American Journal of Botany

American Journal of Botany 97(4): 620–627. 2010.

The chloroplast genome of *Anomochloa marantoidea* (Anomochlooideae; Poaceae) comprises a mixture of grass-like and unique features[1]

Leah M. Morris and Melvin R. Duvall[2]

Department of Biological Sciences, 1425 W. Lincoln Hwy, Northern Illinois University, DeKalb, Illinois USA 60115-2861

Features in the complete plastome of *Anomochloa marantoidea* (Poaceae) were investigated. This species is one of four of Anomochlooideae, the crown node of which diverged before those of any other grass subfamily. The plastome was sequenced from overlapping amplicons using previously designed primers. The plastome of *A. marantoidea* is 138 412 bp long with a typical gene content for Poaceae. Five regions were examined in detail because of prior surveys that identified structural alterations among graminoid Poales. *Anomochloa marantoidea* was found to have an intron in *rpoC1*, unlike other Poaceae. The insertion region of *rpoC2* is unusually short in *A. marantoidea* compared with those of other grasses, but with atypically long subrepeats. Both *ycf1* and *ycf2* are nonfunctional as is typical in grasses, but *A. marantoidea* has a uniquely long ψ*ycf1*. Finally, the *rbcL-psaI* spacer in *A. marantoidea* is atypically short with no evidence of the ψ*rpl23* locus found in all other Poaceae. Some of these features are of noteworthy dissimilarity between *A. marantoidea* and those crown grasses for which entire plastomes have been sequenced. Complete plastome sequences of other Anomochlooideae and outgroups will further advance our understanding of the evolutionary events in the plastome that accompanied graminoid diversification.

Key words: *Anomochloa marantoidea*; Anomochlooideae; plastome; Poaceae; *rpl23* pseudogene; *rpoC1* intron.

The evolution and adaptive radiation of the graminoid Poales (sensu Bremer, 2002) has been the subject of much scrutiny because of the economic and ecological importance of grasses, the dominant lineage in the clade. One aspect of this body of research is the evolution of the graminoid chloroplast genome (plastome). Grasses are especially well represented among the completely or almost completely sequenced plastomes, including species from 15 genera—*Agrostis*, *Hordeum*, and *Sorghum* (Saski et al., 2007); *Bambusa* and *Dendrocalamus* (Wu et al., 2009); *Brachypodium* (Bortiri et al., 2008); *Coix* (Leseberg and Duvall, 2009); *Cryptochloa*, *Microcalamus*, and *Puelia* (Duvall et al., in press); *Lolium* (Diekmann et al., 2008), *Oryza* (Shahid-Masood et al., 2004; Tang et al., 2004), *Saccharum* (Asano et al., 2004), *Triticum* (Ogihara et al., 2002) and *Zea* (Maier et al., 1995). No complete plastomes are yet available from nongrass graminoids, although some plastome characters in Anarthriaceae, Centrolepidaceae, Ecdeiocoleaceae, Flagellariaceae, Joinvilleaceae, and Restionaceae have been surveyed (Doyle et al., 1992; Katayama and Ogihara, 1996; Michelangeli et al., 2003; Marchant and Briggs, 2007).

The graminoid plastome shows distinctive molecular evolutionary characters in its content of loci including pseudogenes, protein coding sequences, and introns. Significant losses of sequences that are generally found in angiosperm plastomes have been observed. For example, *ycf1* and *ycf2* are the largest open reading frames present in land plant plastomes, with putative protein products of 1900–2800 amino acids in length. These loci have not been found in the plastomes of any Poaceae and may be missing from other graminoid Poales, though there are incomplete data for other families of the lineage (Maier et al., 1995; Chang et al., 2005).

As another example of a missing sequence, in the majority of land plants, *rpoC1* is one of the relatively few plastome genes found with an intron. The homoplasious loss of this intron is reported in several lineages (Downie et al., 1996), the largest of which is Poaceae; all grasses for which *rpoC1* has been sequenced have lost the *rpoC1* intron.

Some graminoids are also distinguished by sequences not found in the plastomes of other angiosperms. Grasses have a mutational hotspot in the noncoding region between *rbcL* and *psaI* (Morton and Clegg, 1993). In most Poaceae, this hotspot contains the pseudogene ψ*rpl23*, probably as a result of a nonreciprocal translocation of the functional copy of *rpl23* from one of the inverted regions (Katayama and Ogihara, 1996).

Investigations of the plastome characters of the graminoids are generally confined to a survey of relatively few of the 10 600 described species including representatives of the crown grasses and one or more outgroups. Once identified, a character found in several major lineages, such as the insertion sequence in *rpoC2*, is then associated with the encompassing taxon, in this case Poaceae. While this is a generally reasonable approach, the identification by Clark et al. (1995) that Anomochlooideae are the sister group to other Poaceae suggests the importance of including representatives of this subfamily in any study that purports to define exceptional features of the grass plastome. Here we present a complete plastome sequence for a species of Anomochlooideae to refine our understanding of the origin of plastome mutations in graminoids.

Anomochlooideae has only two known extant genera, *Anomochloa* (one species) and *Streptochaeta* (three species). These two genera share few morphological synapomorphies, and

[1] Manuscript received 31 July 2009; revision accepted 22 January 2010.

The authors thank Lynn Clark, Iowa State University, for donating vouchered plant tissue. This work was supported in part by a Fellowship from Northern Illinois University to L.M.M., the Plant Molecular Biology Center, and the Department of Biological Sciences at Northern Illinois University.

[2] Author for correspondence (e-mail: t80mrd1@wpo.cso.niu.edu), phone: 815-753-7806, fax: 815-753-7855

doi:10.3732/ajb.0900226

their phylogenetic relationships have been poorly understood (Clark and Judziewicz, 1996). Molecular data suggest that the ancestor of *Anomochloa* and *Streptochaeta* diverged prior to all other grasses (Clark et al., 1995; Mathews et al., 2000; GPWG, 2001) so they may be provisionally viewed as sister taxa within Anomochlooideae despite their unusual morphologies (GPWG, 2001).

Anomochloa marantoidea Brogn., is endemic to the Bahia coastal forests of Brazil (Thomas et al., 1998). As one of the Anomochlooideae, *A. marantoidea* occupies a position of special molecular phylogenetic interest, although morphological homologies with Poaceae are difficult to interpret. This species is distinguished morphologically by pulvini at both the apex and base of its pseudopetioles, by keeled midveins on both sides of its broad leaves, and by the absence of ligules at the juncture of pseudopetiole and culm (Judziewicz and Soderstrom, 1989). The floral structures are unusual in having four anthers, rather than the more usual (1–) 3 (−6) found in other grasses, and having "spikelet equivalents" (Clark and Judziewicz, 1996) rather than true spikelets, comparable to those in *Streptochaeta*. This distinction is the basis for the name "spikelet clade," which refers to all Poaceae excluding Anomochlooideae (GPWG, 2001). The overall suite of morphological traits is atypical of Poaceae, contributing to the difficulty of establishing relationships between *A. marantoidea* and the rest of the family on the basis of morphology and anatomy (Judziewicz and Soderstrom, 1989).

In this paper we explore the molecular evolutionary information in the complete plastome of *A. marantoidea*. Comparisons were made with other completed monocot plastomes, which at this point are limited to crown grasses and four nongraminoid monocots—*Dioscorea elephantipes* Engl. (Hansen et al., 2007; GenBank accession NC_009601), *Lemna minor* L. (Mardanov et al., 2008; NC_010109), *Phalaenopsis aphrodite* Rchb. f. (Chang et al., 2005; NC_007499), and *Acorus calamus* L. (Goremykin et al., 2005; NC_007407). Evidence from some previous Southern hybridization surveys of Poales (Cummings et al., 1994; Katayama and Ogihara, 1993, 1996) was also used to better define the molecular evolution of graminoid plastomes.

MATERIALS AND METHODS

DNA extraction and sequencing—Silica-gel-dried leaf samples of *Anomochloa marantoidea* Brogn. were obtained from cultivated plants (voucher: L. Clark #1299 [ISC]). DNA was extracted using a Qiagen DNeasy Plant Mini kit (Qiagen, Valencia, California, USA), with the addition of liquid nitrogen during homogenization to facilitate tissue grinding.

On the basis of data from previously sequenced grass plastomes, the *A. marantoidea* plastome was amplified as a series of regions of ~1 kb each, with primers designed from sequences that are highly conserved. Twenty primer pairs were designed and donated (as aliquots) by Dhingra and Folta (2005), and Leseberg and Duvall (2009) designed approximately 100 others. In the event of the failure of primers for a given region, amplification of that region was then attempted first by combining the upstream forward primer with the reverse primer for the region, or by combining that region's forward primer with the downstream reverse primer, in two separate polymerase chain reaction (PCR) reactions. If both of these approaches were unsuccessful, then one or more new primers were designed from the successfully sequenced flanking region(s) (sequences of these primers available on request).

These regions were amplified using a touchdown PCR program following the "Round I" conditions of Dhingra and Folta (2005), using *Pfu* Turbo DNA polymerase (Stratagene, Carlsbad, California, USA) in most PCR reactions. The Fidelitaq PCR system (USB Corp., Cleveland, Ohio, USA) was also used to amplify a small number of regions that failed under the *Pfu* Turbo PCR conditions. Once successfully amplified, plastome regions were purified with Wizard SV PCR Clean-up System kits (Promega, Madison, Wisconsin, USA). The purified DNA was then sent for automated capillary sequencing in both directions (Macrogen, Seoul, South Korea).

Plastome analysis—Sequenced amplicons from *A. marantoidea* were assembled in the program Gene Inspector vers. 1.6 (Textco, Biosoftware, West Lebanon, New Hampshire, USA). The identities of individual contigs were verified with BLASTN searches (Altschul et al., 1990). Alignment against a reference plastome (*Oryza sativa* L., GenBank accession NC_001320) was performed by a combination of machine alignments, using the CLUSTAL W (Thompson et al., 1994) module embedded in Gene Inspector 1.6, and Geneious Pro vers. 3.8 software (Drummond et al., 2007), with minor manual adjustments. The reference alignment was used to verify the arrangement and position of loci.

The plastome of *A. marantoidea* was then annotated using the program DOGMA (Wyman et al., 2004). Manual modifications were made to the automated alignment based on homologous comparisons against grass plastome references. These modifications were to add the locations of pseudogenes and hypothetical open reading frames to the annotation, adjust intron-exon boundaries, and specify the exact coordinates of locus termini.

RESULTS

The annotated plastome sequence for *A. marantoidea* was deposited in GenBank as accession GQ329703. The size of the complete plastome is 138 412 bp. The large single-copy (LSC) region is 82 274 bp long, the short single-copy (SSC) region is 12 162 bp long, and the inverted repeats (IRA and IRB) are each 21 988 bases long. The general organization of the plastome of *A. marantoidea* largely reflects the highly conserved gene content and gene order of the grass plastome including the three characteristic inversions previously documented in *A. marantoideae* (Michelangeli et al., 2003). However, in other significant respects, the plastome of *A. marantoidea* is unique.

One of the constituent loci of the RNA polymerase operon cluster found in the LSC region is *rpoC1*, which in *A. marantoidea* has 2058 bp of protein-coding sequence. Typical of mostly nongraminoid plastomes, the locus also has an intron, beginning after nucleotide 432 of *rpoC1*, which is 736 bp long. There are no published reports of other Poaceae with an *rpoC1* intron. The intron sequence of *A. marantoideae* has 72–87% nucleotide identity with those of other monocots including species of *Dioscorea*, *Lemna*, *Phalaenopsis*, and *Acorus*.

The *rpoC2* locus, also in the RNA polymerase operon of the LSC region, shows substantial variation in size and sequence among monocots largely because of an insert in the coding region (Igloi et al., 1990; Shimada et al., 1990; Katayama and Ogihara, 1993). After a Southern hybridization survey for the *rpoC2* insert of 55 angiosperms including 37 graminoids with one species each of Joinvilleaceae and Flagellariaceae, two species of Restionaceae, and 33 species of Poaceae, Cummings et al. (1994) concluded that the extra coding sequence was restricted to Poaceae and was comprised largely of 21-bp repeats with a consensus sequence of TATGGAACCCTAGAGGAAGAA. There is an insert in the *rpoC2* locus of *A. marantoidea* following nucleotide position 1 926. Although the inserts were previously sequenced from both *A. marantoidea* and *Streptochaeta angustifolia* Soderstr. (AF064759 and AF064760), they were reported to be "unalignable at the amino acid level" (Barker et al., 1998, p. 338). However, we found that the nucleotide sequence of the insert in *A. marantoidea* aligns to the second half of a typical *rpoC2* insert. This insert in *A. marantoidea* is almost entirely comprised of 39-bp subrepeats (confirming an observation of Barker et al., 1998) with one interruption by a

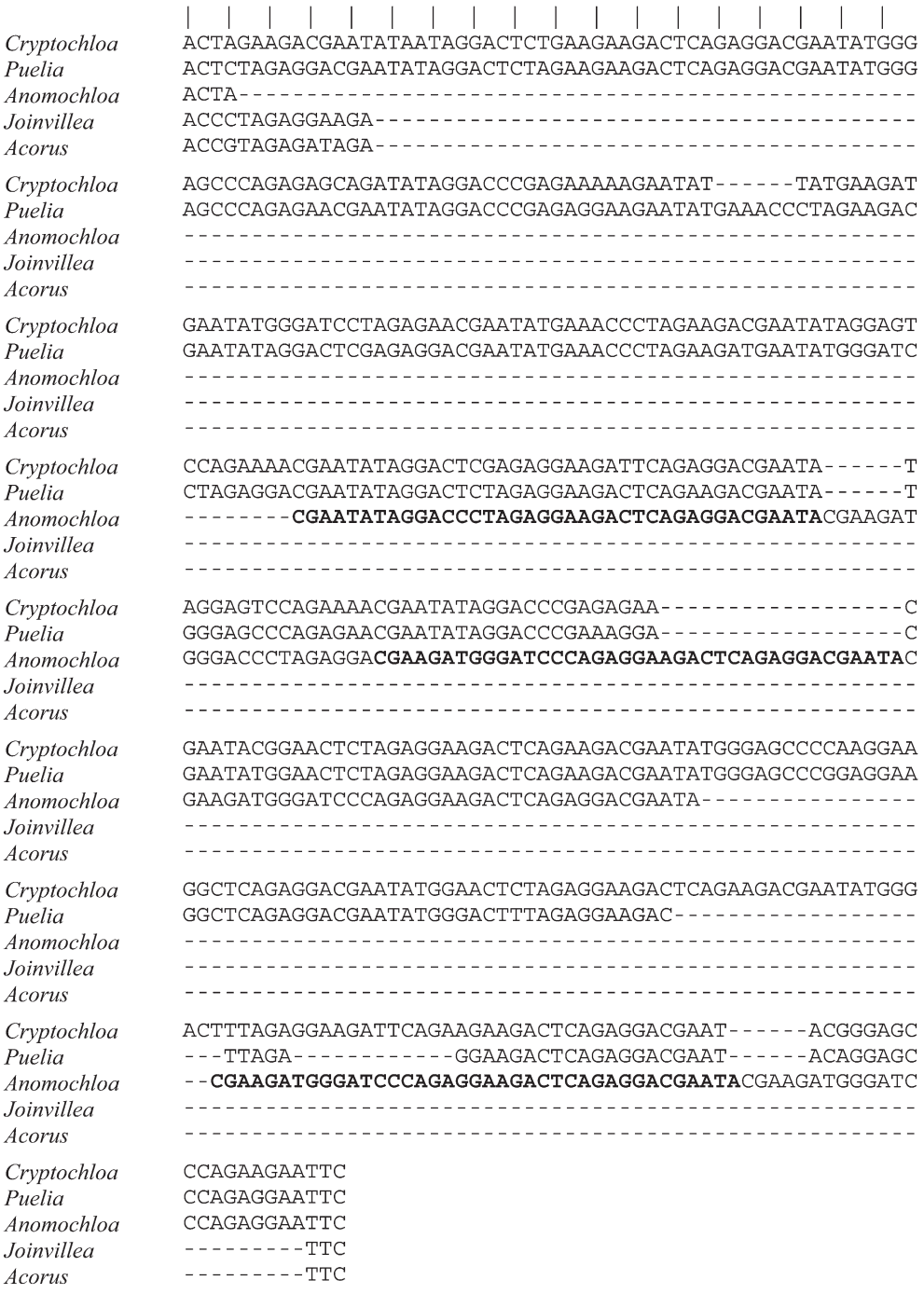

Fig. 1. Aligned *rpoC2* insert region plus short portions of the flanking regions for five monocots. Sequences were aligned following the rule-governed method of Barker et al. (1998) to preserve subrepeat structure. The sequence begins at coordinate 28 933 of the *A. marantoidea* plastome. First codon positions are indicated across the top. Subrepeats are shown alternately in boldface and normal text in *Anomochloa*. *Cryptochloa strictiflora* and *Puelia olyriformis* sequences were obtained from accessions M4608–M4611 at http://www.treebase.org. *Joinvillea plicata* (Hook. f.) Newell & Stone and *Acorus calamus* sequences were from GenBank accessions FJ486257 and NC_007407, respectively.

21-bp subrepeat and a final repeat that is interrupted after base 22 (Fig. 1). The nucleotide subrepeats of *A. marantoidea* are similar to those of *Streptochaeta angustifolia*. The consensus sequence of the larger repeat is CGAA**g**atgg**g**a**t**cc**c**agaggaaga**c**TCAGAGGACGAATA, where the 21-bp consensus subrepeat of Cummings et al. (1994) (lower case) is embedded within this longer consensus sequence of *A. marantoidea* with differences at five sites in bold face. The complete insertion in *A. marantoidea* is only 198 bp, which is the shortest reported *rpoC2* insert sequence.

Anomochloa marantoidea has an unusually short *rbcL-psaI* intergenic spacer (IGS) of 746 bp; the corresponding IGSs in other Poaceae have a mean length of 1112 bp with only one species, *Brachypodium distachyon* (L.) P. Beauv., having a

shorter IGS (Diekmann et al., 2009; FJ261955). Two short segments of this IGS in *A. marantoidea* have over 83% similarity to the *rbcL-psaI* intergenic spacer of other Poaceae. The first 250 bp are homologous to the region just downstream of *rbcL*, and the final 110 bp is homologous to the upstream flanking region of *psaI*. *Anomochloa marantoidea* also has a sequence of 386 bases in the central portion of its IGS. This sequence is of unknown homology; BLASTN searches fail to find any chloroplast targets.

Although Morton and Clegg (1993) documented that *ψrpl23* is commonly found in this region in grasses, a BLASTN search of the IGS of *A. marantoidea* against the *ψrpl23* sequence from *Saccharum officinarum* L. shows "no significant similarity." The absence of *ψrpl23* in *A. marantoidea* coincides with the fact that this region is about 500 bp shorter than that in other grasses. However, even the very short IGS of *Brachypodium distachyon* has a 40-bp fragment with 95% identity to the *ψrpl23* of *Saccharum officinarum*. In this respect, *A. marantoidea* differs from other Poaceae and is instead like species outside the family that have been surveyed (Katayama and Ogihara, 1996).

In the IR regions of the plastome of *A. marantoidea* between the functional *rpl23* and ORF28 loci, there is a 660-bp sequence showing high identity to two fragments near the 3′ terminus of *ycf2* in nongrass monocots, such as *Dioscorea elephantipes*. Grasses are characterized by the loss of much of the *ycf2* sequence; it typically persists only as a set of short open reading frames (ORF34, 38, 46, 139, 241; Maier et al., 1995) variably found in this region of the grass plastome. In this regard, the plastome of *A. marantoidea*, is similar to that of other grasses, and the bulk of its *ψycf2* shows 95% nucleotide identity with ORF241 in *Zea mays* L.

Also in the IRs of *A. marantoidea* between *trnN-GUU* and *rps15*, there is a region of 876 bp beginning at coordinate 101 005 in which there are plastome segments with high homology to the *ycf1* loci in species of nongraminoid monocots. Specifically, there are two subregions 285 and 366 bp long with a minimum of 71% nucleotide identity to *ycf1* in *Dioscorea elephantipes* and other nongraminoid monocots. This sequence is about 15% of the length of a complete copy of the *ycf1* locus, and its structure is further eroded by indels that introduce frameshifts so that in *A. marantoidea* the *ycf1*-like sequence can be clearly considered to be a pseudogene. Like the intact copies in the plastomes of *Dioscorea* and *Lemna*, this *ψycf1* is found in the IR near the SSC boundaries, unlike the cases in *Phalaenopsis* and *Acorus* where *ycf1* is located in the SSC region. A similarly large *ψycf1* is otherwise unknown from Poaceae and to our knowledge has not been identified in the plastomes of other graminoids.

DISCUSSION

The molecular evolutionary characters in the plastome of *A. marantoidea* reflect mutational events taking place during the evolutionary divergence and radiation of graminoids. Some of these events can be placed precisely, while others are more ambiguous either due to uncertain character state assignments or because data are missing from critical taxa. The inferred placements of plastome characters in the graminoid phylogeny are summarized (Fig. 2).

***Implications of the* rpoC1 *intron*—**In the majority of land plants, *rpoC1* is interrupted by a single intron of approximately 700 bases. The loss of this intron is a homoplasy of several lineages of angiosperms (Downie et al., 1996), one of those putatively being Poaceae. All other grasses for which *rpoC1* has been sequenced—at this writing there are 138 *rpoC1* sequences from species and subspecies of Poaceae in GenBank—have been found to have lost this intron.

Katayama and Ogihara (1996) assayed the distribution of the *rpoC1* intron in Poales. They found that in addition to Poaceae, this intron has been lost in Restionaceae, Centrolepidaceae, and Anarthriaceae, but it is retained in Joinvilleaceae, which is more closely related to Poaceae. They were unable to establish its presence or absence in Ecdeiocoleaceae, another family that has been suggested to be in the sister clade to Poaceae (Michelangeli et al., 2003; Marchant and Briggs, 2007) although exact relationships between Poaceae and its immediate outgroups are unresolved. Consequently, the variable distribution of this intron may be viewed in one of two ways. First, there may have been a synapomorphic loss of the intron after the spikelet clade diverged, where its retention in Anomochlooideae is a symplesiomorphy of Poaceae with Joinvilleaceae and Cyperaceae. This hypothesis retains the taxonomic boundaries proposed by the GPWG (2001), but eliminates one diagnostic "grass-specific" molecular character. Alternately, if the synapomorphic loss of the *rpoC1* intron is viewed as a defining characteristic of Poaceae, then Anomochlooideae cannot be considered part of that family.

***Implications of the* rpoC2 *insert*—**The insertion sequence in the *rpoC2*-coding region is a remarkable feature with no known function. It has been considered to be a "grass-specific" character not found in the *rpoC2* loci of other plant families (Cummings et al., 1994; Barker et al., 1998, 2003; Morrone et al., 2007). The size of the overall insert in *A. marantoideae* is unusually small, about half the length of a typical insert as found in *Streptochaeta angustifolia* and crown grasses (Duvall et al., 2001). In other species the insert ranges from 369 bp in Puelioideae [*Puelia olyriformis* (Franch.) Clayton—M. Duvall unpublished data] down to 333 bp in two pooids (*Agrostis stolonifera* L. and *Lolium perenne* L.) yet up to 405 bp in another pooid (*Triticum aestivum* L.) and again from 390 to 456 bp in Panicoideae [*Chasmanthium latifolium* (Michx.) Yates—M. Duvall and C. Yankaitis, unpublished data—and *S. officinarum* L., respectively].

The subrepeats in the *rpoC2* insert of *A. marantoideae* and *S. angustifolia* are also unusual since most are 18 bp longer than is typical for other Poaceae (Cummings et al., 1994; Barker et al., 1998). The 39-bp repeat structure at this point appears to be unique to Anomochlooideae, with no other variation in repeat size evident. One possible explanation of these observations is that the insertion event occurred in the ancestor of Poaceae, after which some unknown further events in either or both Anomochlooideae and the spikelet clade caused subrepeats of either 21 or 39 bp to become prevalent. However, a mechanism for interconverting the complete set of subrepeats from one size to another is not known.

ψrpl23, *and the* rbcL-psaI *mutational hotspot*—Grasses have a mutational hotspot in the noncoding region between *rbcL* and *psaI* (Morton and Clegg, 1993; Clegg et al., 1994). In most grasses, this hotspot contains an *rpl23* pseudogene that ranges from 40–243 bp in length. Shimada and Sugiura (1989) proposed that a portion of the IR was inserted into the LSC, either in one event or in multiple smaller events; the *ψrpl23*

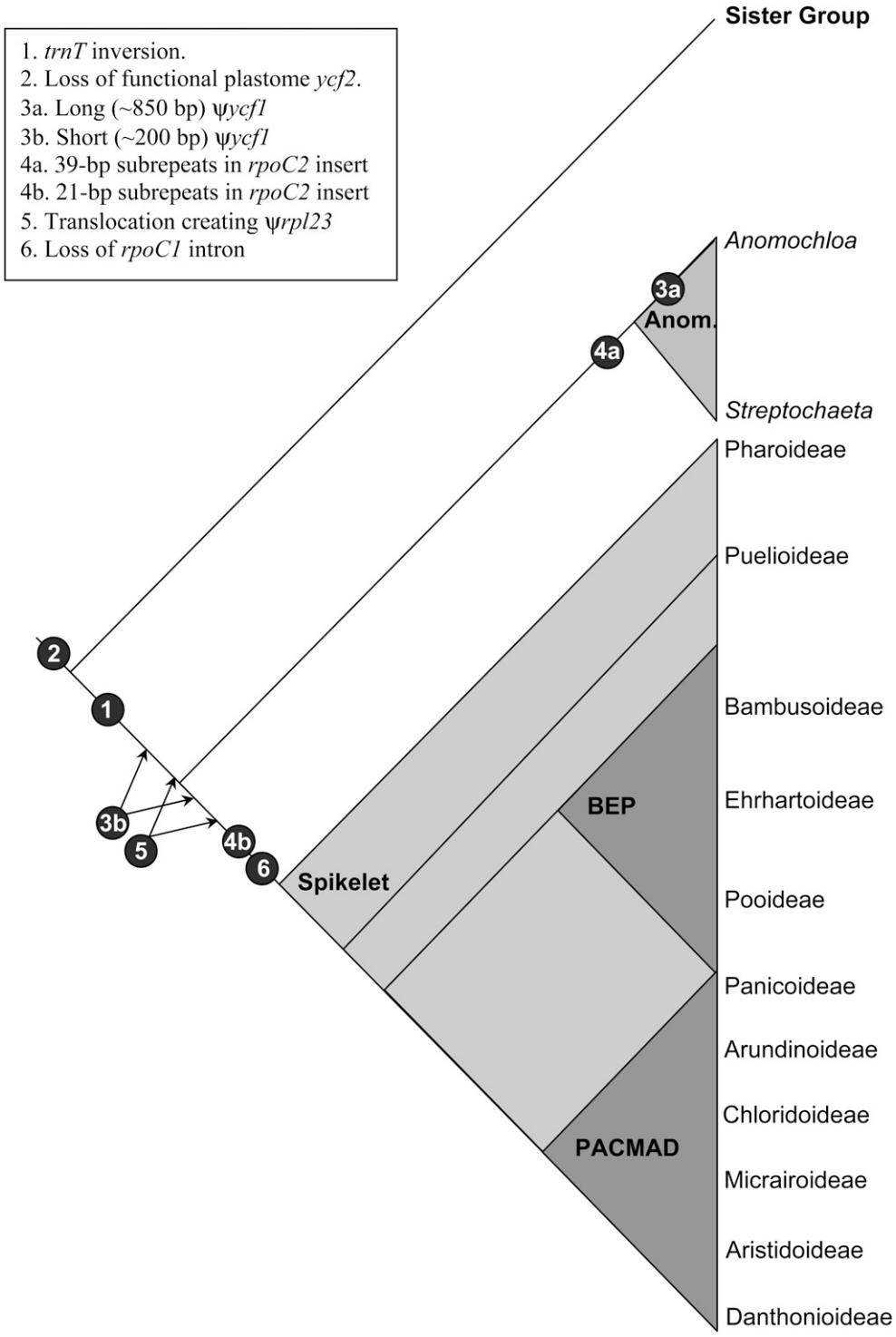

Fig. 2. Placements of plastome characters mapped onto a summary cladogram for Poaceae. The tree topology reflects the current understanding of the phylogeny of the family (GPWG, 2001; Duvall et al., 2007; Bouchenak-Khelladi et al., 2008). Three major subclades of Poaceae, the spikelet, BEP, and PACMAD clades are indicated as are all subfamilies. Because the precise relationships between Poaceae, Ecdeiocoleaceae, and Joinvilleaceae have not been resolved, the immediate outgroup to Poaceae is labeled only "Sister Group." Characters with uncertain character states are indicated with the most likely alternatives (e.g., 3a, 3b, and 4a, 4b). Characters of uncertain placements, largely due to incomplete sampling, are indicated at the most likely positions (arrows).

locus is thought to result from a nonreciprocal translocation of this gene from one copy of the inverted region; though at least two genera, *Festuca* and *Lolium*, show variability in the presence of *ψrpl23* in this IGS (Katayama and Ogihara, 1996). Katayama and Ogihara (1996) were unable to detect any evidence of *ψrpl23* in *Lolium perenne* (which we confirmed in

an examination of the recently released complete plastome sequence of this species—GenBank accession AM777385) or *Festuca rubra* L., though they were able to detect it in *F. arundinacea* Schreb. (which we also confirmed in its complete plastome sequence—FJ466687) and *F. ovina* L. using Southern hybridization techniques. Given the hypervariability of this region in grasses in general (Morton and Clegg, 1993; Clegg et al., 1994), it seems reasonable to suggest that the loss of *rpl23* in the LSC is homoplastic in crown taxa of Poaceae.

In the *rbcL-psaI* spacer of *A. marantoidea*, there are no sequences homologous to either ψ*rpl23* or ORF106, the latter of which is also found in the IGS of *Oryza* and other grasses (Maier et al., 1995). The absence of these sequences in the IGS of *A. marantoidea* is consistent with its substantial shortening, but the unique and unidentifiable central region of the IGS is more difficult to explain. The unique central portion of the *rbcL-psaI* spacer and the absence of ψ*rpl23* in the plastome of *A. marantoidea* contrast with the IGS found in the spikelet clade. The variability of this IGS in grass plastomes allows that the distinctive configuration found in *A. marantoidea* may be due to autapomorphies. However, given the failure to find ψ*rpl23* in a preliminary assembly of the IGS of *Streptochaeta angustifolium*, there is also the possibility that the translocation of *rpl23* happened after the divergence between Anomochlooideae and the spikelet clade and that the LSC copy of ψ*rpl23* was secondarily lost in some pooid species.

ycf2 ***and its loss in selected graminoid Poales***—In most land plants, *ycf2* translates to over 2000 amino acids, and on the basis of limited nucleic acid identity, it has been suggested that this putative protein may be a chloroplast-specific ATPase (Wolfe, 1994). Further, *ycf2* has been shown to be crucial to cell survival (Drescher et al., 2000). Some lineages of graminoids are characterized by the loss of much of the *ycf2* sequence from the plastome; it appears only as the pseudogene ψ*ycf2* (Katayama and Ogihara, 1996). Maier et al. (1995) suggested that the degradation of *ycf2* was secondary to the transfer of its function to the nuclear genome. Katayama and Ogihara (1996) found that among the graminoid Poales—exclusive of Ecdeiocoleaceae, which was not included in their study—the *ycf2* sequence has been retained in Anarthriaceae and Flagellariaceae, but lost in Centrolepidaceae, Joinvilleaceae, Restionaceae, and Poaceae. This pattern of distribution suggests that the transfer of *ycf2* to the nucleus predated the origin of graminoids where retention of a functional copy in the plastome is a symplesiomorphy for the entire group with sporadic homoplasious losses occurring in many lineages. Detailed data for graminoids are limited to the completely sequenced plastomes of other Poaceae where remnants persist as ψ*ycf2* just as they do in *A. marantoidea*.

ψ**ycf1** ***in*** **A. marantoidea**—Similar to *ycf2*, *ycf1* is a large open reading frame present in most land plants, with a large putative protein product (over 1900 amino acids). The function of *ycf1* is disuputed (see De Las Rivas et al., 2002) though the gene product was found to be essential to cell survival in tobacco (Drescher et al., 2000). *Anomochloa marantoidea* has a relatively long ψ*ycf1* locus with sequence similarity to *ycf1* in nongraminoid monocots. There is also a 99-bp ORF25 locus with homology to *ycf1* in the plastome of *A. marantoidea*, which is embedded within its larger ψ*ycf1*. ORF25 is explicitly annotated as a 78-bp open reading frame in *S. officinarum* (GenBank accession NC_006084). We can also report that ORF25 is identifiable in all banked plastomes of Poaceae on a BLASTN search, even though it is not always annotated. However, the sequences constituting the remainder of ψ*ycf1* in *A. marantoidea* otherwise fail to show homology with banked grass plastomes from Poaceae on BLASTN searches, so neither a functional *ycf1* as found in nongraminoid monocots nor ψ*ycf1* as found in *A. marantoidea* are to be found in any other completely sequenced grass plastome. At this point, ψ*ycf1* in *A. marantoidea* appears to represent a unique state, which is neither an intact *ycf1* as found in the plastomes of nongrass monocots nor the short remnant ORF25 in Poaceae. However, given the paucity of data for other graminoids, it is not possible to make definitive statements about the timing of the origin of the longer ψ*ycf1* as it relates to the radiation of the graminoids.

Conclusion—We have described five major molecular characters of evolutionary interest in the complete plastome of *A. marantoidea*. Microstructural characters such as small inversions and indels have been found in other graminoid plastomes, particularly in noncoding regions (Leseberg and Duvall, 2009). However, these are generally more informative within subfamilies and tribes of Poaceae, so their utility in *A. marantoidea* awaits the sequencing of plastomes from other more closely related species.

Evolutionary characters among graminoids, both those reported in this paper and those previously described, fall into three categories. First are those that are considered to be synapomorphic for Poaceae. *Anomochloa marantoidea* shares with other Poaceae the loss of the *ycf2* sequence, which variously remains only as nonfunctional remnants comprising a ψ*ycf2*. This can be considered a synapomorphic loss for Poaceae. There is another major character shared among all grasses surveyed—the *trnT* inversion—which is the only one of three LSC inversions that is exclusive to Poaceae (Doyle et al., 1992; Michelangeli et al., 2003). Finally, the fruit of all grasses is of a distinctive type called the caryopsis, in which the seed coat is fused to a thin pericarp (Judziewicz and Soderstrom, 1989).

Second, there are two characters for which a character state assignment for *A. marantoidea* is somewhat ambiguous because it exhibits a unique condition that differs from both the character state as found in other Poaceae and as it is found among those of its outgroups, respectively. The functional *ycf1*, present in other land plant plastomes, has been largely eradicated from the plastome of Poaceae leaving only the remnant ORF63, which is 192 bp in *Zea mays*. However, a distinctive and larger ψ*ycf1* locus equivalent in length to about 15% of this approximately 5700-bp locus is identifiable in *A. marantoidea*, which also includes the small subregion ORF63. Perhaps the plastome of *A. marantoidea* merely shows the loss of a functional *ycf1* as it occurred in the ancestor of Poaceae. Alternatively, perhaps the substantially larger ψ*ycf1* of *A. marantoidea* represents an autapomorphic event in this species.

The other such character concerns the *rpoC2* insertion sequence. Outgroups of Poaceae do not have this extra coding sequence. However the *rpoC2* locus of *A. marantoidea* has a uniquely short insert. Moreover, in Anomochlooideae it is comprised largely of 39-bp subrepeats, which contrast with the 21-bp subrepeats of the spikelet clade. What is unclear is how the multiple subrepeats were simultaneously interconverted from subrepeats of a different size by slipped-strand mispairing or any other mechanism, and why, given the likelihood that slipped-strand mispairing causes the observed size variation in the region, other Poaceae should have an insertion sequence nearly twice the size of that in *A. marantoidea*. Perhaps

A. marantoidea simply shares the condition of having an *rpoC2* insert with other Poaceae, but the possibility that some combination of parallel events account for the unique features of the insert in either *A. marantoidea* or Anomochlooideae cannot be entirely discounted.

Third, there are a number of characters that distinguish *A. marantoidea* from other Poaceae so that the inclusion of Anomochlooideae in Poaceae obscures the delimitation of the family. One is the *rpoC1* intron. No radiations of angiosperms that are comparable in size to the spikelet clade of Poaceae have been found to lack this intron, although in eudicots it is reported lost in 17 species of *Medicago* (Fabaceae), four species of *Passiflora* (Passifloraceae), at least two species of Goodeniaceae, two tribes of Aizoaceae, and all species of Cactoideae (Cactaceae) (Downie et al., 1996, 1998; Wallace and Cota, 1996; Thiede et al., 2007). Excluding the spikelet clade of grasses, all other monocots including *A. marantoidea* apparently have the *rpoC1* intron. Moreover, unpublished reports suggest additional relevant findings. While Pharoideae (Clark et al., 1995, citing Wallace et al., Iowa State University, unpublished data) and Puelioideae (Duvall et al., in press) lack the intron at least one other species of Anomochlooideae, *Streptochaeta angustifolia*, has an *rpoC1* intron (J. Leebens-Mack, University of Georgia, personal communication). The presence of the intron in both genera supports the putative monophyly of Anomochlooideae, which was unresolved since both genera are so divergent that their phylogenetic association has been suspected to be a long-branch attraction artifact (GPWG, 2001). At the same time, the *rpoC1* intron sets the subfamily apart from other Poaceae, and the loss of the *rpoC1* intron can be mapped to the ancestral lineage of the spikelet clade (Fig. 2).

Another distinctive character is the absence of *ψrpl23* from the IGS downstream of *rbcL*. Searches of a provisional assembly of this IGS in *Streptochaeta angustifolium* (provided by J. Leebens-Mack) failed to find a *ψrpl23* homologue. The IGS sequence in *A. marantoidea* shows homology only to the ends of the spacer in other grasses, while its center is of uncertain homology.

Floral structures are often definitive for families of angiosperms. The distinctive architectures of the spikelet and the floret have long been definitive for Poaceae. The uncertain homologies between the "spikelet equivalents" of Anomochlooideae and other Poaceae may be viewed as a trivial morphological exception. However, in the molecular context of the distinctive features of the plastome of *A. marantoidea*, the differences in floral structures between Anomochlooideae and the large spikelet clade may assume somewhat greater significance.

The complete plastome sequence of *A. marantoidea* presented here enhances our understanding of character state transitions among graminoid Poales. Certain features of this plastome are consistent with expectations for Poaceae based on existing data. Other features emphasize a substantial divergence between Anomochlooideae and the spikelet clade, obscure the boundary between Poaceae and other graminoids, and suggest a reevaluation of the morphology and phylogenetics of the subfamily. The issue of whether Anomochlooideae is better separated from the grasses or included among Poaceae is somewhat academic. However, if a goal of systematics is to maximize the number of synapomorphic morphological and molecular features in recognized taxa, then careful delimitations of those taxa are necessary. Under the current taxonomy, so-called "grass-specific" characters such as the absence of the *rpoC1* intron and *ψrpl23* are now seen to have likely arisen in the stem lineage of the spikelet clade and not in the ancestor of Poaceae (Fig. 2). In any case, complete plastome sequences from other grass taxa such as Streptochaeteae and Pharoideae, and other graminoid families such as Joinvilleaceae and Ecdeiocoleaceae, will prove instructive for our further understanding of this evolutionary radiation.

LITERATURE CITED

ALTSCHUL, S. F., W. GISH, W. MILLER, E. W. MYERS, AND D. J. LIPMAN. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215: 403–410.

ASANO, T., T. TSUDZUKI, S. TAKAHASHI, H. SHIMADA, AND K. KADOWAKI. 2004. Complete nucleotide sequence of the sugarcane (*Saccharum officinarum*) chloroplast genome: A comparative analysis of four monocot chloroplast genomes. *DNA Research* 11: 93–99.

BARKER, N. P., P. H. LINDER, AND E. H. HARLEY. 1998. Sequences of the grass-specific insert in the chloroplast *rpoC2* gene elucidate generic relationships of the Arundinoideae (Poaceae). *Systematic Botany* 23: 327–350.

BARKER, N. P., P. H. LINDER, C. M. MORTON, AND M. LYLE. 2003. The paraphyly of *Cortaderia* (Danthonioideae; Poaceae): Evidence from morphology and chloroplast and nuclear sequence data. *Annals of the Missouri Botanical Garden* 90: 1–24.

BORTIRI, E., D. COLEMAN-DERR, G. LAZO, O. ANDERSON, AND Y. GU. 2008. The complete chloroplast genome sequence of *Brachypodium distachyon*: Sequence comparison and phylogenetic analysis of eight grass plastomes. *BMC Research Notes* 1: 61–68.

BOUCHENAK-KHELLADI, Y., N. SALAMIN, V. SAVOLAINEN, F. FOREST, M. VAN DER BANK, M. W. CHASE, AND T. R. HODKINSON. 2008. Large multi-gene phylogenetic trees of the grasses (Poaceae): Progress towards complete tribal and generic level sampling. *Molecular Phylogenetics and Evolution* 47: 488–505.

BREMER, K. 2002. Gondwanan evolution of the grass alliance of families (Poales). *Evolution* 56: 1374–1387.

CHANG, C.-C., H.-C. LIN, I.-P. LIN, T.-Y. CHOW, H.-H. CHEN, W.-H. CHEN, C.-H. CHENG, ET AL. 2005. The chloroplast genome of *Phalaenopsis aphrodite* (Orchidaceae): Comparative analysis of evolutionary rate with that of grasses and its phylogenetic implications. *Molecular Biology and Evolution* 23: 279–291.

CLARK, L. G., AND E. J. JUDZIEWICZ. 1996. The grass subfamilies Anomochlooideae and Pharoideae (Poaceae). *Taxon* 45: 641–645.

CLARK, L. G., W. ZHANG, AND J. F. WENDEL. 1995. A phylogeny of the grass family (Poaceae) based on *ndhF* sequence data. *Systematic Botany* 20: 436–460.

CLEGG, M. T., B. S. GAUT, G. H. LEARN, AND B. R. MORTON. 1994. Rates and patterns of chloroplast DNA evolution. *Proceedings of the National Academy of Sciences, USA* 91: 6795–6801.

CUMMINGS, M., L. KING, AND E. A. KELLOGG. 1994. Slipped-strand mispairing in a plastid gene: *rpoC2* in grasses (Poaceae). *Molecular Biology and Evolution* 11: 1–8.

DE LAS RIVAS, J., J. J. LOZANO, AND A. R. ORTIZ. 2002. Comparative analysis of chloroplast genomes: Functional annotation, genome-based phylogeny, and deduced evolutionary patterns. *Genome Research* 12: 567–583.

DHINGRA, A., AND K. M. FOLTA. 2005. Amplification, sequencing and annotation of plastomes. *BMC Genomics* 6: 176–189.

DIEKMANN, K., T. R. HODKINSON, E. FRICKE, AND S. BARTH. 2008. An optimized chloroplast DNA extraction protocol for grasses (Poaceae) proves suitable for whole plastid genome sequencing and SNP detection. *PLoS ONE* 3(7): e2813; doi:10.1371/journal.pone.0002813.

DIEKMANN, K., T. R. HODKINSON, K. H. WOLFE, R. VAN DEN BEKEROM, P. J. DIX, AND S. BARTH. 2009. Complete chloroplast genome sequence of a major allogamous forage species, perennial ryegrass (*Lolium perenne* L.). *DNA Research* 16: 165–176.

DOWNIE, S. R., D. S. KATZ-DOWNIE, E. J. ROGERS, H. L. ZUJEWSKI, AND E. SMALL. 1998. Multiple independent losses of the plastid *rpoC1*

intron in *Medicago* (Fabaceae) as inferred from phylogenetic analyses of nuclear ribosomal DNA internal transcribed spacer sequences. *Canadian Journal of Botany* 76: 791–803.

DOWNIE, S. R., E. LLANAS, AND D. S. KATZ-DOWNIE. 1996. Multiple independent losses of the *rpoC1* intron in angiosperm chloroplast DNAs. *Systematic Botany* 21: 135–151.

DOYLE, J. J., J. I. DAVIS, R. I. SORENG, D. GARVIN, AND M. J. ANDERSON. 1992. Chloroplast DNA inversions and the origin of the grass family (Poaceae). *Proceedings of the National Academy of Sciences, USA* 89: 7722–7726.

DRESCHER, A., S. RUF, T. CALSA JR., H. CARRER, AND R. BOCK. 2000. The two largest chloroplast genome-encoded open reading frames of higher plants are essential genes. *Plant Journal* 22: 97–104.

DRUMMOND A. J., B. ASHTON, M. CHEUNG, J. HELED, M. KEARSE, R. MOIR, S. STONES-HAVAS, ET AL. 2007. Geneious v3.0 [computer program]. Website http://www.geneious.com/.

DUVALL, M. R., J. I. DAVIS, L. G. CLARK, J. D. NOLL, D. H. GOLDMAN, AND J. G. SÁNCHEZ-KEN. 2007. Phylogeny of the grasses (Poaceae) revisited. *Aliso* 23: 237–247.

DUVALL, M. R., C. H. LESEBERG, C. P. GRENNAN, AND L. M. MORRIS. In press. Molecular evolution and phylogenetics of complete chloroplast genomes in Poaceae. *In* O. Seberg, G. Petersen, A. S. Barfod, and J. I. Davis [eds.], Proceedings of the Fifth International Symposium on Grass Systematics and Evolution, Copenhagen, Denmark, 2008. Aarhus University Press, Copenhagen, Denmark.

DUVALL, M. R., J. NOLL, AND A. MINN. 2001. Phylogenetics of Paniceae (Poaceae). *American Journal of Botany* 88: 1988–1992.

GOREMYKIN, V. V., B. HOLLAND, K. I. HIRSCH-ERNST, AND F. H. HELLWIG. 2005. Analysis of *Acorus calamus* chloroplast genome and its phylogenetic implications. *Molecular Biology and Evolution* 22: 1813–1822.

GPWG [GRASS PHYLOGENY WORKING GROUP]. 2001. Phylogeny and subfamilial classification of the grasses (Poaceae). *Annals of the Missouri Botanical Garden* 88: 373–457.

HANSEN, D. R., S. G. DASTIDAR, Z. CAI, C. PENAFLOR, J. V. KUEHL, J. L. BOORE, AND ROBERT K. JANSEN. 2007. Phylogenetic and evolutionary implications of complete chloroplast genome sequences of four early-diverging angiosperms: *Buxus* (Buxaceae), *Chloranthus* (Chloranthaceae), *Dioscorea* (Dioscoreaceae), and *Illicium* (Schisandraceae). *Molecular Phylogenetics and Evolution* 45: 547–563.

IGLOI, G., A. MEINKE, I. DORY, AND H. KOSSEL. 1990. Pseudogenes and short repeated sequences in the rice chloroplast genome. *Molecular & General Genetics* 221: 379–394.

JUDZIEWICZ, E. J., AND T. R. SODERSTROM. 1989. Morphological, anatomical, and taxonomic studies in *Anomochloa* and *Streptochaeta* (Poaceae: Bambusoideae). *Smithsonian Contributions to Botany* 68: 1–52.

KATAYAMA, H., AND Y. OGIHARA. 1993. Structural alterations of the chloroplast genome found in grasses are not common in monocots. *Current Genetics* 23: 160–165.

KATAYAMA, H., AND Y. OGIHARA. 1996. Phylogenetic affinities of the grasses to other monocots as revealed by molecular analysis of chloroplast DNA. *Current Genetics* 29: 572–581.

LESEBERG, C. H., AND M. R. DUVALL. 2009. The complete chloroplast genome of *Coix lacryma-jobi* and a comparative molecular evolutionary analysis of plastomes in cereals. *Journal of Molecular Evolution* 69: 311–318.

MAIER, R., K. NECKERMANN, G. IGLOI, AND H. KOSSEL. 1995. Complete sequence of the maize chloroplast genome: Gene content, hotspots of divergence and fine tuning of genetic information by transcript editing. *Journal of Molecular Biology* 251: 614–628.

MARCHANT, A. D., AND B. G. BRIGGS. 2007. Ecdeiocoleaceae and Joinvilleaceae, sisters of Poaceae (Poales): evidence from *rbc*L and *mat*K data. *Telopea* 11: 437–450.

MARDANOV, A. V., N. V. RAVIN, B. B. KUZNETSOV, T. H. SAMIGULLIN, A. S. ANTONOV, T. V. KOLGANOVA, AND K. G. SKYABIN. 2008. Complete sequence of the duckweed (*Lemna minor*) chloroplast genome: Structural organization and phylogenetic relationships to other angiosperms. *Journal of Molecular Evolution* 66: 555–564.

MATHEWS, S., R. TSAI, AND E. KELLOGG. 2000. Phylogenetic structure in the grass family (Poaceae): Evidence from the nuclear gene phytochrome B. *American Journal of Botany* 87: 96–107.

MICHELANGELI, F., J. I. DAVIS, AND D. STEVENSON. 2003. Phylogenetic relationships among Poaceae and related families as inferred from morphology, inversions in the plastid genome, and sequence data from the mitochondrial and plastid genomes. *American Journal of Botany* 90: 93–106.

MORRONE, O., M. A. SCATAGLINI, AND F. O. XULOAGA. 2007. *Cyphonanthus*, a new genus segregated from *Panicum* (Poaceae: Panicoideae: Paniceae) based on morphological, anatomical and molecular data. *Taxon* 56: 521–532.

MORTON, B. R., AND M. T. CLEGG. 1993. A chloroplast DNA mutational hotspot and gene conversion in a noncoding region near *rbcL* in the grass family (Poaceae). *Current Genetics* 24: 357–365.

OGIHARA, Y., K. ISONO, T. KOJIMA, A. ENDO, M. HANAOKA, T. SHIINA, T. TERACHI, ET AL. 2002. Structural features of a wheat plastome as revealed by complete sequencing of chloroplast DNA. *Molecular Genetics and Genomics* 266: 740–746.

SASKI, C., S. LEE, S. FJELLHEIM, C. GUDA, R. JANSEN, J. TOMKINS, O. ROGNLI, ET AL. 2007. Complete chloroplast genome sequences of *Hordeum vulgare*, *Sorghum bicolor* and *Agrostis stolonifera*, and comparative analyses with other grass genomes. *Theoretical and Applied Genetics* 115: 571–590.

SHAHID-MASOOD, M., T. NISHIKAWA, S. FUKUOKA, P. NJENGA, T. TSUDZUKI, AND K. KADOWAKI. 2004. The complete nucleotide sequence of wild rice (*Oryza nivara*) chloroplast genome: First genome wide comparative sequence analysis of wild and cultivated rice. *Gene* 340: 133–139.

SHIMADA, H., M. FUKUTA, M. ISHAKAWA, AND M. SUGIURA. 1990. Rice chloroplast RNA polymerase genes: The absence of an intron in ropC1 and the presence of an extra sequence in rpoC2. *Molecular & General Genetics* 221: 395–402.

SHIMADA, H., AND M. SUGIURA. 1989. Pseudogenes and short repeated sequences in the rice chloroplast genome. *Current Genetics* 16: 293–301.

TANG, J., H. XIA, M. CAO, X. ZHANG, W. ZENG, S. HU, W. TONG, ET AL. 2004. A comparison of rice chloroplast genomes. *Plant Physiology* 135: 412–420.

THIEDE, J., S. A. SCHMIDT, AND B. RUDOLPH. 2007. Phylogenetic implication of the chloroplast *rpo*C1 intron loss in the Aizoaceae (Caryophyllales). *Biochemical Systematics and Ecology* 35: 372–380.

THOMAS, W. W., A. M. V. DE CARVALHO, A. M. A. AMORIM, J. GARRISON, AND A. L. ARBELÁEZ. 1998. Plant endemism in two forests in southern Bahia, Brazil. *Biodiversity and Conservation* 7: 311–322.

THOMPSON, J. D., D. G. HIGGINS, AND T. J. GIBSON. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* 22: 4673–4680.

WALLACE, R. S., AND J. H. COTA. 1996. An intron loss in the chloroplast gene *rpoC1* supports a monophyletic origin for the subfamily Cactoideae of the Cactaceae. *Current Genetics* 29: 275–281.

WOLFE, K. H. 1994. Similarity between putative ATP binding sites in land plant plastid ORF2280 proteins and the FtsH/CDC48 family of ATPases. *Current Genetics* 25: 379–383.

WU, F.-H., D.-P. KAN, S.-B. LEE, H. DANIELL, Y.-W. LEE, C.-C. LIN, AND C.-S. LIN. 2009. Complete nucleotide sequence of *Dendrocalamus latiflorus* and *Bambusa oldhamii* chloroplast genomes. *Tree Physiology* 29: 847–856.

WYMAN, S., J. BOORE, AND R. JANSEN. 2004. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20: 3252–3255.