ELSEVIER

CrossMark

Evolutionary history of chloridoid grasses estimated from 122 nuclear loci

Amanda E. Fisher [a,*], Kristen M. Hasenstab [b], Hester L. Bell [b], Ellen Blaine [c], Amanda L. Ingram [d], J. Travis Columbus [b]

[a] Department of Biological Sciences, California State University Long Beach, Long Beach, CA 90840, USA
[b] Rancho Santa Ana Botanic Garden and Claremont Graduate University, 1500 North College Avenue, Claremont, CA 91711, USA
[c] Department of Computer Science, Stanford University, 353 Serra Mall, Stanford, CA 94305, USA
[d] Department of Biology, Wabash College, Crawfordsville, IN 47933, USA

ARTICLE INFO

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 bi-parentally 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.

© 2016 Elsevier Inc. All rights reserved.

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 C_4 Kranz leaf anatomy [since the re-classification of the C_3 species *Eragrostis walteri* to Arundinoideae (Ingram et al., 2011)]. Several recent papers (Kellogg, 2015; P.M. Peterson et al., 2011; Soreng et al., 2015) treat Centropodieae as a tribe in Chloridoideae, resulting in the inclusion of a C_3 lineage, *Ellisochloa*, within the sub-family. 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 cae-spitose 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 classifica-tions 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 sub-family based on chloroplast and nrITS data recognize five tribes: Centropodieae, Triraphideae, Eragrostideae, Zoysieae, and Cyn-odonteae (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 uni-parentally (maternally) inherited in most species (Birky, 2001, 1995) and haploid; therefore, they are expected to achieve coales-cence 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 phyloge-netic 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 develop-ments 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 suc-cessfully 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 lar-gest tribe in the subfamily, and it has been the most difficult to resolve with chloroplast and nrITS data. We used the hyb-seq data-set to estimate a nuclear phylogeny across the Chloridoideae sub-family 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, Era-grostideae, 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 previ-ously 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 angios-perms. We searched the grass physiology literature for loci impli-cated 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