

# The Complete Chloroplast Genome of *Coix lacryma-jobi* and a Comparative Molecular Evolutionary Analysis of Plastomes in Cereals

Charles H. Leseberg · Melvin R. Duvall

Received: 26 May 2009 / Accepted: 12 August 2009 / Published online: 24 September 2009  
© Springer Science+Business Media, LLC 2009

**Abstract** Graminoid molecular evolution was investigated by chloroplast genome (plastome) scale analyses. A complete plastome from *Coix lacryma-jobi* (Poaceae) and a draft plastome from *Joinvillea plicata* (Joinvilleaceae) were sequenced and analyzed. The draft plastome included conserved protein-coding loci routinely analyzed in previous studies plus one additional locus of demonstrated phylogenetic utility. The methodological approach was to directly sequence overlapping amplicons from known plastome regions. Over 100 pairs of amplification and sequencing primers were designed and positioned to flank overlapping 1,200-base pair fragments around the entire plastome. Newly determined sequences were analyzed with published plastomes from representatives of Panicoideae, Ehrhartoideae, and Pooideae. Considerable variation was found for studies within the family and even within Andropogoneae. Readily interpreted mutation patterns were observed, such as small inversions in hairpin-loop regions and indels, which were common in intergenic spacers. Maximum or near-maximum bootstrap support was observed in all analyses resolving relationships between subfamilies. However, the addition of characters from noncoding regions increased the number of parsimony-informative characters and length-

ened short internal branches (Andropogoneae), better defining intergeneric relationships. Thus, characters in complete plastomes can be used over a wide scope of phylogenetic studies.

**Keywords** *Coix lacryma-jobi* · *Joinvillea plicata* · Graminoid Poales · Plastome · Chloroplast genome · Indel · Inversion

## Introduction

Variation in chloroplast genomes (plastomes) is a source of information for studies of widely different scope. At one extreme, the conservation of plastome structure has advanced our understanding of broad angiosperm relationships (APG II 2003; Chase et al. 1993; Duvall et al. 1993a; Goremykin et al. 2003, 2005; Moore et al. 2007; Graham et al. 2000; Saarela et al. 2007; Shaw et al. 2005, 2007 etc.). Plastomes also include fine-scale variation for the resolution of relationships in species-rich radiations, such as the graminoid Poales (Bremer 2002). Historically, molecular studies of Poaceae targeted a few specific regions—including conserved genes (*ndhF*—Clark et al. 1995; *rbcL*—Duvall and Morton 1996), variable regions (*rpoC2* insert—Barker et al. 1999; Duvall et al. 2001), or introns and other noncoding sequence, such as *rpl16* and *trnL-F* (Cialdella et al. 2007; Jakob and Blattner 2006; Kelchner and Clark 1997; Yang et al. 2007). These studies identified two major lineages of Poaceae referenced by acronyms composed of the initial letters of the included subfamilies: the “PACMAD” clade (panicoid, arundinoid, chloridoid, micrairoid, aristidoid, and danthonioid grasses—recently revised from PACCAD, Duvall et al. 2007) and the “BEP” clade (bambusoid, ehrhartoid, and poid grasses—Kellogg 2000). Although the subfamilial interrelationships

**Electronic supplementary material** The online version of this article (doi:10.1007/s00239-009-9275-9) contains supplementary material, which is available to authorized users.

C. H. Leseberg · M. R. Duvall (✉)  
Department of Biological Sciences, Northern Illinois University,  
DeKalb, IL 60115-2861, USA  
e-mail: mel-duvall@niu.edu

C. H. Leseberg  
e-mail: CHLeseberg@yahoo.com

are largely resolved, many intertribal and intergeneric phylogenetic issues remain.

Plastome sequences are available for nine genera (ten species) of Poaceae. In this paper we report a draft plastome for *Joinvillea plicata* and a complete sequence for *Coix lacryma-jobi*. *Coix lacryma-jobi* is an andropogonoid grass native to tropical Asia that is widely adventive and considered invasive (Mito and Uesugi 2004; Mosango et al. 2001; Shluker 2003; Townsend and Newell 2006; Villaseñor and Espinosa-García 2004). Deliberate introductions are likely since *C. lacryma-jobi* is sometimes cultivated (Arber 1965; Barrau 1965; Clayton and Renvoize 1986; Watson and Dallwitz 1992). Production of this grain has declined in Asia with the introduction of other cereals (Roder 2006). *Coix lacryma-jobi* is also used as an herbal remedy in China (Ruan et al. 2006). Extracts from *C. lacryma-jobi*, thought to enhance the effectiveness of chemotherapy in the treatment of cancer, were the first traditional Chinese herbal remedies to be approved for clinical trials in the U.S. (Normile and Yimin 2003).

The addition of a plastome from *C. lacryma-jobi* better balanced the sampling of existing plastomes between the PACMAD and BEP lineages. Higher-resolution investigations involving microstructural mutations could also be performed because plastome sequences from three other Andropogoneae are published. The sequences from *J. plicata* are useful as outgroup data for plastome-scale research. These results are analyzed to indicate the scope of research applications with plastomes.

## Methods

The general approach was to produce overlapping amplicons at known positions around the grass plastome for

direct sequencing (following Dhingra and Foltá 2005; Mardanov et al. 2008).

## Primer Design

The complete plastome sequences of *Saccharum officinarum* (Asano et al. 2004; Calsa et al. 2004, NC006084), *Zea mays* (Maier et al. 1995, NC001666), *Oryza sativa* (NC001320), and *Triticum aestivum* (Ogihara et al. 2002, NC002762) were aligned (Clustal X—Thompson et al. 1997). Primers were designed manually by choosing highly conserved regions among the four species with an approximate spacing of one kilobase (kb). Primers were positioned to have amplicons overlap by 100–200 base pairs (bp). When possible, priming sites were identical for all species, but preference was ultimately given to *Z. mays* and *S. officinarum*, which are more closely related to *C. lacryma-jobi* (Clayton and Renvoize 1986). Primers were ideally at least 25 bp in length, with a  $T_m$  of approximately 55°C, and without significant self-dimer formation ( $\Delta G < -10$ ). In regions with large indels, *S. officinarum* and *Z. mays* were used as references to maintain the 1 kb-spacing. The amplification of the IR used 40 of 54 IRB primers provided to us (Dhingra and Foltá 2005). The remaining 227 primers are listed (Tables 1, S1–S3).

## Taxa

Sequences from *C. lacryma-jobi* (voucher: *M. Duvall s. n.* 26 May 2006, DEK) and *J. plicata* (*L. Thien 84*, NO) were newly determined and analyzed with those of the four plastomes aligned during primer design plus six others: *Agrostis stolonifera* (Saski et al. 2007; GenBank accession NC\_008591), *Brachypodium distachyon* (Bortiri et al. 2008; EU325680), *Hordeum vulgare* (Saski et al. 2007;

**Table 1** Sequences of alternate graminoid-specific primers used in the LSC region of *J. plicata*

Primer	Begin	End	Sequence
109pRJ	9591	9566	ACAAAAGTTGTACCTGTAAACCAACC
113pFJ	13790	13817	CACCACTTAACATCTATGTCAGCTTTTC
124pFJ	25226	25249	CCCGTTTCTCCTCAACAAATAAAGGC
132pFJ	33764	33789	ATGCGGGTCTTAGCAAAAAAGGATTG
140pFJ	42101	42127	TAAAAACATTAACACTGAACGCCAGAC
142pRJ	42980	42954	GTGGATTGGTGGATTCTAATAGTCGG
165pFJ	65001	65027	GTTCTTGCTATGGGTAAACTATATG
167pFJ	67146	67168	AAGGAAGGAGGATTTTCAATGCGGG
171pFJ	70549	70572	GTATAACATCTGGTGTACCGTTGAC
179pFJ	79809	79833	TAACGGCTGCATTGTGCATCATAGCG
181pRJ	81131	81108	AAGCTCCATTGCGGGGTTTCAGACC
181pFJ	81917	81941	AGTCTACCCTTTCTGATCCATTCGG
183pRJ	83185	83162	ATAACAGATCGTATGGTAGGTCGC

Beginning and ending coordinates are given with reference to the *S. officinarum* reference plastome (GenBank accession: NC\_006084)

NC\_008590), *Lolium perenne* (Diekmann et al. 2008; AM777385), *Oryza nivara* (Shahid-Masood et al. 2004; Tang et al. 2004; NC\_005973), and *Sorghum bicolor* (Saski et al. 2007; NC\_008602).

### DNA Preparation

Total DNA was isolated from fresh leaf tissue of *C. lacryma-jobi* using the Qiagen DNeasy Kit (Qiagen Inc., Valencia, CA). The extract of *J. plicata* was obtained for a previous study. The amount of DNA used for PCR reactions was optimized by testing different dilutions with primer pair IRB19, which amplified the highly conserved 16 S rrn locus (Dhingra and Folta 2005).

### PCR Conditions and Sequencing Strategy

Touchdown PCR was performed following “Round I” conditions of Dhingra and Folta (2005) with elongation times extended to 40–150 s to insure complete amplification. All amplicons were generated using *Pfu* Turbo DNA polymerase (Stratagene Inc., Carlsbad, CA), which reduced PCR enzyme generated errors.

Amplicons were prepared for capillary sequencing using Wizard SV PCR clean-up kits (Promega Corp., Madison, WI). Automated capillary sequencing was performed by MacroGen Inc. (Seoul, South Korea) on each amplicon in both directions giving 2× coverage within the amplicon overlap and the overlapping two adjacent amplicons. Only sequences with MacroGen QV scores (equivalent to Phred scores) of at least 20 were retained. Sequences were manually inspected in Sequencher vers. 4.14 (GenCodes Corp., Ann Arbor, MI). Contigs were formed from overlapping sequences and base-calling discrepancies were resolved by inspection of the electropherograms.

Fragments that initially failed to amplify were sequentially treated as follows. First, the forward primer of the upstream successful amplicon was used with the unsuccessful amplicon’s reverse primer and the downstream successful amplicon’s reverse primer with the unsuccessful forward primer. Second, *Coix* specific primers were designed in the sequenced regions flanking the unsuccessful amplicon and used to amplify the missing sequence. Third, the amplification was repeated with the previous primer combinations using FidelityTaq (USB Corp., Cleveland, OH) following the manufacturer’s standard protocol except that the durations of the extension steps were increased to 3.0 min.

### Phylogenetic Analysis

A set of 61 protein coding genes was analyzed following the precedent in other studies (Cai et al. 2006; Goremykin et al. 2003; Leebens-Mack et al. 2005; Moore et al. 2007).

*ndhF*, which was avoided in other whole plastome studies because of its lack of conservation among angiosperms, was used here for its relatively greater conservation and phylogenetic utility in grasses (Clark et al. 1995). This set of loci was extracted from the completed plastomes and from the draft plastome of *J. plicata*, aligned, and concatenated (Gene Inspector vers. 1.6, Textco BioSoftware, Inc., West Lebanon, NH). Gaps introduced by the alignment were excluded. The maximum parsimony (MP) method, implemented in PAUP\* (vers. 4.0b10 Swofford 2003), was used for branch and bound analyses. Non-parametric MP bootstrap analysis was also performed with 1000 pseudoreplicates (Felsenstein 1985).

### Small Inversion Analysis

Kim and Lee (2005) identified hairpin structures in the plastomes of grasses and other angiosperms. These structures were characterized by inverted repeats in the stem-forming regions ranging from 11 to 24 bp and had high stability as determined by estimates of free energy. The stem-loop structures were found to be associated with inversions ranging from 5 to 50 bp. Inversions of the loop regions took place when recombination occurred between the flanking stem-forming regions. Kim and Lee (2005) found stem-loop structures to be widely conserved in angiosperms including species of *Oryza* and *Triticum*. Homologous regions were thus readily identified by position in the grass plastome and by sequence similarity.

Inversion analysis was conducted on the 11 complete plastomes. We selected six regions, in which the hairpin structure was preserved and for which two or more taxa showed the inverted loop. For each inversion, the loop orientation was determined and arbitrarily coded as “0” (uninverted) or “1” (inverted) for 11 Poaceae. In four cases, because of incomplete sequencing in the outgroup, inversions were scored as missing. *Saccharum officinarum* was used as a reference to obtain coordinates. These binary coded data were included in a combined branch and bound MP bootstrap analysis.

### Indel Analysis

Indel mutations were identified in 11 aligned complete plastomes excluding *J. plicata*. Each potential indel was evaluated on the basis of three criteria. (1) The indel could be unambiguously attributed to a slipped-strand mispairing event, identified by the presence of a perfect or near-perfect repeated sequence. Indels were accepted when some copies of the repeat were imperfect due to subsequent mutations. (2) The indel was two or more nucleotides long, readily distinguished from putative single-base sequencing artifacts. (3) Both character states (insertion and deletion) of the indel

were shared by at least two of the sequences. Indel states were coded as “0” (deletion) or “1” (insertion) and concatenated onto the data matrix for MP analysis. If an indel region was lost because of a larger deletion, or not sequenced, e.g., in *J. plicata*, it was scored as missing (Table S4).

## Results

### Size, Gene Content, and Organization of the *Coix lacryma-jobi* Plastome

The complete *C. lacryma-jobi* plastome was 145,745 bp long (GenBank accession FJ261955). The two IRs were 22,715 bp, a small single copy region was 12,523 bp, and a large single copy region was 82,792 bp. The positions of the boundaries of the IR and the gene content and order were identical to those in *S. officinarum*.

An estimated 70% of the *J. plicata* plastome was completed from which 50 of the 62 protein coding genes were banked as either partial or complete cds (GenBank accessions FJ486219–FJ486269, Leseberg (2009)). Note that three other sequences, *rbcL* (L01471), *ndhF* (U21973), and *rpoC2* (AF001864) were already available. Of the total aligned sequence data, 11.6% were missing for *J. plicata*. While the coding regions were readily aligned with those of grasses, alignments of noncoding regions were more ambiguous and so only the conserved regions of *J. plicata* were reported and analyzed here.

### Indel and Small Inversion Analysis

Seventy-eight indels meeting the specified criteria were found among the 11 Poaceae ranging in size from 2 to

26 bp (mean length, 5 bp; see Supplementary Table S4). Indels were found primarily in intergenic sequences (IGS; 57 indels), 14 were found in introns, and 7 were in coding regions (Table 2).

Six loop regions could be identified on the basis of position, flanking stem sequences, and homologous loop regions in 11 Poaceae except for two instances—the inversion at coordinate position 63 275 in *S. officinarum* was apparently lost as part of a larger deletion in both *Triticum* and *Hordeum* (Table 3).

### Phylogenetic Analysis

Four branch and bound maximum parsimony (MP) analyses were performed with a series of increasingly inclusive data sets. The datasets included [1] a set of 61 protein-coding genes, [2] the *ndhF* coding sequence, [3] 78 indels, and [4] six inversions. MP analyses were performed on subsets [1], [1,2], [1–3], and [1–4] and the results are presented (Table 4). The tree and bootstrap values presented in Fig. 1 were produced from the total combined datasets [1–4]. This analysis resulted in a single MP tree of 8169 steps with a consistency index (excluding uninformative characters) of 0.7639 and a retention index of 0.8513 (Table 4). The ranges of total and informative characters were 42478–44627 and 2681–2985, respectively. The ranges of total summed internal branch lengths for the entire tree and the Andropogoneae were 3557–3972 and 682–776, respectively. All nodes were supported with 100% bootstrap values with the exception of the BEP clade. The support for this internal node increased with the inclusion of more datasets with values of 90, 96, 96, and 97 for datasets [1], [1,2], [1–3], and [1–4], respectively (trees not shown). Within the Andropogoneae, *S. officinarum* and *S. bicolor* were sister to *C. lacryma-jobi*, and this clade of three species was sister to *Z. mays*.

**Table 2** Locations of 78 Indels identified among 11 Poaceae

Location	Number of Indels
<i>matK</i>	2
<i>trnK</i> intron	2
<i>rps16</i> intron	3
<i>trnG</i> intron	2
<i>rpoB</i>	1
<i>rpoC2</i>	2
<i>atpF</i> intron	1
<i>ycf3</i> intron	3
<i>trnV</i> intron	1
<i>petB</i> intron	1
<i>orf251</i>	1
<i>orf69</i>	1
<i>ndhB</i> intron	1
IGS	57

IGS Intergenic spacer

## Discussion

Recent whole plastome phylogenetic studies have been largely restricted to analyses of point mutations in 61 conserved protein-coding genes (Cai et al. 2006; Chang et al. 2006; Goremykin et al. 2003; Leebens-Mack et al. 2005; Moore et al. 2007; Saski et al. 2007; Wu et al. 2009). Here we show in the graminoid radiation the effects of combining this data partition with the *ndhF* coding sequence, hairpin-loop inversions, and indel characters found in coding and noncoding regions. Since maximum likelihood models for indel and inversion data are incompletely developed, the analyses here were confined to MP analyses. The combination of all these types of data increased characters and support at both deep and shallow

**Table 3** Character coding and locations for small inversions in regions originally identified by Kim and Lee (2005) as stable hairpin-loops

Coordinate positions are from the *Saccharum officinarum* reference plastome (GenBank accession: NC\_006084) (–) Indicates this region did not have the inversion sequence due to deletions (?) This region of *Joinvillea* was not available due to incomplete sequencing

Taxa	Position 18867	Position 63275	Position 66394	Position 73165	Position 110262	Position 110755
<i>Lolium</i>	0	1	1	0	0	0
<i>Brachypodium</i>	0	1	1	0	1	0
<i>Agrostis</i>	0	0	1	0	0	0
<i>Triticum</i>	0	–	1	0	1	0
<i>Hordeum</i>	0	–	0	0	1	0
<i>Oryza_nivara</i>	0	0	0	1	0	0
<i>Oryza_sativa</i>	0	0	0	1	1	0
<i>Saccharum</i>	1	1	0	0	1	1
<i>Sorghum</i>	1	1	0	0	0	1
<i>Coix</i>	1	0	0	0	0	1
<i>Zea</i>	0	0	0	0	0	1
<i>Joinvillea</i>	?	0	?	0	?	?

**Table 4** Cumulative statistics for maximum parsimony analyses with successively increasing character data

Dataset used	Total number of characters	Number of parsimony—informative characters	Tree length	CI excluding uninformative characters	RI	Sum of total internal branch lengths	Sum of internal branch lengths, Andropogoneae
61 <sup>a</sup>	42478	2681	7440	0.7658	0.8527	3557	682
61/ <i>ndhF</i>	44544	2902	8059	0.7642	0.8505	3881	730
61/ <i>ndhF</i> /Indel <sup>b</sup>	44621	2979	8158	0.7645	0.8517	3964	773
61/ <i>ndhF</i> /Indel/ Inversions <sup>c</sup>	44627	2985	8169	0.7639	0.8513	3972	776

<sup>a</sup> Following Cai et al. (2006); Goremykin et al. (2003); Leebens-Mack et al. (2005); Moore et al. (2007)

<sup>b</sup> This study

<sup>c</sup> Described by Kim and Lee (2005)

CI Consistency Index, RI Retention Index

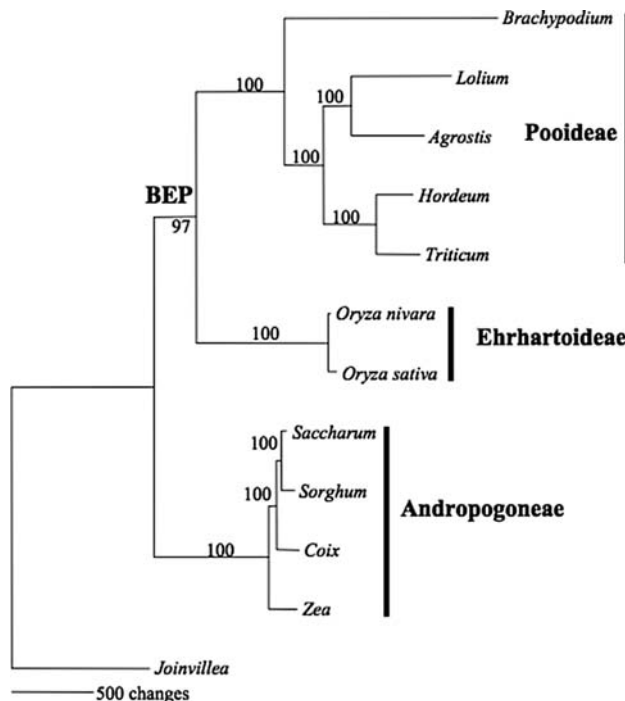
nodes in parsimony analyses. This is especially relevant for intensively sampled taxa where the number of informative substitutions in the conserved genes becomes limiting.

The phylogenetic analyses performed here were largely in agreement with each other and with recent multi-gene analyses (e.g., Bouchenak-Khelladi et al. 2008; Duvall et al. 2007). With the exception of the BEP node in the MP analyses, the levels of support for all nodes in the MP analysis were 100%. The BEP clade has historically been rather weakly supported since the study that first identified it (Clark et al. 1995). In this study, successively adding data from different partitions increased the MP bootstrap value for the BEP clade from 90 to 97% when all data were included (Fig. 1). Character support for the BEP node, as approximated by branch length (ACCTRAN optimization), increased correspondingly with values of 481, 525, 538, and 540 steps for data partitions [1], [1,2], [1–3], and [1–4], respectively. Note that maximum support for the BEP clade in MP trees was obtained when a representative bambusoid was included in the analyses (*Cryptochloa strictiflora*—

Grennan and Duvall, unpublished), so that improved sampling retrieves this clade with stronger support.

*Coix lacryma-jobi* was sister to a clade comprising *S. officinarum* and *S. bicolor* in our analysis consistent with nuclear gene studies (*tb1*—Lukens and Doebley 2001; *PHYB*, *GBSSI*—Mathews et al. 2002). The position of *C. lacryma-jobi* in our MP tree is in agreement with a broadly delimited Andropogoneae (GPWG 2001; Spangler et al. 1999), but contradicts other studies that classified *C. lacryma-jobi* in a putative sister tribe, Maydeae, with *Z. mays* (e.g., Kellogg and Birchler 1993; Watson and Dallwitz 1992). Here, Maydeae is paraphyletic with the remaining Andropogoneae (Fig. 1).

Results for the position of *C. lacryma-jobi* exemplify how data from complete plastomes can be used in comparisons of relatively closely related taxa such as the four Andropogoneae in this study. Maximum parsimony bootstrap values for all nodes in Andropogoneae were at a maximum in the 61-locus tree (not shown). However, the number of parsimony informative characters and the



**Fig. 1** Branch and bound maximum parsimony phylogram. Analysis included 61 conserved protein-coding sequences, the *ndhF* coding sequence, 6 inversions, and 78 indels from 12 graminoid plastomes. The BEP clade (subfamilies Pooideae and Ehrhartoideae) and the Andropogoneae, are indicated. Branches are proportional to the number of changes along the branch. Numbers along the branches are MP bootstrap values

lengths of internal branches were increased by including data from other character partitions. For example, character support for the (*S. officinarum*, *S. bicolor*) clade increased with values of 22, 24, 30, and 31 steps for data partitions [1], [1,2], [1–3], and [1–4], respectively. Similarly, character support for the position of *C. lacryma-jobi* increased with values of 39, 45, 49, and 50 steps for the same series of data partitions. The number of characters supporting the monophyly of Andropogoneae likewise increased from 621, 661, 694, to 695. The use of the other data partitions increased parsimony informative characters by 11.3% across Andropogoneae. Internal branch lengths in Andropogoneae were increased by 7.0, 5.9, and 0.39%, respectively (Table 4). Given the short branches in this clade in the 61-locus tree (mean value, 30.5), the addition of further Andropogoneae is predicted to reduce support and/or decrease phylogenetic resolution when only conserved protein coding loci are analyzed, an apparent limit to plastome-scale phylogenetic analyses as they have been typically conducted. However, the abundance of little-used characters in complete plastomes can add significantly to the information supporting the relationships of closely related genera.

As more plastomes become available, characters that are currently noninformative autapomorphies are expected to

become informative in the same way that the addition of the *C. lacryma-jobi* plastome changed autapomorphies in *Z. mays* into synapomorphies for subsets of Andropogoneae. Indel character 75 is one example (Table S4). Without the *C. lacryma-jobi* sequence the deletion of one four nucleotide direct repeat in *Z. mays* is autapomorphic, and so would have been rejected by our indel criterion number 3, as uninformative. However, since the *C. lacryma-jobi* sequence also has the deletion, this condition can be interpreted to be symplesiomorphic for Andropogoneae in the context of our MP tree, with a subsequent synapomorphic insertion in the (*S. officinarum*, *S. bicolor*) clade. Indels 5, 28, 30, and 51 become similarly informative only after the addition of the *C. lacryma-jobi* sequence to the other sampled Poaceae.

## Conclusion

In this study a novel set of primers is specified, which were designed to create overlapping amplicons in the grass plastome and tested in two graminoids (Tables S1–S3 and LSC-specific primers for *J. plicata* in Table 1). This set of primers has already been successfully used for other grasses including species of *Anomochloa*, *Chasmanthium*, *Cryptochloa*, *Microcalamus*, *Pharus*, *Puelia*, and another outgroup, *Ecdeiocolea* (Ecdeiocoleaceae) (Duvall et al. unpublished). Now that this significant investment of obtaining graminoid-specific primers is accomplished, future sequencing of graminoid plastomes will be facilitated.

*Joinvillea plicata* was chosen as an outgroup for draft plastome sequencing because of the close relationship between Joinvilleaceae and Poaceae (Campbell and Kellogg 1987; Doyle et al. 1992; Duvall et al. 1993a, b; Linder and Rudall 1993; Kellogg 2000; Marchant and Briggs 2007; Michelangeli et al. 2003). This set of conserved protein coding gene sequences is a useful resource for outgroup rooting of Poaceae in phylogenetic studies.

The 78 indels (Table 2) and six inversions (Table 3) in this study were identifiable only on a whole plastome level and 91% were located in non-coding regions. The indel dataset proved quite robust with a total of 75 out of the 78 capable of being scored unambiguously for all 11 species (Table S4). The indel and inversion characters also had low homoplasy. The portion of the data matrix with sequences from 62 loci had a consistency index (excluding uninformative characters) of 0.6441 while that for the 78 indels and six inversions (all characters informative) was 0.8442.

The use of the substitutions in the 61 conserved gene dataset clearly increases phylogenetic information compared to single and multi-gene analyses of the past. However, it is shortsighted to rely solely on these characters in studies of the graminoid radiation when substantial

molecular evolution occurs in non-coding regions identifiable from whole plastome sequences. Other data, such as entire intergenic and intron sequences as well as pseudogenes may also be exploited for potential divergence among closely related species in species-rich evolutionary radiations. The complete plastome can provide considerable character state data to support key nodes in a wide range of evolutionary studies.

**Acknowledgments** We thank Amit Dhingra (University of Florida, Gainesville) for aliquots of primers from the inverted repeat region. We also thank Christie Eissler for assistance with primer design and contig assembly and Michelle Christian for technical assistance. This work was supported by the Plant Molecular Biology Center and by a Venture Grant from the Northern Illinois University Foundation to MRD and by a research award from the Department of Biological Sciences, Northern Illinois University, to CHL.

## References

- APG (Angiosperm phylogeny group) II (2003) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants. *Bot J Linn Soc* 141:399–436
- Arber A (1965) The Gramineae, a study of cereal, bamboo and grass (Reprint by Weinheimer JC). 184. Wheldon and Wesley, Ltd. Stechert-Hafner Service Agency, Inc. Codicote, Herts. New York, pp 480
- Asano T, Tsudzuki T, Takahashi S, Shimada H, Kadowaki K (2004) Complete nucleotide sequence of the sugarcane (*Saccharum officinarum*) chloroplast genome: a comparative analysis of four monocot chloroplast genomes. *DNA Res* 11:93–99
- Barker N, Linder P, Harley E (1999) Sequences of the grass specific insert in the chloroplast *rpoC2* gene elucidate generic relationships of the Arundinoideae (Poaceae). *Syst Bot* 23:327–350
- Barrau J (1965) Witnesses of the past: notes on some food plants of Oceania. *Ethnology* 4(3):282–294
- Bortiri E, Coleman-Derr D, Lazo GR, Anderson OD, Gu YQ (2008) The complete chloroplast genome sequence of *Brachypodium distachyon*: sequence comparison and phylogenetic analysis of eight grass plastomes. *BMC Res Notes* 1:61
- Bouchenak-Khelladi Y, Salamin N, Savolainen V, Forest F, van der Bank M, Chase M, Hodkinson T (2008) Large multi-gene phylogenetic trees of the grasses (Poaceae): progress towards complete tribal and generic level sampling. *Mol Phylogenet Evol* 47:488–505
- Bremer K (2002) Gondwanan evolution of the grass alliance of families (Poales). *Evolution* 56:1374–1387
- Cai Z, Penaflor C, Kuehl J, Leebens-Mack J, Carlson J, de Pamphilis C, Boore J, Jansen R (2006) Complete chloroplast genome sequences of *Drimys*, *Liriodendron*, and *Piper*: implications for the phylogeny of magnoliids. *BMC Evol Biol* 6:77
- Calsa T, Carraro D, Benatti M, Barbosa A, Kitajima J, Carrer H (2004) Structural features and transcript-editing analysis of sugarcane (*Saccharum officinarum* L.) chloroplast genome. *Curr Genet* 46:366–373
- Campbell CS, Kellogg EA (1987) Sister group relationships of the Poaceae. In: Soderstrom TR, Hilu KW, Campbell CS, Barkworth MA (eds) Grass systematics, evolution. Smithsonian Institution Press, Washington, DC, pp 217–224
- Chang C-C, Lin H-C, Lin I-P, Chow T-Y, Chen H-H, Chen W-H, Cheng C-H, Lin C-Y, Liu S-M, Chang C-C, Chaw S-M (2006) The chloroplast genome of *Phalaenopsis aphrodite* (Orchidaceae): comparative analysis of evolutionary rate with that of grasses and its phylogenetic implications. *Mol Biol Evol* 23:279–291
- Chase MH, Soltis DE, Olmstead RG (1993) Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbcL*. *Ann Mo Bot Gard* 80:528–580
- Cialdella AM, Giussani LM, Aagesen L, Zuloaga FO, Morrone O (2007) A phylogeny of Piptochaetium (Poaceae: Pooideae: Stipeae) and related genera based on a combined analysis including *trnL-F*, *rpl16*, and morphology. *Syst Bot* 32:545–559
- Clark LG, Zhang W, Wendel JF (1995) A phylogeny of the grass family (Poaceae) based on *ndhF* sequence data. *Syst Bot* 20:436–460
- Clayton WD, Renvoize SA (1986) Genera Graminum. *Kew Bull Add Ser* 13:1–389
- Dhingra A, Folta K (2005) ASAP: amplification, sequencing and annotation of plastomes. *BMC Genomics* 6:176
- Diekmann K, Hodkinson TR, Fricke E, Barth S (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
- Doyle JJ, Davis JI, Soreng RJ, Garvin D, Anderson MJ (1992) Chloroplast DNA inversions and the origin of the grass family (Poaceae). *PNAS USA* 89:7722–7726
- Duvall MR, Morton B (1996) Molecular phylogenetics of Poaceae: an expanded analysis of *rbcL* sequence data. *Mol Phylogenet Evol* 5:352–358
- Duvall M, Clegg M, Chase M, Clark WD, Kress WJ, Hills H, Eguarte L, Smith J, Gaut B, Zimmer E, Learn G (1993a) Phylogenetic hypotheses for the Monocotyledons constructed from *rbcL* sequence data. *Ann Mo Bot Gard* 80:607–619
- Duvall MR, Learn GE Jr, Eguarte LE, Clegg MT (1993b) Phylogenetic analysis of *rbcL* sequences identifies *Acorus calamus* as the primal extant monocotyledon. *PNAS USA* 90:4641–4644
- Duvall M, Noll J, Minn A (2001) Phylogenetics of Paniceae (Poaceae). *Am J Bot* 88:1988–1992
- Duvall MR, Davis JI, Clark LG, Noll JD, Goldman DH, Sánchez-Ken JG (2007) Phylogeny of the grasses (Poaceae) revisited. *Aliso* 23:237–247
- Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39:783–791
- Goremykin VV, Hirsch-Ernst KI, Wölfl S, Hellwig FH (2003) Analysis of the *Amborella trichopoda* chloroplast genome sequence suggests that *Amborella* is not a basal angiosperm. *Mol Biol Evol* 20:1499–1505
- Goremykin VV, Holland B, Hirsch-Ernst KI, Hellwig FH (2005) Analysis of *Acorus calamus* chloroplast genome and its phylogenetic implications. *Mol Biol Evol* 22(9):1813–1822
- Graham S, Reeves S, Burns A, Olmstead R (2000) Microstructural changes in noncoding chloroplast DNA: interpretation, evolution, and utility of indels and inversions in basal angiosperm phylogenetic inference. *Int J Plant Sci* 161(6 Suppl):S83–S96
- Grass Phylogeny Working Group (2001) Phylogeny and subfamilial classification of the grasses (Poaceae). *Ann Mo Bot Gard* 88:373–457
- Jakob SS, Blattner FR (2006) Chloroplast genealogy of *Hordeum* (Poaceae): Long-term persisting haplotypes, incomplete lineage sorting, regional extinction, and the consequences for phylogenetic inference. *Mol Biol Evol* 23:1602–1612
- Kelchner S, Clark LG (1997) Molecular evolution and phylogenetic utility of the chloroplast *rpl16* intron in *Chusquea* and the Bambusoideae (Poaceae). *Mol Phylogenet Evol* 8:385–397
- Kellogg EA (2000) The grasses: a case study in macroevolution. *Ann Rev Ecol Syst* 31:217–238
- Kellogg EA, Birchler JA (1993) Linking phylogeny and genetics Zea-mays as a tool for phylogenetic studies. *Syst Biol* 42:415–439

- Kim KJ, Lee HL (2005) Widespread occurrence of small inversions in the chloroplast genomes of land plants. *Mol Cells* 19:104–113
- Leebens-Mack J, Raubeson LA, Cui L, Kuehl JV, Fourcade MH, Chumley TW, Boore JL, Jansen RK, de Pamphilis CW (2005) Identifying the basal angiosperm node in chloroplast genome phylogenies: sampling one's way out of the Felsenstein zone. *Mol Biol Evol* 22:1948–1963
- Leseberg CH (2009) Establishing the protein-protein interaction network of tomato MADS-domain proteins, investigating novel tomato mutants, and further plant genomic investigations of evolution and phylogenetics. Ph.D. dissertation, Northern Illinois University, DeKalb, IL, USA
- Linder HP, Rudall PJ (1993) The megagametophyte in Anarthria (Anarthriaceae, Poales) and its implications for the phylogeny of Poales. *Am J Bot* 80:1455–1464
- Lukens LN, Doebley J (2001) Molecular evolution of the teosinte branched gene among maize and related grasses. *Mol Biol Evol* 18:627–638
- Maier R, Neckermann K, Igloi G, Kossel H (1995) Complete sequence of the maize chloroplast genome: gene content, hotspots of divergence and fine tuning of genetic information by transcript editing. *J Mol Biol* 251:614–628
- Marchant AD, Briggs BG (2007) Ecdiocoliaceae and Joinvilleaceae, sisters of Poaceae (Poales): evidence from *rbcl* and *matK* data. *Telopea* 11:437–450
- Mardanov AV, Ravin NV, Kuznetsov BB, Samigullin TH, Antonov AS, Kolganova TV, Skyabin KG (2008) Complete sequence of the duckweed (*Lemna minor*) chloroplast genome: structural organization and phylogenetic relationships to other angiosperms. *J Mol Evol* 66:555–564
- Mathews SY, Spangler RE, Mason-Gamer RJ, Kellogg EA (2002) Phylogeny of Andropogoneae inferred from phytochrome B, GBSSI, and *ndhF*. *Int J Plant Sci* 163(3):441–450
- Michelangeli FA, Davis JI, Stevenson DW (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. *Am J Bot* 90:93–106
- Mito T, Uesugi T (2004) Invasive alien species in Japan: the status quo and the new regulation for prevention of their adverse effects global environmental research. ©AIRIES 8:171–191
- Moore MJ, Bell CD, Soltis PS, Soltis DE (2007) Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. *Proc Natl Acad Sci* 104(49):19363–19368
- Mosango M, Maganyi O, Namaganda M (2001) A Floristic Study of Weed Species of Kampala (Uganda). *Syst Geogr Pl* 71:223–236
- Normile D, Yimin D (2003) The new face of traditional Chinese medicine. *Science* 299:188–190
- Ogihara Y, Isono K, Kojima T, Endo A, Hanaoka M, Shiina T, Terachi T, Utsugi S, Murata M, Mori N, Takumi S, Ikeo K, Gojobori T, Murai R, Murai K, Matsuoka Y, Ohnishi Y, Tajiri H, Tsunewaki K (2002) Structural features of a wheat plastome as revealed by complete sequencing of chloroplast DNA. *Mol Genet Genomics* 266:740–746
- Roder W (2006) Speculations on the importance of job's tears in past agricultural systems of Bhutan. *Econ Bot* 60:187–191
- Ruan WJ, Lai MD, Zhou JG (2006) Anticancer effects of Chinese herbal medicine, science or myth? *J Zhejiang Univ Sci B* 7:1006–1014
- Saarela JM, Rai H, Doyle J, Endress P, Mathews S, Marchant A, Briggs B, Graham S (2007) Hydatellaceae identified as a new branch near the base of the angiosperm phylogenetic tree. *Nature* 446:312–315
- Saski C, Lee S, Fjellheim S, Guda C, Jansen R, Tomkins J, Rognli O, Daniell H, Clarke J (2007) Complete chloroplast genome sequences of *Hordeum vulgare*, *Sorghum bicolor* and *Agrostis stolonifera*, and comparative analyses with other grass genomes. *Theor Appl Genet* 115:571–590
- Shahid-Masood M, Nishikawa T, Fukuoka S, Njenga P, Tsudzuki T, Kadowaki K (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
- Shaw J, Lickey EB, Beck JT, Farmer SB, Liu WS, Miller J, Siripun KC, Winder CT, Schilling EE, Small RL (2005) The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analyses. *Am J Bot* 92:142–166
- Shaw J, Lickey EB, Schilling EE, Small RL (2007) Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. *Am J Bot* 94:275–288
- Shluker AD (2003) State of Hawai'i aquatic invasive species management plan. The Nature Conservancy of Hawai'i, Department of Land and Natural Resources, Division of Aquatic Resources, Honolulu, p 140
- Spangler R, Zaitchik B, Russo E, Kellogg EA (1999) Andropogoneae evolution and generic limits in Sorghum (Poaceae) using *ndhF* sequences. *Syst Bot* 24:267–281
- Swofford DL (2003) PAUP\*: phylogenetic analysis using parsimony (\*and other methods). Version 4. Sinauer Associates, Sunderland, MA
- Tang J, Xia H, Cao M, Zhang X, Zeng W, Hu S, Tong W, Wang J, Wang J, Yu J, Yang H, Zhu L (2004) A comparison of rice chloroplast genomes. *Plant Physiol* 135:412–420
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL\_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25:4876–4882
- Townsend S, Newell D (2006) IABIN invasive species thematic network content building project implement, update and maintain an I3N IAS database in Jamaica, Technical Progress Report. IABIN Focal Point, Institute of Jamaica, Kingston, Jamaica
- Villaseñor J, Espinosa-García F (2004) The alien flowering plants of Mexico. *Divers Distrib* 10:113–123
- Watson L, Dallwitz MJ (1992) The grass genera of the world. CAB International, Wallingford, p 1024
- Wu F-H, Kan D-P, Lee S-B, Daniell H, Lee Y-W, Lin C-C, Lin N-S, Lin C-S (2009) Complete nucleotide sequence of *Dendrocalamus latiflorus* and *Bambusa oldhamii* chloroplast genomes. *Tree Physiol* 29:847–856
- Yang HQ, Peng S, Li DZ (2007) Generic delimitations of Schizotachyum and its allies (Gramineae : Bambusoideae) inferred from GBSSI and trnL-F sequence phylogenies. *Taxon* 56:45–54