MIGRATION OF ENDPOINTS OF TWO GENES RELATIVE TO BOUNDARIES BETWEEN REGIONS OF THE PLASTID GENOME IN THE GRASS FAMILY (POACEAE)1

JERROLD I. DAVIS2,4 AND ROBERT J. SORENG3

1L.H. Bailey Hortorium and Department of Plant Biology, Cornell University, 412 Mann Library, Ithaca, New York 14853-4301 USA; and 2Department of Botany and United States National Herbarium, National Museum of Natural History, Smithsonian Institution, Washington, D.C. 20013-7012 USA

Overlapping genes occur widely in microorganisms and in some plastid genomes, but unique properties are observed when such genes span the boundaries between single-copy and repeat regions. The termini of ndhH and ndhF, situated near opposite ends of the small single-copy region (SSC) in the plastid genomes of grasses (Poaceae), have migrated repeatedly into and out of the adjacent inverted-repeat regions (IR). The two genes are transcribed in the same direction, and the 5′ terminus of ndhH extends into the IR in some species, while the 3′ terminus of ndhF extends into the IR in others. When both genes extend into the IR, portions of the genes overlap and are encoded by the same nucleotide positions. Fine-scale mapping of the SSC-IR junctions across a sample of 92 grasses and outgroups, integrated into a phylogenetic analysis, indicates that the earliest grasses resembled the related taxa Joinvillea (Joinvilleaceae) and Ecdieicolea (Ecdieicoleaceae), with ca. 180 nucleotides of ndhH extending into the IR, and with ndhF confined to the SSC. This structure is maintained in early-diverging grass lineages and in most species of the BEP clade. In the PACMAD clade, ndhH lies completely or nearly completely within the SSC, and ca. 20 nucleotides of ndhF extend into the IR. The nucleotide substitution rate has increased in the PACMAD clade in the portion of ndhH that has migrated into the SSC.

Key words: gene overlap; inverted-repeat region; ndhF, ndhH; nucleotide substitution rate; PACMAD clade; phylogenetics; plastid genome; Poaceae; small single-copy region.

In most embryophytes, the plastid genome is divided into four major regions, the large single-copy region (LSC), small single-copy region (SSC), and two intervening inverted repeat regions (IRs), which are identical in sequence but arranged in reverse order (reviewed by Palmer, 1985; Sugiuara, 1992; Bock, 2007). If a gene spans a boundary (also known as a junction) between the IR and one of the single-copy regions, one copy of the fragment that lies within the IR is contiguous with the rest of the gene in the adjacent single-copy region, and the other copy is adjacent to whatever lies at the other end of the same single-copy region (Fig. 1). Two genes, ndhF and ndhH, are situated near opposite ends of the SSC region in most angiosperms, and each extends into the IR in some grass species. By sequencing across the two SSC-IR junctions, the positions of both genes, relative to the junctions, can be determined. Here we describe the varied occurrence of one or both of these genes spanning the SSC-IR junction in a range of species of the grass family and provide evidence for repeated migrations of portions of these genes across the junctions. Nucleotide substitution rates generally are slower in the IR than in the single-copy regions (e.g., Wolfe et al., 1987; Maier et al., 1995; Muse and Gaut, 1997; Yamane et al., 2006), and we document an acceleration in substitution rate that coincides with the migration of a portion of ndhH from the IR into the SSC in taxa of the PACMAD clade, a major lineage within the grass family, comprising subfamilies Panicoideae, Arundinoideae, Chloridoideae, Micrairoideae, Aristidoideae, and Danthonioideae.

Grass phylogenetics—The grass family (Poaceae) comprises ca. 10000 species (Clayton and Renvoize, 1986; Watson and Dallwitz, 1992), most of them herbaceous annuals and perennials, though the family also includes the woody-textured bamboos. Phylogenetic analyses of the monocots, variously based on molecular and morphological characters, have converged in the placement of two small plant families, Joinvilleaceae and Ecdieicoleaceae, as the closest relatives of the grasses (e.g., Doyle et al., 1992; Linder and Rudall, 1993; Chase et al., 1995, 2006; Linder and Kellogg, 1995; Kellogg and Linder, 1995; Stevenson and Loconte, 1995; Briggs et al., 2000; Bremer, 2002; Michelangeli et al., 2003; Davis et al., 2004; Graham et al., 2006; Marchant and Briggs, 2007; Bouchenak-Khelladi et al., 2008).

Analyses focusing on the grasses have provided an increasingly well-substantiated phylogenetic structure for the family. A major landmark in this effort was the combined analysis of several molecular character sets, plus a set of structural characters, yielding an overall phylogeny and a classification consistent with it as proposed by the Grass Phylogeney Working Group (GPWG, 2001, and prior analyses reviewed therein). The classification proposed by the GPWG recognized 12 subfamilies. Among them were four large subfamilies (Bambusoideae, ...
Chloridoideae, Panicoideae, and Pooidae, each including more than 1000 species) that collectively included ca. 90% of all grass species, plus eight smaller subfamilies and a few anomalous genera that were not assigned to subfamily. Three of the smaller subfamilies were placed as a series of lineages (Anomochlooideae, Pharoideae, and Puelioidae) diverging in succession from a major clade that includes the four large subfamilies and all other grasses. We use the phrase “early-diverging” to refer to the three small subfamilies that diverged early in the history of the family from the lineage that now includes most grass species. Within the major clade, all species fall into one or the other of two large subclades, one comprising subfamilies Panicoideae, Arundinoideae, Chloridoideae, Centothecoideae, Aristidoideae, and Danthonioideae, and designated the PACCAD clade, the other comprising subfamilies Bambusoideae, Ehrhar-toideae, and Pooidae, and designated the BEP clade. Both of these clades had been observed in previous analyses, and prior acronyms for them are PACC (Davis and Soreng, 1993) and BOP (Clark et al., 1995), respectively.

Many additional phylogenetic analyses of the grass family have been conducted since, variously focusing on the overall structure of the family or on particular lineages within it. Within what had been designated the PACCAD clade, representatives of the small subfamily Centothecoideae long have been associated with Panicoideae (e.g., Clark et al., 1995; Soreng and Davis, 1998; Hilu et al., 1999; Mathews et al., 2000; GPWG, 2001; Duvall et al., 2007), and a recent analysis indicated that lineages from these two subfamilies are intermixed to the extent that neither is monophyletic (Sánchez-Ken and Clark, 2007), thus suggesting that taxa from the former Centothecoideae should be subsumed within a broadly defined Panicoideae. A formal classification along these lines had been published earlier (Zuloaga et al., 2003). Another recent development (Sánchez-Ken et al., 2007) has been the reinstatement of Micrairaideae, which includes a putative monophyletic assemblage of Micraira, Eriachne, and other genera previously placed in isolated locations within the PACCAD clade, and sometimes left unassigned to subfamily. With the incorporation of Centothecoideae into Panicoideae and the adoption of Micrairaideae, the PACCAD clade is unchanged in composition, but now is designated the PACMAD clade (Duvall et al., 2007). Thus, a modified version of the GPWG classification, also comprising 12 subfamilies, is recognized today, with the same three small early-diverging subfamilies, plus the nine subfamilies of the BEP and PACMAD clades.

**Plastid genome structure**—The plastid genome of embryophytes is double-stranded, usually ca. 120–160 kb long, and it usually includes more than 120 genes (Palmer, 1985; Bock, 2007; Jansen et al., 2007). Being divided into the LSC, SSC, and two equal and intervening IR regions, it falls within the category of amphihermic genomes (Rayko, 1997). Although one copy of the IR has been lost from some plant lineages (e.g., Lavin et al., 1990), most plastid genomes that have been examined retain both copies. The plastid genome has been employed extensively in plant phylogenetic studies, some of which have principally used variation in nucleotide sequences (e.g., Chase et al., 1993; Soltis et al., 2000; Leebens-Mack et al., 2005; Hansen et al., 2007), while others have examined structural features, such as insertions/deletions (indels), inversions, and variation in the positions of boundaries between the genomic regions (e.g., Doyle et al., 1992; Goulding et al., 1996; Plunkett and Downie, 2000; Perry et al., 2002; Cosner et al., 2004; Stefanović and Olmstead, 2005; Wang et al., 2008).

Although the plastid genome conventionally is mapped as a circular, double-stranded DNA molecule, with fixed relationships between the two LSC, SSC, and two IR regions, and with the latter usually labeled as separate structures (IRα and IRβ), it actually exists in a variety of conformations within living cells, including interlinked aggregations of multiple copies (Palmer, 1983; Bendich, 2004; Oldenburg and Bendich, 2004; Bock, 2007). However, the conventional circular map does summarize many of the heritable structural features of the genome correctly. Among plant lineages that retain both copies of the IR, the sizes and boundaries of these regions, relative to the adjacent LSC and SSC, have been modified in various groups (e.g., Goulding et al., 1996; Aii et al., 1997; Plunkett and Downie, 2000; Perry et al., 2002; Stefanović and Olmstead, 2005; Wang et al., 2008).

Within the grass family, two genes that lie near opposite ends of the SSC region (ndhF and ndhH) sometimes extend into the IR regions (Maier et al., 1990; Ogihara et al., 2002; Davis and Soreng, 2007; Sasaki et al., 2007; Boritto et al., 2008). These two genes are transcribed in the same direction within the SSC region (Fig. 1). Thus, the 5′ term of ndhH is situated near one SSC-IR junction (JSA), and the 3′ term of ndhH is situated near the other (JSB). In Triticum aestivum and Oryza sativa, ndhH extends across the junction, with the 5′ term and more than 150 nucleotides situated within the IR, while ndhF lies fully within the SSC. In Zea mays, just one nucleotide at the 5′ term of ndhH lies within the IR, while ndhF extends across the junction, with the 3′ term and 29 nucleotides situated within the IR. Ogihara et al. (2002) concluded from the similarities between Triticum and Oryza in these characteristics that these two taxa are more closely related to each other than either is to Zea, but in making this suggestion they did not provide evidence that the similarity is synapomorphy. Triticum (BEP clade, subfamily Pooideae) is, in fact, believed to be more closely related to Oryza (BEP clade, subfamily Ehrharthtoideae) than either is to Zea (PACMAD clade, subfamily Panicoideae). However, one result of the current study was the demonstration that in most taxa of the sample, including grasses of early-diverging lineages, and outgroups, ca. 150–250 nucleotides (maximum 400) of ndhH extend into the IR region, the major exception being taxa of the PACMAD clade. Specifically, the occurrence of this portion of ndhH in the IR (as in Triticum and Oryza) is interpreted as the sympleiomorphic state in the grasses, and the migration of this gene region relative to the SSC-IR boundary, resulting in the transfer of most of these sites into the SSC (as in Zea), is interpreted as derived within the PACMAD clade. Meanwhile, the 3′ term of ndhH lies entirely within the SSC in the outgroups and most grasses and is interpreted as having migrated into the IR region in an ancestor of the PACMAD clade. The portion of ndhH that has migrated into the SSC from the IR in the PACMAD clade has experienced a corresponding nucleotide substitution rate acceleration.

When both genes extend into the IR, as in Zea, one or more nucleotide positions just within the IR encode portions of both genes. Within this region, the two genes are encoded on opposite strands, because the reading frame of ndhH extends from the IR into the SSC, while that of ndhF runs in the opposite direction. This situation, with two genes overlapping where they extend from different ends of a single-copy region into the same end of an IR, is a special case of the general phenomenon of gene overlap, which occurs widely in microorganisms (Fukuda et al., 2003) and some plastid genes (e.g., psbD and psbC). In this particular form of gene overlap, the region of overlap is
The taxonomic system adopted here for the grasses, at the subfamily level, is a modification of the 12-subfamily system proposed by the GPWG (2001), excluding Centothecoideae and including Micrairoideae, as described above. Of the 12 subfamilies, all are sampled except Puelioideae, for which a suitable DNA isolation was not available. Genera are assigned to subfamily and tribe (and to subtribe within Poeae) according to the Catalogue of New World Grasses (CNWG) (Judziewicz et al., 2000; Peterson et al., 2001; Soreng et al., 2003; Zuloaga et al., 2003 [and online revision of 18 June 2009 http://mobot.mobot.org/W3T/Search/nwgclass.html; archived web page available on request]). The CNWG classification is comprehensive at the generic level for all grass genera occurring in the Americas and includes miscellaneous additional genera for informational purposes. Most genera in the present analysis are classified in the CNWG system, either explicitly, or implicitly via nomenclatural linkage (e.g., by the inclusion of a tribe whose name is based on that of a genus that does not occur in the Americas). For four genera that were not assigned to subfamilies by the GPWG (2001), we follow the CNWG system (cf. Zhang, 2000; Sánchez-Ken and Clark, 2001, 2007; Duvall et al., 2007; Sánchez-Ken et al., 2007) in placing Streptogyna in Ehrhartoideae, Gynerium in Panicoideae, and Micraira and Eriachne in Micrairoideae. Four other taxa in the present analysis are not classified in the CNWG system, and they are assigned to subfamilies and tribes as follows (cf. Barker, 1997; Barker et al., 1999, 2007; GPWG, 2001; Pirie et al., 2008): Amphipogon (Arundinoideae, Arundineae); Stipagostris (Aristidoideae, Aristideae); Merxmuellera macowani (Danthonioideae, Danthonieae); and Merxmuellera rangei (Chloridoideae, no tribal assignment).

**Materials and Methods**

**Taxon Sample**—The taxon sample includes 90 representative species of Poeae, plus one each of Joinvilleaceae and Ecdieicoleaceae (Appendix 1). The taxonomic system adopted here for the grasses, at the subfamily level, is a modification of the 12-subfamily system proposed by the GPWG (2001), excluding Centothecoideae and including Micrairoideae, as described above. Of the 12 subfamilies, all are sampled except Puelioideae, for which a suitable DNA isolation was not available. Genera are assigned to subfamily and tribe (and to subtribe within Poeae) according to the Catalogue of New World Grasses (CNWG) (Judziewicz et al., 2000; Peterson et al., 2001; Soreng et al., 2003; Zuloaga et al., 2003 [and online revision of 18 June 2009 http://mobot.mobot.org/W3T/Search/nwgclass.html; archived web page available on request]). The CNWG classification is comprehensive at the generic level for all grass genera occurring in the Americas and includes miscellaneous additional genera for informational purposes. Most genera in the present analysis are classified in the CNWG system, either explicitly, or implicitly via nomenclatural linkage (e.g., by the inclusion of a tribe whose name is based on that of a genus that does not occur in the Americas). For four genera that were not assigned to subfamilies by the GPWG (2001), we follow the CNWG system (cf. Zhang, 2000; Sánchez-Ken and Clark, 2001, 2007; Duvall et al., 2007; Sánchez-Ken et al., 2007) in placing Streptogyna in Ehrhartoideae, Gynerium in Panicoideae, and Micraira and Eriachne in Micrairoideae. Four other taxa in the present analysis are not classified in the CNWG system, and they are assigned to subfamilies and tribes as follows (cf. Barker, 1997; Barker et al., 1999, 2007; GPWG, 2001; Pirie et al., 2008): Amphipogon (Arundinoideae, Arundineae); Stipagostris (Aristidoideae, Aristideae); Merxmuellera macowani (Danthonioideae, Danthonieae); and Merxmuellera rangei (Chloridoideae, no tribal assignment).

**DNA Methods**—Two regions of the plastid genome were sequenced, each of them spanning one of the two SSC-IR boundaries (JSA and JSB; Fig. 1). One of the regions ("region 1", rps15-ndhF) extends from a point within rps15, in the IR region, across JSB to a point within and near the 5′ terminus of ndhF, in the SSC region (Fig. 1). The other region ("region 2", rps15-ndhA) extends from the same point within rps15 across JSA to a point near the 5′ terminus of ndhA, at the other end of the SSC region. Nucleotide sequences were obtained from total genomic DNA isolations, using standard PCR and automated cycle sequencing methods. Amplification and sequencing primers are described in
Appendix 2, and positions of the most frequently used amplification primers are indicated in Fig. 1. Because both regions include identical portions of the IR, some primers (e.g., *rps15-80F*) were used in the amplification and sequencing of both regions. Also, for taxa in which the IR region is long enough to include the region corresponding to primer *ndhH-88F* (i.e., approximately the first 113 nucleotides of this gene), this primer sometimes was used in the sequencing of both regions. This condition is met in most taxa in the sample, the major exceptions being those of the PACMAD clade. Successful amplification and sequencing of the two regions provides complete sequences of *ndhH* (length of the complete gene = 1182 nucleotides in the reference plastid genome sequence of *Tritium aestivum*, GenBank accession NC_002762); 122 nucleotides of *rps15* (length of the complete gene = 273 nucleotides in the same reference sequence), and nearly complete sequences of *ndhF* (lacking ca. 50–70 nucleotides from the 5′ terminus; the complete length of *ndhF* is 2220 nucleotides in the same reference sequence).

**Data structure and analysis**—Sequences of *ndhH*, *ndhF*, and *rps15* were generated, aligned manually, and deposited in GenBank (Appendix 1). Nucleotide sites within inferred gaps were encoded as “unknown” and thus treated as missing characters for taxa with deletions. Regions interpreted as only ambiguously alignable were excluded from analyses. Nine structural features of the three genes with parsimony-informative distributions, including two inversions, seven indels, and two characters representing presence/absence of portions of the IR in and out of the IR region were identified and encoded as characters (Tables 1, 2). The indels were scored using simple gap coding (Simmons and Ochoterena, 2000). Nucleotides of *ndhF* within the two inversion regions were included in the analysis by replacing the observed sequence with its reverse complement for each taxon that was interpreted as being inverted (Graham et al., 2000; Soreng et al., 2007). For archival purposes, the data matrix includes the sequences as observed, but with the nucleotides in the inversion regions inactivated, and with these regions duplicated elsewhere in the matrix as active characters, with the observed sequences replaced by the reverse complements for the inverted taxa. The data matrix and explanatory text are available as supplemental materials (Appendices S1–S3, see Supplemental Data with online version of article).

Because both of the sequenced regions extend into *rps15* and because primer *rps15-80F* was used to amplify both of them, the portion of this gene that lies within the sequenced regions was sequenced twice from each taxon, as were all nucleotides between *rps15* and the SSC-IR boundaries. The portions of the two sequenced regions that extend into the IR region were aligned against each other to identify the locations of the two SSC-IR boundaries in each taxon (JSA and JSB, the points at which the sequences first differ from each other), and the portions of the 3′ terminus of *ndhF* and the 5′ terminus of *ndhH* were added to these boundaries (cf. Fig. 2). Contradictory nucleotide sequences were never observed within the twice-sequenced segment of the IR region for any taxon. In some cases, however, one or more sites within one or the other of the two fragments within the IR region was not read clearly, usually in the intergenic spacer region, but in some cases within *rps15*. In the latter cases, the reported *rps15* sequence is a composite of both taxa (even though the direction of the migration of this portion of *ndhF*, which also lies on this branch). For most taxa, just one of the two genes (*ndhF* or *ndhH*) or neither of them extends into the IR region. In a few cases, however, both of these genes extend into the IR region, and in these cases there are nucleotide sites that lie within both genes and thus are homologous between the two genes (Fig. 2). The two genes are encoded in reverse order in the IR, so the sense strand of one is the antisense strand of the other, but nonetheless, any nucleotide that lies within the IR region in both genes is homologous between the two genes. To avoid including these nucleotides twice in the analysis (i.e., once within the *ndhF* sequence of a taxon, and once within the *ndhH* sequence of the same taxon), the duplicated nucleotides were excluded from the *ndhF* sequences prior to analysis. This exclusion affected a total of 70 cells of the data matrix, and of these, seven are in characters that are active and parsimony informative in the data matrix. To determine whether the exclusion of these seven cells affected the results of the principal analysis (see below), two additional analyses were conducted, one of them with these cells included in the *ndhF* sequences but excluded from the *ndhH* sequences, and the other with these cells included in the sequences of both genes (i.e., included twice in the data matrix). The sets of trees obtained by both of these analyses were identical to those obtained by the principal analysis, and no further reference to the other analyses is made. The archived data matrix includes 46 characters, for which 70 cells are nonempty, immediately following *ndhF*. These cells represent the sites interpreted as homologous between *ndhF* and *ndhH*, and the corresponding cells in the *ndhH* portion of the matrix are scored “N”.

Parsimony analysis was conducted with the program TNT, version 1.1 (Goloboff et al., 2008), with all characters weighted equally and coded as nonadditive, missing nucleotides in indel regions coded as unknown, clades interpreted as resolved in individual trees only if supported by under all possible optimizations (using the command “collapse 3”), and with *Joinvillea* specified as the outgroup for the analysis. Four trees held at the origin of each branch (and the branch immediately following it) were removed from the matrix prior to analysis, and all reported tree lengths and consistency indices are based on only the parsimony-informative portion of the matrix. The data matrix was analyzed by invoking 1000 replicate search initiations with random taxon addition sequences, with 20 trees held per replicate and subjected to exhaustive TBR swapping and then to 200 ratchet iterations (Nixon, 1999), using 5% probabilities for character upweighting and downweighting. All trees obtained by these searches were pooled and subjected to exhaustive TBR swapping with up to 1 000 000 trees held in memory. A separate analysis was conducted using just the nucleotide sequence characters. Strict-consensus jackknife support (JS) (Farris et al., 1996; Soreng and Davis, 1998; Davis et al., 2004) was calculated from 10 000 replicates using a deletion frequency of 37%, with each jackknife replicate conducted using the same search procedures as used in the basic analysis, except that 10 search initiations were conducted for each jackknife replicate, rather than 1000, and up to 1000 trees were held during the final tree-bisection-reconnection (TBR) swapping phase, rather than 1 000 000. The program WinClada version 1.03 (Nixon, 2002) was used to edit data, determine tree lengths and related indices, examine character transformations on the resulting trees, and generate figures.

Optimizations of six structural characters are illustrated on a consensus tree (Fig. 3). The decision to present character optimizations on the consensus tree was made only after it was verified that the all transformations of these characters are identical in position and number among all most-parsimonious trees and the consensus tree. Optimization of one of these characters is ambiguous in one region of the tree (character 16, Tables 1, 2, near *Arundo* in Fig. 3; either two parallel gains or a gain and a loss), and this ambiguity is shared among all most-parsimonious trees and the consensus tree; all other optimizations of this and other characters mapped in Fig. 3 are unambiguous.

**Nucleotide substitution rates**—To determine whether the inferred shift in the relative positions of *ndhH* and the SSC-IR boundary is associated with a change in the nucleotide substitution rate in this region, we compared the number of inferred steps in this gene region among taxa of the PACMAD clade to the number of steps among other taxa. This comparison was conducted by enumeration of substitutions in various gene regions, with a focus on the portion of *ndhH* that migrated from the IR region into the SSC region in the PACMAD clade. This was conducted in a cladistic framework, without reference to a model or other method to account for unobserved steps. Steps were counted on one transformation (or one parsimony-informative character) for the combined data set, as optimized under accelerated transformation, along all branches within the PACMAD clade, and compared to the corresponding sum for all other branches in the tree, except for the branch leading to the PACMAD clade. This branch was excluded from both of these categories, because transformations along it cannot be assigned either to the IR or SSC region (i.e., each substitution that is observed in this branch is a composite of both transformations that occurred during the migration of this portion of *ndhF*, which also lies on this branch). Corresponding comparisons also were made between numbers of steps within and outside of the PACMAD clade (always excluding the branch that leads to the PACMAD clade) for three other gene regions, each of which lies either within the IR region or the SSC region in all or nearly all taxa. Ambiguously aligned nucleotides, which had been excluded from cladistic analyses, also were excluded from these comparisons, but all other variable sites were included, regardless of whether they were parsimony informative, so that the sums would include autapomorphic steps. The four regions compared, and their aligned lengths, are as follows: (1) *ndhH* sites 5–154, 150 nucleotides—This region lies entirely within the SSC region in all taxa of the PACMAD clade and entirely within the IR region in all but seven taxa outside the PACMAD clade (entirely within the SSC region in *Celtica, Pleuroepongon*, and *Olyra*, and partially within the SSC region in *Lithachne, Brachypodium*, *Brylkinia*, and *Moliniera*). (2) *ndhH*, from site 301 to the 3′ terminus, 894 nucleotides—This region lies entirely within the SSC region in all taxa of the PACMAD clade and in all except four taxa outside the PACMAD clade (*Cynosurus* and the three species of *Bromus*); in each of the latter four taxa, no more than 100 nucleotides of this region lie within the IR region. (3) *rps15*, from site 311 to 313, 72 nucleotides excluding nine sites from the two inversions, and excluding unambiguously aligned regions and 17 additional sites near the 3′ terminus, which lie within the IR region in some taxa, 2105 nucleotides—This region lies entirely within the SSC region in all taxa examined. (4) *rps15*, from site 152 to the 3′ terminus, 122 nucleotides—This region lies entirely within the IR region in all taxa examined.
Fig. 2. Nucleotide sequences for both DNA strands at the two SSC-IR junctions of the plastid genomes of six grass taxa (cf. Appendix 1, Fig. 1). IR regions lie to the left of each junction (JSA and JSB), and the SSC region to the right. Boldface, capital letters signify encoding regions (sense strand) near the 5′ terminus of \textit{ndhF}, also labeled at right as reading across the boundary into the SSC region, and underlined capital letters signify encoding regions (sense strand) near the 3′ terminus of \textit{ndhF}, also labeled at right as reading across the boundary into the IR region.

**RESULTS**

**Data structure**—The combined sequences of the three genes sum to 3438 aligned sites, of which 907 (26.4%) are parsimony-informative. In addition to the nucleotide sequence characters, the inclusion of nine informative structural characters (Tables 1–3) brought the total number of informative characters in the matrix to 916. Details concerning the distribution of these characters among the taxa in the sample, and on cladograms, are provided below.

**Grass phylogenetics**—Analysis of the combined matrix of three gene sequences and nine structural characters yielded 12 most-parsimonious trees of length 4024, CI 0.35, and RI 0.69 (Table 3, Fig. 3). Trees of this length were obtained by each of the 1000 searches that were conducted prior to the final phase of TBR swapping. The same set of trees was obtained when the analysis included only the nucleotide sequence characters; further discussion (e.g., support values) is based on results obtained with nucleotide and structural characters. Monophyly was not tested for Pharoideae and Aristidoideae, as each was sampled only once. Of the nine grass subfamilies represented by at least two species, eight are resolved as monophyletic with the current taxon sampling. In four of these cases, JS is between 98 and 100%, and among the other four, support for monophyly of Danthonioideae is 77%, and for each of the remaining, three it is less than 60%. In two of these cases, the support for a core group is strong, but an additional accession is weakly associated with the core group, so support for the subfamily itself is low. One of these cases is Chloridoideae, which has 51% JS, with \textit{Merxmuellera rangei} placed as the sister of the rest of the subfamily, and with 100% support for the clade that includes all
Table 1. Descriptions, encoding rules, number of steps, and consistency of nine structural characters of the plastid genome and nine nucleotides within two inversion regions. Locations of inversions and indels refer to the \textit{ndhF} and \textit{ndhH} sequences of the reference sequences of \textit{Triticum aestivum} (GenBank accession NC_002762) and may differ by a few nucleotides under alternative alignments.

| Character no. (no. of steps, CI, and RI on most-parasimoomous trees for structural characters), character description | Character 0 (6, 0.16, 0.44), presence/absence of one or more nucleotides of \textit{ndhH} in the IR region (cf. Figs. 1, 2): 0 = absence, 1 = presence. Number of nucleotides in the IR region is provided in Table 2 and as the first of two numbers (hyphen = 0) after each taxon name in Fig. 3. Character 1 (4, 0.25, 0.88): presence/absence of one or more nucleotides of \textit{ndhF} in the IR region (cf. Figs. 1, 2): 0 = absence, 1 = presence. Number of nucleotides in the IR region is provided for each taxon in Fig. 2 and as the second of two numbers (hyphen = 0) after each taxon name in Fig. 3. Parenthesized numbers in Fig. 3 denote presence and number of nucleotides inserted (+) or deleted ( − ), relative to the reference sequence of \textit{Triticum aestivum}, within the portion of \textit{ndhF} that extends into the IR region. For example, in \textit{Arundo} ("17–3"), 17 nucleotides of \textit{ndhF} lie within the IR region, and there is a 3-bp deletion within this portion of the gene, relative to \textit{Triticum}, so the portion of \textit{ndhF} within the IR region in \textit{Arundo} is interpreted as being homologous with a 20-nucleotide portion in \textit{Triticum}. Character 2 (2, 0.50, 0.05), \textit{ndhF} inversion of sites 1918–1920: 0 (uninverted) = GTA (55 taxa) or a sequence differing from this at no more than one site, and from the inverted sequence at all three sites: ATA (1), CTA (4), GCA (1), GGA (1), and GTT (1). 1 (inverted) = TAC (23 taxa) or a sequence differing from this at no more than one site and from the inverted sequence at all three sites: TAT (1). Six taxa differ from one of the main sequences at one site and from the other at two sites, and they are scored unknown for the inversion and for the nucleotides in the \textit{ndhF} sequence: TTA (4), TTC (1), GAA (1). Characters 3–5: sequence in region of \textit{ndhF} inversion 1 or for taxa interpreted as having the inversion (state 1 of character 2), the reverse complement of the observed sequence. Taxa scored as unknown for character 2 are also scored as unknown for these three characters. Character 6 (1, 1.00, 1.00): \textit{ndhF} inversion of sites 1932–1937: 0 (uninverted) = GAAAAA (46 taxa) or a sequence differing from this at no more than three sites and from the inverted sequence at no fewer than five sites: AAAAAA (14), CAAAAA (13), CCAAAA (5), GAAAAA (1), GGAACA (1), TAAAAA (9), AAAAAG (1), 1 (inverted) = TTTTTT (2 taxa) or a sequence differing from this at no more than three sites and from the uninvrouted sequence at no fewer than five sites (none). Characters 7–12. Sequence in region of \textit{ndhF} inversion 2, or for taxa interpreted as having the inversion (state 1 of character 6), the reverse complement of the observed sequence. Character 13 (2, 0.50, 0.00): Insertion of 3 nucleotides between \textit{ndhF} sites 1414 and 1415. 0 = 3 nucleotides absent; 1 = 3 nucleotides present. Character 14 (2, 0.50, 0.00): Insertion of 3 nucleotides between \textit{ndhF} sites 1576 and 1577. 0 = 3 nucleotides absent; 1 = 3 nucleotides present. Character 15 (4, 0.25, 0.66): One vs. two copies of 15 nucleotides between \textit{ndhF} sites 1693 and 1719 (location and extent of indel duplication is approximate, as multiple reasonable alignments are possible; cf. Fig. 4). 0 = 15 nucleotides absent, 1 = 15 nucleotides present. Character 16 (5, 0.20, 0.75): Deletion of \textit{ndhF} nucleotides 2209–2211. 0 = 3 nucleotides absent; 1 = 3 nucleotides present. Character 17 (2, 0.50, 0.00): Insertion of 6 nucleotides between \textit{ndhH} sites 19 and 20. 0 = 6 nucleotides absent; 1 = 6 nucleotides present. |

other elements of the subfamily. The other is Ehrhartioideae, which has 59% support, with \textit{Streptogyne} placed as the sister of the rest of the subfamily and with 100% support for the clade that includes all other elements of the subfamily. The third subfamily with weak support is Anomochlooidae (49%). The only subfamily represented by more than one element and not resolved as monophyletic is Arundinoideae. In this case, a monophyletic Micrairoideae (with 100% support) is nested within a clade that also includes the three representatives of Arundinoideae. Support for this overall grouping is 66%, and it is less than 50% for each of the internal nodes within the paraphyletic arrangement of the three genera of Arundinoideae.

Higher-level relationships are as follows: The grass family is monophyletic in all 12 trees (JS = 100%; Fig. 3). Within the family, Anomochlooidae is resolved as the sister of a clade that includes all other grasses (JS = 78% for the latter), and within the latter group, Pharoideae is sister of a clade that includes the remaining grasses (JS = 100% for this clade), and in which the PACMAD and BEP clades are both monophyletic and placed as sister taxa. JS for the PACMAD clade is 100%; within it, three major subclades are detected, with relationships unresolved among the three. The first of these clades includes Micrairoideae and the three representatives of Arundinoideae (as described above), the second includes Aristidoideae, Danthonioideae, and Chloridoideae, and the third is the Panicoideae (including elements previously assigned to Centothecoideae). Support is generally weak for relationships among subfamilies within the PACMAD clade, with the strongest support (66%) for the clade that includes Micrairoideae and the three representatives of Arundinoideae. Within the PACMAD clade, six of the seven tribes that were sampled by more than one taxon are monophyletic; the seventh, Arundineae, the solitary tribe in subfamily Arundinoideae, is not. JS for the BEP clade is 76%.

Within this clade, Ehrhartioideae (including \textit{Streptogyne}) is resolved as the sister of Bambusoideae and Pooideae, with 57% support for the clade that includes the latter two subfamilies. Nine tribes within the BEP clade are each sampled more than once, and all except two are resolved as monophyletic. The exceptions are Bambuseae (paraphyletic, with Olyreae nested within) and Bromaeae (paraphyletic, with Triticaceae nested within). Further description of relationships within Pooideae is deferred to a forthcoming analysis in which this group is sampled in greater depth.

\textbf{Structural features of the plastid genome}—Examples of six representative structures of the two SSC-IR boundaries are illustrated in Fig. 2, which includes fine-scale maps of the 5′ terminus of \textit{ndhH} and the 3′ terminus of \textit{ndhF}, and positions of the endpoints of these genes relative to the two SSC-IR boundaries. The range of observed structures include presence of a portion of \textit{ndhF} (\textit{Sporobolus}), \textit{ndhH} (\textit{Triticum}), both (\textit{Olyra}, \textit{Eragrostis}, \textit{Brachypodium}, \textit{Ehrhartia}), or neither (not depicted in Fig. 2) within the IR region (Tables 1, 2). In \textit{Ehrhartia}, both genes extend into the IR region (272 nucleotides of \textit{ndhH} and 43 nucleotides of \textit{ndhF}), and thus 43 nucleotides of the two genes are homologous between the two genes. Among the 82 taxa in which \textit{ndhH} extends into the IR region, the length of the portion of this gene that lies in the IR region ranges from one nucleotide to 400 (\textit{Cynosurus}).

When presence/absence of a portion of \textit{ndhH} in the IR region is optimized on trees obtained from the cladistic analysis, presence of the 5′ terminus of \textit{ndhH} within the IR region is determined to be a plesiomorphy of the grasses (Fig. 3, character 0). Within the grasses, there are five independent transitions to the state in which \textit{ndhH} lies entirely within the SSC region, plus one reversion to the plesiomorphic state. Three of the five transitions
Table 2.  Scores for characters described in Table 1. Numbers of nucleotides of ndhH and ndhF that extend into the IR region are indicated for taxa with state 1 for characters 0 and 1, respectively (cf. Table 1). Inversion presence/absence characters (characters 2 and 6) and nucleotide sequences within inverted regions (characters 3–5 and 7–12) are in boldface for taxa scored as having the inversions; the sequences provided here for the inverted regions of these taxa are the reverse complements of the sequences actually observed in the coding strand of ndhF. Taxa are arranged by family, subfamily, and tribe (cf. Appendix I).

<table>
<thead>
<tr>
<th>Taxon</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>345</th>
<th>6</th>
<th>1</th>
<th>11</th>
<th>1111</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>0</th>
<th>12</th>
<th>34567</th>
</tr>
</thead>
<tbody>
<tr>
<td>Joinvillea</td>
<td>1</td>
<td>192</td>
<td>0</td>
<td>0</td>
<td>GTA</td>
<td>0</td>
<td>GAAAA</td>
<td>10010</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ecediocolea</td>
<td>1</td>
<td>179</td>
<td>0</td>
<td>0</td>
<td>GCA</td>
<td>0</td>
<td>GAACA</td>
<td>01010</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anomochloa</td>
<td>1</td>
<td>179</td>
<td>0</td>
<td>0</td>
<td>GTA</td>
<td>0</td>
<td>CAAAA</td>
<td>00010</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Streptochaeta</td>
<td>1</td>
<td>178</td>
<td>0</td>
<td>0</td>
<td>GTA</td>
<td>0</td>
<td>AAAAA</td>
<td>00100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pharus</td>
<td>1</td>
<td>281</td>
<td>0</td>
<td>0</td>
<td>GTA</td>
<td>0</td>
<td>CAAAA</td>
<td>00100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amphipogon</td>
<td>0</td>
<td>1</td>
<td>15</td>
<td>0</td>
<td>GTA</td>
<td>0</td>
<td>CAAAA</td>
<td>00100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arundo</td>
<td>1</td>
<td>1</td>
<td>17</td>
<td>?</td>
<td>???</td>
<td>0</td>
<td>CAAAA</td>
<td>00100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Molinia</td>
<td>0</td>
<td>0</td>
<td>12</td>
<td>0</td>
<td>GTA</td>
<td>0</td>
<td>AAAAA</td>
<td>00100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eriachne mucronata</td>
<td>0</td>
<td>1</td>
<td>19</td>
<td>0</td>
<td>GTA</td>
<td>0</td>
<td>AAAAA</td>
<td>00100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eriachne phulchella</td>
<td>1</td>
<td>1</td>
<td>21</td>
<td>0</td>
<td>GTA</td>
<td>0</td>
<td>CAAAA</td>
<td>00100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Micraira</td>
<td>0</td>
<td>116</td>
<td>0</td>
<td>0</td>
<td>GTA</td>
<td>0</td>
<td>AAAAA</td>
<td>00100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stipagrostis</td>
<td>1</td>
<td>1</td>
<td>20</td>
<td>?</td>
<td>???</td>
<td>0</td>
<td>CAAAA</td>
<td>00100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Danthonia</td>
<td>1</td>
<td>1</td>
<td>32</td>
<td>0</td>
<td>GTA</td>
<td>0</td>
<td>CAAAA</td>
<td>00100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mertensuella macowanii</td>
<td>1</td>
<td>1</td>
<td>20</td>
<td>0</td>
<td>GTA</td>
<td>0</td>
<td>CAAAA</td>
<td>00100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mertensuella rangei</td>
<td>1</td>
<td>1</td>
<td>32</td>
<td>0</td>
<td>GTA</td>
<td>0</td>
<td>CAAAA</td>
<td>00100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distichlis</td>
<td>1</td>
<td>1</td>
<td>20</td>
<td>0</td>
<td>GTA</td>
<td>0</td>
<td>CAAAA</td>
<td>00100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eragrostis</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>23</td>
<td>0</td>
<td>GTA</td>
<td>0</td>
<td>CAAAA</td>
<td>00100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uniola</td>
<td>1</td>
<td>1</td>
<td>12</td>
<td>0</td>
<td>GTA</td>
<td>0</td>
<td>CAAAA</td>
<td>00010</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spartina</td>
<td>0</td>
<td>1</td>
<td>19</td>
<td>0</td>
<td>GTA</td>
<td>0</td>
<td>CAAAA</td>
<td>00100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sporobolus</td>
<td>0</td>
<td>1</td>
<td>19</td>
<td>?</td>
<td>???</td>
<td>0</td>
<td>CAAAA</td>
<td>00100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zoysia</td>
<td>0</td>
<td>1</td>
<td>19</td>
<td>?</td>
<td>???</td>
<td>0</td>
<td>CAAAA</td>
<td>00100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chamnamuthium</td>
<td>1</td>
<td>1</td>
<td>20</td>
<td>0</td>
<td>GTA</td>
<td>0</td>
<td>TAAAA</td>
<td>00000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thysanolaena</td>
<td>1</td>
<td>4</td>
<td>23</td>
<td>?</td>
<td>???</td>
<td>0</td>
<td>CAAAA</td>
<td>00000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gynium</td>
<td>0</td>
<td>1</td>
<td>14</td>
<td>0</td>
<td>GTA</td>
<td>0</td>
<td>AAAAA</td>
<td>00000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Panicum</td>
<td>1</td>
<td>1</td>
<td>32</td>
<td>0</td>
<td>GTA</td>
<td>0</td>
<td>AAAAA</td>
<td>00000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pennisetum</td>
<td>1</td>
<td>1</td>
<td>32</td>
<td>0</td>
<td>GTA</td>
<td>0</td>
<td>AAAAA</td>
<td>00000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Saccharum</td>
<td>1</td>
<td>1</td>
<td>32</td>
<td>0</td>
<td>GTA</td>
<td>0</td>
<td>AAAAA</td>
<td>00000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sorghum</td>
<td>1</td>
<td>1</td>
<td>32</td>
<td>0</td>
<td>GTA</td>
<td>0</td>
<td>AAAAA</td>
<td>00000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zea</td>
<td>1</td>
<td>1</td>
<td>32</td>
<td>0</td>
<td>GTA</td>
<td>0</td>
<td>AAAAA</td>
<td>00000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chusquea</td>
<td>1</td>
<td>186</td>
<td>0</td>
<td>0</td>
<td>GTA</td>
<td>0</td>
<td>GAAAA</td>
<td>00110</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Guadua</td>
<td>1</td>
<td>186</td>
<td>0</td>
<td>0</td>
<td>GTA</td>
<td>0</td>
<td>GAAAA</td>
<td>00110</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phyllostachys</td>
<td>1</td>
<td>187</td>
<td>0</td>
<td>0</td>
<td>GTA</td>
<td>0</td>
<td>GAAAA</td>
<td>00110</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pseudosasa</td>
<td>1</td>
<td>187</td>
<td>0</td>
<td>0</td>
<td>GTA</td>
<td>0</td>
<td>GAAAA</td>
<td>00110</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baeopersicloa</td>
<td>1</td>
<td>187</td>
<td>0</td>
<td>0</td>
<td>GTA</td>
<td>0</td>
<td>GAAAA</td>
<td>00110</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eremis</td>
<td>1</td>
<td>252</td>
<td>0</td>
<td>0</td>
<td>GTA</td>
<td>0</td>
<td>CAAAA</td>
<td>00110</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lithachne</td>
<td>1</td>
<td>151</td>
<td>0</td>
<td>0</td>
<td>GTA</td>
<td>0</td>
<td>GAAAAG</td>
<td>00110</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Olyra</td>
<td>1</td>
<td>1</td>
<td>17</td>
<td>0</td>
<td>GTA</td>
<td>0</td>
<td>GAAAA</td>
<td>00110</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pariana</td>
<td>1</td>
<td>252</td>
<td>0</td>
<td>0</td>
<td>GTA</td>
<td>0</td>
<td>GAAAA</td>
<td>00110</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Streptogyna</td>
<td>1</td>
<td>181</td>
<td>0</td>
<td>0</td>
<td>GTA</td>
<td>0</td>
<td>GAAAA</td>
<td>00110</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ehrharta</td>
<td>1</td>
<td>272</td>
<td>1</td>
<td>43</td>
<td>0</td>
<td>GTA</td>
<td>0</td>
<td>GAAAA</td>
<td>00110</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leersia</td>
<td>1</td>
<td>181</td>
<td>0</td>
<td>0</td>
<td>GTA</td>
<td>0</td>
<td>GAAAAG</td>
<td>00110</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oryza nivara</td>
<td>1</td>
<td>163</td>
<td>0</td>
<td>0</td>
<td>GTA</td>
<td>0</td>
<td>GAAAA</td>
<td>00110</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oryza sativa</td>
<td>1</td>
<td>163</td>
<td>0</td>
<td>0</td>
<td>GTA</td>
<td>0</td>
<td>GAAAA</td>
<td>00110</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brachyelytrum</td>
<td>1</td>
<td>223</td>
<td>0</td>
<td>0</td>
<td>GTA</td>
<td>0</td>
<td>GAAAA</td>
<td>00110</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Narthuas</td>
<td>1</td>
<td>174</td>
<td>0</td>
<td>?</td>
<td>???</td>
<td>0</td>
<td>TAAAA</td>
<td>00010</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Legeum</td>
<td>1</td>
<td>212</td>
<td>0</td>
<td>0</td>
<td>GTA</td>
<td>0</td>
<td>GAAAA</td>
<td>00110</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anisopogon</td>
<td>1</td>
<td>177</td>
<td>0</td>
<td>0</td>
<td>GTA</td>
<td>0</td>
<td>GAAAA</td>
<td>00110</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dathiea</td>
<td>1</td>
<td>181</td>
<td>0</td>
<td>0</td>
<td>GTA</td>
<td>0</td>
<td>GAAAA</td>
<td>00110</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phaenosperma</td>
<td>1</td>
<td>181</td>
<td>0</td>
<td>0</td>
<td>GTA</td>
<td>0</td>
<td>GAAAA</td>
<td>00110</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sinochasea</td>
<td>1</td>
<td>181</td>
<td>0</td>
<td>0</td>
<td>GTA</td>
<td>0</td>
<td>GAAAA</td>
<td>00110</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Achnatherum</td>
<td>1</td>
<td>181</td>
<td>0</td>
<td>0</td>
<td>GTA</td>
<td>0</td>
<td>GAAAA</td>
<td>00110</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ampelodesmos</td>
<td>1</td>
<td>181</td>
<td>0</td>
<td>0</td>
<td>GTA</td>
<td>0</td>
<td>GAAAA</td>
<td>00110</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Celtica</td>
<td>1</td>
<td>181</td>
<td>0</td>
<td>0</td>
<td>GTA</td>
<td>0</td>
<td>GAAAA</td>
<td>00110</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hesperostipa</td>
<td>1</td>
<td>182</td>
<td>0</td>
<td>0</td>
<td>GTA</td>
<td>0</td>
<td>GAAAA</td>
<td>00110</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nassella pulchra</td>
<td>1</td>
<td>181</td>
<td>0</td>
<td>0</td>
<td>GTA</td>
<td>0</td>
<td>GAAAA</td>
<td>00110</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nassella viridula</td>
<td>1</td>
<td>192</td>
<td>0</td>
<td>0</td>
<td>GTA</td>
<td>0</td>
<td>GAAAA</td>
<td>00110</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Orcopsis</td>
<td>1</td>
<td>182</td>
<td>0</td>
<td>0</td>
<td>GTA</td>
<td>0</td>
<td>TAAAA</td>
<td>00110</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pipitatherum</td>
<td>1</td>
<td>174</td>
<td>0</td>
<td>0</td>
<td>GTA</td>
<td>0</td>
<td>GAAAA</td>
<td>00110</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stipa</td>
<td>1</td>
<td>192</td>
<td>0</td>
<td>0</td>
<td>GTA</td>
<td>0</td>
<td>GAAAA</td>
<td>00110</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Timonuria</td>
<td>1</td>
<td>172</td>
<td>0</td>
<td>0</td>
<td>GTA</td>
<td>0</td>
<td>GAAAA</td>
<td>00110</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trikerja</td>
<td>1</td>
<td>181</td>
<td>0</td>
<td>0</td>
<td>GTA</td>
<td>0</td>
<td>GAAAA</td>
<td>00110</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brylkinia</td>
<td>1</td>
<td>149</td>
<td>0</td>
<td>0</td>
<td>GTA</td>
<td>0</td>
<td>GAAAA</td>
<td>00110</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
lies entirely within the SSC region are in lineage to diverge within the grasses from the clade that in-
among taxa (Table 2, Fig. 3). In the outgroups and in the first portion of presence of a portion of ropogon mucronata therefore differs in this feature from its sister taxon, gion, and then back into it, in deae clade, the 5 tree. Within one of these groups, the Arunidinoideae/Micrairoi-
within each of the three major clades resolved in the consensus occur in the PACMAD clade, each of them occurring once in which the 5\textsuperscript{t} terminus of ndhH migrates out of the IR region occur in the PACMAD clade, each of them occurring once within each of the three major clades resolved in the consensus tree. Within one of these groups, the Arunidinoideae/Micrairoi-
deae clade, the 5 tree. Within one of these groups, the Arunidinoideae/Micrairoi-
within each of the three major clades resolved in the consensus occur in the PACMAD clade, each of them occurring once

### Table 2. Continued.

<table>
<thead>
<tr>
<th>Taxon</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>345</th>
<th>6</th>
<th>11 1</th>
<th>11111</th>
<th>34567</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glyceria</td>
<td>168</td>
<td>0</td>
<td>0</td>
<td>GTA</td>
<td>0</td>
<td>GAAAAA</td>
<td>00110</td>
<td></td>
</tr>
<tr>
<td>Melica</td>
<td>208</td>
<td>0</td>
<td>0</td>
<td>GTA</td>
<td>0</td>
<td>GAAAAA</td>
<td>00110</td>
<td></td>
</tr>
<tr>
<td>Pleurogenon</td>
<td>0</td>
<td>0</td>
<td>GTA</td>
<td>0</td>
<td>GAAAAA</td>
<td>00110</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schizachne</td>
<td>175</td>
<td>0</td>
<td>0</td>
<td>GTA</td>
<td>0</td>
<td>GAAAAA</td>
<td>00110</td>
<td></td>
</tr>
<tr>
<td>Diarrhena</td>
<td>181</td>
<td>0</td>
<td>0</td>
<td>GTA</td>
<td>0</td>
<td>GAAAAA</td>
<td>00110</td>
<td></td>
</tr>
<tr>
<td>Brachydropodium</td>
<td>42</td>
<td>11</td>
<td>GTA</td>
<td>0</td>
<td>CAAAAA</td>
<td>00110</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bromus inermis</td>
<td>314</td>
<td>1</td>
<td>GTA</td>
<td>0</td>
<td>TAAAAA</td>
<td>00110</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bromus korotkiji</td>
<td>314</td>
<td>1</td>
<td>GTA</td>
<td>0</td>
<td>TAAAAA</td>
<td>00110</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bromus suksdorfi</td>
<td>317</td>
<td>1</td>
<td>GTA</td>
<td>0</td>
<td>TAAAAA</td>
<td>00110</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Littledealea</td>
<td>187</td>
<td>0</td>
<td>1</td>
<td>GTA</td>
<td>0</td>
<td>GAAAAA</td>
<td>00110</td>
<td></td>
</tr>
<tr>
<td>Elymus</td>
<td>207</td>
<td>0</td>
<td>1</td>
<td>GTA</td>
<td>0</td>
<td>TAAAAA</td>
<td>00110</td>
<td></td>
</tr>
<tr>
<td>Hordeum</td>
<td>216</td>
<td>0</td>
<td>0</td>
<td>GTA</td>
<td>0</td>
<td>TAAAAA</td>
<td>00110</td>
<td></td>
</tr>
<tr>
<td>Triticum</td>
<td>207</td>
<td>0</td>
<td>1</td>
<td>GTA</td>
<td>0</td>
<td>TAAAAA</td>
<td>00110</td>
<td></td>
</tr>
<tr>
<td>Agrostis</td>
<td>174</td>
<td>0</td>
<td>1</td>
<td>GTA</td>
<td>0</td>
<td>GAAAAA</td>
<td>00110</td>
<td></td>
</tr>
<tr>
<td>Avena</td>
<td>181</td>
<td>0</td>
<td>1</td>
<td>GTA</td>
<td>1</td>
<td>GAAAAA</td>
<td>00110</td>
<td></td>
</tr>
<tr>
<td>Trisetum</td>
<td>181</td>
<td>0</td>
<td>1</td>
<td>GTA</td>
<td>1</td>
<td>GAAAAA</td>
<td>00110</td>
<td></td>
</tr>
<tr>
<td>Briza</td>
<td>181</td>
<td>0</td>
<td>1</td>
<td>ATA</td>
<td>0</td>
<td>GAAAAA</td>
<td>00110</td>
<td></td>
</tr>
<tr>
<td>Torreyochloa</td>
<td>181</td>
<td>0</td>
<td>1</td>
<td>GTA</td>
<td>0</td>
<td>GAAAAA</td>
<td>00110</td>
<td></td>
</tr>
<tr>
<td>Aira</td>
<td>181</td>
<td>0</td>
<td>1</td>
<td>GTA</td>
<td>0</td>
<td>GAAAAA</td>
<td>00110</td>
<td></td>
</tr>
<tr>
<td>Deschampsia</td>
<td>206</td>
<td>0</td>
<td>1</td>
<td>GTA</td>
<td>0</td>
<td>GAAAAA</td>
<td>00110</td>
<td></td>
</tr>
<tr>
<td>Molineriella</td>
<td>26</td>
<td>0</td>
<td>1</td>
<td>GTA</td>
<td>0</td>
<td>GAAAAA</td>
<td>00110</td>
<td></td>
</tr>
<tr>
<td>Phleum</td>
<td>175</td>
<td>0</td>
<td>1</td>
<td>GTA</td>
<td>0</td>
<td>GAAAAA</td>
<td>00110</td>
<td></td>
</tr>
<tr>
<td>Cynosurus</td>
<td>400</td>
<td>0</td>
<td>1</td>
<td>GTA</td>
<td>0</td>
<td>GAAAAA</td>
<td>10110</td>
<td></td>
</tr>
<tr>
<td>Dactylis</td>
<td>169</td>
<td>0</td>
<td>1</td>
<td>GTA</td>
<td>0</td>
<td>GAAAAA</td>
<td>00110</td>
<td></td>
</tr>
<tr>
<td>Holcus</td>
<td>157</td>
<td>0</td>
<td>1</td>
<td>GTA</td>
<td>0</td>
<td>GAAAAA</td>
<td>00110</td>
<td></td>
</tr>
<tr>
<td>Festuca</td>
<td>169</td>
<td>0</td>
<td>1</td>
<td>GTA</td>
<td>0</td>
<td>GAAAAA</td>
<td>00110</td>
<td></td>
</tr>
<tr>
<td>Parapholis</td>
<td>206</td>
<td>0</td>
<td>1</td>
<td>GTA</td>
<td>0</td>
<td>GAAAAA</td>
<td>00110</td>
<td></td>
</tr>
<tr>
<td>Poa</td>
<td>192</td>
<td>0</td>
<td>1</td>
<td>GTA</td>
<td>0</td>
<td>GAAAAA</td>
<td>00110</td>
<td></td>
</tr>
<tr>
<td>Puccinellia</td>
<td>170</td>
<td>0</td>
<td>1</td>
<td>GTA</td>
<td>0</td>
<td>GAAAAA</td>
<td>00110</td>
<td></td>
</tr>
<tr>
<td>Sclerochloa</td>
<td>156</td>
<td>0</td>
<td>1</td>
<td>GTA</td>
<td>0</td>
<td>GAAAAA</td>
<td>00110</td>
<td></td>
</tr>
</tbody>
</table>

in which the 5\textsuperscript{t} terminus of ndhH migrates out of the IR region occurs in the PACMAD clade, each of them occurring once within each of the three major clades resolved in the consensus tree. Within one of these groups, the Arunidinoideae/Micrairoi-deae clade, the 5\textsuperscript{t} terminus of ndhH migrates out of the IR region, and then back into it, in Eriachne pulchella, which therefore differs in this feature from its sister taxon, Eriachne mucronata. The other two transitions to the state in which ndhH lies entirely within the SSC region are in Celtica and Pleurogenon, both of which are in the Pooidae.

Although there are only six transitions in the presence/absence of a portion of ndhH within the IR region, the size of the portion of ndhH that lies within the IR region varies widely among taxa (Table 2, Fig. 3). In the outgroups and in the first lineage to diverge within the grasses from the clade that includes all others (Anomochlooideae), 179–192 nucleotides of ndhH lie within the IR region. The number for Ecediocolea (179) is the same as that for Anomochloa. In the next-diverging grass lineage, Pharus, about 100 additional nucleotides lie within the IR region. Only 1–4 nucleotides of ndhH lie within the IR region in taxa of the PACMAD clade, and as already noted, these nucleotides migrate into the SSC region three times within this group. Within the BEP clade, numbers initially are similar to those of the early-diverging grasses, with occasional cases of increase (e.g., to 272 in Ehrhartia, to more than 300 in the various species of Bromus, and to 400 in Cynosurus) and decrease (e.g., to 149 in Brylkinia, 42 in Brachypodium). In the two instances in the BEP clade in which ndhH migrates entirely out of the IR region (Celtica and Pleurogenon), the sister of the taxon in which this has occurred does not have an unusually

### Table 3. Characteristics of partitions of the data matrix analyzed in this study. Aligned length of ndhF excludes regions in which alignment is ambiguous and includes nine nucleotide sites from two inversion regions, with sequences from inverted taxa reinverted for analysis (see Materials and Methods).

<table>
<thead>
<tr>
<th>Character partition</th>
<th>Aligned length, nucleotides</th>
<th>No. of steps</th>
<th>RI</th>
<th>CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>ndhH nucleotides</td>
<td>1194</td>
<td>1282–1284</td>
<td>0.32</td>
<td>0.66</td>
</tr>
<tr>
<td>rps13 nucleotides</td>
<td>122</td>
<td>49–50</td>
<td>0.58–0.59</td>
<td>0.90</td>
</tr>
<tr>
<td>ndhF nucleotides</td>
<td>2122</td>
<td>2662–2665</td>
<td>0.35–0.36</td>
<td>0.70</td>
</tr>
<tr>
<td>Total nucleotides, three genes</td>
<td>3438</td>
<td>3996</td>
<td>0.35</td>
<td>0.69</td>
</tr>
<tr>
<td>Structural characters</td>
<td>n/a</td>
<td>28</td>
<td>0.32</td>
<td>0.78</td>
</tr>
<tr>
<td>Total, all characters</td>
<td>n/a</td>
<td>4024</td>
<td>0.35</td>
<td>0.69</td>
</tr>
</tbody>
</table>
low number of sites in the IR region (174 in *Piptatherum*, 168 in *Glyceria*).

Among the 27 taxa in which a portion of *ndhF* lies within the IR region, the length of this portion is one nucleotide in *Brachypodium* (Table 2) and between 12 and 32 nucleotides in all other taxa except *Ehrharta* (Fig. 2, Table 2), in which 43 nucleotides of *ndhF* lie within the IR region. These nucleotides are the reverse complement of a region in *ndhH* (on the other strand) that is present in all sampled taxa and that also extends into the IR region in *Ehrharta*. The alignment of *ndhF* sequences, with the stop codons of *Ehrharta*, *Triticum*, and other taxa (usually TAA) treated as homologous, suggests that *Ehrharta* has a 36-nucleotide insertion relative to *Triticum*, just prior to the stop codon. However, the homology of the 43-nucleotide portion of *ndhF* of *Ehrharta* (comprising the stop codon, the preceding 36 nucleotides, and the four nucleotides that precede them) with a region in *ndhH* suggests that seven nucleotides at the 3′ end of *ndhF* actually were lost, and that the apparent insertion of 36 nucleotides of *ndhF* in *Ehrharta* represents an extension of the 3′ end of the coding region from the SSC-IR boundary to the first available stop codon inside the IR region.

When the presence/absence of a portion of *ndhF* in the IR region is optimized on trees obtained from the cladistic analysis, the absence of any portion of *ndhF* in the IR region is determined to be a plesiomorphy of the grasses (Fig. 3, character 1). Within the grasses, migration of the 3′ terminus of *ndhF* into the IR region is interpreted as a synapomorphy of the PACMAD clade, and no reversals are observed. Within the PACMAD clade, the length of the portion of *ndhF* that lies in the IR region ranges from 12 to 32 nucleotides. Three additional migrations of the 3′ terminus of *ndhF* into the IR region are inferred, in *Ehrharta*, *Olyra*, and *Brachypodium*. The lengths of the portions of *ndhF* that lie within the IR region in these taxa range from 1 to 43 nucleotides, and as noted, the 43 nucleotides of *ndhF* in *Ehrharta* include seven that are homologous with those in *Triticum*, plus an autapomorphic insertion of 36.

Two structural characters in the analysis are inversions within *ndhF*, one of them three nucleotides in length, the other six in length (Table 1, characters 2 and 6, respectively). Both inversion sites are surrounded by short inverted repeat sequences, indicative of a hairpin structure (Kelchner and Wendel, 1996). The inverted state of the three-nucleotide inversion arises once in all trees, in the sister group of *Diarrhena*, with a reversion to the plesiomorphic state occurring in *Hordeum* (Fig. 3, character 2). The six-nucleotide inversion arises once, as a synapomorphy of the clade that includes *Avena* and *Trisetum* (Fig. 3, character 6), and no reversions are observed.

The five remaining parsimony-informative structural characters are indels. Three of these (characters 13, 14, and 17; Tables 1 and 2; not mapped in Fig. 3) involve distantly related taxa (e.g., *Joinvillea* and *Cynosurus* for character 13) and are interpreted as parallelisms. The three-nucleotide indel near the 3′ terminus of *ndhF*, located within the portion of this gene that lies within the IR region in some taxa, was encoded as character 16 (Tables 1 and 2). As with the 36-nucleotide insertion in *Ehrharta*, variation in character 16 affects the length of the portion of *ndhF* that extends into the IR region (Fig. 3, numbers in parentheses). There are five steps in this character in all most-parsimonious trees and in the consensus tree (Fig. 3, character 16). The undeleted state is interpreted as plesiomorphic for the grasses. The deleted state arises independently in *Olyra* and at the origin of the PACMAD clade. Within the PACMAD clade there is a reversion to the undeleted state in Chloridoideae. Two additional steps occur within the Arundinoideae/Micrairoideae clade, either a reversion to the undeleted state followed by a secondary reversion to the deleted state (as mapped in Fig. 3, under accelerated optimization), or parallel reversions to the undeleted state in *Molinia* and *Amphipogon* (under delayed optimization).

A 15-nucleotide indel (character 15) has two forms, one of which appears to represent a tandem duplication of 15 nucleotides (Figs. 3, 4). The unduplicated state occurs in 10 taxa (Tables 1, 2), and one potential alignment of this region is depicted in Fig. 4 for these 10 taxa and several representative taxa with the putatively duplicated state. This region is one of the four regions of ambiguous alignment within *ndhF* that were excluded from the cladistic analysis on the basis of ambiguity of alignment. The 15-nucleotide segment that is apparently duplicated in some taxa can be recognized as consisting of two portions, the first nine nucleotides of which is a relatively constant motif (GATAATGG and slight variants) with a second more divergent six-nucleotide motif (in taxa with two copies there appear to be two variants of this, one based on ATATAGG variants, the other on ATAGCG and variants). The six nucleotides that follow the nine-nucleotide motif in five of the taxa that have one copy of the 15-nucleotide region (*Joinvillea*, *Ecdetocolea*, *Anomochloa*, *Streptocheta*, and *Pharus*) resemble the version of the two six-nucleotide motifs closest to the 5′ end of the gene in duplicated taxa. In the other five taxa with single copies (*Leersia*, two species of *Oryza*, *Nardus*, and *Celtico*), the six nucleotides that follow the nine-nucleotide motif resemble the second of the two six-nucleotide motifs. Thus, it is possible to recognize two different forms of the unduplicated state without reference to a phylogeny. Because of the general ambiguity of alignment in this region, only two states were recognized for this character (unduplicated and duplicated), and all 10 taxa with a single 15-nucleotide copy in this region were scored identically. Optimization of this character on the 12 most-parsimonious cladograms (and consensus tree) implies four steps (Fig. 3, character 15, i.e., one duplication and three subsequent deletions in disjunct parts of the tree). The single-copy state is shared by the two outgroups and three early-diverging grasses and is interpreted as a plesiomorphy of the grasses. The five taxa that share this state correspond to one of the two groups that can be recognized a priori (Fig. 4). The tandem duplication state originates as a synapomorphy of the clade that includes the PACMAD and BEP clades (the leftmost instance of character 15 noted in Fig. 3), and the single-copy state then reorients independently as a deletion of one of the duplicated 15 nucleotide regions (i.e., the more 3′ copy) in three groups within this large clade (*Oryzeae*, *Nardus*, and *Celtico*). If the two groups of taxa with deletions alignable in different positions had been scored as having different states and the same trees were obtained, then the unduplicated state that is plesiomorphic in the grasses would still have been lost once because of duplication, and the deleted state that later arose within the grasses would have had three separate origins.

**Nucleotide substitution rates**—In the 894-nucleotide portion of *ndhH* that lies within the SSC region in nearly all sampled taxa, there are 1174 steps in the randomly selected tree that was used to compare nucleotide substitution rates or 1.31 steps per character (Table 4). For this region, the number of steps outside the PACMAD clade (0.88 steps per character) is about three times the number within the PACMAD clade (0.32). In
Fig. 3. Strict consensus of 12 most-parsimonious trees for 90 grass species and two nongrass outgroups, as resolved by combined analysis of three gene sequences and nine structural characters of the plastid genome (see text). Family names are indicated for the two outgroup taxa, and assignments of grasses to tribe and (for tribe Poeae) subtribe are signified by letter codes as specified in Appendix 1, where species names are provided. Within the grass family, two major clades (PACMAD and BEP) and 11 subfamilies are indicated by lines and labels at right. Two nonparenthesized numbers beside each taxon name signify the number of nucleotides of \( \text{ndhH} \) and \( \text{ndhF} \), respectively, that extend into the IR region of the plastid genome (see text); hyphens denote 0 nucleotides. Parenthesized numbers denote the presence of an insertion (+) or deletion (−), relative to the reference sequence of \( \text{Triticum aestivum} \).
the 2105-nucleotide portion of \textit{ndhf} that lies within the SSC region in all taxa, there are 2944 steps in the tree, or 1.40 steps per character; as with the portion of \textit{ndhH} that lies predominantly within the SSC region, the number of steps in this region outside the PACMAD clade is about three times the number within this clade (1.03 vs. 0.35 steps per character, respectively). In the sequenced portion of \textit{rps15} (122 nucleotides in length), which lies within the IR region in all sampled taxa, there are 66 steps in the tree, or 0.54 steps per character, and the number of steps outside the PACMAD clade is about five times the number within (0.44 vs. 0.09 steps per character, respectively). Finally, in the 150-nucleotide region of \textit{ndhH} that lies entirely or almost entirely within the IR region in most non-PACMAD taxa in the sample, and in the SSC region in all members of the PACMAD clade, there are 96 steps in the tree that was examined, an average of 0.64 steps per character, with about one-half of these steps outside the PACMAD clade, and the other half within the PACMAD clade (0.31 vs. 0.33 steps per character, respectively).

**DISCUSSION**

**Structural features of the plastid genome**—Positions of the endpoints of \textit{ndhf} and \textit{ndhH}, relative to the SSC-IR junction, have a 3-bp deletion, relative to the SSC-IR junctions (characters 0 and 1, respectively) [Tables 1, 2; Fig. 3], of any length within the IR region of the plastid genome were among the structural characters included in the cladistic analysis (characters 0 and 1, respectively) [Tables 1, 2; Fig. 3], of any length within the IR region of the plastid genome were among the structural characters included in the cladistic analysis (characters 0 and 1, respectively) [Tables 1, 2; Fig. 3], of any length within the IR region of the plastid genome were among the structural characters included in the cladistic analysis (characters 0 and 1, respectively) [Tables 1, 2; Fig. 3], of any length within the IR region of the plastid genome were among the structural characters included in the cladistic analysis (characters 0 and 1, respectively) [Tables 1, 2; Fig. 3], of any length within the IR region of the plastid genome were among the structural characters included in the cladistic analysis (characters 0 and 1, respectively) [Tables 1, 2; Fig. 3]. In light of these similarities, the possibility was considered that a laboratory or clerical error might have led to the erroneous labeling of sequences from a representative of the PACMAD clade as having been collected from \textit{Olyra}. Alternatively, the possibility was considered that the actual plastid sequence that occurs in \textit{Olyra} might have been derived from a species of the PACMAD clade via horizontal gene transfer. A third possibility was that one or both of the sequences determined for \textit{Olyra} might be a laboratory
Table 4. Nucleotide substitution rates in four regions of the plastid genome, within and outside of the PACMAD clade. The first two gene regions in the table are in the SSC region of the plastid genome in all or nearly all sampled taxa; the third gene region is in the IR region of the plastid genome in all sampled taxa; the fourth gene region is in the SSC region in all taxa of the PACMAD clade and in the IR region in most other taxa.

<table>
<thead>
<tr>
<th>Gene region; no. of aligned sites</th>
<th>Total steps (total in branch to PACMAD clade; ratio, no. of steps to no. of sites)</th>
<th>No. of steps within PACMAD clade; ratio, no. of steps to no. of sites</th>
<th>No. of steps outside PACMAD clade; ratio, no. of steps to no. of sites</th>
<th>Ratio, no. of steps outside PACMAD clade to no. of steps within it</th>
</tr>
</thead>
<tbody>
<tr>
<td>ndhH site 301 to 3′ terminus; 894</td>
<td>1174 (100); 1.31</td>
<td>287; 0.32</td>
<td>787; 0.88</td>
<td>2.74</td>
</tr>
<tr>
<td>ndhH, in part (see text); 2105</td>
<td>2944 (24); 1.40</td>
<td>747; 0.35</td>
<td>2173; 1.03</td>
<td>2.91</td>
</tr>
<tr>
<td>rps15, sequenced portion; 122</td>
<td>66 (1); 0.54</td>
<td>11; 0.09</td>
<td>54; 0.44</td>
<td>4.91</td>
</tr>
<tr>
<td>ndhH sites 5–54; 150</td>
<td>96 (0); 0.64</td>
<td>49; 0.33</td>
<td>47; 0.31</td>
<td>0.96</td>
</tr>
</tbody>
</table>
and the Bambusoideae was placed as sister of Pooideae in a monophyletic BEP clade. Analysis of the 92 ndhH sequences yielded a consensus tree in which Olyra was the sister of Lithachne within a monophyletic Bambusoideae, and the terminal branch for Olyra was not unusually long. Hence, the available evidence is consistent with a conclusion that the ndhf and ndhH sequences obtained from Olyra for this study are not chimeras and that the phylogenetic affinities of each gene are with those of other taxa of the Olyreae and other elements of Bambusoideae, not with those of the PACMAD clade. Consequently, the structural similarities between the plastid genome of Olyra and those of Arundo and other taxa of the PACMAD clade are interpreted as having arisen independently.

The various migrations of gene termini relative to the two SSC-IR junctions are also of significance in terms of the phenomenon of gene overlap. When both ndhf and ndhH extend into the IR, one or more nucleotides in the IR simultaneously encode portions of both genes (Figs. 1, 2). As noted, the extent of this overlap is limited by the size of the smallest gene segment that extends into the IR. Within the present taxon sample, gene overlap arises at the point of origin of the PACMAD clade, where the 3′ terminus of ndhf enters the IR, which already includes the 5′ terminus of ndhH. At about this point in the history of the clade, most of the portion of ndhH previously included in the IR migrates into the SSC, leaving 1–4 nucleotides overlapping with ndhf. The overlap is lost multiple times within the clade, whenever the last few nucleotides of ndhH migrate into the SSC, and it is regained once in the PACMAD clade, in Eriachne pulchella.

Elsewhere in the taxon sample, there are three additional origins of gene overlap, in Ehrharta, Olyra, and Brachypodium. Olyra differs from its closest relatives in much the same way that taxa of the PACMAD clade differ from the early-diverging grass lineages and most of the BEP clade. As in the PACMAD clade, almost the entire portion of ndhH that once was situated in the IR has migrated into the SSC, and a portion of ndhf, less than 30 nucleotides in length, has migrated into the IR. As with the PACMAD clade, the occurrence of these events on the same branch of a cladogram does not indicate whether they occurred simultaneously or in succession, and in the latter case, which event may have occurred first. Thus, the co-occurrence of this pair of autapomorphic features in Olyra suggests either that a single mutational event mediated the origin of both features or that it is maladaptive for more than a few nucleotides of these two genes to overlap in this manner, possibly because it constrains the evolution of each. Under the latter interpretation, degrees of overlap of more than a few nucleotides, when they do arise, are soon eliminated by the migration of all or nearly all of one gene or the other out of the IR.

In Brachypodium, a similar pattern of overlap exists, but in this case it is ndhf that extends the shortest distance (one nucleotide) into the IR. The structure in B. pinnatum, as sampled here, differs substantially from those of its closest relatives in the Pooideae, and also from that of its close relative, B. distachyon (GenBank NC_011032; Bortiri et al., 2008), which also resembles other taxa of Pooideae, in having 209 nucleotides of ndhH in the IR, and ndhf confined to the SSC. Thus, both B. distachyon and Olyra latifolia exhibit substantially modified positions of the endpoints of ndhH and ndhf, with both genes extending into the IR, but with only one nucleotide of overlap. These patterns, along with the limited degree of overlap observed in taxa of the PACMAD clade, suggest that however an overlap may arise, it is maladaptive if too extensive and soon is lost as one gene or the other migrates back into the SSC.

The situation in Ehrharta, however, represents an exception to this pattern, with 43 nucleotides of the IR encoding portions of two different genes, on opposite DNA strands (and in opposite directions), including the termination codon of ndhf. Of these 43 nucleotides, 36 are interpreted as an insertion, but the entire 43-nucleotide region is homologous with the corresponding DNA strand in ndhH (Fig. 2), and no insertion is evident in that gene. In other words, the “insertion” in ndhf arose not by the expansion of a DNA region, but by the loss of a stop codon in ndhf (either through a point mutation or a small rearrangement that did not affect the length of the gene), resulting in the lengthening of the coding region to a point previously downstream from the 3′ terminus, where another stop codon was encountered.

Distributions of the three- and six-nucleotide inversions in ndhf (characters 2 and 6, respectively [Tables 1 and 2; Fig. 3]), as determined by the present analysis, correspond to those described previously (Davis and Soreng, 2007; Soreng et al., 2007). The three-nucleotide inversion is a synapomorphy of the clade within Pooideae that is the sister of Diarrhena (consisting of Bromeae, Triticeae, Poeae, and other small tribes), and there is a single reversion to the uninvverted state, in Hordeum. Within the clade that has the three-nucleotide inversion, the six-nucleotide inversion is a synapomorphy of the two representatives of Poeae subtribe Aveninae.

The 15-nucleotide indel in ndhf (character 15 [Tables 1, 2, Fig. 3]) was determined to be homoplasious when first reported by Clark et al. (1995), who observed that the deleted state occurs in close relatives of the grasses and early-diverging grass lineages, as well as in Oryzeae. Thus, the authors interpreted the deleted state as plesiomorphic in the grasses, with an insertion marking the clade that includes taxa now conventionally placed in the PACMAD and BEP clades, and with a reversion to the deleted state marking the Oryzeae. As suggested by the alignment presented here (Fig. 4), it is reasonable to recognize differences on an a priori basis between two forms of the deleted state. Under this determination, the deleted state in taxa of the BEP clade could be scored as a different character or state than for taxa outside the BEP clade. However, results of the present phylogenetic analysis, consistent with those supported by other analyses, still would suggest that the state that occurs in the BEP clade has arisen three times in parallel within this group (Fig. 3). These repeated insertion/deletion events may be attributable to slipped-strand mispairing, as has been inferred in similar instances (e.g., Levinson and Gutman, 1987; Cummings et al., 1994; Kelchner, 2000; Dertien and Duvall, 2009).

The remaining structural characters in the present analysis are three additional indels (characters 13, 14, and 17), each of which differentiates two taxa from disparate regions of the tree from the rest of the sample and thus exhibits two steps and an RI of 0 in the analysis (Tables 1, 2; Fig. 3). These results, like those for the positions of endpoints of ndhf and ndhH relative to the SSC-IR junctions, as well as for the three-nucleotide inversion in ndhf and the 15-nucleotide indel in ndhf, tend to confirm the propensity of structural mutations to arise independently, in particular locations of the genome, often in identical positions, and to exhibit subsequent reversals to their plesiomorphic states (e.g., Kelchner and Wendel, 1996; Graham et al., 2000; Tsumura et al., 2000; Kim and Lee, 2005; Bain and Jansen, 2006). Thus, although structural characters of this sort often provide useful phylogenetic evidence (e.g., Luo et al.,...
2006), they also are often individually homoplasious, like nucleotide sequence characters and even morphological characters, and are best interpreted in the context of a phylogeny based on a wide range of characters.

Nucleotide substitution rates—Comparison of the relative rates of nucleotide substitution within and outside the PACMAD clade, as determined for four gene regions, indicates that the two regions that lie within the SSC region in all or nearly all taxa, one of them a portion of ndhF, the other a portion of ndhH, evolve at similar rates (averages of 1.40 and 1.31 steps per nucleotide, respectively, across the same set of taxa; Table 4). Also, in both of these cases, the number of steps within the PACMAD clade is about one-third the number occurring elsewhere in the tree. The number of taxa in the PACMAD clade also is about one-third that of the number of taxa elsewhere in the tree, but the latter group is also a paraphyletic assemblage, and includes deeper branches in the tree, so these numbers do not provide absolute measures of evolutionary rates, but they do allow for comparisons of relative rates among gene regions across the same portions of the tree.

The portion of rps15 that was examined lies entirely within the IR region in all taxa and exhibits about one-third the total number of steps per character as the two gene regions that lie in the SSC region of the genome. This general pattern is consistent with previous observations that substitution rates in the IR regions of the plastid genome are substantially lower than those in the single-copy regions (e.g., Wolfe et al., 1987; Maier et al., 1995; Muse and Gaut, 1997; Yamane et al., 2006). The observed number of steps in this portion of rps15 within the PACMAD clade is only one-fifth of the number outside this clade, while the corresponding ratio is about one-third for the two gene regions that lie in the SSC region of the genome. However, this region of rps15 is only 122 nucleotides in length, so the relatively low observed rate for the PACMAD clade, relative to the rest of the tree, may not be an accurate indication of general substitution rates for genes in the IR region, or even for this gene. Leaving aside this matter of precision, the observation of substantially fewer steps within the PACMAD clade than outside of it is generally consistent with the patterns observed for the gene regions that lie within the SSC region of the genome.

A different pattern is observed for the portion of ndhH that lies in the SSC region in taxa of the PACMAD clade and in the IR region in taxa outside the PACMAD clade. Like the portion of rps15 that was examined, this gene region is relatively small (150 nucleotides), so the observed rates may not be precise indicators of actual substitution rates. The number of steps in this gene region among taxa of the PACMAD clade is about equal to the number of steps among taxa outside the PACMAD clade. With about 0.33 steps per nucleotide site within the PACMAD clade, the substitution rate for this portion of ndhH is comparable to the rates observed within the PACMAD clade for the two other gene regions that lie in the SSC region of the genome (0.32 and 0.35). Conversely, with about 0.31 steps per nucleotide outside the PACMAD clade, the substitution rate for this portion of ndhH is comparable to the rate observed outside the PACMAD clade for the other gene region that lies in the IR region of the genome (0.44). Thus, relative substitution rates of the various gene regions correspond to a general pattern in which those that lie within the IR region of the genome evolve more slowly than those that lie within the SSC region of the genome. Muse and Gaut (1997) noted that the migration of a gene from the IR into a single-copy region was likely to lead to acceleration of the substitution rate in that gene. As demonstrated here, migration of even a small portion of a gene from the IR into the SSC can be associated with an increase in nucleotide substitution rate in the portion of the gene that migrated.

LITERATURE CITED


National Herbarium family (Poaceae) as inferred from chloroplast DNA restriction site diverging angiosperms: Dioscorea inferred, from sequences. Cladistics for phylogenetic analysis. plastid data set.


### Appendix 1. Taxa sampled for DNA sequences, plant accession information (with herbarium codes, Holmgren and Holmgren, 1998), and GenBank accession numbers for ndhF, ndhH, rps15. Taxonomic scheme (see section Grass phylogenetics) includes parenthesized codes for tribes of Poaceae (three letters), and subtribes of tribe Poeae (four letters). For eight taxa with an asterisk, gene sequences were obtained from published plastid genome sequences.

<table>
<thead>
<tr>
<th>Taxonomic scheme</th>
<th>Species</th>
<th>Voucher (Herbarium)</th>
<th>ndhF</th>
<th>ndhH</th>
<th>rps15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Joinvilleae</td>
<td>Joinvillea gaudichaudiana</td>
<td>J.I. Davis 751 (BH)</td>
<td>GU222696</td>
<td>GU222836</td>
<td>GU222752</td>
</tr>
<tr>
<td></td>
<td>Ecdiocolae monostachya F. Muell.</td>
<td>J.G. Conran et al. 938 (PERTH, ADU)</td>
<td>AY622313</td>
<td>GU222837</td>
<td>GU222753</td>
</tr>
<tr>
<td>Poaceae</td>
<td>Anomochloideae (ano) Anomochloa marantoidaea Brongn.</td>
<td>J.I. Davis 753 (BH)</td>
<td>GU222697</td>
<td>GU222838</td>
<td>GU222754</td>
</tr>
<tr>
<td></td>
<td>Streptochaeteeae (str) Streptocheta sordiroa Hack.</td>
<td>J.I. Davis 5952 (US)</td>
<td>AY622318</td>
<td>GU222839</td>
<td>GU222755</td>
</tr>
<tr>
<td></td>
<td>Pharoideae</td>
<td>Pharus latifolius L.</td>
<td>J.I. Davis; R.J. Soreng; no voucher</td>
<td>GU222698</td>
<td>GU222840</td>
</tr>
<tr>
<td></td>
<td>Arundineae (aru)</td>
<td>Arundo donax L.</td>
<td>M. Crisp 278 (CANB)</td>
<td>GU222699</td>
<td>GU222841</td>
</tr>
<tr>
<td></td>
<td>Molinea caerulea (L.) Moench</td>
<td>R.J. Soreng 3305; no voucher</td>
<td>GU222715</td>
<td>GU222878</td>
<td>GU222759</td>
</tr>
<tr>
<td></td>
<td>Micraioideae</td>
<td>Micraira subulifolia F. Muell.</td>
<td>S.W.L. Jacobs 8770 (NSW)</td>
<td>GU222694</td>
<td>GU222860</td>
</tr>
<tr>
<td></td>
<td>Arundinarieae</td>
<td>Arundo donax L.</td>
<td>M. Crisp 278 (CANB)</td>
<td>GU222691</td>
<td>GU222861</td>
</tr>
<tr>
<td></td>
<td>Ehrharteae</td>
<td>Ehrharta calycina Sm.</td>
<td>J.I. Davis; no voucher</td>
<td>GU222719</td>
<td>GU222862</td>
</tr>
<tr>
<td></td>
<td>Ehrhartoideae</td>
<td>Ehrhartoideae</td>
<td>J.I. Davis; no voucher</td>
<td>GU222701</td>
<td>GU222853</td>
</tr>
<tr>
<td></td>
<td>Danthoniaeae</td>
<td>Danthonia californica Bol.</td>
<td>grown from USDA Plant Intr. Sta. 232247; J.I. Davis 763 (BH)</td>
<td>GU222712</td>
<td>GU222854</td>
</tr>
<tr>
<td></td>
<td>Danthoniaeae (dan)</td>
<td>Danthonia californica Bol.</td>
<td>grown from USDA Plant Intr. Sta. 232247; J.I. Davis 763 (BH)</td>
<td>GU222712</td>
<td>GU222854</td>
</tr>
<tr>
<td></td>
<td>Chloridoideae</td>
<td>Mesuella laevis (Pilg.) Conert</td>
<td>N.P. Barker 960 (GRA)</td>
<td>GU222704</td>
<td>GU222846</td>
</tr>
<tr>
<td></td>
<td>Distichlis stricta (Torr.) R.F. Thorne</td>
<td>grown from commercial seed; J.I. Davis 771 (BH)</td>
<td>GU222708</td>
<td>GU222850</td>
<td>GU222766</td>
</tr>
<tr>
<td></td>
<td>Eragrostideae (era) Eragrostis tef (Zucc.) Trotter</td>
<td>J.I. Davis; no voucher</td>
<td>GU222707</td>
<td>GU222849</td>
<td>GU222765</td>
</tr>
<tr>
<td></td>
<td>Zoysieae (zoy)</td>
<td>Zoysia japonica (L.)Subjects</td>
<td>grown from USDA Plant Intr. Sta. 421520; R.J. Soreng s.n. (BH)</td>
<td>GU222709</td>
<td>GU222851</td>
</tr>
<tr>
<td></td>
<td>Paniceae (pan)</td>
<td>Panicum virgatum</td>
<td>&amp; R.J. Soreng s.n. (BH)</td>
<td>GU222702</td>
<td>GU222844</td>
</tr>
<tr>
<td></td>
<td>Paniceae (pan) Panicum virgatum</td>
<td>J.I. Davis; no voucher</td>
<td>GU222710</td>
<td>GU222852</td>
<td>GU222768</td>
</tr>
<tr>
<td></td>
<td>Centotheceae (cen) Chasmanthium nitidum (Balldwin) H.O. Yates</td>
<td>J.K. Wipff &amp; S.D. Jones 2075 (TAES)</td>
<td>GU222699</td>
<td>GU222841</td>
<td>GU222757</td>
</tr>
<tr>
<td></td>
<td>Thysanolaeneae (thy)</td>
<td>Thysanolaena maxima (Roxb.) Kunze</td>
<td>grown from USDA Plant Intr. Sta. 421520; R.J. Soreng s.n. (BH)</td>
<td>GU222703</td>
<td>GU222845</td>
</tr>
<tr>
<td></td>
<td>Gynerieae (gyn)</td>
<td>Gynerium sagittatum (Aubl.) P. Beauv.</td>
<td>grown from USDA Plant Intr. Sta. 421520; R.J. Soreng s.n. (BH)</td>
<td>GU222702</td>
<td>GU222844</td>
</tr>
<tr>
<td></td>
<td>Paniceae (pan)</td>
<td>Panicum virgatum</td>
<td>J.C. LaDuke; no voucher</td>
<td>GU222706</td>
<td>GU222848</td>
</tr>
<tr>
<td></td>
<td>Pennisetum alopecroideus (L.) Spreng.</td>
<td>grown from USDA Plant Intr. Sta. 421520; R.J. Soreng s.n. (BH)</td>
<td>GU222702</td>
<td>GU222844</td>
<td>GU222760</td>
</tr>
<tr>
<td></td>
<td>Andropogoneae (and) *Saccarum officinarum L.</td>
<td>n/a</td>
<td>NC_006084</td>
<td>NC_006084</td>
<td>NC_006084</td>
</tr>
<tr>
<td></td>
<td>*Sorghum bicolor (L.)</td>
<td>Moench</td>
<td>n/a</td>
<td>NC_006062</td>
<td>NC_006062</td>
</tr>
<tr>
<td></td>
<td>*Zea mays L.</td>
<td>n/a</td>
<td>NC_001666</td>
<td>NC_001666</td>
<td>NC_001666</td>
</tr>
<tr>
<td>Bambusoideae</td>
<td>Bambuseae (bam) Phyllostachys Siebold &amp; Zucc. sp.</td>
<td>J.I. Davis 773 (BH)</td>
<td>GU222722</td>
<td>GU222865</td>
<td>GU222781</td>
</tr>
<tr>
<td></td>
<td>Chusquea aff. subulata L.G. Clark</td>
<td>grown from USDA Plant Intr. Sta. 153; or US National Herbarium Greenhouse 286; no voucher</td>
<td>GU222775</td>
<td>GU222870</td>
<td>GU222786</td>
</tr>
<tr>
<td></td>
<td>Guada angustifolia Kunth</td>
<td>grown from USDA Plant Intr. Sta. 421520; R.J. Soreng s.n. (BH)</td>
<td>GU222723</td>
<td>GU222866</td>
<td>GU222782</td>
</tr>
<tr>
<td></td>
<td>Elyreae (oly) Buergeriarchaobia bambusoides Pilg. Eremitis Döll sp.</td>
<td>grown from USDA Plant Intr. Sta. 208983; R.J. Soreng s.n. (BH)</td>
<td>GU222719</td>
<td>GU222862</td>
<td>GU222778</td>
</tr>
<tr>
<td>Ehrhartoideae</td>
<td>Streptogyne americana C.E. Hubb.</td>
<td>J.I. Davis; no voucher</td>
<td>GU222721</td>
<td>GU222864</td>
<td>GU222780</td>
</tr>
<tr>
<td></td>
<td>Ehrharta calycina Sm.</td>
<td>grown from USDA Plant Intr. Sta. 208983; R.J. Soreng s.n. (BH)</td>
<td>GU222719</td>
<td>GU222862</td>
<td>GU222778</td>
</tr>
<tr>
<td>Taxonomic scheme</td>
<td>Species</td>
<td>Voucher (Herbarium)</td>
<td>ndhF</td>
<td>ndhH</td>
<td>rps15</td>
</tr>
<tr>
<td>------------------------</td>
<td>--------------------------------------</td>
<td>---------------------</td>
<td>---------------</td>
<td>---------------</td>
<td>--------------</td>
</tr>
<tr>
<td><strong>Oryzeae (ory)</strong></td>
<td>Leersia virginica Wild.</td>
<td>R.J. Soreng 3399a (BH)</td>
<td>GU222720</td>
<td>GU222863</td>
<td>GU222779</td>
</tr>
<tr>
<td></td>
<td><em>Oryza nivara</em> Sharma &amp; Shastry</td>
<td>n/a</td>
<td>NC_005973</td>
<td>NC_005973</td>
<td>NC_005973</td>
</tr>
<tr>
<td></td>
<td><em>Oryza sativa</em> L.</td>
<td>n/a</td>
<td>NC_001320</td>
<td>NC_001320</td>
<td>NC_001320</td>
</tr>
<tr>
<td><strong>Pooidae</strong></td>
<td><strong>Brachylyceraceae (brl)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Brachyelytrum erectum (Schreb.)</td>
<td>R.J. Soreng 3427a (BH)</td>
<td>GU222731</td>
<td>GU222874</td>
<td>GU222790</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P. Beauv.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Narceae (nar)</strong></td>
<td>Nardus stricta L.</td>
<td>E. Roy &amp; C. Schiers s.n. (1988, B)</td>
<td>GU222733</td>
<td>GU222876</td>
<td>GU222792</td>
</tr>
<tr>
<td></td>
<td><strong>Lygeaceae (lyg)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lygeum spartum L.</td>
<td>R.J. Soreng 3427a (BH)</td>
<td>GU222736</td>
<td>GU222879</td>
<td>GU222795</td>
</tr>
<tr>
<td></td>
<td>Anisopogon avenaceus R. Br.</td>
<td>R.J. Soreng 5358 (US)</td>
<td>GU222737</td>
<td>GU222880</td>
<td>GU222796</td>
</tr>
<tr>
<td></td>
<td>Duthiea brachypodium (P. Candargy) Keng &amp; Keng f.</td>
<td>L.G. Clark 1292 (ISC)</td>
<td>GU222734</td>
<td>GU222877</td>
<td>GU222793</td>
</tr>
<tr>
<td></td>
<td>Phaenosperma globosa Benth.</td>
<td>R.J. Soreng 5644 (US)</td>
<td>GU222735</td>
<td>GU222878</td>
<td>GU222794</td>
</tr>
<tr>
<td><strong>Stipeae (sti)</strong></td>
<td>Achnatherum occidentalis</td>
<td>R.J. Soreng 7418 (US)</td>
<td>GU222739</td>
<td>GU222882</td>
<td>GU222798</td>
</tr>
<tr>
<td></td>
<td>Ampelodesmos mauritanica</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(Poir.) T. Durand &amp; Schinz</td>
<td>R.J. Soreng 3427a (BH)</td>
<td>GU222746</td>
<td>GU222890</td>
<td>GU22806</td>
</tr>
<tr>
<td><strong>Brylkinieae (bry)</strong></td>
<td>Brylkinia caudata (Munro)</td>
<td>Gao Hui 167 (SZ)</td>
<td>GU222750</td>
<td>GU222896</td>
<td>GU22812</td>
</tr>
<tr>
<td></td>
<td>F. Schmidt</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Meliceae (mel)</strong></td>
<td>Glyceria grandis S. Watson</td>
<td>J.I. Davis &amp; R.J. Soreng; no voucher</td>
<td>AY622314</td>
<td>GU222894</td>
<td>GU22810</td>
</tr>
<tr>
<td></td>
<td>Melica cupanii Guss.</td>
<td></td>
<td>AY622315</td>
<td>GU222892</td>
<td>GU22808</td>
</tr>
<tr>
<td></td>
<td>Pleurospogon refractus (A. Gray) Benth.</td>
<td>R.J. Soreng 3381 (BH)</td>
<td>GU222745</td>
<td>GU222895</td>
<td>GU22811</td>
</tr>
<tr>
<td></td>
<td>Schizachne purpurascens (Torr.) Swallen</td>
<td>R.J. Soreng 3348 (BH)</td>
<td>GU222748</td>
<td>GU222893</td>
<td>GU22809</td>
</tr>
<tr>
<td><strong>Diarrheneae (dia)</strong></td>
<td>Diarrhena obovata (Gleason)</td>
<td>J.I. Davis 756 (BH)</td>
<td>DQ786833</td>
<td>GU222897</td>
<td>GU22813</td>
</tr>
<tr>
<td></td>
<td>Brandenberg</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Brachypodieae (brp)</strong></td>
<td>Brachypodium pinnatum (L.) P. Beav.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bromus inermis Leyss.</td>
<td>J.I. Davis 760 (BH)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bromus koroskii Drobot</td>
<td>R.J. Soreng 5160 (US)</td>
<td>GU222751</td>
<td>GU222903</td>
<td>GU22819</td>
</tr>
<tr>
<td></td>
<td>Bromus suksdorfi Vasey</td>
<td>R.J. Soreng 7412 (US)</td>
<td>DQ786822</td>
<td>GU222902</td>
<td>GU22818</td>
</tr>
<tr>
<td></td>
<td>Littledalea tibetica Hemsli.</td>
<td>R.J. Soreng 5487; 5490; 5494 (US)</td>
<td>DQ786852</td>
<td>GU222899</td>
<td>GU22815</td>
</tr>
<tr>
<td><strong>Triticeae (tri)</strong></td>
<td>Elymus trachycaulis (Link)</td>
<td>R.J. Soreng 4291b (BH)</td>
<td>DQ786838</td>
<td>GU222900</td>
<td>GU22816</td>
</tr>
<tr>
<td></td>
<td><em>Hordeum vulgare</em> L. subsp. vulgar</td>
<td>n/a</td>
<td>NC_008590</td>
<td>NC_008590</td>
<td>NC_008590</td>
</tr>
<tr>
<td></td>
<td><em>Triticum aestivum</em> L.</td>
<td>n/a</td>
<td>NC_002762</td>
<td>NC_002762</td>
<td>NC_002762</td>
</tr>
<tr>
<td><strong>Poaceae (poe)</strong></td>
<td>Agrostisidae (agro)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Agrostis stolonifera</em> L.</td>
<td>n/a</td>
<td>NC_008591</td>
<td>NC_008591</td>
<td>NC_008591</td>
</tr>
<tr>
<td></td>
<td>Aveninae (aven)</td>
<td></td>
<td>DQ786814</td>
<td>GU222904</td>
<td>GU22820</td>
</tr>
<tr>
<td></td>
<td><em>Avena sativa</em> L. <em>ASTRO</em></td>
<td></td>
<td>DQ786874</td>
<td>GU222905</td>
<td>GU22821</td>
</tr>
<tr>
<td></td>
<td>Trisetum cernuum subsp. cansens</td>
<td>R.J. Soreng 3383a (BH)</td>
<td>DQ786874</td>
<td>GU222905</td>
<td>GU22821</td>
</tr>
<tr>
<td></td>
<td><em>Briza minor</em> L.</td>
<td></td>
<td>DQ786820</td>
<td>GU222907</td>
<td>GU22823</td>
</tr>
<tr>
<td><strong>Brizinae (briz)</strong></td>
<td>Briza minor L.</td>
<td></td>
<td>DQ786872</td>
<td>GU222906</td>
<td>GU22822</td>
</tr>
<tr>
<td><strong>Astyloidea (torr)</strong></td>
<td>Torreyochloa pauciflora (J. Presl)</td>
<td>J.I. Davis 761 (BH)</td>
<td>DQ786820</td>
<td>GU222907</td>
<td>GU22823</td>
</tr>
<tr>
<td></td>
<td>G.L. Church</td>
<td></td>
<td>DQ786872</td>
<td>GU222906</td>
<td>GU22822</td>
</tr>
<tr>
<td><strong>Airinae (airi)</strong></td>
<td>Aira caryophylla L.</td>
<td>R.J. Soreng 5953b (US)</td>
<td>DQ786806</td>
<td>GU222908</td>
<td>GU22824</td>
</tr>
<tr>
<td></td>
<td>Deschampsia cespitosa (L.)</td>
<td>R.J. Soreng 7417 (US)</td>
<td>DQ786831</td>
<td>GU222913</td>
<td>GU22829</td>
</tr>
<tr>
<td></td>
<td>P. Beav. subsp. cespitosa</td>
<td></td>
<td>DQ786820</td>
<td>GU222907</td>
<td>GU22823</td>
</tr>
</tbody>
</table>
Appendix 1. Continued.

<table>
<thead>
<tr>
<th>Taxonomic scheme</th>
<th>Species</th>
<th>Voucher (Herbarium)</th>
<th>ndhF</th>
<th>ndhH</th>
<th>rps15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alopecurinae (alop)</td>
<td>Molineriella laevis (Brot.) Rouy</td>
<td>R.J. Soreng 3613 (BH)</td>
<td>DQ786857 GU222915 GU222831</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cynosurinae (cyno)</td>
<td>Phleum pratense L.</td>
<td>R.J. Soreng 4293 (BH)</td>
<td>DQ786860 GU222911 GU222827</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dactylidinae (dact)</td>
<td>Cynosurus cristatus L.</td>
<td>grown from RBG, Kew, seed bank 39006 (K)</td>
<td>DQ786829 GU222918 GU222834</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Holcineae (holc)</td>
<td>Holcus annuus C.A. Mey.</td>
<td>R.J. Soreng 3642 (BH)</td>
<td>DQ786849 GU222914 GU222830</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lolini (loli)</td>
<td>Festuca rubra L.</td>
<td>R.J. Soreng 4293 (BH)</td>
<td>DQ786839 GU222916 GU222832</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parapholiinae (para)</td>
<td>Parapholis incurva (L.) C.E. Hubb.</td>
<td>grown from RBG, Kew, seed bank 24867 (K)</td>
<td>DQ786859 GU222919 GU222835</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poiae (poin)</td>
<td>Puccinellia distans (Jacq.) Parl.</td>
<td>J.I. Davis 755 (BH)</td>
<td>DQ786866 GU222909 GU222825</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Puccinellinae (puc)</td>
<td>Puccinellia distans (Jacq.) Parl.</td>
<td>J.I. Davis 755 (BH)</td>
<td>DQ786866 GU222909 GU222825</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sclerochloa dura</td>
<td>Sclerochloa dura (L.) P. Beauv.</td>
<td>R.J. Soreng 3862 (BH)</td>
<td>DQ786869 GU222910 GU222826</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Appendix 2. Primers used to amplify and sequence two regions of the plastid genome (cf. text and Fig. 1). Each primer name consists of (1) the name of the gene to which it is specific; (2) the numerical position within the primer binding region of the nucleotide closest to the 5′ end of the corresponding gene (regardless of the direction in which the primer reads), in the plastid genome sequence of *Triticum aestivum* (GenBank accession number NC_002762); and 3) either F (forward) or R (reverse), designating the direction in which the priming function proceeds, relative to the direction in which the gene is transcribed. Primers were developed by the authors, except as indicated.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Region 1 (ndhF and rps15)</td>
<td></td>
</tr>
<tr>
<td><em>ndhF</em>-1F (Olmstead and Sweere, 1994)</td>
<td>5′ atg gaa caK aca tat Saat tat gc 3′</td>
</tr>
<tr>
<td><em>ndhF</em>-45F</td>
<td>5′ act tcc agg tat tat gtc aat ggg Rtt t 3′</td>
</tr>
<tr>
<td><em>ndhF</em>-274F (modified from Olmstead and Sweere, 1994)</td>
<td>5′ ctt act tct att atg tta ata cta at 3′</td>
</tr>
<tr>
<td><em>ndhF</em>-309F</td>
<td>5′ Wgg aaY Yat ggt tct tat tta tag tga c 3′</td>
</tr>
<tr>
<td><em>ndhF</em>-532F</td>
<td>5′ gcK ttt Dta act aat cgt gta ggg ga 3′</td>
</tr>
<tr>
<td><em>ndhF</em>-818F</td>
<td>5′ gaa ttt tic tiv tag ctc gag tY YtY ttc 3′</td>
</tr>
<tr>
<td><em>ndhF</em>-933F</td>
<td>5′ tea Rag ago tat taa aag Ytt age c 3′</td>
</tr>
<tr>
<td><em>ndhF</em>-972F</td>
<td>5′ att ggg tta tat gat gtt agc tct agg t 3′</td>
</tr>
<tr>
<td><em>ndhF</em>-1117R (Olmstead and Sweere, 1994)</td>
<td>5′ gta ttt ttc ttV tag ctc gag ttY ttc 3′</td>
</tr>
<tr>
<td><em>ndhF</em>-1373R</td>
<td>5′ att ggg tta tat gat gtt agc tct agg t 3′</td>
</tr>
<tr>
<td><em>ndhF</em>-1421F</td>
<td>5′ gta ttt ttc ttV tag ctc gag tY YtY ttc 3′</td>
</tr>
<tr>
<td><em>ndhF</em>-1811F</td>
<td>5′ act tta act gcV ttG tat atg ttt cg 3′</td>
</tr>
<tr>
<td><em>ndhF</em>-1968R</td>
<td>5′ tta act tta act gcV ttG tat atg ttt cg 3′</td>
</tr>
<tr>
<td><em>ndhF</em>-2122F (Olmstead and Sweere, 1994)</td>
<td>5′ gta ttt ttc ttV tag ctc gag tY YtY ttc 3′</td>
</tr>
<tr>
<td><em>ndhF</em>-2101F</td>
<td>5′ gaa ttt ttc ttV tag ctc gag tY YtY ttc 3′</td>
</tr>
<tr>
<td><em>ndhF</em>-978F</td>
<td>5′ tea Rag ago tat taa aag Ytt age c 3′</td>
</tr>
<tr>
<td><em>ndhF</em>-1194F</td>
<td>5′ app ggg tta tat gat gtt agc tct agg t 3′</td>
</tr>
<tr>
<td><em>ndhF</em>-1318F (modified from Olmstead and Sweere, 1994)</td>
<td>5′ gta ttt ttc ttV tag ctc gag tY YtY ttc 3′</td>
</tr>
<tr>
<td><em>ndhF</em>-1117R</td>
<td>5′ gta ttt ttc ttV tag ctc gag tY YtY ttc 3′</td>
</tr>
<tr>
<td><em>ndhH</em>-680F</td>
<td>5′ gta ttt ttc ttV tag ctc gag tY YtY ttc 3′</td>
</tr>
<tr>
<td><em>rps15</em>-808F</td>
<td>5′ gta ttt ttc ttV tag ctc gag tY YtY ttc 3′</td>
</tr>
<tr>
<td>Region 2 (ndhH and rps15)</td>
<td></td>
</tr>
<tr>
<td><em>ndhA</em>-59R</td>
<td>5′ atc Sat atc agg cca tat act tct tta at 3′</td>
</tr>
<tr>
<td><em>ndhH</em>-684R</td>
<td>5′ atc Sat atc agg cca tat act tct tta at 3′</td>
</tr>
<tr>
<td><em>ndhH</em>-88F</td>
<td>5′ gta ttt ttc ttV tag ctc gag tY YtY ttc 3′</td>
</tr>
<tr>
<td><em>rps15</em>-808F</td>
<td>5′ gta ttt ttc ttV tag ctc gag tY YtY ttc 3′</td>
</tr>
</tbody>
</table>