Slipped-Strand Mispairing in a Plastid Gene: *rpo*C2 in Grasses (Poaceae)

Michael P. Cummings, Lynn Mertens King,* and Elizabeth A. Kellogg†*

*Museum of Comparative Zoology, Harvard University; and †Arnold Arboretum, Harvard University

An exception to the generally conservative nature of plastid gene evolution is the gene coding for the β″ subunit of RNA polymerase, *rpo*C2. Previous work by others has shown that maize and rice have an insertion in the coding region of *rpo*C2, relative to spinach and tobacco. To assess the distribution of this extra coding sequence, we surveyed a broad phylogenetic sample comprising 55 species from 17 angiosperm families by using Southern hybridization. The extra coding sequence is restricted to the grasses (Poaceae). DNA sequence analysis of 11 species from all five subfamilies within the grass family demonstrates that the extra sequence in the coding region of *rpo*C2 is a repetitive array that exhibits more than a twofold increase in nucleotide substitution, as well as a large number of insertion/deletion events, relative to the adjacent flanking sequences. The structure of the array suggests that slipped-strand mispairing causes the repeated motifs and adds to the mechanisms through which the coding sequence of plastid genes are known to evolve. Phylogenetic analyses based on the sequence data from grass species support several relationships previously suggested by morphological work, but they are ambiguous about broad relationships within the family.

Introduction

The plastid genomes of higher plants are generally conserved in length, structure, and gene complement (Curtis and Clegg 1984; Clegg et al. 1991). At the DNA sequence level, plastid genes display a low rate of nucleotide substitution, compared with plant nuclear genes (Curtis and Clegg 1984; Wolfe et al. 1987; Clegg et al. 1991). A striking exception to this conservation is the gene encoding the β″ subunit of RNA polymerase, *rpo*C2. In comparison with tobacco (*Nicotiana tabacum*) and spinach (*Spinacia oleracea*) (Shinozaki et al. 1986; Hudson et al. 1988), *rpo*C2 of maize (*Zea mays*) and rice (*Oryza sativa*) has an extra coding sequence (Igloi et al. 1990; Shimada et al. 1990). This extra coding sequence occurs near the center of the gene and increases its length by ~9.5% in rice and ~10.6% in maize; it appears to have a repetitive structure.

As a necessary part of the plastid-gene transcription apparatus, the *rpo* genes and their encoded products, the subunits of RNA polymerase, have been the subject of much study (reviewed in Igloi and Kössel 1992). Four subunits—α, β, β′, and β″—are encoded by the genes *rpo*A, *rpo*B, *rpo*C1, and *rpo*C2, respectively, which are located in the large single-copy region of the plastid genome in angiosperms (Hudson et al. 1988; Hu and Bogorad 1990; Igloi and Kössel 1992). Transcript-mapping experiments in spinach have shown *rpo*B, *rpo*C1, and *rpo*C2, which are arranged in this order and are separated by short intergenic spacers, to be cotranscribed (Hudson et al. 1988). In maize, protein sequencing of purified polypeptides has confirmed that 38-, 120-, and 180-kD polypeptides are encoded by *rpo*A, *rpo*B, and *rpo*C2, respectively (Hu and Bogorad 1990). These polypeptides and a 135-kD polypeptide presumably encoded for by *rpo*C1 have transcription-initiation activity (Hu and Bogorad 1990).

A combination of Southern blot hybridization and DNA sequence analysis was used to characterize the variation and to determine the phylogenetic distribution of the additional sequence region of *rpo*C2. The goals of the study were to understand the molecular mechanisms that generate sequence variation in the usually highly conserved plastid genome and to investigate the use of this additional coding sequence for addressing questions about phylogenetic relationships within the grass family.

Material and Methods

DNA Manipulations

Total cellular DNA representing 55 species from 17 families (see the Appendix) either was prepared from

Key words: molecular evolution, RNA polymerase, systematics.

Address for correspondence and reprints: Michael P. Cummings, Department of Integrative Biology, University of California, Berkeley, California 94720.

Mol. Biol. Evol. 11(1):1–8. 1994.

0737-4038/94/1101-0001$02.00

fresh tissue by using methods described elsewhere (King and Schaal 1990) or was provided by several colleagues. Approximately 500 ng of DNA was digested with *Pst*I or *Eco*RI, and the restriction fragments were separated by electrophoresis using 0.8% or 0.9% agarose gels and a 1 × Tris-borate ethylenediaminetetraacetic acid buffer system. The gels were denatured, neutralized, and blotted on nylon membranes according to standard procedures (Maniatis et al. 1982, pp. 382–386). The blots were baked for 2 h at 80°C in vacuum, prehybridized for 2 h at 65°C in hybridization solution (Church and Gilbert 1984), and hybridized at 45°C overnight in fresh hybridization solution. An oligonucleotide probe was designed using the published sequences for the extra coding region of rice and maize (Igloi et al. 1990; Shimada et al. 1990). From these aligned sequences the following sequence was chosen for the probe: 5′-GATCCTAGAGGACGAATATGGACTCTAGAGAAAGACTCAGGAAGAATATGGGAGCCCAGAGAACG-3′. This sequence does not correspond exactly to either the rice or maize sequence but is similar enough to provide a strong hybridization signal under the conditions used in the study. The probe was end-labeled using T4 polynucleotide kinase, and the blots were washed for 30 min at 45°C in 2 × sodium chloride sodium citrate, 0.1% sodium dodecyl sulfate and for 30 min at 45°C in 1 × sodium chloride sodium citrate, 0.1% sodium dodecyl sulfate and were rinsed and exposed to Kodak XAR film at −70°C.

Oligonucleotide primers for polymerase chain reaction (PCR) amplifications were based on previously determined sequences of tobacco, spinach, maize, and rice (Shinozaki et al. 1986; Hudson et al. 1988; Igloi et al. 1990; Shimada et al. 1990) and were designed to include the extra coding sequence and adjacent flanking sequence (56–63 bp 5′ and 45–46 bp 3′). The specific primer sequences used were 5′-CGGAATTCTTTTACGTAGAAATACTA-3′ and 5′-CGGTCGACTTGTTCCTCGATGCTCAA-3′. PCR amplifications were performed in 25-μl reactions with 0.1–1 μg of genomic DNA, 200 μM of each dNTP, 50 pmol of each primer, 2–3 units of Vent DNA polymerase, and 2.5 μl of buffer supplied by the manufacturer (New England BioLabs). To prevent contamination, the reaction mix was treated with ultraviolet light prior to the addition of genomic DNA and enzyme (Sarkar and Sommer 1990), and negative control reactions (without added template DNA) were included with each set of amplifications. Temperature cycling was performed on an Ericomp thermocycler with the following profile: 94°C for 2 min initial denaturation and then 30–50 cycles of 93°C for 30 s, 47°C for 2 min, and 72°C for 30 s.

PCR products were agarose gel purified and digested with *Eco*RI and *Sal*I (recognition sites for these restriction enzymes were incorporated into the primer sequences). Fragments were purified by adhesion to glass powder (Vogelstein and Gillespie 1979) and were ligated into appropriately digested M13mp18 and mp19 vectors, and DNA sequences were determined (Ausubel et al. 1992, pp. 7.4.1–7.4.17) for at least two clones of each of 11 grass species.

Phylogenetic Analyses

Preliminary sequence manipulations were done using the program EUGENE (Lark Sequencing Technologies). DNA sequences were aligned using the multiple-alignment algorithm of the program CLUSTAL V (Higgins et al. 1992) with a fixed-gap penalty of 35, a floating-gap penalty of 35, and transitions and transversions equally weighted. These settings were chosen after many alignments were explored using a broad range of parameters. The aligned sequences, including both the extra coding region and adjacent flanking sequence but excluding the bases in the primers, formed the data for phylogenetic analysis, which was conducted using the program PAUP (Swofford 1992). Each nucleotide position was scored as a uniformly weighted character, with gaps scored as missing data; character-state transitions were uniformly weighted. The branch-and-bound algorithm was used both to find the most parsimonious tree and for bootstrap analysis.

Phylogenetic analysis based on parsimony assumes strict homology of the aligned nucleotides for DNA sequence data. Since the extra coding region of *rpo*C2 comprises tandem repeats, strict homology of specific repeat units is difficult to determine across all sequences in a multiple alignment. Therefore a second phylogenetic analysis was conducted that did not require the determination of strict positional homology across all the sequences simultaneously. For this analysis, all pairwise alignments were performed using the multiple-alignment algorithm of CLUSTAL V (Higgins et al. 1992). A fixed-gap penalty of 10, a floating-gap penalty of 10, and differential weighting of transitions and transversions were used for alignments involving pairs of grass species. Because of the difference in sequence lengths of spinach and tobacco, compared with those of the grass species, pairwise alignments involving either of these dicots used a fixed-gap penalty of 40 and a floating-gap penalty of 40, with transitions and transversions differentially weighted. Distance values were determined from the pairwise alignments by ignoring gapped positions and using the two-parameter method of Kimura (1980). A phylogenetic tree was constructed from the distances ($\times 10^4$) by using the neighbor-joining method (Saitou and Nei 1987) as modified by Studier and Keppler (1988); a program was supplied by J. A. Studier. All trees were rooted using tobacco and spinach, the only

angiosperm sequences available outside the grasses, as outgroups; for the neighbor-joining tree, this was equivalent to midpoint rooting.

Results

Distribution of the Additional Sequence in the Coding Region

A total of 55 species from 17 families including both monocots and dicots was screened, using Southern blot hybridization, for the presence of the insert (see the Appendix). The survey of species from this wide phylogenetic spectrum demonstrates that the extra sequence in the coding region is restricted to the grasses (Poaceae). Examination of the immediate outgroups to the grasses, Joinvilleaceae (*Joinvillea* sp.) and Restionaceae (*Elegia capensis* and *Rhodocoma gigantea*), shows no evidence of the additional coding sequence. A broad sample of 32 species representing all five subfamilies within the grasses shows that the additional sequence in the coding region is distributed throughout the family. A subset of the same samples was surveyed with PCR, and the results were concordant with the Southern blot hybridizations.

Structural Organization of the Additional Sequence in the Coding Region

Nucleotide sequence of the additional coding region was determined for 11 species. An alignment of these sequences, along with those previously determined for rice and maize, and the corresponding flanking regions from spinach and tobacco, are given in figure 1. The extra coding region ranges in length from 336 bp in *Briza maxima* to 438 bp in *Zea mays*. The original analysis of the rice sequence (Shimada et al. 1990) indicated the length of the additional coding sequence to be 381 bp; however, the alignments in the present analysis show it to be 393 bp.

The extra coding region is composed of tandem repeats in a complex arrangement, with some partially repeated units. The most common repeat unit length is 21 bp, although some individual repeats are shorter or longer. The plurality for the basic repeat unit is 5′-TATGGAACCCTAGAGGAAGAA-3′. *Briza* has ~14 repeat units, and maize has ~19 units. The repeat unit is present immediately 3′ of the extra region and is also present in spinach and tobacco.

Phylogenetic Analyses

Phylogenetic analysis based on the multiple alignment resulted in a single most parsimonious tree (fig. 2*A*). A bootstrap analysis showed that, while the monophyly of some groups is fairly robust, the relationships of many clades are tentative (fig. 2*B*). Additional bootstrap analyses of all other alignments examined, based on a wide range of gap weights, showed the same relationships supported as are shown in figure 2*B*. The neighbor-joining tree (fig. 2*C*) based on distances from the pairwise alignments differed in several ways from the most parsimonious tree, including the placement of *Zea* and *Pennisetum* as sister taxa.

Discussion

The Additional Sequence in *rpo*C2 Is Restricted to the Poaceae

The distribution pattern suggests that the origin of the additional coding sequence postdates the divergence of grasses from their immediate outgroups but predates the divergence of the family itself. Its distribution thus parallels morphological characters such as the caryopsis and the scutellum and lodicules (see Campbell and Kellogg 1987), and, at the molecular level, a small (~2-kb) inversion in the plastid genome (Doyle et al. 1992).

The Additional Sequence in *rpo*C2 Is Highly Variable

Relative to other plastid sequences, which in general are evolutionarily conserved (Curtis and Clegg 1984; Wolfe et al. 1987; Zurawski and Clegg 1987; Clegg et al. 1991), the extra coding region of *rpo*C2 is quite variable between species. This is apparent when variation in the additional sequence region is compared with that in the flanking sequences. Estimates of the proportion of nucleotide differences between all pairs of sequences, with the extra coding and flanking regions analyzed separately (summarized in table 1), show that the extra region exhibits more than a twofold increase in nucleotide substitutions.

In addition to an elevated rate of nucleotide substitution, this region exhibits a high level of insertion/deletion events (fig. 1). Although the number and pattern of gaps in the alignment (indicating insertion/deletion events) is dependent on the method and weights used to obtain the alignment, all alignments obtained using a wide range of gap weights had multiple nonoverlapping gaps, as well as overlapping gaps of varied lengths. Both of these features suggest that numerous insertion/deletion events have occurred. These insertion/deletion events cause an observed range of >30% in length for the extra coding sequence, across the species. Another measure of this sequence-length variation resulting from insertion/deletion events is the coefficient of variation, given in table 1, which shows the extra sequence region to be approximately 10-fold more variable than the flanking region.

Mechanism of Evolution

On the basis of the repetitive structure of the additional coding sequence of *rpo*C2, slipped-strand mis-

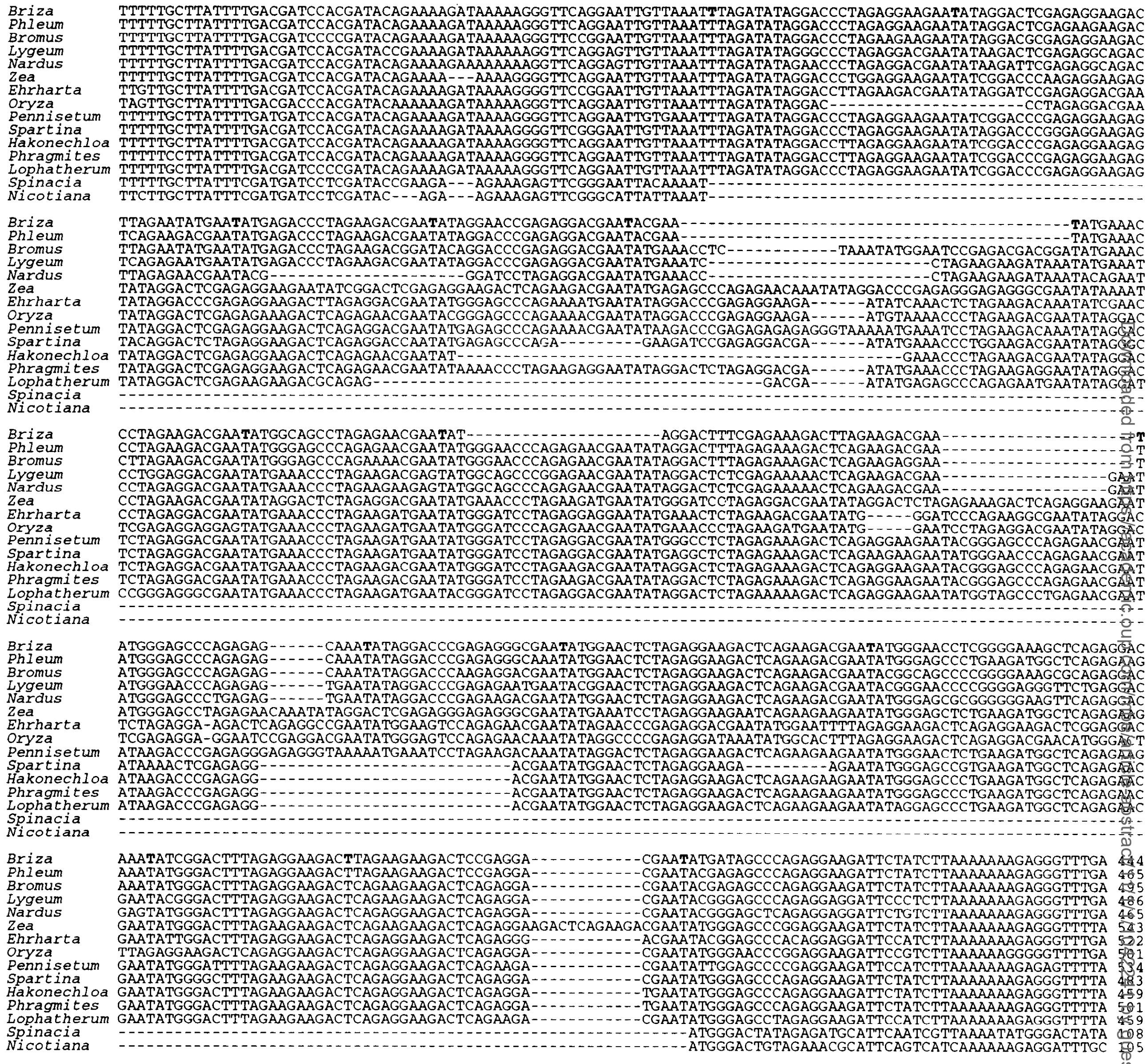

FIG. 1.—Multiple sequence alignment of the extra coding region and adjacent sequence in the plastid gene *rpo*C2. Gapped positions are indicated with a dash (-), bold nucleotides denote approximate start of repeat units in *Briza*, and numerals at end of sequence designate sequence length.

pairing appears to be the mechanism producing much of the variation observed in this region. Slipped-strand mispairing, also known as replication slippage, occurs when double-stranded DNA denatures and reanneals during replication, in such a way that (*a*) the strands are displaced with respect to their original positions and (*b*) subsequent repair results in gain or loss of bases (reviewed in Levinson and Gutman 1987). Levinson and Gutman (1987) have identified sequence patterns that are associated with slipped-strand mispairing, and several of these features are observed in the extra coding region of *rpo*C2, providing evidence to support slipped-strand mispairing. These features include (1) sequence motifs repeated in tandem—as is shown in figure 1 and discussed above, the extra coding sequence is a tandem repeat array, with a repeat unit length of 21 bp being most common; (2) nearby repeat units differing at few positions—many repeats within an array in the extra sequence region differ by only one or a few nucleotides, with most differences representing transition events; and (3) polypurine or polypyrimidine tracts—the extra coding sequence exhibits a higher purine composition compared with the flanking sequence, as shown in table 1. Figure 1 shows that the extra coding sequences are com-

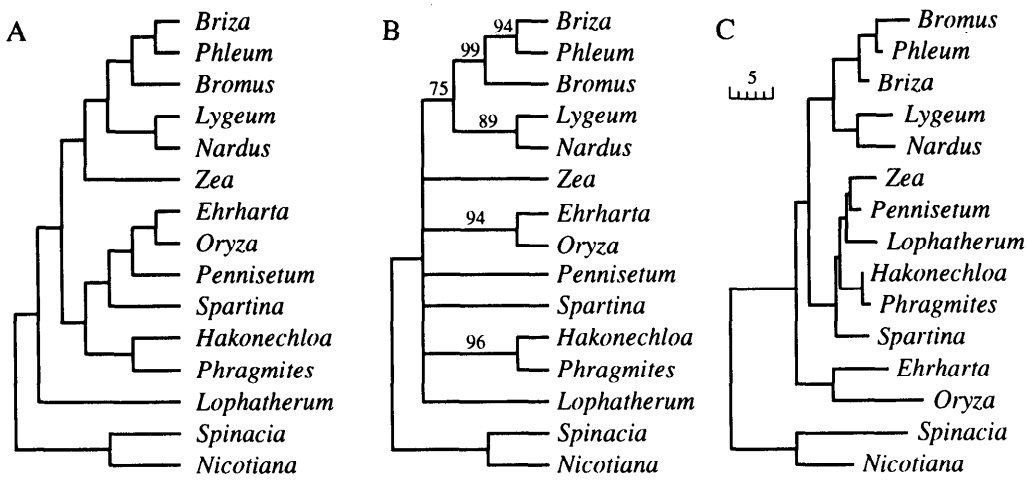

FIG. 2.—*A*, Most parsimonious phylogenetic tree, based on the multiple-sequence alignment (fig. 1). The tree length is 378, and the consistency index (excluding autapomorphies) is 0.706. *B*, Bootstrap (majority-rule) consensus tree, based on 10,000 resampling replicates. The tree length is 403, and the consistency index (excluding autapomorphies) is 0.663. Numerals adjacent to the branches show the percentage of bootstrap replicates containing that clade. *C*, Neighbor-joining tree for the extra coding region of *rpo*C2, based on pairwise distances. Horizontal branch lengths are proportional to distance, and the scale bar represents percent divergence.

posed of many runs of purines and pyrimidines; for example, when reduced to purines (R) and pyrimidines (Y), the most common repeat unit is 5′-YRYRRRRYYYYRRRRRRRRR-3′.

The origin of the basic repeat unit appears to be sequence just 3′ of the extra sequence region, which is also found in spinach and tobacco (fig. 1). The spinach and tobacco sequences are very similar (0.57–0.62 when A, C, G, and T are used; 0.86–0.90 when Y and R are used) to the basic repeat unit of the extra coding region.

Other studies have shown that short tandem repeats occur in regions of the plastid genome (Zurawski et al. 1984; Zurawski and Clegg 1987; Blasko et al. 1988; Wolfson et al. 1991), and slipped-strand mispairing has been suggested to be involved in the evolution of these repeats (Clegg et al. 1991; Wolfson et al. 1991). However, we know of no other reports of a highly variable tandem repeat sequence *within* a known plastid gene. Although the extra coding region shows an increased frequency of substitution and insertion/deletion events, there is no evidence that it affects the function of the encoded subunit of RNA polymerase.

Previous analysis of the extra coding sequence in maize led to speculation that the repetitive region might serve a functional role such as DNA binding, protein-protein interaction, or transcriptional activation (Igloi et al. 1990; Igloi and Kössel 1992). However, the observations in the present study appear to be inconsistent with these speculations: the high level of variation in this region, as a consequence of both nucleotide substitution and insertion/deletion events, indicates a relative lack of functional constraints. The restricted phylogenetic distribution demonstrates, at least for taxa outside the grasses, that the extra sequence is not required for

Table 1
Comparison of Sequence Characteristics for Extra Coding Region and Adjacent Flanking Sequence in Grasses

	MEAN ± SD			
	Pairwise Distance[a]	Sequence Length[b]	Proportion of Purines[c]	Transition Rate[d]
Flanking Region	0.083 ± 0.042	107.769 ± 0.832 (0.8%)	0.592 ± 0.010	0.045 ± 0.026
Extra Region	0.180 ± 0.079	381.231 ± 31.236 (8.2%)	0.699 ± 0.008	0.111 ± 0.041

[a] Estimated no. of substitutions (by the two-parameter model of Kimura [1980])/length of sequence compared.
[b] Data in parentheses are coefficient of variation.
[c] No. of purines/length of sequence.
[d] No. of transitions/length of sequence compared.

proper function of the gene product. For maize, the predicted size of the peptide encoded by *rpo*C2, including the extra coding region, is 176 kD (Igloi et al. 1990), which agrees well with the observed size of 180 kD (Kidd and Bogorad 1980). All available evidence suggests that the extra coding region is part of the mature β″ subunit. Further, the observation that the additional coding sequence does not affect the reading frame of the flanking sequence suggests that it is effectively a neutral feature, albeit a large and complex one.

Phylogenetic Implications

We know of no previous molecular study of the grass family that samples systematically all of the five major subfamilies of the Poaceae. Analysis of *rbc*L, the plastid gene encoding the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase, from three of the subfamilies—Bambusoideae, Panicoideae, and Pooideae (Doebley et al. 1990)—found a tree similar to a previously proposed morphologically based tree (Kellogg and Campbell 1987); however, there was insufficient variation in the *rbc*L sequences to allow us to be confident of the results. Using ribosomal RNA, Hamby and Zimmer (1992) examined the same three subfamilies and found a slightly different topology. The major differences among the trees based on ribosomal RNA, *rbc*L, and morphology are (1) the position of the root, whether the Pooideae or the Bambusoideae is basal, and (2) the relationship of bamboo and rice. However, because of the differences in the taxa sampled for the *rbc*L- and the rRNA-based studies, these differences in topology are impossible to evaluate.

Within the family, the present *rpo*C2 sequence data support a broad circumscription of the pooid clade, including the subfamily Pooideae, with both *Nardus* and *Lygeum* as basal taxa. This delimitation has been suggested elsewhere (Clayton and Renvoize 1986, p. 79) and has been defended on cladistic grounds (Kellogg and Campbell 1987). A cladogram based on morphological characters (Kellogg and Campbell 1987) also linked *Ehrharta* with *Oryza* and *Hakonechloa* with *Phragmites;* both clades are supported in all analyses conducted in the present study, including the alternative alignments that we explored.

There is ample taxonomic evidence, both morphological (Clayton and Renvoize 1986, pp. 16 and 256; Kellogg and Campbell 1987; Watson and Dallwitz 1988, fiche 1) and molecular (Kellogg and Campbell 1987), to link *Zea* and *Pennisetum* in the monophyletic subfamily Panicoideae as shown in the neighbor-joining tree (fig. 2*C*). The maximum-parsimony tree (fig. 2*A*), however, shows the two taxa to be unrelated. Perhaps inclusion of another member of the Andropogoneae might clarify the position of maize.

The rooting at *Lophatherum,* shown in figure 2*A,* is unusual and, to our knowledge, has not been suggested by any other data set, morphological or molecular. The bootstrap tree (fig. 2*B*), however, indicates that the parsimony analysis provides little support for any rooting. The neighbor-joining tree (fig. 2*B*) places the root at the bambusoid clade (*Oryza*/*Ehrharta*), consistent with trees based on rRNA (Hamby and Zimmer 1992) but different from trees based on either *rbc*L (Doebley et al. 1990) or morphology (Kellogg and Campbell 1987). The outgroups are very distant from the grasses, though, and this position of the root is equivalent to midpoint rooting. The basal branching patterns of the grass family, as well as the relationships among the subfamilies, thus remain ambiguous. It is possible that the extra coding region of *rpo*C2 is a useful indicator of relationships within subfamilies of the grasses but is too variable and too short to resolve distant branches.

The evolution of the additional coding sequence in *rpo*C2 through slipped-strand mispairing is similar to that of noncoding sequences in the plastid genome (Clegg et al. 1991; Wolfson et al. 1991). This additional coding sequence in the middle of *rpo*C2 demonstrates that, in certain circumstances, portions of protein-coding genes can tolerate increased amino acid replacement and exhibit tremendous variation.

Sequence Availability

The original sequences reported here have been deposited in GenBank under accession numbers L25376–L25386.

Acknowledgments

We thank G. Adelson, C. Campbell, J. Shore, E. Waters, C. Wimpee, and E. Zimmer for providing some plant DNA used in this study; the Royal Botanic Garden, Kew, especially S. A. Renvoize, for plant material; and J. A. Studier for providing a copy of his program. This work was supported by a grant from the National Science Foundation to E.A.K. and by a grant from the National Institutes of Health to R. C. Lewontin.

APPENDIX

Taxa Surveyed for Extra Region in *rpo*C2

Commelinaceae	*Tradescantia* sp.
Cyperaceae	*Cyperus alternifolius*
Flagellariaceae	*Flagellaria indica*
Joinvilleaceae	*Joinvillea* sp.
Juncaceae	*Juncus tenuis*
Poaceae	Arundinoideae—*Arundo donax, Chionochloa rigida, Cortaderia selloana* var. *pumila, Hakonechloa macra* "Aureola," *Lophatherum gracile, Lygeum spartum, Merx-*

	muellera macowanii, Molinia litoralis, Phragmites australis, and *Rytidosperma unarede;* Bambusoideae—*Bambusa* sp., *Ehrharta stipoides, Microlaena stipoides,* and *Oryza sativa;* Chloridoideae—*Muhlenbergia schreberi, Spartina alterniflora,* and *Sporobolus airoides;* Panicoideae—*Enteropogon seychellarum, Oplismenus* sp., *Pennisetum* sp., *Saccharum officinarum, Tripsacum dactyloides, Zea mays* ssp. *mays,* and *Z. mays* ssp. *mexicana;* and Pooideae—*Briza maxima, Bromus tectorum, Elytrigia repens, Lolium perenne, Nardus stricta, Phalaris canariensis, Phleum pratense, Secale cereale,* and *Triticum aestivum*
Restionaceae	*Elegia capensis* and *Rhodocoma gigantea*
Liliaceae	*Uvularia sessilifolia*
Orchidaceae	*Cephalanthera austinae, Corallorhiza maculata,* and *Cypripedium acaule*
Asclepiadaceae	*Asclepias syriaca*
Asteraceae	*Rubeckia missouriensis* and *Taraxacum officinale*
Brassicaceae	*Arabidopsis thaliana, Brassica nigra,* and *Hesperis matronalis*
Droseraceae	*Drosera rotundifolia*
Lennoaceae	*Pholisma arenarium*
Monotropaceae	*Monotropa uniflora*
Nymphaeaceae	*Victoria* sp.
Primulaceae	*Trientalis borealis*
Turneraceae	*Turnera ulmifolia*

LITERATURE CITED

AUSUBEL, F. M., R. BRENT, R. E. KINGSTON, D. D. MOORE, J. G. SEIDMAN, J. A. SMITH, and K. STRUHL. 1992. Current protocols in molecular biology. Greene/Wiley Interscience, New York.

BLASKO, K., S. A. KAPLAN, K. G. HIGGINS, R. WOLFSON, and B. B. SEARS. 1988. Variation in copy number of a 24-base pair tandem repeat in the chloroplast DNA of *Oenothera hookeri* strain Johansen. Curr. Genet. **14**:287–292.

CAMPBELL, C. S., and E. A. KELLOGG. 1987. Sister group relationships of the Poaceae, Pp. 217–224 *in* T. R. SODERSTROM, K. W. HILU, C. S. CAMPBELL, and M. E. BARKWORTH, eds. Grass systematics and evolution. Smithsonian Institution, Washington, D.C.

CHURCH, G. M., and W. GILBERT. 1984. Genomic sequencing. Proc. Natl. Acad. Sci. USA **81**:1991–1995.

CLAYTON, W. D., and S. A. RENVOIZE. 1986. Genera graminum. Her Majesty's Stationary Office, London.

CLEGG, M. T., G. H. LEARN, and E. M. GOLENBERG. 1991. Molecular evolution of chloroplast DNA, Pp. 135–149 *in* R. K. SELANDER, A. G. CLARK, and T. S. WHITTAM, eds. Evolution at the molecular level. Sinauer, Sunderland, Mass.

CURTIS, S. E., and M. T. CLEGG. 1984. Molecular evolution of chloroplast DNA sequences. Mol. Biol. Evol. **1**:291–301.

DOEBLEY, J., M. DURBIN, E. M. GOLENBERG, M. T. CLEGG, and D. P. MA. 1990. Evolutionary analysis of the large subunit of carboxylase (*rbc*L) nucleotide sequence among the grasses (Gramineae). Evolution **44**:1097–1108.

DOYLE, J. J., J. I. DAVIS, R. J. SORENG, D. GARVIN, and M. J. ANDERSON. 1992. Chloroplast DNA inversions and the origin of the grass family (Gramineae). Proc. Natl. Acad. Sci. USA **89**:7722–7726.

HAMBY, R. K., and E. A. ZIMMER. 1992. Ribosomal RNA as a phylogenetic tool in plant systematics, Pp. 50–91 *in* P. S. SOLTIS, D. E. SOLTIS, and J. J. DOYLE, eds. Molecular systematics of plants. Chapman & Hall, New York.

HIGGINS, D. G., A. J. BLEASBY, and R. FUCHS. 1992. CLUSTAL V: improved software for multiple sequence alignment. Comput. Appl. Biosci. **8**:189–191.

HU, J., and L. BOGORAD. 1990. Maize chloroplast RNA polymerase: the 180-, 120-, and 38-kilodalton polypeptides are encoded in chloroplast genes. Proc. Natl. Acad. Sci. USA **87**:1531–1535.

HUDSON, G. S., T. A. HOLTON, P. R. WHITFELD, and W. BOTTOMLEY. 1988. Spinach chloroplast *rpo*BC genes encode three subunits of the chloroplast RNA polymerase. J. Mol. Biol. **200**:639–654.

IGLOI, G. L., and H. KÖSSEL. 1992. The transcriptional apparatus of chloroplasts. Crit. Rev. Plant Sci. **10**:525–558.

IGLOI, G. L., A. MEINKE, I. DÖRY, and H. KÖSSEL. 1990. Nucleotide sequence of the maize chloroplast *rpo*B/C1/C2 operon: comparison between the derived protein primary structures from various organisms with respect to functional domains. Mol. Gen. Genet. **221**:379–394.

KELLOGG, E. A., and C. S. CAMPBELL. 1987. Phylogenetic analysis of the Gramineae. Pp. 310–322 *in* T. R. SODERSTROM, K. W. HILU, C. S. CAMPBELL, and M. E. BARKWORTH, eds. Grass systematics and evolution. Smithsonian Institution, Washington, D.C.

KIDD, G. H., and L. BOGORAD. 1980. A facile procedure for purifying maize chloroplast RNA polymerase from whole cell homogenates. Biochim. Biophys. Acta **609**:14–30.

KING, L. M., and B. A. SCHAAL. 1990. Genotypic variation within asexual lineages of *Taraxacum officinale.* Proc. Natl. Acad. Sci. USA **87**:998–1002.

KIMURA, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. **16**:111–120.

LEVINSON, G., and G. A. GUTMAN. 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol. Biol. Evol. **4**:203–221.

MANIATIS, T., E. F. FRITSCH, and J. SAMBROOK. 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

SAITOU, N., and M. NEI. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**:406–425.

SARKAR, G., and S. S. SOMMER. 1990. Shedding light on PCR contamination. Nature **343**:27.

SHIMADA, H., M. FUKUTA, M. ISHIKAWA, and M. SUGIURA. 1990. Rice chloroplast RNA polymerase genes: the absence of an intron in *rpo*C1 and the presence of an extra sequence in *rpo*C. Mol. Gen. Genet. **221**:395–402.

SHINOZAKI, K., M. OHME, M. TANAKA, T. WAKASUGI, N. HAYASHIDA, T. MATSUBAYASHI, N. ZAITA, J. CHUNWONGSE, J. OBOKATA, K. YAMAGUCHI-SHINOZAKI, C. OHTO, K. TORAZAWA, B. Y. MENG, M. SUGITA, H. DENO, T. KAMOGASHIRA, K. YAMADA, J. KUSUDA, F. TAKAIWA, A. KATO, N. TOHDOH, H. SHIMADA, and M. SUGIURA. 1986. The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. EMBO J. **5**:2043–2049.

STUDIER, J. A., and K. J. KEPPLER. 1988. A note on the neighbor-joining algorithm of Saitou and Nei. Mol. Biol. Evol. **5**:729–731.

SWOFFORD, D. L. 1992. PAUP: Phylogenetic analysis using parsimony, portable version (Unix) 3.0r+4 (prerelease 0.4). Illinois Natural History Survey, Champaign.

VOGELSTEIN, B., and D. GILLESPIE. 1979. Preparative and analytical purification of DNA from agarose. Proc. Natl. Acad. Sci. USA **76**:615–619.

WATSON, L., and M. D. DALLWITZ. 1988. Grass genera of the world. Australian National University, Canberra.

WOLFE, K. H., W.-H. LI, and P. M. SHARP. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc. Natl. Acad. Sci. USA **84**:9054–9058.

WOLFSON, R., K. G. HIGGINS, and B. B. SEARS. 1991. Evidence for replication slippage in the evolution of *Oenothera* chloroplast DNA. Mol. Biol. Evol. **8**:709–720.

ZURAWSKI, G., and M. T. CLEGG. 1987. Evolution of higher-plant chloroplast DNA-encoded genes: implications for structure-function and phylogenetic studies. Annu. Rev. Plant Physiol. **38**:391–418.

ZURAWSKI, G., M. T. CLEGG, and A. H. D. BROWN. 1984. The nature of nucleotide sequence divergence between barley and maize chloroplast DNA. Genetics **106**:735–749.

ELIZABETH A. ZIMMER, reviewing editor

Received December 22, 1992

Accepted August 25, 1993