny (Columbus et al., 2007; Hilu and Alice, 2001; Mathews et al., 2000; Peterson et al., 2010a; Roodt-Wilding and Spies, 2006) have relied heavily on chloroplast and nrITS molecular loci but have failed to resolve key relationships within tribes. In plants the genes and noncoding regions of the chloroplast genome have been frequently used for phylogenetic analyses. Chloroplasts are assumed to be uniparentally (maternally) inherited in most species (Birky, 2001, 1995) and haploid; therefore, they are expected to achieve coalescence

more quickly than biparentally inherited nuclear loci (Nichols, 2001). Nuclear loci, on the other hand, may provide a more accurate estimate of the evolutionary history of species because loci on different chromosomes will assort independently from each other during meiosis and sampling many unlinked loci reduces the chance of sampling error for the estimate of phylogenetic history (Philippe et al., 2011). Additionally, most plant nuclear loci accumulate mutations at a faster rate than chloroplast DNA (Clegg et al., 1994) and therefore, nuclear loci may provide a means to resolve relationships when branches in a tree derived from chloroplast data are short (soft polytomies; Barrett et al., 2013; Delsuc et al., 2005; Rokas et al., 2003). High throughput next-generation sequencing (NGS) combined with recent developments in targeted capture methods or hyb-seq (Gnirke et al., 2009) enable researchers to sequence hundreds of targeted nuclear loci from across a genome (Grover et al., 2012; Mamanova et al., 2010). These sequences can be used to assemble large datasets for phylogenetics at costs comparable to cloning and Sanger sequencing a few nuclear regions (Cronn et al., 2012; Lemmon and Lemmon, 2013).

We sought to test chloroplast-based hypotheses of relationships among chloridoid grass species with a dataset of nuclear loci. We used hyb-seq (Gnirke et al., 2009) to selectively amplify 177 nuclear loci from 47 species. The hyb-seq NGS method consists of baits that are hybridized to genomic DNA to capture and sequence specific regions of the genome (reviewed in Egan et al., 2012; Lemmon and Lemmon, 2013). This approach has been successfully used in several plant groups to sequence large numbers of loci for phylogeny estimates (de Sousa et al., 2014; Mandel et al., 2014; Salmon et al., 2012; Stephens et al., 2015; Stull et al., 2013; Tennessen et al., 2013; Weitemier et al., 2014). Our sampling emphasized tribe Cynodonteae (25 species) because this is the largest tribe in the subfamily, and it has been the most difficult to resolve with chloroplast and nrITS data. We used the hyb-seq dataset to estimate a nuclear phylogeny across the Chloridoideae subfamily using a coalescent multi-locus species tree approach (Liu et al., 2010).

2. Material and methods

2.1 Sampling

Taxon sampling was guided by previous phylogenetic estimates of PACMAD (Duvall et al., 2007; GPWG II, 2011; GPWG I, 2001) and chloridoid lineages of grasses (Columbus et al., 2007; Peterson et al., 2007). We sequenced 47 species of PACMAD grasses (Table 1). Our outgroup sample included representatives of Aristidoideae, Arundinoideae, Danthonioideae, and Micrairoideae. *Centropodia* and *Ellisochloa* have been included as the earliest-diverging lineage in Chloridoideae in recent treatments of the grass family (Kellogg, 2015; Soreng et al., 2015) and are represented here by two species. We sampled 41 species across the four tribes, Triraphideae, Eragrostideae, Zoysieae, and Cynodonteae. We supplemented our hyb-seq derived sequences with sequences from the published genomes of *Eragrostis tef*, *Sorghum bicolor* (Panicoideae), *Zea mays* (Panicoideae), and *Oryza sativa* subsp. *japonica* (Oryzoideae in the BOP lineage).

2.2 Bait design for hyb-seq

We chose target loci based on several criteria. We aimed for each nuclear locus to have >85% coding sequence similarity across PACMAD and <25% G content, following the recommendations of MycroArray (Ann Arbor, MI). We chose several loci that had previously been used for phylogenetic analyses in grasses, for example *GBSSI* (Ingram and Doyle, 2004), *rpb2* (Denton et al., 1998), *CEL1*, *CEL2*, *pabp1* (Triplett et al., 2012), and *phyB* (Mathews et al., 2000). We also chose 15 genes based on Duarte et al. (2010) that are putatively single-copy nuclear genes shared across angiosperms. We searched the grass physiology literature for loci implicated in developmental pathways for processes or characteristics of interest to us. These included genes related to drought tolerance (e.g., dehydrins, glycine and proline-rich proteins), salt tolerance (e.g., *carboxypeptidase1*, *HKT1*, *salt overly sensitive1*), sexual system (e.g., *antherear1*, *tasselseed2*), and floral organ

identification genes (e.g., *AGL6*, *barrenstalk*, *fruitfull*, *LHS1*, *pistillata*). We ultimately selected a total of 177 target loci. We inferred that we sampled across the nuclear genome based on the location of the 177 loci in the annotated genomes of maize and rice and gene synteny across grass chromosomes.

Each target locus required one to many sequences to serve as the template for hyb-seq bait production. We searched for sequences in the *Oryza*, *Zea*, and *Sorghum* annotated genomes in Phytozome v 9.1 (<http://www.phytozome.net>) or PLAZA 2.5 (<http://bioinformatics.psb.ugent.be/plaza/>) and then BLASTed the sequence against the Genbank nr and EST databases, restricting the search to the PACMAD lineage. We also searched for PACMAD sequences for these loci in Phytozome, PLAZA, MaizeGDB (<http://www.maizegdb.org>), and TIGR (Childs et al., 2007). Putative orthologs across grasses were aligned in Geneious v. 6.1.6 to assess sequence similarity and G content across PACMAD. If a chloridoid sequence was available for the locus, it served as the bait template. If there was no chloridoid sequence we used one to several sequences from across PACMAD, usually including a *Sorghum* sequence.

The sequence set was analyzed in RepeatMasker (Smit et al., 1996) to remove regions that were repetitive and had low complexity. We sent the resulting file to MycroArray (Ann Arbor, MI) to produce a series of baits with 4x tiling across the bait template.

2.3 Library preparation and next-generation sequencing

Leaves from living plants grown at Rancho Santa Ana Botanic Garden or field-collected leaves dried in silica gel were used for DNA extractions. Total genomic DNA was extracted with a modified CTAB protocol (Doyle and Doyle, 1987) that uses an ethanol precipitation step. All extractions were treated with 1 µl RNase. Total genomic DNA extractions were visualized on an agarose gel and quantified with a Nanovue Plus (GE Healthcare, Pittsburgh, PA) to analyze the concentration

and purity of the DNA. The DNA extracts were then shipped to the Center for Genome Research and Biocomputing (CGRB) at Oregon State University (<http://www.cgrb.oregonstate.edu>).

Library construction at CGRB followed the MycroArray MyBait kit, Illumina protocols, and the protocol described in Tennessen et al. (2013). For each sample, 1 µg of DNA was fragmented with a Diagenode BioRuptor Sonicator for 13 mins. To prepare for multiplex sequencing, indexed adaptors (Bio Scientific, Austin, TX) were ligated to genomic fragments. The libraries of 47 samples were quantified on a Qubit fluorometer (Life Technologies, Eugene, OR). Then 30-90 ng of each library were mixed and genomic libraries were hybridized to the MycroArray probes at one-sixth the volume of the manufacturer's recommendation.

After hybridization and purification, the hybridization reactions were combined so that 24 libraries were sequenced per lane of 101 bp paired-end sequencing with an Illumina Hi-Seq 2000 at the CGRB.

2.4 Sequence assembly

Pairs of zipped FASTQ files from the CGRB sequencing facility were filtered using PRINSEQ v. 0.20.4 (www.edwards.sdsu.edu/cgi-bin/prinseq/). The raw data were reviewed to assess the mean quality score of each file and percentage of unknown bases per read. Sequences were discarded if the length was <20 bp or the mean quality score was <20. Bases with phred scores <20 were trimmed from the 5' and/or 3' end with a sliding window size of 1 and a step size of 1. Paired and singleton fragments passing all filters (the Prinseq good file) were exported as FASTQ files.

Filtered sequences (singletons and paired) were imported into Geneious v. 6.1.6 or v. 7.0 on a mid-2011 iMac with 16 GB RAM. The "Set Paired Reads" command was used to link forward and reverse paired reads. The expected distance

between paired reads was calculated as 100-120 b below the sample-specific library fragment size.

The dataset was likely to include many plastome sequences, so the filtered sequences for each species were first assembled to a reference plastome (unpublished data made available by M. R. Duvall, Northern Illinois University) using the Geneious assembler fastest setting and up to 25 iterations. We found that 0.6-10% of the fragment pool for each species assembled to the reference plastome. The fragments that were not assembled during the plastome search became the default fragment set for subsequent nuclear gene searches.

Within Geneious we created a folder containing a bait template file for each targeted exon and the filtered fragments for a single query species. The Map to Reference function was used to assemble the species fragment pools to each exon so that assemblies of individual exons were generated in separate files (L. Ripma pers. comm.). The mapper options used were: Geneious mapper; Sensitivity: Low Sensitivity/Fastest; Fine Tuning: Iterate up to 10 times, option to assemble each sequence list separately. A 50% consensus sequence was saved for each exon for each species. For some loci, the initial assembly seemed to contain multiple sequences, consistent with the simultaneous assembly of paralogous loci. We trialed an iterative approach of 2-5 rounds of increasingly strict assembly to assemble each gene copy. This time-consuming approach was not a viable option for all 122 loci for 47 newly sequenced species. We decided to rely on 50% consensus sequences for a preliminary estimate of the nuclear signal for the following reasons: 1) the sequencing depth for most loci was extremely high; 2) usually one copy seemed to dominate the fragment pool and this copy would be recovered in a 50% consensus sequence; 3) the large number of loci in the analysis would help to mask misleading phylogenetic signal in the data caused by including paralogous loci in the analysis; and 4) the consensus sequences gave reasonable results that largely support previous estimates based on chloroplast loci.

Outgroup sequences were located through CoGe (<https://genomeevolution.org/CoGe/>) by BLASTing the bait template for each exon against the following published genomes: *Eragrostis tef*—Group of Dr. Zerihum Tadele unmasked v1.1.2; *Sorghum bicolor*—JGI: v1.4 unmasked; *Zea mays*—MaizeSequence.org: refgen_v2 (filtered gene set annotations) v2.01 unmasked; and *Oryza sativa* subsp. *japonica*—MSU Rice Genome Annotation: v7 unmasked. Query sequences in CoGe were the exon sequences in batches of 20-50 exons, using an e-value of 0.0001, word size of 8, gap cost of existence: 5, extension: 2, and yes filter query sequence. We exported the Top Hits file, which saves the best sequence based on the high-scoring sequence pair metric (HSP) for each input sequence. Queries that returned large numbers of hits were re-analyzed with a more stringent e-value. Query sequences less than 100 bases long tended to find no hits.

2.5 Alignment of consensus sequences

The Geneious output was a single FASTA file for each species containing the consensus sequence for every recovered exon. In order to align the sequences for a phylogenetic analysis the consensus sequences for all 51 species were parsed into files for each exon. Combine_Files_Delete_Spaces.py (A.J. Harris, <http://plant-biogeography.webs.com/programsandscripts.htm>) was used to concatenate all of the consensus sequence files and the bait exon sequences into a single file. Several python scripts were written (available at <https://github.com/AEFisher/NGS-scripts.git>) to further prepare the data for alignment. TEF.py (Tool to Extract sequence Fragments; named for the chloridoid grain tef) searched the combined file for the sequences for a given exon and saved them to a FASTA file. AlignHelper.py was used to ensure each gene file contained 52 sequences (51 species and one exon reference sequence) and added a line for any missing species with the name of the missing species and a string of question marks as long as the longest sequence in the file. The script processDNAta.py was then used to assess the recovery success for each exon sampled in the study by reading the AlignHelper exon files and creating a

.csv file with presence/absence data for each species, percentage AT and GC in each file, and median sequence length for each exon.

FASTA files for each exon were imported into Geneious and aligned using the MAFFT plugin (Katoh and Standley, 2013). Alignments were then hand-trimmed to the length of the exon, coverage permitting. Lengths of sequence <100b that were difficult to align were replaced with Ns. Large deletions were replaced with question marks. The alignments of 55 targeted loci had high variability and/or many length mutations and so the MAFFT alignment was poor. Generally, we considered an alignment to be poor quality if the aligned pairwise identity was <60% and we removed these loci from subsequent analyses. The good alignments were then "stripped" so that columns in the alignment were removed if they contained at least one gap. This strategy has been used to align plastomes for phylogenetic analyses (Jones et al., 2014; Wysocki et al., 2014) and tends to remove regions prone to insertions or deletions or other areas of the alignment that are ambiguous. Annotated exon alignments for 122 loci were then concatenated in Geneious.

2.6 Phylogenetic analyses

A neighbor-net network of the concatenated alignment was constructed in SplitsTree v4.13.1 (Huson and Bryant, 2006) with constant sites removed. Outgroups were sequentially removed and the network was re-estimated. Networks were also constructed for each locus and the network fit statistic was recorded.

The dataset was partitioned in several ways (Table S1). The optimal data partitions based on shared model parameters were estimated in PartitionFinder (Lanfear et al., 2012) using RaXML (GTR+G and GTR+I+G models) and PhyML (12 models: JC, F81, F81+G, HKY, HKY+G, HKY+I+G, TVM, TVM+G, TVM+I+G, GTR, GTR+G, GTR+I+G). Additionally, eight partitions were designated *a priori* based on putative biological functions of the nuclear loci.

Saturation plots were used as in Kelchner et al. (2012) to assess whether the phylogenetic information in the dataset was saturated. Saturation plots were made by calculating pair-wise uncorrected-p distances and estimating GTR+I+G model-corrected distances in *PAUP v4.0a136 (Swofford, 2003) and graphing them in Excel.

A Bayesian inference (BI) analysis of the 122 locus dataset used MrBayes v3.2.2 (Huelsenbeck and Ronquist, 2001) on CIPRES. The analysis was run for 5 million generations using an *a priori* choice of the GTR+I+G model of nucleotide evolution (Huelsenbeck and Rannala, 2004) with the parameters of the model estimated separately for 92 character sets identified by the RaXML implementation of PartitionFinder. Another BI analysis was run without the Panicoideae species (*Sorghum bicolor* and *Zea mays*) to assess their effect on PACMAD branching relationships. A third BI analysis of 56 housekeeping genes also used partitions identified by PartitionFinder, a GTR+I+G model and 5 million generations. BI analyses were also performed on each of the eight biological function partitions. Each biological function dataset was partitioned by gene and the MCMC chain was run for one million generations.

2.7 Coalescent species tree analyses

Although it is increasingly cost effective to generate large nuclear datasets, most phylogenetic software has been designed for smaller datasets and computational times can become burdensome as the number of taxa and dataset size increase (de Koning et al., 2010; Stamatakis and Ott, 2008). In particular, coalescent-based phylogenetic methods that incorporate hybridization and incomplete lineage sorting (ILS; Carstens and Knowles, 2007; Heled and Drummond, 2010; Kubatko et al., 2009; Liu, 2008; Maddison and Knowles, 2006) are generally not feasible for large datasets. For this reason we used MP-EST as an alternative to BEST, BUCKy, or *BEAST. ML trees were estimated in RAxML HPC AVX v8 (Stamatakis, 2014) for each locus using the script "raxml_launch_serially.sh" (A.

Narechania). RAxML gene trees for the 122 locus and 56 locus housekeeping datasets were used to estimate a species tree under a coalescent model of evolution using MP-EST (Liu et al., 2010) on the STRAW webserver (Shaw et al., 2013).

2.8 Alternative hypothesis testing

The 56 locus housekeeping dataset was used to produce a maximum likelihood estimate (MLE) in RAxML HPC AVX v8 (Stamatakis, 2014) with a user-starting-tree (-f d) and a partitioned GTR+I+G model. Constraint trees were used to test alternative hypotheses of relationships based on the current classification of Chloridoideae (e.g., monophyly of *Sporobolus*) or the chloroplast phylogeny (e.g., *Distichlis* sister to *Bouteloua*). The MLE tree and the constraint tree were then evaluated in RAxML with an SH test (Shimodaira and Hasegawa, 1999), a conservative test of whether the alternative hypothesis can be rejected given the model (GTR+I+G) and dataset (56 nuclear loci).

2.9 Concatenation analysis

A RADICAL analysis (Narechania et al., 2012) was used to assess how locus concatenation and overall dataset size contributed to support for particular branches in the tree topology. RADICAL randomly concatenates loci into datasets of varying sizes and estimates ML trees with RAxML. These trees are then compared to a reference tree, in this case the 56 locus MLE housekeeping tree. RADICAL calculates a statistic called an average occurrence value based on the frequency that a node in the reference tree appears in the set of ML trees for each dataset size. The fixation point of a node is defined as the number of randomly concatenated loci that are required for a node to always appear in the resulting ML trees. The degradation point of a node is defined as the minimum number of loci that can be randomly combined to yield a tree that does not contain a node. The RADICAL analysis was run for 10 chains of 121 locus combinations on the University of Florida high performance computing cluster.

3. Results

3.1 Success of nuclear capture

The Illumina sequence data contained the targeted nuclear loci, as well as a skim of the genome that included high-copy elements such as the plastome. 416 exons representing 122 nuclear loci were successfully recovered for all 47 newly sequenced species (Fig. 1A). The matrix is available as TreeBase study number S19162. The median number of recovered exons was 534 per species (out of a total of 614). Sequence recovery was not correlated with concentration of the initial total genomic DNA extraction (Fig. 1B).

A total of 333 exons, representing 122 loci, were able to produce satisfactory alignments, for a dataset of 105,486 nucleotides after columns containing gaps were stripped. Each locus alignment contained 1-19 exons with an average of 3 exons per locus (locus details available in the supplementary material). The trimmed alignments ranged from 124 b (*VRN1*) - 3512 b (*EP2EX7 Erect Panicle*) with a median length of 846 b. The concatenated alignment had 24.4% identical sites, 85.2% pairwise identity, and an overall GC content of 54.2%. The GC content of the species ranged from 39% (*Eragrostis tef*) to 56% (*Triraphis*; Table S2). The alignment contained only 0.8% missing data. *Eriachne* had the most missing data (3,008 b) and *Zea* had the least missing data (164 b).

The dataset consisted of nuclear loci from across the genome (Table S1). We sampled 56 housekeeping genes, 27 floral architecture loci, 17 genes related to water and salt stress, 6 photosynthesis genes, 6 vegetative development genes, 5 photoperiod genes, 3 phyllotaxy genes, 3 genes related to sexual system, 2 genes related to heavy metal tolerance, and 1 vernalization and cold tolerance gene. Four of the genes were included in two biological function categories.

3.2 Data conflict

Saturation plots of the 122 locus dataset and partitions based on putative biological function (Fig. 2) suggest that only a small portion of the species pair-wise comparisons are substantially different when a complex model of nucleotide evolution (GTR+I+G) is applied. Many of the model-corrected comparisons do not substantially diverge from the 1:1 ratio line implying that signal saturation in the dataset will not adversely affect phylogeny estimation.

PartitionFinder using RAxML and assessing parameters for GTR models found 109 optimal partitions for the 122 loci. GTR+I+G was the best model for every partition, except Glycine Rich Protein SDG137C, which was assigned to GTR+G (Table S1). PartitionFinder using PhyML and considering 12 models found 94 optimal partitions for 122 loci (Fig. 3). Most of the partitions were assigned to GTR+I+G, but 19 partitions were found to be best characterized by other models, including five partitions with the relatively simple model, HKY+I+G (Table S1).

A neighbor-net network of the 122 locus dataset revealed two main sources of conflict in the chloridoid ingroup that are visible as cycles in the network (arrows in Fig. 4). *Jouvea* shares most characters with species in the New World clade of Cynodonteae, but it also shares some characters with Eragrostideae and the outgroups. Likewise, *Kaokochloa* shares most characters with Eragrostideae and some characters with species in the Zoysieae clade, excluding *Urochondra*. The amount of conflict in the network is low, in comparison with the overall signal (network fit = 99.046).

The effect of outgroups on the topology was evaluated using taxon removal network analyses (Graham et al., 2002; Kelchner and Bamboo Phylogeny Group, 2012). There were no changes to the ingroup relationships in any of the networks when outgroups were removed, but removing the two panicoids (*Sorghum* and *Zea*)

resulted in *Oryza* connecting with Aristidoideae (networks available in supplementary material).

3.3 Phylogenetic estimates

BI estimates of the 56 housekeeping locus and 122 locus datasets recovered a monophyletic Chloridoideae and four tribes with strong support (BI estimates available in supplementary material). The BI analysis excluding Panicoideae demonstrates that there were no topological differences within Chloridoideae when the two panicoid species were removed. Comparison of the 56 and 122 locus BI trees including all 51 species reveals 13 nodes in conflict. Ten out of 13 of the conflicting nodes are in the Cynodonteae and most have 1.00 posterior probability (PP) support in both BI estimates.

The MP-EST species tree of the 56 housekeeping locus dataset recovered a monophyletic Chloridoideae and four tribes with strong support (Fig. 5). Many of the relationships were congruent with the BI estimate, but 12 nodes in the MP-EST species tree were novel (nodes marked with --/-- in Fig. 5). The MP-EST species tree of the 122 locus dataset (supplementary material) recovered a similar topology, but with four conflicting nodes in Cynodonteae and one conflicting node in the outgroups in comparison with the 56 locus MP-EST tree (circled nodes).

BI analyses of eight data partitions based on putative biological function resulted in estimates of varying resolution and support (supplementary material). Generally, partitions with more loci had higher support and more closely resembled the 56 or 122 loci BI estimates. There were several relationships with high support that were only found when loci with the same biological function were concatenated. For example, the BI analysis of 27 floral architecture loci found a moderately supported sister relationship between *Triodia* and *Tripogon* (0.89 PP). The five photoperiod loci tree recovered *Ctenium* as a close relative of *Orcuttia* and *Triodia* (0.64 PP), *Perotis* sister to *Dactyloctenium* and *Tragus* (0.98 PP), and *Farrago*

allied with *Dinebra*, *Microchloa*, *Chloris* and *Astrebala* (0.98 PP). The six photosynthesis loci place *Orcuttia* in a clade with *Perotis*, *Farrago*, and *Ctenium* (1.0 PP). Three sexual system loci yield sister relationships between *Bouteloua gracilis* and *Jouvea* (0.72 PP), *Urochondra* and *Triodia* (0.92 PP), and *Aristida* and *Triraphis* (0.98 PP). The six vegetative development loci estimate a tree with *Tridens* sister to *Triodia* (0.86 PP). Finally, the 17 loci water and salt stress dataset place *Microchloa* sister to a clade that includes *Orcuttia*, *Perotis*, *Farrago* and *Ctenium* (0.76 PP).

3.4 Alternative hypothesis testing

SH tests comparing the 56 locus MLE and 15 alternative hypotheses allowed us to further test the strength of the phylogenetic signal (Table S3). The SH test was able to reject relationships previously inferred from analyses of chloroplast data that were not found in any of our nuclear analyses. However, the SH test with the 56 locus MLE tree failed to reject relationships that had strong support in our estimates of the 122 locus dataset.

3.5 Concatenation analysis

The RADICAL analysis provides a context for understanding how data concatenation and dataset size affect the phylogenetic estimate by randomly sampling loci and estimating ML trees. We ran RADICAL for 10 chains for a total of 1210 ML estimates from a pool of randomly concatenated loci. Some nodes were always present, even when a small number of loci were sampled (Fig. 6A). For example, monophyly of Chloridoideae could be established by sampling any three loci, and Eragrostideae were monophyletic when any four loci were sampled (Fig. 6A, 6B). Some nodes were only present when a large number of loci were sampled. The branch leading to Cynodonteae was present in all trees only after 76 loci were sampled. Five Cynodonteae nodes remained unstable in ML analyses, even after sampling 122 loci. We plotted the frequency of observing individual nodes as dataset size increased and found that some nodes exhibited different "behaviors" as

loci were randomly added (Fig. 6B-E). As more loci were concatenated, the branch leading to Panicoideae steadily increased in frequency as the sister group to the rest of PACMAD (Fig. 6C). The placement of Micrairoioideae and Aristidoideae branches remained volatile no matter the size of the dataset. Unsurprisingly, nodes within Chloridoideae that did not achieve fixation in the RADICAL analysis were also the nodes that were unstable in the MP-EST, BI, and ML estimates of the 122 and 56 locus datasets.

4. Discussion

4.1 Utility of hyb-seq data for phylogenetics

This study is the first phylogenetic estimate of the chloridoid subfamily of grasses using data from more than 100 loci across the nuclear genome. We developed baits from PACMAD grass sequences that recovered ca. 70% of the targeted loci, resulting in alignments for 122 nuclear loci. Phylogenetic analyses revealed high support values for most branches and generally corroborated relationships among four accepted tribes (Triraphideae, Eragrostideae, Zoysieae, Cynodonteae), however we found a number of novel relationships in the coalescent analysis of the nuclear dataset that have not been recovered with chloroplast or nrITS data.

Much of the data that have been used to infer evolutionary relationships in grasses, and in plants in general, have come from a small number of loci in the chloroplast genome (e.g., Chase et al., 1993; Columbus et al., 2007; GPWG II, 2011; Peterson et al., 2010a). Analyses of the complete set of chloroplast genes are increasingly common (Barrett et al., 2013; Burke et al., 2014; Givnish et al., 2010; Jones et al., 2014; Leseberg and Duvall, 2009; Qiu et al., 2006; Ruhfel et al., 2014) and have also been used to evaluate chloridoid phylogeny (Duvall et al., 2016). Chloroplast sequences have been the preferred data source for estimates of plant

phylogeny for a variety of reasons. Chloroplast loci tend to amplify more reliably than nuclear loci and are often present as a single copy in the genome, while there may be multiple copies of nuclear loci due to gene duplication or polyploidization. Comparisons of nuclear and organellar data can be used to evaluate evolutionary patterns of hybridization, introgression, and ILS (Linder and Rieseberg, 2004; Maddison, 1997). The nrITS region has often been used as the representative nuclear locus in plant evolution studies, but this may be problematic in some cases. There may be hundreds of nrITS copies in the nuclear genome subject to a homogenization process that can obscure the evolutionary signal (Álvarez and Wendel, 2003; Feliner and Rosselló, 2007). A nuclear locus that is present in the genome as a single copy may allow for more robust comparisons with organellar hypotheses of phylogeny, but an ideal dataset would contain many nuclear loci to accommodate differences in gene history and rate variation that potentially affect estimates of species phylogeny (Ness et al., 2011; G. Peterson et al., 2011; Whittall et al., 2006).

However, there may be drawbacks to using large nuclear datasets in a phylogenetic analysis. Determining the orthology of sequences for each locus is perhaps the most difficult problem in comparison to chloroplast datasets, since many nuclear genes are part of gene families that may have experienced duplication and loss (Lemmon and Lemmon, 2013; Ness et al., 2011). Comparing these paralogous sequences within an alignment can lead to incorrect estimates of phylogeny (Maddison, 1997). One way to determine orthology is with a gene tree analysis that samples all of the members and copies of a gene family found in multiple species (e.g., Malcomber and Kellogg, 2005), but high similarity between DNA sequences has also been considered evidence for orthology (e.g., Salmon et al., 2012). Paralogous loci may have caused some of the high amounts of gene tree conflict seen in other studies with large nuclear datasets (e.g., Carstens and Knowles, 2007; Philippe et al., 2011) and may be causing some of the conflict among gene trees in our dataset and merits further testing with each locus.

An additional potential drawback to phylogenomic datasets is that bootstrap and posterior probability support values are inflated when datasets have large numbers of variable characters (Jeffroy et al., 2006; Parks et al., 2012) so that high support values at a node do not preclude that there are alternative signals in the data that would also receive high support from bootstrapping or posterior probabilities (Delsuc et al., 2005; Kubatko and Degnan, 2007; Philippe et al., 2011). Instead of relying on branch support, it may be better to consider whether a relationship in a phylogenomic tree is robust to different partitions of the data and different analytical methods. We found 14 conflicting nodes with high branch support when using different phylogenetic frameworks and 5-13 conflicting nodes using different data partitions. We discuss these conflicts in more detail below.

We chose the housekeeping partition of the dataset *a priori* to track the phylogenetic history of chloridoids because the putative biological functions of the loci sampled in the 56 locus dataset (Table S1) suggest that they have housekeeping functions and are less likely to be under positive selection. Several studies have found that using genes under selection for phylogenetic analyses results in erroneous estimates (Erixon and Oxelman, 2008; Guisinger et al., 2008). We analyzed the 56 locus dataset with a partitioned model in MrBayes (Huelsenbeck and Ronquist, 2001) and RAxML (Stamatakis, 2014), but these algorithms do not take into account the potential for different gene histories. Concatenating large nuclear datasets without consideration of coalescent processes has been shown to lead to incorrect species tree estimates (Degnan and Rosenberg, 2006; Heled and Drummond, 2010; Kubatko and Degnan, 2007; Lemmon and Lemmon, 2013; Pyron et al., 2014), especially when there are short branches in the tree (Kubatko and Degnan, 2007; Maddison and Knowles, 2006). Incomplete coalescence is less likely to be a problem when estimating relationships among older lineages, such as between subfamilies or tribes, because lineage-specific alleles are more likely to be fixed (Maddison, 1997; Wendel and Doyle, 1998). However, ILS signals can persist even at older nodes of the tree, especially when there have been rapid radiations (Edwards et al., 2005; Mirarab et al., 2014). Whether this is the case in Cynodontae

remains to be tested with a molecular dating analysis. A computationally efficient method for estimating the species tree from a large set of nuclear gene trees under the coalescent is MP-EST (maximum pseudo-likelihood estimate; Liu et al., 2010) as well as other new methods (Mirarab and Warnow, 2015). An important limitation to coalescent methods is that they do not model historic or current gene flow between species. In the chloridoids we do not expect that hybridization is occurring between genera, but the results of the current study may help to guide future work in chloridoids that sample at the species or population level to test for hybridization. To accommodate the potential for ILS we analyzed the 56 locus dataset under a coalescent framework in MP-EST (Liu et al., 2010). We focus on the results of this tree below, noting differences with previous studies based on chloroplast and nrITS data (Fig. 5).

4.2 Phylogenetic relationships

The chloridoid nuclear phylogeny (Fig. 5) contains many branches with strong support that corroborate earlier chloroplast-based estimates, while other branches in the nuclear tree are novel hypotheses. Our data corroborate the monophyly of Chloridoideae, the presence of four clades equivalent to accepted tribes, and relationships among the tribes. When different partitions of the data were sampled (56 versus 122 loci), the resulting topologies were incongruent and strongly supported, suggesting conflicting signals in the nuclear data. These conflicts were all within the outgroups or within the tribes and most were among genera in Cynodonteae (Fig. 5). Several branches in the tree did not stabilize until more than 100 loci were sampled (Fig. 6), indicating that there are alternative signals in the data that could provide support for different tree topologies. This conflict between data partitions was expected based on the results of other studies (de Sousa et al., 2014) and the conflict apparent in the network analysis (Fig. 4). Comparison of the results from different phylogenetic methods and the concordance analysis provided more detailed evidence of the species causing the discordance. We used networks, branch support values, SH tests, coalescent analyses, a RADICAL

concordance analysis, and model-based and *a priori* biological function-based data partitioning to test the nature of the conflict in the data.

4.2.1 *Centropodieae*

The MP-EST species trees of the 56 and 122 locus datasets contained a polytomy between Chloridoideae, Arundinoideae, Danthonioideae, and *Centropodia* + *Ellisochloa* (Fig. 5). The BI 122 locus dataset found Chloridoideae + *Centropodia*/*Ellisochloa* sister to Danthonioideae (supplementary material) just as in GPWG II (2011) and Duvall et al.'s (2016) plastome study, while GPWG I (2001) found Chloridoideae + *Centropodia*/*Ellisochloa* sister to Arundinoideae. Although chloroplast estimates have converged on PACMAD relationships with high support (Cotton et al., 2015; GPWG II, 2011; Spriggs et al., 2014), it remains to be seen if denser sampling across PACMAD with nuclear loci will resolve the sister to Chloridoideae. Our results suggest that the inclusion of *Centropodia* and *Ellisochloa* as a tribe (Centropodieae) of Chloridoideae may be an artifact of the chloroplast signal. Importantly, uncertainty among the outgroups did not affect relationships within Chloridoideae in network (Fig. 4) or BI analyses (supplementary material).

4.2.2 *Relationships in Chloridoideae*

Triraphis is sister to the rest of the chloridoids and relationships among tribes (Triraphideae sister to Eragrostideae + [Zoysieae + Cynodonteae]) are congruent with previous studies that have used fewer than six chloroplast loci (e.g., Columbus et al., 2007; GPWG II, 2011; Peterson et al., 2010a) and single nuclear loci (Columbus et al., 2007; Mathews et al., 2000; Peterson et al., 2010a). An analysis of chloridoid plastomes that samples a subset of these species (Duvall et al., 2016) is congruent with the 56 locus MP-EST tree, except for the placement of *Distichlis*.

4.2.3 *Relationships in Eragrostideae*

Eragrostideae tend to have spikelets with multiple florets and many of these species are adapted to arid habitats (Ingram and Doyle, 2004; Peterson et al., 2007). Relationships in the tribe with 56 and 122 nuclear loci are congruent with chloroplast and nrITS estimates (Columbus et al., 2007; Peterson et al., 2010b). The Cotteinae of Peterson et al. (2010a) is sister to the rest of the tribe and includes *Kaokochloa*. An undescribed species of Uniolinae and *Fingerhuthia* are sister to a paraphyletic *Eragrostis*. *Eragrostis* is by far the largest genus in Chloridoideae with ca. 350 species and our nuclear data suggests that species in several small genera (e.g., *Diandrochloa*, *Pogonarthria*) should be reclassified in *Eragrostis*, as other authors have advocated (Ingram and Doyle, 2004; Peterson et al., 2010a).

Suprisingly, *Sporobolus subtilis* resolved in Eragrostideae rather than in Zoysieae (Fig. 4, 5) and an SH test strongly rejected a tree with *S. subtilis* constrained to Zoysieae (Table S3). This species is a perennial with an open panicle that is distributed across eastern and southern Africa and Madagascar. *Sporobolus subtilis* has spikelets with a single floret, a character that aligns it with *Sporobolus*. However, it also has a three-nerved lemma (as in *Eragrostis*, but also found in some species of *Sporobolus* as well as *Urochondra*) and a well-developed rachilla extension, as in some members of Cynodonteae (Ortiz-Diaz and Culham, 2000). Further work is needed to sample chloroplast and nuclear loci from additional collections of this species to confirm its placement in Eragrostideae.

4.2.4 Relationships in Zoysieae

Zoysieae species tend to have spikelets with a single floret (except *Pogoneura* which has 2-3 florets per spikelet), single-nerved lemmas and free pericarps (Peterson et al., 2007). *Urochondra* is sister to the rest of Zoysieae (Fig. 5). Our results support the recent expansion of *Sporobolus* (Peterson et al., 2014b) to include *Calamovilfa*, *Crypsis*, *Pogoneura*, and *Spartina* based on evidence from chloroplast and nuclear loci (Columbus et al., 2007; Fortune et al., 2007; Ortiz-Diaz

and Culham, 2000; Peterson et al., 2014b; Peterson et al., 2010a). *Sporobolus* now contains ca. 220 species.

The relationship between *Sporobolus* (*Calamovilfa*) *arcuatus*, *S. heterolepis*, and *S. maritimus* is an example of conflicting relationships with high support values when different phylogenetic methods are used. The ML and MP-EST analyses found *S. arcuatus* + *S. (Spartina) maritimus* with 100% ML bootstrap, while the BI analysis found *S. arcuatus* + *S. heterolepis* with 1.00 BI PP. An SH test could not reject *S. arcuatus* + *S. maritimus* (Table S3). Peterson et al. (2014) found that *S. arcuatus* + *S. heterolepis* are together sister to *S. maritimus* with nrITS and chloroplast data. GPWG II (2011) found *S. arcuatus* (as *Calamovilfa*) + a *Spartina* clade (four species sampled) with chloroplast data. Columbus et al. (2007) found *S. arcuatus* (as *Calamovilfa*) + *S. indicus* with chloroplast data and sister to *S. pectinatus* (as *Spartina*) with nrITS. Denser species sampling across *Sporobolus* (ca. 220 spp.) is needed to better understand the evolutionary processes in *Sporobolus* s.l. Conflict in our nuclear data suggests there may be lingering ILS or hybridization occurring in the genus.

4.2.5 Relationships in Cynodonteae

We found strong support for a monophyletic Cynodonteae. This is a morphologically diverse tribe with more than 60 genera that have been recently studied with chloroplast and nrITS data (Peterson et al., 2015, 2014a, 2014b, 2012, 2010b). Cynodonteae relationships in our 56 locus MP-EST tree (Fig. 5) are congruent with Duvall et al.'s (2016) plastome estimate, but they only sampled eight species in the tribe. Peterson et al. 's (2010a) topology using plastid and nrITS data resulted in lineages similar to, but incongruent with our MP-EST topology. Most of the phylogenetic conflict in our nuclear dataset occurs in Cynodonteae, as evidenced by incongruent topologies from the MP-EST, BI and ML estimates and several branches with low support values (Fig. 5). SH tests with the nuclear data could not reject many of the alternative hypotheses of relationships that we tested (all SH

results are available in the supplementary material). These branches required high numbers of loci to stabilize their positions in the RADICAL analysis (Fig. 6A) and they also had the most volatile reactions to dataset concatenation (Fig. 6B-E).

There are two main Cynodonteae lineages in our 56 locus MP-EST analysis (Fig. 5), but the composition of the lineages is not robust to different phylogenetic methods. For example, in the MP-EST 56 locus species tree (Fig. 5) the *Triodia* clade is sister to the rest of the clade, but in the 56 locus BI and the 122 locus MP-EST trees the *Triodia* clade is sister to the New World clade with strong support (supplementary material). As another example, *Ctenium*, *Farrago*, and *Perotis* (CFP clade) resolve as sister to the *Chloris* clade in the MP-EST 56 locus species tree, BI tree and chloroplast and nrITS estimates (GPWG II, 2011; Peterson et al., 2014a, 2010a), while they are sister to the New World clade in the 122 locus BI tree (supplementary material).

In the New World clade we sampled two species each of the large, predominantly New World genera *Bouteloua* (60 spp.) and *Muhlenbergia* (176 spp.) and species in each genus were reciprocally monophyletic in every analysis except for the analysis of the sexual system genes (below). A random sampling of any three nuclear loci were able to resolve these genera as monophyletic, indicating strong support for these relationships across the nuclear genome (Fig. 6). We found that several relationships within the New World clade (Fig. 5) were not robust to different phylogenetic methods or dataset partitions. *Hilaria* and *Tridens* are sister in the 56 locus MP-EST tree (Fig. 5), while *Tridens* is sister to the rest of the New World clade in the 122 locus BI tree (supplementary material) and *Hilaria*, *Tridens*, and *Distichlis* form a grade at the base of the New World clade in the 122 locus MP-EST tree (supplementary material). Almost every possible topology has been estimated for these species with chloroplast and nuclear data, suggesting strong conflict in the evolutionary signal and the potential for ILS. Most of the relationships in the New World clade of the MP-EST 56 locus tree (Fig. 5) are congruent with Peterson et al.'s (2010a) chloroplast estimate. The exceptions are the sister

relationship between *Tridens* + *Hilaria* and the branching order of *Jouvea* and *Distichlis*.

The Andean-centered genus *Munroa* and the Californian endemic *Swallenia* are sister with strong support (Fig. 5). These genera have membranous lemmas with ciliate margins and narrow, condensed inflorescences (Bell, 2013; Clayton and Renvoize, 1986; Peterson et al., 2010a). Our 56 locus MP-EST tree resolves *Munroa* + *Swallenia* sister to *Muhlenbergia*, but it's not entirely clear from our data whether they are more closely related to *Muhlenbergia* or *Bouteloua*, nor how *Jouvea* is related to this group. Peterson et al.'s (2010a, 2010b) chloroplast and nrITS data produced a topology congruent with our 56 locus MP-EST tree (Fig. 5), while Columbus et al. (2007) found *Muhlenbergia* sister to *Swallenia* + (*Bouteloua* + *Munroa*) with nrITS. This clade includes *Hilaria* and *Distichlis* in the chloroplast tree of Columbus et al. (2007).

In our MP-EST and 56 locus BI trees *Distichlis* is sister to the rest of the New World clade, except for *Hilaria* and *Tridens* (Fig. 5), while *Distichlis* is sister to *Hilaria* in the 122 locus BI tree (supplementary material). Chloroplast and nrITS data place *Distichlis* sister to *Bouteloua* (Peterson et al., 2010a) or sister to a larger clade (Columbus et al., 2007). Clayton and Renvoize (1986) hypothesized a close relationship between *Jouvea*, *Swallenia*, and *Distichlis* based on leaf morphology and reduced inflorescences but these characters appear to be homoplasious within the New World clade. We recovered a close relationship between *Tridens* and *Hilaria* that has not been previously found with chloroplast or nrITS data, but this relationship is not robust to changing analytical methods or data. Denser sampling in the Cynodonteae may help to stabilize the placement of *Tridens*.

The *Chloris* clade (Fig. 5) consistently contains *Chloris*, *Microchloa*, *Astrebla*, *Dinebra*, *Acrachne*, *Dactyloctenium*, and *Tragus* in all of our analyses. In contrast, *Tragus* (a pantropical genus) has been found to be more closely related to the New World clade with chloroplast (GPWG II, 2011; Peterson et al., 2010a) and nrITS data

(Columbus et al., 2007; Peterson et al., 2010a). The sister group to the *Chloris* clade was not robust to different analytical methods or data partitions and most of the relationships recovered in the *Chloris* clade of the MP-EST 56 locus tree are incongruent with Peterson et al.'s (2014, 2010a) chloroplast hypothesis, with the exception of the CFP clade (Fig. 5).

We were surprised that the western U.S. vernal-pool genus *Orcuttia* resolved as sister to the Australian genus *Triodia* (Fig. 5). Chloroplast data place *Orcuttia* as a close relative to the pantropical genus *Ctenium* and the paleotropical *Perotis* (GPWG II, 2011; Peterson et al., 2010a), while nrITS data suggests a close relationship between *Orcuttia*, *Ctenium* and *Tridens* (Columbus et al., 2007). We were able to reject the hypothesis that *Orcuttia* is sister to *Ctenium* with our nuclear data (Table S3).

Tripogon is sister to the CFP clade + *Chloris* clade in the 56 locus MP-EST tree (Fig. 5), but it has a different placement in each of our other analyses. Studies using chloroplast data have found that *Tripogon*, *Oropetium*, and *Melanocenchris* are sister to the rest of the *Chloris* clade (GPWG II, 2011), sister to *Dinebra xerophila* (as *Leptochloa fusca*; Columbus et al., 2007), or sister to the New World clade (Peterson et al., 2010a). Interestingly, we were able to reject the placement of *Tripogon* as sister to the *Chloris* clade with an SH test. Clayton and Renvoize (1986) placed *Tripogon* close to *Oropetium* and *Indopoa*, and near *Leptochloa*. We did not sample these genera and perhaps inclusion of *Oropetium* or *Leptochloa* would help to stabilize the placement of *Tripogon* in future analyses of nuclear data.

4.3 Loci putatively involved in biological functions

We found several unusual tree topologies when sampling loci putatively involved in particular biological pathways. Conflicts with the species phylogeny estimate may be caused by convergent evolution (supplementary material) and these trees can serve as the initial evidence for additional inquiries into the genetic

basis for adaptation in chloridoids, such as drought tolerance, salinity and metal tolerance, sexual system, and photosynthesis pathways.

5. Conclusions and recommendations

We estimated the evolutionary history of chloridoid grasses from the perspective of the nuclear genome and compared this estimate with those based on the chloroplast genome and nrITS. This phylogenetic estimate utilized sequences generated from > 100 nuclear loci and explored novel strategies for evaluating and analyzing large data sets. We recovered clades corresponding to four tribes and many relationships within tribes that corroborate chloroplast estimates. The sampled nuclear loci largely corroborate chloroplast findings (with several exceptions), providing greater confidence that the tribal relationships are representative of the species phylogeny and not only the history of the chloroplast. However, most relationships in Cynodonteae were incongruent with chloroplast hypotheses, suggesting incomplete lineage sorting and, potentially, gene flow. We present the coalescent estimate of phylogeny using 56 nuclear loci as our best estimate of chloridoid phylogeny to date. Given the large amount of conflict we found in Cynodonteae, future studies will require a broader sampling at the generic and species level to test if the intra-tribal phylogenetic hypotheses proposed here are robust.

Our sequencing methods were conservative in that we had excellent coverage for the targeted nuclear loci, other high-copy nuclear elements (nrITS), and the plastome. We recommend that more than 24 samples may be successfully multiplexed per Illumina Hi-Seq lane.

We also recommend that future hyb-seq projects on grass phylogeny sequence longer reads than the 101 bases we sequenced with the Illumina Hi-Seq platform. Combining Illumina short-reads with long-reads, such as those offered by

Pacific Biosciences (California, USA), would be particularly beneficial for
differentiating multiple copies of the same gene in polyploid grasses.

Acknowledgements

We thank the many curators and assistants who helped with fieldwork and
herbarium collections. Gregory Boggy at MycroArray advised us on the hyb-seq
probe design. Aaron Liston and the staff at the Oregon State University Center for
Genome Research and Biocomputing provided helpful advice to design the hyb-seq
study and analyze the sequence data. Mel Duvall (Northern Illinois University)
provided access to unpublished chloridoid plastomes for sequence filtering.
Elizabeth Kellogg (Donald Danforth Plant Science Center) suggested several loci to
target. Madelaine Bartlett (University of Massachusetts at Amherst) advised us on
the functional categories used to partition the dataset. Thank you to CIPRES for
access to analysis software and computing power. CoGe developer Matt Bomhoff
(University of Arizona) kindly developed the Top Hits export feature for this study.
The RADICAL analyses were run on The University of Florida computing cluster.
Thank you to Grant Godden (Michigan State University) and Apurva Narechania
(American Museum of Natural History) for help with RADICAL. J. Mark Porter
(Rancho Santa Ana Botanic Garden), Lee Ripma (Ripma Biological, San Diego, CA),
and Kevin Weitemeier (Oregon State University) offered helpful advice for handling
the data and analyses. Funding for specimen collection and sequencing by NSF grant
DEB 0920147 to J. Travis Columbus and DEB 0921203 to Amanda Ingram.

References

- Ainouche, M.L., Baumel, A., Salmon, A., Yannic, G., 2003. Hybridization, polyploidy and speciation in *Spartina* (Poaceae). *New Phytol.* 161, 165–172.
- Álvarez, I., Wendel, J.F., 2003. Ribosomal ITS sequences and plant phylogenetic inference. *Mol. Phylogenet. Evol.* 29, 417–434.
- Barrett, C.F., Davis, J.I., Leebens-Mack, J., Conran, J.G., Stevenson, D.W., 2013. Plastid genomes and deep relationships among the commelinid monocot angiosperms. *Cladistics* 29, 65–87.
- Bell, H., 2013. Genetic diversity in *Swallenia alexandrae* (Poaceae, Chloridoideae), a narrow endemic from the Eureka Dunes (Inyo County, California). *Aliso* 31, 25–23.
- Bell, H.L., Columbus, J.T., 2008. Proposal for an expanded *Distichlis* (Poaceae, Chloridoideae): support from molecular, morphological, and anatomical characters. *Syst. Bot.* 33, 536–551.
- Birky, C.W.J., 2001. The inheritance of genes in mitochondria and chloroplasts: laws, mechanisms, and models. *Annu. Rev. Genet.* 35, 125–148.
- Birky, C.W.J., 1995. Uniparental inheritance of mitochondrial and chloroplast genes: mechanisms and evolution. *Proc. Natl. Acad. Sci. USA* 92, 11331–11338.
- 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. *Bot. J. Linn. Soc.* 162, 543–557.
- Burke, S. V., Clark, L.G., Triplett, J.K., Grennan, C.P., Duvall, M.R., 2014. Biogeography and phylogenomics of New World Bambusoideae (Poaceae), revisited. *Am. J. Bot.* 101, 886–891.
- Carstens, B.C., Knowles, L.L., 2007. Estimating species phylogeny from gene-tree probabilities despite incomplete lineage sorting: an example from *Melanoplus* grasshoppers. *Syst. Biol.* 56, 400–411.
- Chase, M., Soltis, D., Olmstead, R., Morgan, D., Les, D., Mishler, B., Duvall, M., Price, R., Hills, H., Qiu, Y., Kron, K., Rettig, J., Conti, E., Palmer, J., Manhart, J., Sytsma, K., Michaels, H., Kress, W., Donoghue, M., Clark, W., Hedren, M., Gaut, B., Jansen, R., Kim, K., Wimpee, C., Smith, J., Furnier, G., Straus, S., Xiang, Q., Plunkett, G., Soltis, P., Swensen, S., Eguiarte, L., Learn, G., Barret, S., Graham, S., Dayanandan, S., Albert, V., 1993. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbcL*. *Ann. Missouri Bot. Gard.* 80, 528–580.
- Childs, K.L., Hamilton, J.P., Zhu, W., Ly, E., Cheung, F., Wu, H., Rabinowicz, P.D., Town, C.D., Buell, C.R., Chan, A.P., 2007. The TIGR Plant Transcript Assemblies database. *Nucleic Acids Res.* 35, D846–851.
- Christin, P.-A., Spriggs, E., Osborne, C.P., Strömberg, C.E., Salamin, N., Edwards, E.J., 2014. Molecular dating, evolutionary rates, and the age of the grasses. *Syst. Biol.* 63, 153–165.

Christin, P.-A., Besnard, G., Samaritani, E., Duvall, M.R., Hodkinson, T.R., Savolainen, V., Salamin, N., 2008. Oligocene CO₂ decline promoted C₄ photosynthesis in grasses. *Curr. Biol.* 18, 37–43.

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

Clayton, W.D., Renvoize, S.A., 1986. Genera Graminum, grasses of the world, Kew Bulletin Additional Series XIII. Kew Bull Additional Series XIII.

Clegg, M.T., Gaut, B.S., Learn Jr, G.H., Morton, B.R., 1994. Rates and patterns of chloroplast DNA evolution. *Proc. Natl. Acad. Sci. USA* 91, 6795–6801.

Columbus, J.T., Cerros-Tlatilpa, R., Kinney, M.S., Siqueiros-Delgado, M.E., Bell, H.L., Griffith, M.P., Refulio-Rodriguez, N.F., 2007. Phylogenetics of Chloridoideae (Gramineae): a preliminary study based on nuclear ribosomal internal transcribed spacer and chloroplast *trnL-F* sequences. *Aliso* 23, 565–579.

Columbus, J.T., Kinney, M.S., Delgado, M.E.S., Porter, J.M., Jacobs, S.W.L., Everett, J., 2000. Phylogenetics of *Bouteloua* and relatives (Gramineae: Chloridoideae): cladistic parsimony analysis of internal transcribed spacer (nrDNA) and *trnL-F* (cpDNA) sequences. *Grasses Syst. Evol.* 189–194.

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

Cronn, R., Knaus, B.J., Liston, A., Maughan, P.J., Parks, M., Syring, J. V, Udall, J., 2012. Targeted enrichment strategies for next-generation plant biology. *Am. J. Bot.* 99, 291–311.

de Koning, A.P.J., Gu, W., Pollock, D.D., 2010. Rapid likelihood analysis on large phylogenies using partial sampling of substitution histories. *Mol. Biol. Evol.* 27, 249–265.

de Sousa, F., Bertrand, Y.J.K., Nylinder, S., Oxelman, B., Eriksson, J.S., Pfeil, B.E., 2014. Phylogenetic properties of 50 nuclear loci in *Medicago* (Leguminosae) generated using multiplexed sequence capture and next-generation sequencing. *PLoS One* 9, e109704.

Degnan, J.H., Rosenberg, N.A., 2006. Discordance of species trees with their most likely gene trees. *PLoS Genet.* 2, e68.

Delsuc, F., Brinkmann, H., Philippe, H., 2005. Phylogenomics and the reconstruction of the tree of life. *Nat. Rev. Genet.* 6, 361–375.

Denton, A.L., McConaughy, B.L., Hall, B.D., 1998. Usefulness of RNA polymerase II coding sequences for estimation of green plant phylogeny. *Mol. Biol. Evol.* 15, 1082–1085.

Doyle, J.J., Doyle, J.L., 1987. Genomic plant DNA preparation from fresh tissue-CTAB method. *Phytochem. Bull.* 19, 11–15.

Duarte, J., Wall, P.K., Edger, P., Landherr, L., Ma, H., Pires, J.C., Leebens-Mack, J., DePamphilis, C., 2010. Identification of shared single copy nuclear genes in

993 *Arabidopsis*, *Populus*, *Vitis* and *Oryza* and their phylogenetic utility across
994 various taxonomic levels. BMC Evol. Biol. 10, 1.

995 Duvall, M.R., Davis, J.I., Clark, L.G., Noll, J.D., Goldman, D.H., Sanchez-Ken, J.G., 2007.
996 Phylogeny of the grasses (Poaceae) revisited. Aliso 23, 237–247.

997 Duvall, M.R., Fisher, A.E., Columbus, J.T., Ingram, A.L., Wysocki, W.P., Burke, S. V.,
998 Clark, L.G., Kelchner, S.A., 2016. Phylogenomics and plastome evolution of the
999 chloridoid grasses (Chloridoideae: Poaceae). Int. J. Plant Sci. 177, 235–246.

1000 Edwards, S. V, Bryan Jennings, W., Shedlock, A.M., 2005. Phylogenetics of modern
1001 birds in the era of genomics. Proc. Biol. Sci. 272, 979–992.

1002 Egan, A.N., Schlueter, J., Spooner, D.M., 2012. Applications of next-generation
1003 sequencing in plant biology. Am. J. Bot. 99, 175–185.

1004 Erixon, P., Oxelman, B., 2008. Whole-gene positive selection, elevated synonymous
1005 substitution rates, duplication, and indel evolution of the chloroplast *clpP1*
1006 gene. PLoS One 3, e1386.

1007 Feliner, G.N., Rosselló, J.A., 2007. Better the devil you know? Guidelines for insightful
1008 utilization of nrDNA ITS in species-level evolutionary studies in plants. Mol.
1009 Phylogenet. Evol. 44, 911–919.

1010 Fortune, P.M., Schierenbeck, K.A., Ainouche, A.K., Jacquemin, J., Wendel, J.F.,
1011 Ainouche, M.L., 2007. Evolutionary dynamice of Waxy and the origin of
1012 hexaploid *Spartina* species (Poaceae). Mol. Phylogenet. Evol. 43, 1040–1055.

1013 Givnish, T.J., Ames, M., McNeal, J.R., McKain, M.R., Steele, P.R., DePamphilis, C.W.,
1014 Graham, S.W., Pires, J.C., Stevenson, D.W., Zomlefer, W.B., Briggs, B.G., Duvall,
1015 M.R., Moore, M.J., Heaney, J.M., Soltis, D.E., Soltis, P.S., Thiele, K., Leebens-Mack,
1016 J.H., 2010. Assembling the tree of the monocotyledons: plastome sequence
1017 phylogeny and evolution of Poales 1. Ann. Missouri Bot. Gard. 97, 584–616.

1018 Gnirke, A., Melnikov, A., Maguire, J., Rogov, P., LeProust, E.M., Brockman, W., Fennell,
1019 T., Giannoukos, G., Fisher, S., Russ, C., 2009. Solution hybrid selection with ultra-
1020 long oligonucleotides for massively parallel targeted sequencing. Nat.
1021 Biotechnol. 27, 182–189.

1022 GPWG I, 2001. Phylogeny and subfamilial classification of the grasses (Poaceae).
1023 Ann. Missouri Bot. Gard. 88, 373–457.

1024 GPWG II, 2011. New grass phylogeny resolves deep evolutionary relationships and
1025 discovers C₄ origins. New Phytol. 193, 304–312.

1026 Graham, S.W., Olmstead, R.G., Barrett, S.C.H., 2002. Rooting phylogenetic trees with
1027 distant outgroups: a case study from the commelinoid monocots. Mol. Biol.
1028 Evol. 19, 1769–1781.

1029 Grover, C.E., Salmon, A., Wendel, J.F., 2012. Targeted sequence capture as a powerful
1030 tool for evolutionary analysis. Am. J. Bot. 99, 312–319.

1031 Guisinger, M.M., Kuehl, J. V, Boore, J.L., Jansen, R.K., 2008. Genome-wide analyses of
1032 Geraniaceae plastid DNA reveal unprecedented patterns of increased
1033 nucleotide substitutions. Proc. Natl. Acad. Sci. 105, 18424–18429.

- 1034 Heled, J., Drummond, A., 2010. Bayesian inference of species trees from multilocus
1035 data. *Mol. Biol. Evol.* 27, 570–580.
- 1036 Hilu, K.W., Alice, L.A., 2001. A phylogeny of Chloridoideae (Poaceae) based on *matK*
1037 sequences. *Syst. Bot.* 26, 386–405.
- 1038 Hilu, K.W., Alice, L.A., Liang, H., 1999. Phylogeny of Poaceae inferred from *matK*
1039 sequences. *Ann. Missouri Bot. Gard.* 86, 835–851.
- 1040 Huelsenbeck, J.P., Rannala, B., 2004. Frequentist properties of Bayesian posterior
1041 probabilities of phylogenetic trees under simple and complex substitution
1042 models. *Syst. Biol.* 53, 904–913.
- 1043 Huelsenbeck, J.P., Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogeny.
1044 *Bioinformatics* 17, 754–755.
- 1045 Huson, D.H., Bryant, D., 2006. Application of phylogenetic networks in evolutionary
1046 studies. *Mol. Biol. Evol.* 23, 254–267.
- 1047 Ingram, A.L., Christin, P.A., Osborne, C.P., 2011. Molecular phylogenies disprove a
1048 hypothesized C₄ reversion in *Eragrostis walteri* (Poaceae). *Ann. Bot.* 107, 321.
- 1049 Ingram, A.L., Doyle, J.A., 2004. Is *Eragrostis* (Poaceae) monophyletic? Insights from
1050 nuclear and plastid sequence data. *Syst. Bot.* 29, 545–552.
- 1051 Ingram, A.L., Doyle, J.A., 2003. *Eragrostis* (Poaceae): monophyly and infrageneric
1052 classification, in: Third International Conference on the Comparative Biology of
1053 the Monocotyledons. Aliso, Claremont, CA, USA, pp. 595–604.
- 1054 Jeffroy, O., Brinkmann, H., Delsuc, F., Philippe, H., 2006. Phylogenomics: the
1055 beginning of incongruence? *Trends Genet.* 22, 225–231.
- 1056 Jones, S.S., Burke, S. V., Duvall, M.R., 2014. Phylogenomics, molecular evolution, and
1057 estimated ages of lineages from the deep phylogeny of Poaceae. *Plant Syst. Evol.*
1058 300, 1421–1436.
- 1059 Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software
1060 version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–
1061 780.
- 1062 Kelchner, S.A., Bamboo Phylogeny Group, 2012. Higher level phylogenetic
1063 relationships within the bamboos (Poaceae: Bambusoideae) based on five
1064 plastid markers. *Mol. Phylogenet. Evol.* 67, 404–413.
- 1065 Kellogg, E.K., 2015. Flowering Plants. Monocots: Poaceae volume 13. Springer
- 1066 Kinney, M.S., Columbus, J.T., Friar, E.A., 2008. Unisexual flower, spikelet, and
1067 inflorescence development in monoecious/dioecious *Bouteloua dimorpha*
1068 (Poaceae, Chloridoideae). *Am. J. Bot.* 95, 123–132.
- 1069 Kinney, M.S., Columbus, J.T., Friar, E.A., 2003. Molecular evolution of the maize sex-
1070 determining gene TASSELSEED2 in *Bouteloua* (Poaceae). *Mol. Phylogenet. Evol.*
1071 29, 519–528.
- 1072 Kubatko, L.S., Carstens, B.C., Knowles, L.L., 2009. STEM: species tree estimation
1073 using maximum likelihood for gene trees under coalescence. *Bioinformatics* 25,

1074 971–973.

1075 Kubatko, L.S., Degnan, J.H., 2007. Inconsistency of phylogenetic estimates from
1076 concatenated data under coalescence. *Syst. Biol.* 56, 17–24.

1077 Lemmon, E.M., Lemmon, A.R., 2013. High-throughput genomic data in systematics
1078 and phylogenetics. *Annu. Rev. Ecol. Evol. Syst.* 44, 99–121.

1079 Lanfear, R., Calcott, B., Ho, S.Y., Guindon, S. 2012. PartitionFinder: combined
1080 selection of partitioning schemes and substitution models for phylogenetic
1081 analyses. *Mol. Biol. Evol.* 29, 1695–1701.

1082 Leseberg, C.H., Duvall, M.R., 2009. The complete chloroplast genome of *Coix lacryma-*
1083 *jobi* and a comparative molecular evolutionary analysis of plastomes in cereals.
1084 *J. Mol. Evol.* 69, 311–318.

1085 Linder, C.R., Rieseberg, L.H., 2004. Reconstructing patterns of reticulate evolution in
1086 plants. *Am. J. Bot.* 91, 1700–1708.

1087 Liu, L., 2008. BEST: Bayesian estimation of species trees under the coalescent model.
1088 *Bioinformatics* 24, 2542–2543.

1089 Liu, L., Yu, L., Edwards, S. V, 2010. A maximum pseudo-likelihood approach for
1090 estimating species trees under the coalescent model. *BMC Evol. Biol.* 10, 302.

1091 Liu, Q., Triplett, J.K., Wen, J., Peterson, P.M., 2011. Allotetraploid origin and
1092 divergence in *Eleusine* (Chloridoideae, Poaceae): evidence from low-copy
1093 nuclear gene phylogenies and a plastid gene chronogram. *Ann. Bot.* 108, 1287–
1094 1298.

1095 Liu, Q., Zhao, N.X., Hao, G., 2005. Inflorescence structures and evolution in subfamily
1096 Chloridoideae (Gramineae). *Plant Syst. Evol.* 251, 183–198.

1097 Maddison, W.P., 1997. Gene trees in species trees. *Syst. Biol.* 46, 523.

1098 Maddison, W.P., Knowles, L.L., 2006. Inferring phylogeny despite incomplete lineage
1099 sorting. *Syst. Biol.* 55, 21–30.

1100 Malcomber, S.T., Kellogg, E.A., 2005. SEPALLATA gene diversification: brave new
1101 whorls. *Trends Plant Sci.* 10, 427–435.

1102 Mamanova, L., Coffey, A.J., Scott, C.E., Kozarewa, I., Turner, E.H., Kumar, A., Howard,
1103 E., Shendure, J., Turner, D.J., 2010. Target-enrichment strategies for next-
1104 generation sequencing. *Nat. Methods* 7, 111–118.

1105 Mandel, J.R., Dikow, R.B., Funk, V.A., Masalia, R.R., Staton, S.E., Kozik, A., Michelmore,
1106 R.W., Rieseberg, L.H., Burke, J.M., 2014. A target enrichment method for
1107 gathering phylogenetic information from hundreds of loci: an example from the
1108 Compositae. *Appl. Plant Sci.* 2, 1300085.

1109 Mathews, S., Tsai, R.C., Kellogg, E.A., 2000. Phylogenetic structure in the grass family
1110 (Poaceae): Evidence from the nuclear gene phytochrome B. *Am. J. Bot.* 87, 96–
1111 107.

1112 Mirarab, S., Bayzid, M.S., Boussau, B., Warnow, T., 2014. Statistical binning enables
1113 an accurate coalescent-based estimation of the avian tree. *Science* 346,

- 1114 1250463.
- 1115 Mirarab, S., Warnow, T., 2015. ASTRAL-II: coalescent-based species tree estimation
1116 with many hundreds of taxa and thousands of genes. *Bioinformatics* 31, i44–
1117 i52.
- 1118 Narechania, A., Baker, R.H., Sit, R., Kolokotronis, S.-O., DeSalle, R., Planet, P.J., 2012.
1119 Random addition concatenation analysis: a novel approach to the exploration of
1120 phylogenomic signal reveals strong agreement between core and shell genomic
1121 partitions in the cyanobacteria. *Genome Biol. Evol.* 4, 30–43.
- 1122 Ness, R.W., Graham, S.W., Barrett, S.C.H., 2011. Reconciling gene and genome
1123 duplication events: using multiple nuclear gene families to infer the phylogeny
1124 of the aquatic plant family Pontederiaceae. *Mol. Biol. Evol.* 28, 3009–3018.
- 1125 Nichols, R., 2001. Gene trees and species trees are not the same. *Trends Ecol. Evol.*
1126 16, 358–364.
- 1127 Ortiz-Diaz, J.J., Culham, A., 2000. Phylogenetic relationships of the genus *Sporobolus*
1128 (Poaceae: Eragrostideae) based on nuclear ribosomal DNA ITS sequences, in:
1129 Jacobs, S.W.L., Everett, J. (Eds.), *Grasses: Systematics and Evolution*. CSIRO
1130 Publishing, Collingwood, Victoria, Australia, pp. 184–188.
- 1131 Parks, M., Cronn, R., Liston, A., 2012. Separating the wheat from the chaff: mitigating
1132 the effects of noise in a plastome phylogenomic data set from *Pinus* L.
1133 (Pinaceae). *BMC Evol. Biol.* 12, 100.
- 1134 Peterson, G., Aagesen, L., Seberg, O., Larsen, I.H., 2011. When is enough , enough in
1135 phylogenetics? A case in point from *Hordeum* (Poaceae). *Cladistics* 27, 428–446.
- 1136 Peterson, P.M., Columbus, J.T., Pennington, S.J., 2007. Classification and
1137 biogeography of new world grasses: Chloridoideae. *Aliso* 23, 580–594.
- 1138 Peterson, P.M., Romaschenko, K., Arrieta, Y.H., 2015. Phylogeny and subgeneric
1139 classification of *Bouteloua* with a new species, *B. herrera-arrietae* (Poaceae:
1140 Chloridoideae: Cynodonteae: Boutelouinae). *J. Syst. Evol.* 53, 351–366.
- 1141 Peterson, P.M., Romaschenko, K., Arrieta, Y.H., 2014a. A molecular phylogeny and
1142 classification of the Cteniinae, Farragininae, Gouiniinae, Gymnopogoninae,
1143 Perotidinae, and Trichoneurinae (Poaceae: Chloridoideae: Cynodonteae) 63,
1144 275–286.
- 1145 Peterson, P.M., Romaschenko, K., Arrieta, Y.H., Saarela, J.M., 2014b. A molecular
1146 phylogeny and new subgeneric classification of *Sporobolus* (Poaceae:
1147 Chloridoideae: Sporobolinae), *Taxon*.
- 1148 Peterson, P.M., Romaschenko, K., Barker, N.P., Linder, H.P., 2011. Centropodieae and
1149 *Ellisochloa*, a new tribe and genus in Chloridoideae (Poaceae). *Taxon* 60, 1113–
1150 1122.
- 1151 Peterson, P.M., Romaschenko, K., Johnson, G., 2010a. A classification of the
1152 Chloridoideae (Poaceae) based on multi-gene phylogenetic trees. *Mol.*
1153 *Phylogenet. Evol.* 55, 580–598.
- 1154 Peterson, P.M., Romaschenko, K., Johnson, G., 2010b. A phylogeny and classification

- 1155 of the Muhlenbergiinae (Poaceae: Chloridoideae: Cynodonteae) based on
1156 plastid and nuclear DNA sequences. *Am. J. Bot.* 97, 1532–54.
- 1157 Peterson, P.M., Romaschenko, K., Snow, N., Johnson, G., 2012. A molecular phylogeny
1158 and classification of *Leptochloa* (Poaceae: Chloridoideae: Chlorideae) sensu lato
1159 and related genera. *Ann. Bot.* 109, 1317–1330.
- 1160 Philippe, H., Brinkmann, H., Lavrov, D. V., Littlewood, D.T.J., Manuel, M., Wörheide, G.,
1161 Baurain, D., 2011. Resolving difficult phylogenetic questions: why more
1162 sequences are not enough. *PLoS Biol.* 9, e1000602.
- 1163 Pyron, R.A., Hendry, C.R., Chou, V.M., Lemmon, E.M., Lemmon, A.R., Burbrink, F.T.,
1164 2014. Effectiveness of phylogenomic data and coalescent species-tree methods
1165 for resolving difficult nodes in the phylogeny of advanced snakes (Serpentes:
1166 Caenophidia). *Mol. Phylogenet. Evol.* 81, 221–231.
- 1167 Qiu, Y.-L., Li, L., Wang, B., Chen, Z., Knoop, V., Groth-Malonek, M., Dombrowska, O.,
1168 Lee, J., Kent, L., Rest, J., Estabrook, G.F., Hendry, T.A., Taylor, D.W., Testa, C.M.,
1169 Ambros, M., Crandall-Stotler, B., Duff, R.J., Stech, M., Frey, W., Quandt, D., Davis,
1170 C.C., 2006. The deepest divergences in land plants inferred from phylogenomic
1171 evidence. *Proc. Natl. Acad. Sci. U. S. A.* 103, 15511–15516.
- 1172 Rokas, A., Williams, B.L., King, N., Carroll, S.B., 2003. Genome-scale approaches to
1173 resolving incongruence in molecular phylogenies. *Nature* 425, 798–804.
- 1174 Roodt-Wilding, R., Spies, J.J., 2006. Phylogenetic relationships in southern African
1175 chloridoid grasses (Poaceae) based on nuclear and chloroplast sequence data.
1176 *Syst. Biodivers.* 4, 401–415.
- 1177 Ruhfel, B.R., Gitzendanner, M.A., Soltis, P.S., Soltis, D.E., Burleigh, J.G., 2014. From
1178 algae to angiosperms-inferring the phylogeny of green plants (Viridiplantae)
1179 from 360 plastid genomes. *BMC Evol. Biol.* 14, 23.
- 1180 Salmon, A., Udall, J.A., Jeddelloh, J.A., Wendel, J., 2012. Targeted capture of
1181 homoeologous coding and noncoding sequence in polyploid cotton. *G3 Genes,*
1182 *Genomes, Genet.* 2, 921–930.
- 1183 Shaw, T.I., Ruan, Z., Glenn, T.C., Liu, L., 2013. STRAW: Species TRee Analysis Web
1184 server. *Nucleic Acids Res.* 41, W238–41.
- 1185 Shimodaira, H., Hasegawa, M., 1999. Multiple comparisons of log-likelihoods with
1186 applications to phylogenetic inference. *Mol. Biol. Evol.* 16, 1114–1116.
- 1187 Siqueiros-Delgado, M.E., Ainouche, M., Columbus, J.T., Ainouche, A., 2013. Phylogeny
1188 of the *Bouteloua curtipendula* complex (Poaceae: Chloridoideae) based on
1189 nuclear ribosomal and plastid DNA sequences from diploid taxa. *Syst. Bot.* 38,
1190 379–389.
- 1191 Smit, A., Hubley, R., Green, P., 1996. RepeatMasker Open-3.0.
1192 www.repeatmasker.org.
- 1193 Soreng, R.J., Peterson, P.M., Romaschenko, K., Davidse, G., Zuloaga, F.O., Judziewicz,
1194 E.J., Filgueiras, T.S., Davis, J.I., Morrone, O., 2015. A worldwide phylogenetic
1195 classification of the Poaceae (Gramineae). *J. Syst. Evol.* 53, 117–137.

1196 Spriggs, E.L., Christin, P.-A., Edwards, E.J., 2014. C₄ photosynthesis promoted species
1197 diversification during the Miocene grassland expansion. PLoS One 9, e97722.

1198 Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-
1199 analysis of large phylogenies. Bioinformatics 30, 1312–1313.

1200 Stamatakis, A., Ott, M., 2008. Efficient computation of the phylogenetic likelihood
1201 function on multi-gene alignments and multi-core architectures. Philos. Trans.
1202 R. Soc. Lond. B. Biol. Sci. 363, 3977–84.

1203 Stephens, J.D., Rogers, W.L., Heyduk, K., Cruse-Sanders, J.M., Determann, R.O., Glenn,
1204 T.C., Malmberg, R.L., 2015. Resolving phylogenetic relationships of the recently
1205 radiated carnivorous plant genus *Sarracenia* using target enrichment. Mol.
1206 Phylogenet. Evol. 85, 76–87.

1207 Stull, G.W., Moore, M.J., Mandala, V.S., Douglas, N.A., Kates, H.-R., Qi, X., Brockington,
1208 S.F., Soltis, P.S., Soltis, D.E., Gitzendanner, M.A., 2013. A targeted enrichment
1209 strategy for massively parallel sequencing of angiosperm plastid genomes.
1210 Appl. Plant Sci. 1, 1200497.

1211 Swofford, D.L., 2003. PAUP* phylogenetic analysis using parsimony (*and other
1212 methods) version 4.

1213 Tennessen, J.A., Govindarajulu, R., Liston, A., Ashman, T.-L., 2013. Targeted sequence
1214 capture provides insight into genome structure and genetics of male sterility in
1215 a gynodioecious diploid strawberry, *Fragaria vesca* ssp. *bracteata* (Rosaceae).
1216 G3 (Bethesda). 3, 1341–51.

1217 Triplett, J.K., Wang, Y., Zhong, J., Kellogg, E.A., 2012. Five nuclear loci resolve the
1218 polyploid history of switchgrass (*Panicum virgatum* L.) and relatives. PLoS One
1219 7, e38702.

1220 Weitemier, K., Straub, S.C.K., Cronn, R.C., Fishbein, M., Schmickl, R., McDonnell, A.,
1221 Liston, A., 2014. Hyb-Seq: combining target enrichment and genome skimming
1222 for plant phylogenomics. Appl. Plant Sci. 2, 1400042.

1223 Wendel, J.F., Doyle, J.J., 1998. Phylogenetic incongruence: window into genome
1224 history and molecular evolution, in: Soltis, D.E., Soltis, P.S. (Eds.), Molecular
1225 Systematics of Plants II. Springer US, pp. 265–296.

1226 Whittall, J.B., Medina-Marino, A., Zimmer, E.A., Hodges, S.A., 2006. Generating single-
1227 copy nuclear gene data for a recent adaptive radiation. Mol. Phylogenet. Evol.
1228 39, 124–134.

1229 Wysocki, W.P., Clark, L.G., Kelchner, S.A., Burke, S. V., Pires, J.C., Edger, P.P., Mayfield,
1230 D.R., Triplett, J.K., Columbus, J.T., Ingram, A.L., Duvall, M.R., 2014. A multi-step
1231 comparison of short-read full plastome sequence assembly methods in grasses.
1232 Taxon 63, 899–910.

1233

1234

Figure Captions

Fig. 1. Overview of sequence recovery for the hyb-seq next-generation sequencing dataset and four taxa whose sequences were obtained from CoGe. A) Pattern of sequence recovery for 614 targeted exons (X axis) and 51 species (Y axis). Recovered sequences are green. Missing sequences are pink. 416 exons were successfully sequenced for 100% of the species. Six exons were not recovered for any sequenced species. B) Sequence recovery and initial DNA extraction concentration for each species. A similar number of exons (total bar height) were recovered for each species. The fewest exons were recovered from *Jouvea pilosa* (475) and the most exons were recovered from *Monachather paradoxus* (549). Exon recovery did not correlate with the concentration of the original DNA extraction (red bar height).

Fig. 2. Saturation plots of the 122 gene dataset and biological function partitions. Saturation is estimated by plotting uncorrected (y axis) and model-corrected (x axis, GTR+I+G) pairwise differences. The black line is a 1:1 ratio for comparison. None of the datasets displayed substantial variation from the 1:1 line.

Fig. 3. Best-fit PhyML models estimated for 94 optimum partitions in PartitionFinder (Lanfear et al., 2012). GTR+I+G was chosen as the best-fit model for 80% of the partitions.

Fig. 4. Neighbor-net network of the 122 locus nuclear dataset. Two major areas of conflict are indicated by the arrows and lead to *Jouvea pilosa* and *Kaokochloa nigrirostris*.

Fig. 5. MP-EST species tree from the 56 housekeeping genes dataset. Phylogram in upper left corner and cladogram with branch support values. Unmarked or * branches have 1.00 BI PP and 100% ML bootstrap (BS) values while other branches have the values indicated (BI PP/ML BS). Branches with ML BS or BI PP values not present (--/--) were unique to the MP-EST analysis. The gray box indicates the PACMAD outgroups. **A-AQ** inflorescence and spikelet diversity across chloridoids. Image letter corresponds to letter following the species name in the tree. Photographs by J.T. Columbus.

Fig. 6. Results of the RADICAL concordance analysis with random sampling of 1-122 genes. A) RADICAL concordance analysis support for the branches on the 56 locus BI consensus tree. RADICAL calculates a “Fixation point” that is the number of loci required for a node to always appear in the tree topology for a dataset of that size or a “Degradation point” that is the minimum number of loci for which a node does not occur in a tree (values marked with an *). Nodes with a degradation value rather than a fixation point were not always estimated with random combinations of genes. See Narechania et al. (2011) for a detailed description of these statistics. Nodes without a value (=) did not reach fixation in the analysis. Gray numbers below the branch refer to the Cynodontae graph (E). Graphs B-E represent the average level

of occurrence of a node across the 10 randomizations for each dataset size (the X axis).

1315 Table 1. Species sample and vouchers. All vouchers are deposited at RSA-POM
 1316 (Claremont, California, USA).
 1317

Subfamily	Species	Sample voucher
Chloridoideae	<i>Acrachne racemosa</i> (B. Heyne ex Roem. & Schult.) Ohwi	Columbus 5534
Chloridoideae	<i>Aeluropus pungens</i> (M. Bieb.) K. Koch <i>var. pungens</i>	American-Iranian Botanical Delegation s.n. (TUH 34011)
Aristidoideae	<i>Aristida pallens</i> Cav. <i>var. geminata</i> Caro	Columbus 3100
Chloridoideae	<i>Astrebla pectinata</i> (Lindl.) F. Muell. ex Benth.	Columbus 5147
Chloridoideae	<i>Bouteloua dactyloides</i> (Nutt.) Columbus (female)	Columbus 3228
Chloridoideae	<i>Bouteloua gracilis</i> (Kunth) Lag. ex Griffiths	Columbus 3219
Danthonioideae	<i>Capeochloa cincta</i> (Nees) N.P. Barker & H.P. Linder subsp. <i>cincta</i>	Columbus 5581
Unplaced	<i>Centropodia glauca</i> (Nees) Cope	Columbus 5589
Chloridoideae	<i>Chloris gayana</i> Kunth	Columbus 5531
Chloridoideae	<i>Cottea pappophoroides</i> Kunth	Columbus 4917
Chloridoideae	<i>Ctenium concinnum</i> Nees	Columbus 5513
Chloridoideae	<i>Dactyloctenium aegyptium</i> (L.) Willd.	Columbus 5886
Chloridoideae	<i>Diandrochloa namaquensis</i> (Nees ex Schrad.) De Winter <i>var. namaquensis</i>	Ingram 705
Chloridoideae	<i>Dinebra retroflexa</i> (Vahl) Panz. <i>var. retroflexa</i>	Columbus 5108
Chloridoideae	<i>Distichlis bajaensis</i> H.L. Bell	Bell 458E
Unplaced	<i>Ellisochloa rangei</i> (Pilg.) P.M. Peterson & N.P. Barker	Columbus 5642
Chloridoideae	<i>Eragrostis sessilispica</i> Buckley	Columbus 3328
Micrairoideae	<i>Eriachne pallescens</i> R. Br. <i>var. pallescens</i>	Columbus 5156
Chloridoideae	<i>Farrago racemosa</i> Clayton	Columbus 5767
Chloridoideae	<i>Fingerhuthia cf. africana</i>	Columbus 5681
Chloridoideae	<i>Hilaria rigida</i> (Thurb.) Benth. ex Scribn.	Columbus 3588
Chloridoideae	<i>Jouvea pilosa</i> (J. Presl.) Scribn. (female)	Columbus 4760
Chloridoideae	<i>Kaokochloa nigrirostris</i> De Winter	Columbus 5679
Chloridoideae	<i>Microchloa caffra</i> Nees	Columbus 5463
Arundinoideae	<i>Monachather paradoxa</i> Steud.	Columbus 5120
Chloridoideae	<i>Muhlenbergia emersleyi</i> Vasey	Columbus 5451
Chloridoideae	<i>Muhlenbergia paniculata</i> (Nutt.)	Columbus 3224

	Columbus	
Chloridoideae	<i>Munroa squarrosa</i> (Nutt.) Torr.	Columbus 3293
Chloridoideae	<i>Odysea paucinervis</i> (Nees) Stapf	Columbus 5582
Chloridoideae	<i>Orcuttia tenuis</i> Hitchc.	Columbus 5738
Chloridoideae	<i>Perotis hildebrandtii</i> Mez	Columbus 5739
Chloridoideae	<i>Pogonarthria squarrosa</i> (Roem. & Schult.) Pilg.	Ingram 635
Chloridoideae	<i>Sporobolus arcuatus</i> (K.E. Rodgers) P.M. Peterson [<i>Calamovilfa arcuata</i>]	Columbus 5442
Chloridoideae	<i>Sporobolus coromandelianus</i> (Retz.) Kunth	Columbus 5603
Chloridoideae	<i>Sporobolus heterolepis</i> (A. Gray) A. Gray	Columbus 5428
Chloridoideae	<i>Sporobolus maritimus</i> (Curtis) P.M. Peterson & Saarela [<i>Spartina maritima</i>]	Columbus 5575
Chloridoideae	<i>Sporobolus schoenoides</i> (L.) P.M. Peterson [<i>Crypsis schoenoides</i>]	Columbus 5726
Chloridoideae	<i>Sporobolus stapfianus</i> Gand.	Columbus 5574
Chloridoideae	<i>Sporobolus subtilis</i> Kunth	Columbus 5564
Chloridoideae	<i>Swallenia alexandrae</i> (Swallen) Soderstr. & H.F. Decker	Bell 255-41
Chloridoideae	<i>Tragus berteronianus</i> Schult.	Columbus 5702
Chloridoideae	<i>Tridens brasiliensis</i> (Nees ex Steud.) Parodi	Columbus 4816
Chloridoideae	<i>Triodia pungens</i> R. Br.	Columbus 5236
Chloridoideae	<i>Tripogon cf. major</i>	Columbus 5788
Chloridoideae	<i>Triraphis mollis</i> R. Br.	Columbus 5123
Chloridoideae	Uniolinae sp.	Simon 4500
Chloridoideae	<i>Urochondra setulosa</i> (Trin.) C.E. Hubb.	Columbus 6007

1318

1319

Figure1

Figure2

Figure3

Figure4

Figure5

Figure6