A 313 plastome phylogenomic analysis of Pooideae: Exploring relationships among the largest subfamily of grasses

Lauren M. Orton a, *, Patricia Barbera b, Matthew P. Nissenbaum a, Paul M. Peterson c, Alejandro Quintanar d, Robert J. Soreng e, Melvin R. Duvall f, g

a Plant Molecular and Bioinformatics Center, Biological Sciences, Northern Illinois University, 1425 W. Lincoln Hwy, DeKalb, IL 60115-2861, USA
b Department of Africa and Madagascar, Missouri Botanical Garden, St. Louis, MO 63110, USA
c Department of Botany, National Museum of Natural History, Smithsonian Institution, Washington DC 20013-7012, USA
d Herbario MA, Unidad de Herbarios, Real Jardín Botánico de Madrid CSIC, 28014 Madrid, Spain
e Institute for the Study of the Environment, Sustainability and Energy, Northern Illinois University, 1425 W. Lincoln Hwy, DeKalb, IL 60115-2861, USA

* Corresponding author.
E-mail address: lauren.orton@outlook.com (L.M. Orton).

1. Introduction

From the late Cretaceous origins of Poaceae Barnhart, to modern day agriculture, the grass family is well known for the impacts it has had in both conquering diverse landscapes and its economic significance as biofuel or food sources for wide varieties of animal species, including humans (Bouchenak-Khelladi et al., 2010; Orton et al., 2017, 2019). The Poaceae comprise 12 subfamilies: Anomochlooideae Pilg. ex Potztal, Aristidoideae Caro, Arundinoideae ex Kunth ex Beilschm., Bambusoideae Luers., Chloridoideae Kunth ex Beilschm., Danthonioideae H.P.Linder & N.P.Barker, Micrairoideae Pilg., Oryzoideae Kunth ex Beilschm., Panicoideae A.Braun, Pharoideae L.G.Clark & Judz., Pooideae Benth., and Puelioideae L.G.Clark & al. Among these, two large clades are recognized (Kellogg, 2015; Soreng et al., 2015, 2017; Saarela et al., 2017): the BOP (Bambusoideae, Oryzoideae, and Pooideae) clade, and the PACMAD (Panicoideae, Aristidoideae, Chloridoideae, Micrairoideae, Arundinoideae, Danthonioideae) clade. The BOP is comprised of C3 photosynthetic cool-season grasses that have wide distributions across temperate regions in primarily open to closed habitats. The PACMAD is comprised of both C3 and C4 photosynthetic species with cosmopolitan distributions and greatly varying habitats (Cotton et al., 2015; Gallaher et al., 2019). While relationships of these clades have been extensively studied, many of these studies utilized single or multi-loci data. More recently, complete genomic datasets (in particular, complete plastomes) focused on the complex subfamilial relationships in the BOP and PACMAD clades have been increasingly prevalent in...
These complicated phylogenetic histories are oftentimes resolved with the addition of data to fill absences in previously underrepresented groups. In the case of the Poaceae, a vast number of species have no complete plastid genome (plastome) available in data repositories. Poaceae are comprised of roughly 12,000 species and 780 genera (Kellogg, 2015; Christenhua & Byung, 2016; Soreng et al., 2017). The representation of grass plastomes on NCBI’s GenBank (https://www.ncbi.nlm.nih.gov) is roughly 1,450 of the nearly 12,000 known species.

The subfamily Pooideae is the largest of the grass subfamilies, a member of the BOP clade, and is a group with a number of economically significant crop, pasture, and lawn grasses. Pooideae arguably has the widest impact, due to its size, diversity, and systematic/phylogenetic complexity, among grass subfamilies (Kellogg, 2015). Even with approximately 4,000 species and more than 200 genera (Saarela et al., 2015; Soreng et al., 2015, 2017; Orton et al., 2019), Pooideae complete plastomes are represented on NCBI’s GenBank by only about 300 species, many of which are multiple sequences of cultivars or varieties belonging to a single species. This underrepresentation across the grass family and its subfamilies has presented challenges for reconstructing plastome phylogenomic histories.

Additionally, fossil specimens used for divergence date estimation calibration points among Poaceae, and especially from Pooideae, are but not seen in nuclear phylogenies (Quintanar et al., 2007; Saarela et al., 2016a,b; Hand et al., 2013; Jones et al., 2014; Cotton et al., 2015; Saarela et al., 2015, 2018; Wysocki et al., 2015; Attigala et al., 2016; Duvall et al., 2016; Orton et al., 2017, 2019; Gallaher et al., 2019). Previous studies have also incorporated morphological data to provide an additional avenue of exploration with regard to relationships among and within these clades (Kellogg, 2015; Saarela et al., 2015; Soreng et al., 2015, 2017).

Here we present the largest analysis of grass plastomes to date, focusing on the Poaceae with a targeted interest in resolving the complex and often reticulate relationships of the Pooideae subfamily, as well as interpreting relationships across the BOP clade. The goal of our work is to explore how plastome level data in large-scale studies will provide added resolution to reticulate phylogenomic relationships in groups plagued by complex phylogenomic histories, determine how alignment gaps impact phylogenetic reconstruction, and in particular, how they impact topology. We also compare our results with previous divergence estimation analyses for Pooideae, in which we expand upon both the molecular sampling (by including both coding and noncoding regions from complete plastomes) and fossil calibrations (by including additional accepted grass fossils).

2. Materials and methods

2.1. Sampling and DNA extraction

A total of 47 plastomes were targeted for DNA sequencing for this study. These species were chosen to bolster underrepresented tribes/subtribes and to seek additional resolution of phylogenomic relationships. Silica dried leaf tissues were collected or received from collaborators for these species (Supplemental Table 1). DNA was manually extracted through homogenization of tissue in liquid nitrogen and application of the DNeasy Plant Mini Kit protocol (Qiagen, Valencia, CA, USA). Total genomic DNAs in the extracts were quantified using the Qubit assay (Invitrogen, ThermoFisher Scientific, Wilmington, DE, USA), and diluted to 2.5 ng/µl in 20 µl sterile water.

Illumina sequencing libraries were prepared using the Nextera Prep Kit (Illumina Inc., San Diego, CA, USA; Caruccio, 2011), and of the 47 species 26 were sequenced from single end libraries, 16 were sequenced from paired end libraries, and 5 were previously sequenced by collaborators using MiSeq for mate-pair libraries. The DNA Clean and Concentrator kit (Zymo Research, Irvine, CA, USA) was used to purify DNA. Standard manufacturer protocols for each respective sample preparation kit were used for library preparation. Samples were then sent to the Iowa State University Core DNA Facility, Ames, IA, USA, for sequencing on the HiSeq 2500.

2.2. Plastome assembly and annotation

Once sequence read libraries were received, plastomes were assembled using de novo methods based on Wysocki et al. (2014). A proprietary Python language (van Rossum, 1995) based pipeline was used to streamline workflow. Programs included in the pipeline were: DynamicTrim v2.1 of the SolexaQA software suite (Cox et al., 2010), which performed initial quality trimming on reads using default settings. CutAdapt removed any remaining Illumina adapter sequences (Martin, 2011). LengthSort v2.1 (Cox et al., 2010) removed any sequences shorter than 25 bp in length. CDHit-EST of the CDHit package (Fu et al., 2012) identified and removed redundant sequences (parameter settings: sequence identity threshold (–c 1)). Assembly of contigs was done using SPAdes v. 3.8.1 software suite (Bankevich et al., 2012) (kmer parameter settings: -k 19,25,31,37,43,49,55,61,67,73) for all species, except Echinaria capitata, which was assembled under an older protocol using Velvet de novo assembler (Zerbino & Birney, 2008) following the methods as outlined in Wysocki et al. (2014). Remaining contigs were scaffolded using the anchored conserved region extension method (Wysocki et al., 2014) which uses highly conserved regions in grass plastomes to orient and scaffold contigs. Any remaining gaps between contigs were manually resolved through in silico genome walking, a method that locates regions of overlap in the quality processed reads to build across gapped regions until a complete plastome is produced.
Assembly verification was done by mapping quality-processed reads to the assembled plastome via Geneious Pro v. 11.0.5 software (Biomatters Ltd, Auckland, NZ; Kearse et al., 2012), and any assembly errors were identified and manually resolved. Following the methods of Burke et al. (2012), inverted repeat boundaries were located using NCBI’s BLASTn utility by performing a local alignment of the plastome against itself, locating regions of plus/minus transitions, and annotated using Geneious software. Remaining annotations were identified and completed by aligning the newly completed plastome to a closely related Pooidae species in Geneious software and transferring annotations to the newly completed plastome (Wysocki et al., 2014). Coding sequences were adjusted when necessary to preserve stop codons and correctly position coding sequence boundaries to preserve reading frames.

### 2.3. Phylogenomic analyses

Phylogenomic reconstruction was completed on a stripped alignment matrix of the 313 taxa that was 78,002 bp in length (see below and Methods 2.6 for additional descriptions of how gaps in the alignment matrix were stripped from each alignment column). The 313 taxa included in this analysis were sampled to proportionally represent grass subfamilies based on availability of species both in GenBank and those for which complete plastomes were sequenced here (Supplemental Table 2). The alignment matrix was created using the MAFFT plugin available in Geneious v11.0.5 software (Katoh and Standley, 2013) following removal of Inverted Repeat A (IRA), to avoid redundant representation of inverted repeat sites. The alignment was then stripped in Geneious v11.0.5 by using the “mask alignment” tool and sequentially removing an increasing threshold percentage of gaps per alignment column (“stripping threshold”, Duval et al., 2020).

The Saarela et al. (2018) study conducted genome partition analyses of grass plastomes on a dataset with 250 taxa across Poaceae. In that study, 14 partitions ([rbcL, ndhF, matK], plastome coding sequences, plastome non-coding sequences, and complete plastome data) with gapped sites only, positively selected sites only, or nucleotides selected sites, or both gapped and positively selected sites were analyzed in matrices that ranged from 3,476 bp to 197,529 bp. The results indicated that best practices in grass phylogenomics include utilizing complete plastome data while excluding ambiguously aligned regions by stripping alignment matrices.

The GTR + I + G model was chosen based on best-fit results of a jModelTest (Posada, 2008; Darriba et al., 2012) using the Akaike information criterion (Akaike, 1974). A maximum likelihood (ML) analysis was conducted using the RAxML – HPC2 v8.11.11 program (Stamatakis, 2014) available through XSEDE on the CIPRES science gateway (Miller et al., 2010). The GTRGAMMA model was selected and a ML best tree was produced and bootstrap analysis conducted. A consensus tree was generated from 1,000 bootstrap replicates using the Consense tool from the Phylop v3.66 software suite (Felsenstein, 2005), available on CIPRES. Trees were visualized using TreeGraph2 (Stöver and Müller, 2010) and FigTree v1.4.2 (Rambaut, 2014).

### 2.4. Bayesian evolutionary analysis sampling trees (BEAST2) divergence date analysis

A divergence date analysis was conducted using BEAST v2.4.8 (Bouckaert et al., 2014) available on CIPRES. Parameters and priors were set using the complementary BEAUTI program provided with the BEAST2 software package. Fossil dates were used as calibration points with the upper bound set to 110 Ma, which was the estimated age of the spikelet fossil identified in Poinar et al. (2015), to constrain the node age to the oldest identified and dated grass fossil, and prevent an unconstrained upper bound skewing the ages substantially older. The lower bound was set to the fossil age. These methods of calibration, which use fossil dates only to set minimum ages on the stem nodes of clades, make the fewest assumptions about the largely unknown relationships between the ages of fossils and the ages of the groups to which they belong (Christin et al., 2014). See Table 1 for priors and fossil calibrations related to specific nodes. An uncorrelated relaxed log normal clock model and uniform distribution across priors (Christin et al., 2014) were chosen along with the GTR site model, while also including a gamma category count of six, 0.743488 estimated shape parameter, and 0.500577 estimated proportion of invariant sites. The estimated shape parameter and proportion of invariant sites were determined from the ML analysis and values input as BEAST2 program parameters. This clock model and prior distribution were selected based on Christin et al. (2014) where it was elaborated that an uncorrelated relaxed lognormal clock reduces the effects of substantial variability in evolutionary rates while a uniform distribution allows for fossil ranges in calibrations by providing upper and lower boundaries for the fossil age. The Yule tree prior was utilized as it more readily assesses relationships between single representatives across different species (Beast2 Tree Priors and Dating: https://www.beast2.org/tree-priors-and-dating/).

Additional parameter settings included: four MCMC chains of 500,000,000 generations, and trees logged every 100,000 generations. Trees were summarized using TreeAnnotator (Rambaut, 2014) and burn-in was set at 25%. A chronogram was created using R packages: Strap (Bell and Lloyd, 2015) and IPS (Heil & et al., 2019). The R package RWTY (Warren et al., 2017) was used to assess convergence.

The resulting tree and its associated posterior probability (PP)

<table>
<thead>
<tr>
<th>Citation</th>
<th>Fossil Calibration Point</th>
<th>Lower bound (age, Ma)</th>
<th>Fossil</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Spikelet in amber</td>
<td>Spikelet clade (all grasses minus Anomochloideae Pilg. ex Potzelt)</td>
<td>97</td>
<td>Poinar et al. (2015)</td>
</tr>
<tr>
<td>2</td>
<td>Oryzeae phyto liths</td>
<td>Oryzeae Dumort.</td>
<td>66</td>
<td>Prasad et al. (2011)</td>
</tr>
<tr>
<td>3</td>
<td>Stpa florissumi (Knojw.) MacGinittie</td>
<td>Stipeae Dumort.</td>
<td>34</td>
<td>MacGinittie (1953)</td>
</tr>
<tr>
<td>5</td>
<td>Festuca cf. amethystina L.</td>
<td>Festuca L.</td>
<td>26</td>
<td>Thomasson (1985)</td>
</tr>
<tr>
<td>6</td>
<td>Lygnm sp.</td>
<td>Lygnm Loefl.</td>
<td>16</td>
<td>Juchnievicz (1975)</td>
</tr>
<tr>
<td>7</td>
<td>Nassella spp.</td>
<td>Nasseula (Trin.) E. Drev.</td>
<td>8</td>
<td>Jimenez-Moreno, Faquettie, &amp; Suc (2010)</td>
</tr>
<tr>
<td>8</td>
<td>Cusquea GSSCs</td>
<td>Cusquea Kunth</td>
<td>35</td>
<td>Thomasson (1980)</td>
</tr>
<tr>
<td>9</td>
<td>Leersia stleftseemandsnoeffuerENSORH.</td>
<td>Leersia Sw.</td>
<td>30</td>
<td>Stromberg (2005)</td>
</tr>
<tr>
<td>12</td>
<td>Dichanthelium sp.</td>
<td>Dichanthelium (Hitchc. &amp; Chase) Gould = Thyridolepis S.T.Blake</td>
<td>8</td>
<td>Thomasson (1978)</td>
</tr>
<tr>
<td>13</td>
<td>Setaria sp.</td>
<td>Conleus L. + Setaria P.Beauv.</td>
<td>7</td>
<td>Elias (1942)</td>
</tr>
</tbody>
</table>

**Table 1** Fossil calibration points and associated information.

<table>
<thead>
<tr>
<th>Citation</th>
<th>Fossil Calibration Point</th>
<th>Lower bound (age, Ma)</th>
<th>Fossil</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Spikelet in amber</td>
<td>Spikelet clade (all grasses minus Anomochloideae Pilg. ex Potzelt)</td>
<td>97</td>
<td>Poinar et al. (2015)</td>
</tr>
<tr>
<td>2</td>
<td>Oryzeae phyto liths</td>
<td>Oryzeae Dumort.</td>
<td>66</td>
<td>Prasad et al. (2011)</td>
</tr>
<tr>
<td>3</td>
<td>Stpa florissumi (Knojw.) MacGinittie</td>
<td>Stipeae Dumort.</td>
<td>34</td>
<td>MacGinittie (1953)</td>
</tr>
<tr>
<td>5</td>
<td>Festuca cf. amethystina L.</td>
<td>Festuca L.</td>
<td>26</td>
<td>Thomasson (1985)</td>
</tr>
<tr>
<td>6</td>
<td>Lygnm sp.</td>
<td>Lygnm Loefl.</td>
<td>16</td>
<td>Juchnievicz (1975)</td>
</tr>
<tr>
<td>7</td>
<td>Nassella spp.</td>
<td>Nasseula (Trin.) E. Drev.</td>
<td>8</td>
<td>Jimenez-Moreno, Faquettie, &amp; Suc (2010)</td>
</tr>
<tr>
<td>8</td>
<td>Cusquea GSSCs</td>
<td>Cusquea Kunth</td>
<td>35</td>
<td>Thomasson (1980)</td>
</tr>
<tr>
<td>9</td>
<td>Leersia stleftseemandsnoeffuerENSORH.</td>
<td>Leersia Sw.</td>
<td>30</td>
<td>Stromberg (2005)</td>
</tr>
<tr>
<td>12</td>
<td>Dichanthelium sp.</td>
<td>Dichanthelium (Hitchc. &amp; Chase) Gould = Thyridolepis S.T.Blake</td>
<td>8</td>
<td>Thomasson (1978)</td>
</tr>
<tr>
<td>13</td>
<td>Setaria sp.</td>
<td>Conleus L. + Setaria P.Beauv.</td>
<td>7</td>
<td>Elias (1942)</td>
</tr>
</tbody>
</table>

**Table 1** Fossil calibration points, fossil and source information for all calibrations used. Citations listed in separate reference list. GSSCs = grass silica short cells.
values, served as the Bayesian Inference (BI) tree for this study. The ML and BI trees are available as Supplemental Figs. 1 and 2, respectively. Support values were noted in both ML (MLBV support less than 100% shown) and BI tree includes all PP values.

2.5. Neighbor-Net analysis

A neighbor-Net analysis was conducted using the Splits-Tree program (Huson, 1998). Additionally, a phi test (Bruen et al., 2006) of recombination was done through the Splits-Tree program and a delta score with Q-Residual score was also calculated for the 313 taxa matrix to explore the level of reticulation across the dataset. All default parameters were used.

2.6. Alignment gaps analysis

Analyses of the effect of alignment gaps on resolution and topology were conducted by creating stripped nucleotide matrices from the “mask alignment” tool available in Geneious v.11.0.5. The “mask alignment” tool removes columns from the alignment that contain at least a specified percent of gaps. This tool only allows integer values (i.e. 5%, but not 5.5%). Masking thresholds were created for the first 15% of gapped positions stripped (1-15%), as well as for 0% (no gaps remaining), 25%, 50%, 75% and for 100% (all gaps remaining in the matrix), following the methods of (Duvall et al., 2020). As with phylogeny reconstruction, 1Rα was removed to avoid redundancy in inverted repeat site representation. Relative percentages of AT richness, matrix mean pairwise identity, and total number of gaps per stripping threshold alignment were catalogued in 10,000 bp segments using Geneious v11.0.5 statistics, and noted. Data were visualized using Circa (http://omgenomics.com/circa). Maximum likelihood analyses were done identically to the previously described methods (section 2.3), using RAxML-HPC2 on XSEDE for each alignment (20 total alignment matrices with varying stripping thresholds applied). Relationships among Pooideae species were manually noted and any differences in topologies across stripping thresholds were described in detail. Graphical representation of substantial topological inconsistencies was completed in a manner similar to the tree displayed in Carlsen et al. (2018). Support values were also examined to determine any relationships with low support that coincided with topological inconsistencies across stripping thresholds. Further comparisons to the Neighbor-Net tree were conducted to explore evidence of reticulation, similar to methods used in Mitchell et al. (2017) to explore hybridization as a potential hypothesis for differing tree topologies. Trees were visualized using FigTree v1.4.2 (Rambaut, 2014). These 20 alignment matrices are available via Dryad (https://doi.org/10.5061/dryad.ph8cz8w9p0).

3. Results

3.1. Plastome phylogenomic analysis

This study utilized plastome level data for 313 taxa with representatives comprising: 164 Pooideae, 40 Bambusoideae, 15 Oryroideae, 44 Panicoideae, 20 Chloridoideae, six Danthonioideae, six Arundinoideae, five Micrairoideae, six Aristidoideae, one Pueloideae, four Pharoideae, and two Anomochlooideae. Of the 164 Pooideae plastomes, 47 were sequenced and assembled for this study; and of those, 42 are species without a previously published plastome (Supplemental Tables 1 and 2). Fifty-five Micrairoideae, six Aristidoideae, one Pueloideae, four Pharoideae, and the spikelet clade. See Table 2 for mean clade ages and 95% HPD intervals of BOP, PACMAD, and Pooideae supertribes and tribes. The results of this analysis indicate an origin for Pooideae around 84.8 Ma (76.3–93.3 Ma; 95% HPD interval) and all major supertribes having diverged by 30 Ma, except Triticodae which diverged at roughly 25 Ma (1.8–49.3 Ma; 95% HPD interval).

3.3. Neighbor-Net analysis

The neighbor-net analysis produced a separation network from the 313 plastome alignment. Contradictory patterns emerge among species within the early diverging tribes: Brachyelytreae Ohwi, Phaeospermateae Renvoize & Clayton, and Diarrheneae C.S.Campb. Similarly, these tribes also returned low support for their positions in the separation network bootstrap analysis (Brachyelytreae: 26.3; Phaeospermateae: 27.2; Diarrheneae: 27.1) (Supplemental Fig. 3). The Phi test did not find statistically significant evidence of recombination (p = 1.0). The delta score for the dataset was: 0.07428, and Q-residual score: 4.524E-4. These values indicate that there is minimal reticulation, and the data behave in a “tree-like” fashion.

3.4. Alignment gaps analysis

Twenty matrices and their resulting ML tree topologies were exhaustively catalogued through manual comparisons across all 20 matrices and trees to determine any topological inconsistencies based on different stripping thresholds. Matrices ranged in size from 78,002 sites (no gaps included) to 211,879 sites (all gaps included). Percent AT richness among the matrices consistently increased from 59.9% (no

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Results of BEAST divergence estimation analysis.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Crown Node</strong></td>
<td><strong>95% HPD lower, Ma</strong></td>
</tr>
<tr>
<td>Pooideae</td>
<td>16.0</td>
</tr>
<tr>
<td>BOP</td>
<td>89.8</td>
</tr>
<tr>
<td>Phaeospermateae Renvoize</td>
<td>53.4</td>
</tr>
<tr>
<td>Clayton</td>
<td>38.3</td>
</tr>
<tr>
<td>Melicodae Soreng</td>
<td>3.5</td>
</tr>
<tr>
<td>Stipodae Linn</td>
<td>34.0</td>
</tr>
<tr>
<td>Diarrheneae C.S.Campb.</td>
<td>7.0</td>
</tr>
<tr>
<td>Triticodae D.Macfarl. &amp; L. Watson</td>
<td>0.3</td>
</tr>
<tr>
<td>Pooideae Linn</td>
<td>1.8</td>
</tr>
</tbody>
</table>

T2. Results of BEAST divergence estimation analysis. Pooideae taxonomic classifications are predominantly sensu Soreng et al. (2017), refer to the Discussion for in-depth analyses, classifications, and taxonomic authorities of Pooideae taxa included in this study. 95% HPD lower and upper ranges are listed as well as mean age. All times in Ma (mega-annum).
gaps) to 62.5% (all gaps). Mean pairwise identity across the matrices consistently decreased from 97.0% (no gaps) to 88.7% (all gaps) (Fig. 2; Table 3). Additionally, visualization of these data in Fig. 2 clearly show the level of variability across alignments that contain gaps versus an alignment that contains no gaps. Coupled with tree topology comparisons, it becomes clear that interpretations of relationships can be clouded, particularly among taxa with potential hybridization or complex evolutionary histories, when alignment gaps are present in varying amounts across an analysis. All representative supertribes, tribes, supertribes, and subtribes were exhaustively analyzed and differing topologies were noted (Fig. 1, Supplemental Table 3).

4. Discussion

4.1. PACMAD

4.1.1. Phylogeny overview

This study reiterates the complex phylogenomic interpretation of the Aristidoideae/Panicoideae relationship and its relation to the remaining CMAD subfamilies as elaborated in Duvall et al. (2020). All individual subfamilies are recognized as clades. However, sampling of the PACMAD clade here may have influenced the sister group topology of Aristidoideae/Panicoideae and the remaining subfamilies (CMAD). Here, the best tree returned the “Panicoideae sister” topology, while the bootstrap tree returned the “Aristidoideae sister” topology (Burke et al., 2016b; Saarela et al., 2018).

4.2. BOP

4.2.1. Phylogeny overview

The BOP clade, comprised of Bambusoideae, Oryzoideae, and Pooideae, is also returned as monophyletic. In turn, each of the subfamilies is monophyletic with (Oryzoideae, (Bambusoideae, Pooideae)). This larger clade includes a number of economically and ecologically significant species (Wysocki et al., 2015; Saarela et al., 2018; Orton et al., 2019).

### Table 3

Descriptive information for stripping thresholds.

<table>
<thead>
<tr>
<th>Stripping threshold (%)</th>
<th>Matrix length (bp)</th>
<th>AT (%)</th>
<th>Mean Pairwise Identity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 (no gaps)</td>
<td>78,002</td>
<td>59.9</td>
<td>97.0</td>
</tr>
<tr>
<td>1</td>
<td>89,911</td>
<td>60.7</td>
<td>96.5</td>
</tr>
<tr>
<td>2</td>
<td>94,194</td>
<td>61.0</td>
<td>96.2</td>
</tr>
<tr>
<td>3</td>
<td>97,398</td>
<td>61.2</td>
<td>95.9</td>
</tr>
<tr>
<td>4</td>
<td>99,720</td>
<td>61.4</td>
<td>95.6</td>
</tr>
<tr>
<td>5</td>
<td>101,398</td>
<td>61.5</td>
<td>95.4</td>
</tr>
<tr>
<td>6</td>
<td>102,645</td>
<td>61.6</td>
<td>95.3</td>
</tr>
<tr>
<td>7</td>
<td>103,868</td>
<td>61.7</td>
<td>95.1</td>
</tr>
<tr>
<td>8</td>
<td>104,874</td>
<td>61.8</td>
<td>94.9</td>
</tr>
<tr>
<td>9</td>
<td>105,897</td>
<td>61.8</td>
<td>94.7</td>
</tr>
<tr>
<td>10</td>
<td>106,822</td>
<td>61.9</td>
<td>94.6</td>
</tr>
<tr>
<td>11</td>
<td>107,370</td>
<td>61.9</td>
<td>94.5</td>
</tr>
<tr>
<td>12</td>
<td>107,776</td>
<td>61.9</td>
<td>94.4</td>
</tr>
<tr>
<td>13</td>
<td>108,247</td>
<td>62.0</td>
<td>94.3</td>
</tr>
<tr>
<td>14</td>
<td>108,632</td>
<td>62.0</td>
<td>94.2</td>
</tr>
<tr>
<td>15</td>
<td>109,003</td>
<td>62.0</td>
<td>94.1</td>
</tr>
<tr>
<td>25</td>
<td>111,634</td>
<td>62.1</td>
<td>93.3</td>
</tr>
<tr>
<td>50</td>
<td>115,178</td>
<td>62.3</td>
<td>92.0</td>
</tr>
<tr>
<td>75</td>
<td>119,086</td>
<td>62.4</td>
<td>90.5</td>
</tr>
<tr>
<td>100 (all gaps included)</td>
<td>211,879</td>
<td>62.5</td>
<td>88.7</td>
</tr>
</tbody>
</table>

T3. Descriptive statistics for each stripping threshold. Matrix length (in base pairs), percent AT richness, and mean pairwise identities are given.

4.3. Pooideae

4.3.1. Overview

The subfamily Pooideae is comprised of 15 tribes and 26 subtribes (Soreng et al., 2017), of which all but one subtribe and one tribe are represented here: Calothecinae Soreng and Duthieeae Röser & Jul. Schneider. Kellogg (2015) recognizes 10 tribes and 15 subtribes. Here we follow the more recent classifications by Soreng et al. (2017) and Tkach et al. (2020); the latter work added the subtribes: Hypsechoilinae Röser & Tkach, Antinorinae Röser & Tkach, Helichtochilinae Röser & Tkach, Avenulinae Röser & Tkach, Brixochilinae Röser & Tkach. While Helichtochloa is included as a representative of Helichtochilinae, the remaining subtribe additions from Tkach et al. (2020) comprise single genera with one or two species and are not included in this study. However, we also accept many supergeneric taxa listed in Kellogg (2015). It should be noted, that the Calothecinae Soreng has only one recognized genus (Chascolytrum Desv.), which accounts for nine synonymous genera (Nogueira da Silva et al., 2020), and Duthieeae (despite having eight recognized genera), is a relatively new tribe in Pooideae, initially proposed by Pilger (1954) as a subtribe and more recently elevated to a tribe in 2011 (Schneider et al., 2011). Therefore, collection of specimens and complete plastome sequencing has lagged behind longer-established Pooideae for both Calothecinae and Duthieae.

There are a number of well documented complex and reticulate relationships across the tribes of Pooideae, which have been repeatedly revised (Kellogg, 2015; Saarela et al., 2015; Soreng et al., 2015, 2017; Orton et al., 2019) as phylogenetic data of more species becomes readily available. Here we have increased sampling among Pooideae by 47 plastomes, and explored the effects of alignment gaps on topology in large-scale data of 313 taxa. By utilizing plastome data, expanding sampling, and exploring effects of alignment gaps, we have better clarified or confirmed previously unresolved or poorly supported relationships across Pooideae.

4.4. Neighbor-Net

The SplitsTree network, created through a neighbor-net analysis, depicts the relationships between taxa and determines conflicting patterns that emerge. As this study is one of the largest plastome separation networks, it is no surprise that there are a number of conflicts seen among taxa. The results of the delta and Q-residuals indicate the data cluster in a tree-like fashion and there is a high level of resolution among the included taxa, and no detectable reticulation despite evidence to suggest otherwise from previous studies (Mason-Gamer & Kellogg, 1996; Petersen et al., 2006; Mason-Gamer et al., 2010a,b; Fan et al., 2013). However, we have noted specific instances of known or suspected reticulate relationships or hybridizing complexes where there are identifiable points of conflict among taxa or clades (see below for descriptions, and Supplemental Fig. 3) despite the analysis returning no substantial evidence of reticulation per the neighbor-net analysis. These conflicts overwhelmingly appear among taxa with previously noted reticulate relationships or known hybridizing complexes, and were isolated to more narrow taxonomic classifications (tribe/subtribe levels such as Poeae, Diarrheneae, Triticeae, and within Stipeae/Ampelodesmeae). These results may also be a product of the uniparentally inherited, more stable plastome data as opposed to highly variable nuclear loci that often portray clear evidence of reticulation.

4.5. Alignment gaps

The twenty matrices and associated ML trees were extensively catalogued. This analysis expanded upon the levels of stripping thresholds seen in Duvall et al. (2020), particularly in the 1-15% stripping thresholds. As noted by Duvall et al. (2020), there are significant topology shifts within the top 11% of stripping thresholds, and the general trend in AT richness increases as more gapped positions are included in
Fig. 1. Mirrored tree depicting topological inconsistencies between alignments with all gaps removed (left side) and all gaps retained (right side). Positional lines are drawn across trees to show differences in the positions of taxa. Pie charts are situated on the positional lines to indicate percentage representation of topologies. Charts with only two topologies indicated are situated closer to the majority represented topology (i.e. if the majority represented topology was the tree derived from the alignment with all gaps remaining, the chart is placed closest to that tree). Pie charts with more than two topological inconsistencies represented are enlarged to the right of the mirror tree to indicate topology representation and stripping threshold (percentages labeled outside of each pie chart slice). Misidentified specimens are marked with an (*). Additional information on species misidentification (seed accessions, etc.) can be found in Supplemental Tables 1 and 2.
the analysis. Here, we have explored both AT richness and mean pairwise identity, as well as topological inconsistencies across stripping thresholds. These results are consistent with Duvall et al. (2020) with regard to AT richness, as the overall trend shows decreasing AT richness as the number of alignment gaps included in the analysis decreases. There are two stripping thresholds that have 61.8% AT richness (8–9% stripping thresholds) and three that have 61.9% AT richness (10–12% stripping thresholds) and three with 62% AT richness (13–15%).

One hypothesis for the overall decrease seen in AT richness across the dataset (from all gaps remaining to no gaps remaining) may be due to poly-A and/or poly-T sequences in the matrices. Duvall et al. (2020) noted that a relatively substantial portion of the gapped positions removed per stripping threshold corresponded to these poly-A and T sequences, and as the stripping thresholds increased in the percentage of gapped positions removed, the more of these ambiguous poly-A and T regions were removed. Likewise, Illumina kits have a documented bias in AT rich regions (Burke et al., 2016a,b).

Mean pairwise identity across the dataset behaves the opposite of AT richness. The highest mean pairwise identity across stripping thresholds occurs when no gaps remain in an alignment, and the lowest mean pairwise identity occurs when all gaps remain in an alignment. As gaps are increasingly removed from the alignment matrices, fewer ambiguously aligned regions remain and therefore, mean pairwise identity increases.

Total number of gaps per stripping threshold alignment were catalogued in 10,000 bp segments and noted. Gaps fluctuated greatly across the 10,000 bp segments in each stripping threshold alignment. For example, in the 4% stripping threshold alignment numbers of gaps per segment range from 10,899 between bases 1–10,000; 13,376 gaps between segment bases 10,001–20,000; to 7,452 gaps between segment bases 20,001–30,000. The fluctuations in gaps per segment are most noticeable when all gaps remain in the stripping threshold alignment. Here we see numbers of gaps ranging in the hundreds of thousands to millions of gaps among the segments (Fig. 2). This variability has shown to have a substantial effect on topological inconsistencies as many programs used for nucleotide alignment are unable to unambiguously interpret regions with complex distributions of gapped positions, which may then interfere with phylogenetic inference.

The topologies of each stripping threshold were extensively compared and catalogued here by tribe or supertribe level (Fig. 1, Supplemental Table 3), and relationships summarized by taxonomic grouping (see below for in depth text descriptions of specific relationships). In the case of Poeae, relationships were also compared at the subtribe level. Relationships were also compared and contrasted between the assessments of Kellogg (2015), Soreng et al. (2017), and Tkach et al. (2020). While this general taxonomic overview falls in line with the more recent classification done by Soreng et al. (2017), we do recognize minor differences between assessments.

4.5.1. Brachyelytreae Ohwi

Represented in its entirety by one genus and three species (Soreng et al., 2017), and in this study, one species represents the tribe (Brachyelytrum aristosum (Michx.) P. Beauv. ex Trel.). The Brachyelytreae has remained sister to all other Pooideae in all plastid phylogenetic analyses of the subfamily (Kellogg, 2015; Soreng et al., 2015, 2017; Saarela et al., 2015, 2018). As the earliest diverging of Pooideae, its
position is maximally supported at all stripping thresholds.

4.5.2. *Nardaeae* W.D.J.Koch and *Lygeaeae* J.Presl (*Nardodae* Soreng)

Both *Nardus stricta* L. and *Lygeum spartum* L. belong to the supertribe Nardodae, and are maximally supported as sister taxa in our phylogenomic analyses as well as Soreng et al. (2017) and Saarela et al. (2018). In Kellogg (2015), Lygeaeae is combined with Nardaeae.

4.5.3. *Phaenospermateae* Renvoise & Clayton

Along with Brachelytreae and Nardodae, Phaenospermateae is another of the early diverging tribes of Pooidae. The only representative of the Phaenospermateae (*Phaenosperma globosum* Munro ex Benth.) is present in this study and its position sister to Nardodae is maximally supported across all stripping thresholds. Previous studies (Saarela et al., 2018; Schubert et al., 2019) have also returned this position. Both Kellogg (2015) and Soreng et al. (2015) included Duthieae as a synonym in the Phaenospermateae which then contained eight genera and 14 spp. in the previous circumscription. However, the revised Soreng et al. (2017) study separated Duthieae from the Phaenospermateae. Additionally, there is evidence of hybridization between Phaenosperma and two or more tribes (Blaner et al., 2014; Hochbichl et al., 2015).

4.5.4. *Melicaceae* Link ex Endl. and *Brylykineae* tateoka (*Meliaceae* Soreng)

In Melicaceae, *Brylykina F.Schmidt* (Brylykineae) is sister to Meliceae. All relationships in Melicaceae are maximally supported (100% MLBV) across all gap analysis alignment stripping thresholds. Brylykineae was placed in Meliceae in Kellogg (2015), considered as a possible subtribe in Melicaceae in the previous Soreng et al. (2015) classification that was later updated in the Soreng et al. (2017) assessment, which places Brylykineae as sister to Meliceae.

4.5.5. *Stipeae* Dumort. and *Amelodoseae* Tzitzin (*Stipodeae* L.Liu)

The Stipeae, along with the Amelodoseae, form the supertribe Stipodeae. Interestingly, the monophyly of Stipeae is interrupted by *Amelodesmos mauritanicus* (Poir.). T.Durand & Schinz in our analyses and Romaschenko et al. (2012), and although the Amelodoseae is a member of Stipodeae, the position of *A. mauritanicus* is placed within the core group of Stipeae, but poorly supported regardless of stripping threshold. This position is in the clade formed by *Oryzopsis asperifolia* Michx. and *Stipa roylei* (Nees) Duthie. In the alignment gap analysis, this position regularly rotated between (A. mauritanicus, *Oryzopsis* S. roylei) and (S. roylei, (A. mauritanicus, *Oryzopsis*)) or (Oryzopsis, A. mauritanicus, S. roylei). Suspected hybridization between Amelodesaeae and Stipeae was first described by Romaschenko et al. (2014) and is noted in Soreng et al. (2015, 2017), while also being listed in the “Worldwide Classification of Poaceae (Gramineae)” at www.tropicos.org, albeit representatives of Duthieae were included in their analysis and are absent here. Additional hybridization of Amelodesmos between other tribes (Duthieae or Phaenospermateae) is evidenced in shared nuclear genes (Romaschenko et al., 2014). However, Kellogg (2015) includes Amelodesmos in the Stipeae, which is consistent with our results. Yet, morphologically, the Amelodesmos spikelet differs substantially from Stipeae and cannot be reconciled within the Stipeae as a result, and therefore, maintaining Amelodesmos as a separate tribe, but also a member of Stipodeae respects the conflict between morphological and molecular results. Alternative hypotheses of this relationship include the possibility that these tribes diverged rather recently, or that mutation rates are different in the two tribes. Overall, it appears that these tribal representatives are quite closely related.

There also seems to be no particular identifying pattern to the placement of A. mauritanicus in Stipeae, as its position as sister to either S. roylei or Oryzopsis continued to alternate throughout the gap analysis, and there was no definite trend in relation to stripping thresholds, and topological alternation occurred between thresholds 25%, 50%, and 75%, 100%.

Several taxa within Stipeae in our tree (Fig. 1) are non-monoophyletic, while others are monophyletic (this is well noted in several traditional circumscriptions) some of which can be attributed to nomenclatural changes, and others which have been revised to resolve the non-monophyletic relationships (Romaschenko et al., 2012, 2014). In general, the genus *Stipa* L. is monophyletic (*sensu stricto*) but also non-monophyletic (*sensu lato*) with consideration of those species that have been renamed (Columbus & Smith, 2010; Baldwin et al., 2012; Peterson et al., 2019): i.e. synonymous names, *Ericoxyla hymenoides* (Roem. & Schult.) Rydb. = *Oryzopsis hymenoides* (Roem. & Schult.) Ricker ex Piper = *Stipa hymenoides* (Roem. & Schult.) Ricker ex Piper; *Jarava ictiu Ruiz & Pav. = Stipa ictiu (Ruiz & Pav.) Kunth; *Orthoraphium roylei Nees = Stipa roylei* (Nees) Duthie; (https://www.ipni.org). Other members of Stipeae are recognized as monophyletic (i.e. ‘Nassella clade’, core group of *Stipa*) when all gaps are removed. However, there is a lack of resolution across analyses (noted as “NR”—not resolved in Supplemental Fig. 1) for three nodes within the core *Stipa* representatives. Stipeae is also split into two smaller clades, to which *Putis racemosa* (Sm.) Romasch, *P.M. Peterson & Soreng* is sister. Romaschenko et al. (2012, 2014), note that Putis Ohwi, a genus that was previously “resurrected,” is not a particularly close relative of the remaining Stipeae; Kellogg (2015) also notes the unique morphological trait of short rhizomes present in *Putis*.

Kellogg (2015) acknowledges the complexities between Nassella (Trin.) E.Desv. and *Amelichloa* Arriaga & Barkworth and that there is an argument for combining both genera, Soreng et al. (2015, p. 122, 2017, p. 284) noted *Amelichloa* is “(nested within *Nassella*, but an intergeneric hybrid origin has not been ruled out)”. This description is echoed in topological position shifts between *Amelichloa, Nassella hyalina* (Nees) Barkworth, and *Nassella neesiana* (Trin. & Rupr.) Barkworth across stripping thresholds. Here, *Amelichloa* is placed sister to the *Nassella* clade when all gaps are removed from the alignment, in agreement with Romaschenko et al. (2012).


Only one species represents the Diarrheneae in this analysis: *Diarrhena obovata* (Gleason) Brandenburg. There are five total species and two genera (two *Diarrhena* and three *Neomolinia*) in Diarrheneae. Here, Diarrheneae is sister to Brachypodieae + (Triticodae + Poodeae) and is maximally supported (100% Maximum Likelihood Bootstrap Value: MLBV) at this position when all gaps remain in the alignment. However, as gaps are removed, the support decreases from 88.6% MLBV (1% stripping threshold) to 54% MLBV (0% stripping threshold/all gaps removed). Additionally, topology across the stripping thresholds is inconsistent. Among stripping thresholds ranging from 1 – 100% (all gaps present) stripping thresholds, the relationship of Diarrheneae and Brachypodieae is described as above. However, when all gaps are removed, Diarrheneae is sister to Brachypodieae only (Fig. 1; Supplemental Table 3).

4.5.7. Brachypodieae Harz.

The Brachypodieae forms a clade of the four representative species of *Brachypodium* P.Beauv. All relationships are supported at 100% (MLBV) across all gap stripping threshold analyses.


The Triticodae supertribe forms a clade comprised of Littledalaeae, Bromeoideae, and Triticaceae. However, Triticaceae is paraphyletic with Bromoideae, while Littledalaeae remains monophyletic as sister to the Triticodae. Within Triticaceae, evidence of the Triticum L.-Aeglops L. hybridizing complex is minimal despite its prevalence among studies of this tribe which have noted reticulate relationships between both genera (Mason-Gamer & Kellogg, 1996; Petersen et al., 2006; Kawahara, 2009); it is likely that increased sampling would expose more evidence of hybridization/reticulation. Here, we see one clear instance of non-monophyly within Triticaceae (Triticum species): *T. monococcum* subsp.
urartu (Thumanjan ex Gandilyan) Á.Löve & D.Löve and T. monococcum L. were found to be sister to the core Aegilops clade (Fig. 1).

Among the remaining genera included in the Triticeae, Elymus L. is split between Triticeae (Elymus lbinaticus (Hack.) Melderis; a synonym of Pseudoroegneria tauri subsp. lbinatica (Hack.) Á.Löve) and Agrostidinae (Elymus antiquus (Nevski) Tzvelev; synonym: Roegneria antiqua (Nevski) B. S. Sun), but here we have genetic evidence that E. antiquus is likely a misidentified specimen. A comparison of pairwise alignments returned 99.9% identity between E. antiquus and Agrostis stolonifera L., only 92.8% between E. lbinaticus and A. stolonifera, and only 93.0% between E. antiquus and E. lbinaticus. Our “E. antiquus” plastome has greater nucleotide identity with Agrostidinae than with the Triticeae, suggesting that the two accessions identified as Elymus in our study are not truly congeneric representatives of Elymus. E. antiquus is almost certainly a misidentified or contaminated specimen (see GRIN for seed accession used: (https://npgsweb.ars-grin.gov/gringlobal/accession detail.aspx?id=1066862).

Non-monophyletic relationships are also found among representatives of Pseudoroegneria (Nevski) Á.Löve, where P. spicata is synonymous of Elymus spicatus (Pursh) Gould which is, in turn, synonymous of Agropyron spicatum (Pursh) Scribn. & J.G. Sm. However, P. gracilima (Nevski) Á.Löve is an unresolved species name (https://www.theplantlist.org) and interestingly, is nested as (P. gracilima, Thinopyrum pycnanthum [Godr.] Barkworth, Elytrigia lolioides)) while (P. spicata, Connorochloa tenuis [Buchanan] Barkworth, S.W.L.Jacobs & H.Q.Zhang) and also Elymus lbinaticus are sister to this clade. Although these genera are non-monophyletic, this topology is not altogether unexpected based on nomenclatural changes. Many genera in Triticeae have synonyms in Elymus (see example above for Pseudoroegneria). The position of P. gracilima contributes to the non-monophyly of the genus. As P. gracilima is an accepted name we must consider alternate hypotheses for its position. These hypotheses include, 1) USDA misidentification of seed stock, and more likely 2) reclassification of Pseudoroegneria member species may have contributed to these unexpected relationships. Furthermore, Mason-Gamer et al. (2010a,b) identified evidence of hybridization events between Pseudoroegneria, Hordeum L. and Elymus, and even suggested that Pseudoroegneria and Hordeum are potential progenitors of the polyploids present among Elymus species. Multiple studies also explore reticulation among Elymus (Mason-Gamer et al., 2010a,b; Fan et al., 2013).

4.5.9. Poodeae L.Liu, tribe Poeae R.Br.

Many of the relationships described in Orton et al. (2019) are confirmed in this study. There are 48 additional Poeae species included in this study expanding on the sampling in Orton et al. (2019), which focused on the membership of Group 1 and 2 Poeae species. This split between Groups 1 and 2 first noted in Soreng et al. (1990), is also seen in this analysis. Of the additional 48 species, 27 are newly published (Supplemental Table 1). There are 8 Poeae subtribes classified in the PPAM clade by Soreng et al. (2017): Coleanthinae Rouy, Poinae Dumort., Millinae Dumort., Phloeinae Dumort., Beckmanninae Nevski, Cinnaeae Caruel, Alopecurinae Dumort., and Venentatinae Holub ex L.J. Gillespie, Cabi & Soreng. Tkach et al. (2020) introduced new subtribes to the PPAM clade: Brizochloinae Röser & Tkach and Aveninae Röser & Tkach. Furthermore, Kellogg (2015) classified a number of these subtribes, as well as others that are recognized in Soreng et al. (2017), within closely related subtribes. These differences between the Kellogg (2015) and Soreng et al. (2017) classifications are noted. Recent work by Tkach et al. (2020) is discussed as well.

4.5.10. Brizinae Tzvelev

Brizinae (sensu Soreng et al., 2017) is maximally supported as sister to Echinopogoninae Soreng when all gaps remain in the alignment. However, when no gaps remain in the alignment support falls to 96% (MLBV) at the node uniting Brizinae and Echinopogoninae + Agrostidinae. Including additional species to represent both Brizinae and Echinopogoninae in future research may clarify and better support these relationships. Brizinae is included in Agrostidinae in the Kellogg (2015) classification, and Echinopogoninae is not recognized in either Kellogg (2015) or Soreng et al. (2015), and was first introduced in Soreng et al. (2017). In the latter classification, Agrostidinae, Echinopogoninae and Brizinae (and Calothecinae, not included in the present study) are included in supersubtribe Agrostidodinae Soreng.

4.5.11. Agrostidinae Fr.

As elaborated in Orton et al. (2019), there are non-monophyletic relationships among representatives of Agrostidinae. In the 2019 analysis, Orton et al. found the monophyly of Agrostis L. species with the single exception that Polygogon fugax Nees ex Steud. (synonym: Nowadworskiya fugax (Nees ex Steud.) Nevski) was included in this clade. Here, with expanded taxa included in this analysis, we have replicated this likely reticulate relationship between Polygogon Desf. and Agrostis. However, additional taxonomic sampling and inclusion of nuclear data will likely provide more definite assessments. In the Kellogg (2015) classification, the tribe Agrostidinae includes both subtribes Brizinae and Calothecinae (Calothecinae was not included in Kellogg’s 2015 assessment, but the associated genera were included in Agrostidinae), which are recognized in Soreng et al. (2015, 2017) as independent subtribe designations and also members of the Agrostidodinae super-subtribe (see Brizinae above).


Recently, Barberá et al. (2017, 2019, 2020) transferred some groups of Trisetum to the genus Koeleria based on morphological and molecular evidence, and proposed a new genus (Sibiriostrisum Barberá, Soreng, Romasch., A. Quintanar & P.M. Peterson) while resurrecting another (Acrospelion Besser) to further reflect the relationships in Aveninae. Here we transfer Trisetum glaciale to Acrospelion based on the results of Barberá et al. (2017, 2018, 2020).


4.5.13. Millinae Dumort. & Phlieneae Dumort.

Between our stripping thresholds from zero gaps remaining in the alignment to 25%, (Millium effusum L. + Phleum alpinum L.) is sister to the Poa clade. However, in our thresholds of 50%, 75%, and all gaps removed, the representatives of Millinae and Phleinae are instead sister to Beckmanninae Nevski. This differs from Soreng et al. (2017), in which Phleinae is sister to a clade comprised of Beckmanniinae plus three other subtribes (Cinnaeae, Alopecurinae, and Ventenatinae). Millinae is sister to a clade comprised of all of the latter subtribes plus Poinae Dumort. Millinae and Phleinae are unresolved in the multi-focus plastid phylogeny of Tkach et al. (2020).

4.5.14. Beckmanninae Nevski

The lack of resolution between the sister relationship of Beckmanninae and (Arcytagrotis Griseb. + Cinnaeae) + (Alopecurinae + Ventenatinae) is potentially due to the inclusion of Arcytagrotis (incerta sedis in the Pooideae). Soreng et al. (2017) classifies Beckmanninae as a member of the Pooideae supersubtribe, while Kellogg (2015) places Beckmannia Host as a member of subtribe Poinae.

4.5.15. Airinae Fr.

Nomenclatural changes have clarified some relationships in the Airinae. The renaming of Deschampsia flexuosa (L.) Trin. to Avenella flexuosa (L.) Drejer has since resolved a suspected paraphyletic relationship among Aristaveninae F.Albers & Butzin and Airinae among our sampled taxa. Both Holcinae Dumort. and Aristaveninae are placed in
Airinae by Kellogg (2015). Our phylogenomic results are somewhat in disagreement with this placement given the more recent circumscriptions included in Tkach et al. (2020).

4.5.16. Aristaveninae F. Albers & Butzin

The Aristaveninae is represented in this study by a single species: Deschampsia antarctica E. Desv., which is supported as sister to Scolochloa festucaea (Willd.) Link L. + Loliodaeae Soreng at 99% MLBV. The Orton et al. (2019) study also determined D. antarctica was sister to S. festucaea + Loliodaeae.

4.5.17. Holcinae Damort. & Helictochloinae Rüser & Tkach

Holcus lanatus L. (Holcinae) is sister to Helictochloa hookeri (Scribn.) Romero Zarco (Helictochloinae) (86% MLBV), but due to a lack of additional representatives from this sub-tribe, an assessment of monophyly is limited. However, in Orton et al. (2019) this same topology was recovered in an analysis of Group 1 and Group 2 Poae species. Additionally, Kellogg (2015) placed Holcinae within Airinae and Soreng et al. (2017) placed Helictochloa in Airinae. Helictochloa hookeri is included in its own subtribe (with Molinierella Rouy) as noted in Tkach et al. (2020), but no sister relationship with Airinae or Holcinae was detected.

4.5.18. Scolochloinae Tzvel

Scolochloa festucaea (Willd.) Link. is the sister of the Loliodaeae subtribe, and nested within the supersubtribe Loliodinae at maximal support. This position, which differs from Soreng et al. (2017), was newly discovered in the Orton et al. (2019) study on Loliodinae at maximal support. This position, which differs from Soreng (2017) and Kellogg (2015), as well as Gillespie et al. (2016b), which estimated the origins of the spikelet clade to be roughly 100 Ma. This evidence places the origins of Poaceae deeper into the Cretaceous than many earlier estimates.

Our results pertaining to Pooidae origins are in partial agreement with Schubert et al. (2019) where it was estimated that Pooidae diverged roughly 61 – 77 Ma, yet our timeline indicates a slightly older Cretaceous origin for Pooidae circa 76.3 – 93.3 Ma (Campanian – Turonian ages) (Fig. 3).

The atmosphere of the late Cretaceous contained more than double the CO2 levels of modern-day (Andrews et al., 1995; DeConto et al., 1998, 1999), and temperatures were steadily increasing by this time. However, evidence suggests that there were isolated south pole glaciation events occurring during the Turonian age, which may have contributed to periods of constant or decreased temperatures in more southern latitudes (Bormann et al., 2008). While modern Pooidae are comprised of cool season grasses, early lineages had only limited cold tolerance (Schubert et al., 2019). We speculate that later cooling trends of the Paleogene near the Eocene/Oligocene transition between 33 and 35 Ma, which exerted evolutionary pressures as a result of climate cooling, contributed to the later diversifications seen at the supertribe levels (e.g., Pooidae, Nardodae, Stipodae) (Minaya et al., 2017; Pimentel et al., 2017; Schubert et al., 2019).

Nearly all Pooidae supertribes had diverged by 30 Mya, excluding Triticodae which is estimated to have diverged around 25 Ma. This younger age may be an artifact of phylogenomic signal from the wheat cultivar species skewing the age younger due to more recent selective breeding for desired genes, or also due to the inclusion of annuals which have high mutation rates and short generation times (Kim et al., 2009).

Interestingly, grasses likely flourished during this time of greater cooling as they are particularly robust and effective in dispersing long distances, establishing or colonizing diverse habitats, as well as having the ability to adapt to rapidly changing environments, which makes them successful invasive species (Linder et al., 2018; Minaya et al., 2017). In particular, the cool season grasses, in which many lineages evolved frost tolerance, would have navigated the cooling climate more readily than the tropical varieties of grasses and other plants (Linder et al., 2017; Zhong et al., 2018; Schubert et al., 2019). This is reflected in many modern distributions of pooid grasses with ranges primarily in the northern hemisphere and some species inhabiting southern temperate regions (Soreng et al., 2015, 2017; Zhong et al., 2018; Watson et al., 1992 onwards).

Pooideae is a cosmopolitan subfamily, but predominantly distributed in North America and temperate Eurasia (Watson et al., 1992 onwards;
Kellogg, 2015; Soreng et al., 2017; Zhong et al., 2018), with largely open habitat adapted species. There are a number of species exhibiting diverse habitats ranging from mountainous open areas to sandy coastal soils to species that thrive in shade tolerant savannas or woodlands (Watson et al., 1992 onwards). It is likely that most early pooid grasses dispersed northward through the southern region of Laurasia in Pangea and some pushed into southern temperate regions from initial ancestral origins in Gondwana (Brown & Smith, 1972; Minaya et al., 2017; Schubert et al., 2019).

Substantial continental movement continued through the Paleogene, and evidence of ancient dispersal routes for pooid grasses has brought about more recent and pointed studies aimed at exploring dispersal at the tribal or generic level (Blattner, 2006; Inda et al., 2008; Birch et al., 2014). Based on this research, and known prehistoric continental drift, it can be assumed that in general, many of the ancient pooids later migrated north into what is modern Europe and through areas now found in the Mediterranean and southwestern Asian corridor (once portions of Laurasia). Ancient pooids also expanded into temperate southern regions of what is now sub-Saharan Africa during the Miocene - Pliocene (Blattner, 2006; Inda et al., 2008; Minaya et al., 2017).

Eventually, the northward expanding groups moved to Australasia from Eurasian regions and separately to the New World through distribution along migratory routes into (what is now) Asia (Minaya et al., 2017). During the Pliocene, some New World bound pooids began migrating via land bridges into North America and continued expanding into areas of what is now Central and South America before transitioning into Australasian regions through long-distance dispersal among intermediate ancient island chains (Blattner, 2006; Jakob et al., 2007; Inda et al., 2008; Birch et al., 2014; Wen et al., 2016; Minaya et al., 2017). As continued continental drift led to dispersal barriers and vicariance, many groups began to diversify in secluded habitats creating disjunct but cosmopolitan distributions (Blattner, 2006).

5. Conclusion

Pooidae exhibit many complex and often reticulate relationships among and within member tribes and subtribes. Here we demonstrated that large-scale datasets using plastome scale information provide added resolution to many sparsely represented groups through the addition of genetic data. There still remains a deep deficit with regard to Poaceae
plastome level data available in sequence repositories. However, this study has explored the many benefits of utilizing the maximum level of genetic information that exists in plastomes.

Sampling strategies, as well as including or excluding alignment gaps, also greatly affect the quality of datasets and resulting analytical outcomes. Likewise, alignment gaps are a byproduct of sampling when introduced into a dataset. Here we have presented evidence that including alignment gaps can produce alternate topologies, which likely introduces artifacts from ambiguously aligned low-complexity regions of DNA, and should be avoided unless absolutely necessary. However, in comparing phylogenies produced from multiple stripping thresholds to explore the effect of alignment gaps on these analyses, we have seen that known/suspected reticulate relationships of hybridizing complexes behave in unexpected or unusual ways. This suggests that the underlying evolutionary complexities of biparentally inherited organismal histories are at least somewhat detectable, even in the phylogenies of unparentally inherited plastomes.

Finally, the divergence estimation analysis determined a late Cretaceous origin for Poaceae at approximately 76.3–93.4 Ma. While this is slightly older than other recent estimates, it is in line with the hypothesis that divergences among lineages of grasses may be older than commonly estimated, as there is a lack of fossil evidence.

Future directions seek to include greater sampling in underrepresented taxonomic groups, particularly Duthieeae and Calothecinae, and include nuclear data to compare/contrast phylogenetic signals of both plastid and nucleus, as well as further explore possible hybridizing complexes.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ympev.2021.107110.

References


Kawahara, T., 2009. Molecular phylogeny among
Kilian, B.,
Columbus, J.T., Smith Jr, J.P., 2010. Nomenclatural changes for some grasses in
pp. 81
https://doi.org/10.1016/j.ympev.2015.03.010.
determination and comparative analysis for members of the
Kilian, B., Xue, F., lung, L., Kopp, M., Hackel, J. 2019. A key to the
125. https://doi.org/10.3897/phytokeys.126.34096.
https://doi.org/10.1093/molbev/mms172.
https://doi.org/10.1093/molbev/mms172.
https://doi.org/10.1093/molbev/mms172.


Table 1 References


