MOLECULAR PHYLOGENETICS AND EVOLUTION
Vol. 6, No. 1, August, pp. 1–18, 1996
ARTICLE NO. 0053

Phylogenetic Analysis of Apiaceae Subfamily Apioideae Using Nucleotide Sequences from the Chloroplast *rpo*C1 Intron

STEPHEN R. DOWNIE,[1] DEBORAH S. KATZ-DOWNIE, AND KYUNG-JIN CHO

Department of Plant Biology, University of Illinois, Urbana, Illinois 61801

Received May 15, 1995; revised September 25, 1995

Phylogenetic relationships among 25 members of Apiaceae (Umbelliferae) subfamily Apioideae, representing 7 of the 8 tribes and 8 of the 10 subtribes traditionally recognized in the subfamily, and 5 outgroups from Apiaceae subfamilies Hydrocotyloideae and Saniculoideae and allied families Araliaceae and Pittosporaceae have been inferred from nucleotide sequence variation in the intron of the chloroplast gene RNA polymerase C1 (*rpo*C1). Sequence divergence values in pairwise comparisons of unambiguous positions among all taxa ranged from 0 to 11.3% of nucleotides and averaged 3.8%. Trees derived from *rpo*C1 intron sequences estimated using maximum parsimony or maximum likelihood methods are of essentially similar topology, and indicate that (1) subfamily Apioideae, with *Heteromorpha* as its most basal element, is monophyletic and is a sister-group to *Eryngium*, the only representative examined of Apiaceae subfamily Saniculoideae, (2) *Aralia* (Araliaceae) arises from within a paraphyletic Apiaceae subfamily Hydrocotyloideae (represented by *Centella* and *Hydrocotyle*) and this clade is a sister-group to Apioideae+Saniculoideae, (3) there is a major phylogenetic division within Apioideae (excluding the basal *Heteromorpha*), with one clade comprising the genus *Smyrnium* and those taxa traditionally grouped in tribes Dauceae, Scandiceae, and Laserpitieae, and the other clade comprising all other examined taxa, and (4) relationships within each of these 2 major clades of Apioideae are largely equivocal owing to the low levels of nucleotide sequence divergence observed. Although *rpo*C1 intron sequences can provide valuable characters for addressing phylogenetic relationships among the outgroups and distantly related members of Apioideae, they have little power for resolving relationships among closely related taxa.

INTRODUCTION

DNA sequence comparisons of chloroplast genes have proven extremely useful in estimating phylogenetic relationships among diverse plant species, at a variety of taxonomic levels (reviewed most recently in Olmstead and Palmer, 1994). Although *rbc*L has emerged as the chloroplast gene-of-choice for phylogenetic inference, its rate of nucleotide substitution may be too slow to provide sufficient characters to resolve relationships at lower (inter- and intrageneric) taxonomic levels (Doebley *et al.*, 1990; Kim *et al.*, 1992; Soltis *et al.*, 1993; Xiang *et al.*, 1993; Plunkett *et al.*, in press). Instead, comparisons of more rapidly evolving protein-coding genes, such as *ndh*F (Olmstead and Sweere, 1994) and *mat*K (Johnson and Soltis, 1994; Steele and Vilgalys, 1994), may provide the characters necessary to resolve relationships among closely related taxa.

Sequence studies of the chloroplast genome have concentrated primarily on protein-coding regions. However, noncoding regions of the chloroplast genome (such as the intergenic spacers) tend to evolve more rapidly than do coding regions, both in nucleotide substitutions and in the accumulation of insertion and deletion mutations (Curtis and Clegg, 1984; Wolfe *et al.*, 1987; Palmer, 1991; Clegg and Zurawski, 1992; Clegg *et al.*, 1991, 1994). To date, the spacer regions between genes *rbc*L and *atp*B (Golenberg *et al.*, 1993; Manen *et al.*, 1994; Savolainen *et al.*, 1994), between genes *rbc*L and *psa*I (Morton and Clegg, 1993), and between genes *trn*T and *trn*F (Taberlet *et al.*, 1991; Fangan *et al.*, 1994; Gielly and Taberlet, 1994; van Ham *et al.*, 1994) have been investigated for phylogenetic purposes. Additionally, comparative restriction site analyses of noncoding sequences, using polymerase chain reaction (PCR)-amplified products that have been digested with frequently cutting enzymes, have also increased the utility of the chloroplast genome in providing characters for resolution of relationships at lower taxonomic levels (Arnold *et al.*, 1991; Liston, 1992; Rieseberg *et al.*, 1992).

Introns are those transcribed sequences that are excised during the processing of the primary transcript and, like intergenic spacers, are nontranslated. Of the 79 putative protein-coding and 30 tRNA genes in the tobacco chloroplast genome (Shinozaki *et al.*, 1986;

[1] To whom correspondence should be addressed. Fax: 217-244-7246. E-mail: stephen_downie@qms1.life.uiuc.edu.

1055-7903/96 $18.00

Wolfe, 1991), 18 are interrupted by introns (2 of these genes, ORF 168 and *clp*P, each contain 2 introns, and 1 gene, *rps*12, is thought to be *trans*-spliced). Six of these introns split genes that encode tRNAs; the remaining split protein-coding genes. To date, the use of intron sequences in phylogenetic studies has been limited, being restricted to those of only 2 chloroplast tRNA genes (i.e., *trn*L and *trn*V). Pairwise comparisons of *rbc*L and 2 noncoding sequences (namely, the *trn*L intron and the intergenic spacer between the *trn*L 3′ exon and *trn*F) for 4 dicots and 6 monocots have illustrated that these noncoding regions evolve, on average, more than 3 times faster than *rbc*L, and that the intron in *trn*L evolves at a rate that is approximately the same as that of the intergenic spacer (Gielly and Taberlet, 1994). Analysis of *trn*L intron sequences from 7 species of gentians (Gentianaceae) yielded a robust phylogeny fully concordant with conventional taxonomic treatments of the group and demonstrated the utility of tRNA intron sequences for resolving relationships at the intrageneric level. Gielly and Taberlet (1994) noted, however, the difficulty in aligning intron sequences obtained from different genera, even those from within the same family, because of the high incidence of insertion and deletion events. Similar results were observed by Fangan *et al.* (1994) in which the analysis of *trn*L intron sequences from 12 species (representing a diversity of monocots and dicots, and *Picea* and *Marchantia*) showed that two-thirds of all positions (383 of 582 sites) could not be aligned unambiguously except for the most closely related groups of species. The remaining one-third of positions, however, were highly conserved among all taxa. Comparisons of *trn*L and *trn*V intron sequences from a diversity of angiosperm species reveal that the 6 different (stem-and-loop) domains of these introns evolve at different rates, and that those small and scattered regions of high sequence conservation are likely to be essential for the formation of secondary structure necessary for intron-associated functions (Clegg *et al.*, 1986; Learn *et al.*, 1992; Fangan *et al.*, 1994). Thus, it is seen (and often generally perceived) that chloroplast intron sequences, like noncoding intergenic spacers, are rapidly evolving, both in their high rates of nucleotide substitution and in their propensity to accumulate many insertion/deletion mutations (indel events). Because of the presumed difficulty in aligning the sequences and the likely saturation of nonconserved positions by multiple substitutions, these introns are thought not to contribute usable characters for phylogenetic inference, particularly at higher taxonomic levels (Doyle, 1993; Olmstead and Palmer, 1994).

This study is the first to focus explicitly on an intron sequence from a non-tRNA gene for phylogenetic inference and shows that introns in protein-coding chloroplast genes, specifically the intron splitting the gene RNA polymerase C1 (*rpo*C1), can contribute valuable characters for inferring hypotheses of relationship. The chloroplast *rpo*C region, homologous to the β′ subunit of *Escherichia coli* RNA polymerase, is divided into two genes, *rpo*C1 and *rpo*C2, and is located within the large single-copy region of the plastid genome in most angiosperms (Cozens and Walker, 1986; Hudson *et al.*, 1988; Fig. 1). These genes, along with *rpo*B, are cotranscribed as a single operon and encode three subunits of the chloroplast RNA polymerase (Hudson *et al.*, 1988).

The family Apiaceae (e.g., carrots, parsnips, celery, caraway, coriander, fennel, dill, and parsley) comprises about 300 genera and some 3000 species and, although largely confined to temperate regions, is cosmopolitan in distribution (Cronquist, 1981). It is one of the best known families of flowering plants, because of its characteristic inflorescences and fruits and the distinctive chemistry, reflected in the odor, flavor, and even toxicity of many of its members (Heywood, 1993). The division of Apiaceae into 3 subfamilies (Hydrocotyloideae, Saniculoideae, and Apioideae) and 12 tribes, proposed almost a century ago, remains the predominant system of classification for the family (Drude, 1898). However, the phylogenetic relationships among the genera of Apioideae, the largest and most taxonomically complex subfamily, are not clear. This confusion stems from the undue reliance placed on many subtle morphological and anatomical differences of their fruits, and the numerous instances of parallel and convergent evolution that have likely occurred during the evolution of the group (Downie and Katz-Downie, 1996).

This study was undertaken with two broad goals in mind: (1) To evaluate the usefulness of chloroplast DNA (cpDNA) *rpo*C1 intron sequences in providing a sufficient number of reliable characters for inferring phylogeny; and (2) to formulate more precise hypotheses about relationships among the diverse clades comprising Apiaceae subfamily Apioideae using these characters. The relationships proposed here will be evaluated primarily against the frequently cited system of Apiaceae classification proposed by Drude (1898) and in light of the results obtained from a recent systematic investigation of the subfamily using nuclear ribosomal DNA internal transcribed spacer (ITS) sequences (Downie and Katz-Downie, 1996).

MATERIALS AND METHODS

Ingroup taxa. The 25 accessions of Apioideae included in this investigation (Table 1) were chosen for 3 primary reasons. (1) They represent 7 of the 8 tribes and 8 of the 10 subtribes of Apioideae recognized by Drude (1898). (2) Many were the subject of a phylogenetic study based on nuclear ribosomal DNA ITS sequences with which the results from this study will be compared (Downie and Katz-Downie, 1996). (3) On the basis of these ITS results, they represent groups of taxa that are both closely and distantly related within the subfamily. Thus, the utility of *rpo*C1 intron sequences in providing characters to infer phylogeny at various

TABLE 1

Accessions Examined for Chloroplast DNA *rpo*C1 Intron Sequence Variation

Taxon	Distribution[a]	Source and voucher
Apiaceae subfamily Apioideae		
Aethusa cynapium L.	O	France, Cult. Jardin Botanique de Caen (No. 1424); *Downie 337* (ILL)
Anethum graveolens L.	O	France, Cult. Jardin Botanique de Caen (No. 1980); *Downie 326* (ILL)
Angelica archangelica L.	O	Finland, Cult. University of Joensuu Botanical Garden (No. 33); *Downie 79* (ILL)
Anthriscus cerefolium (L.) Hoffm.	O	Spain, Cult. Real Jardín Botánico (No. 1305); *Downie 35* (ILL)
Apium graveolens L.	O	France, Cult. Conservatoire et Jardins Botaniques de Nancy; *Downie 258* (ILL)
Arracacia brandegei Coult. & Rose	N	Mexico, Baja California del Sur, *Breedlove 43405* [=*Constance 2045* (UC)]
Arracacia nelsonii Coult. & Rose	N	Mexico, Oaxaca, *Breedlove 72434* [=*Constance 2410* (UC)]
Conium maculatum L.	O	Germany, Cult. Johannes Gutenberg University (No. 1099); *Downie 63* (ILL)
Coriandrum sativum L.	O	Germany, Cult. Johannes Gutenberg University (No. 1100); *Downie 65* (ILL)
Coulterophytum laxum Robins.	N	Mexico, Michoacán, *Iltis 298 & Cochrane* [=*Constance 1650* (UC)]
Daucus carota L.	O	Germany, Cult. University of Oldenburg Botanic Garden (No. 547); *Downie 164* (ILL)
Heracleum lanatum Michx.	N	U. S. A., California, Muir Woods; *Downie 579* (ILL)
Heracleum rigens DC.	O	India, Karnataka, Mullengiri-Bababudan Hills, Chixmagalur District, *Mukherjee s.n.* [=*Constance 2274* (UC)]
Heteromorpha arborescens (Thumb.) Cham. & Schlechtd.	O	Spain, Cult. Real Jardín Botánico (No. 1330); *Downie 42* (ILL)
Laserpitium siler L.	O	Germany, Cult. Johannes Gutenberg University (No. 1112); *Downie 71* (ILL)
Myrrhidendron donnell-smithii Coult. & Rose	N	Costa Rica, San José Prov., *Grantham and Parsons 0433-90*, Cult. UC Botanical Garden, Berkeley (No. 90.2637)
Orlaya grandiflora (L.) Hoffm.	O	France, Cult. Jardin Botanique de Caen (No. 1474); *Downie 309* (ILL)
Pastinaca sativa L.	O	Germany, Cult. Johannes Gutenberg University (No. 1597); *Downie 70* (ILL)
Pimpinella peregrina L.	O	Germany, Cult. Akademie der Wissenschaften (No. 29/90); *Downie 19* (ILL)
Prionosciadium turneri Const. & Affolt.	N	Mexico, Colima, *Turner s.n.* [=*Constance 2053* (UC)]
Rhodosciadium argutum (Rose) Math. & Const.	N	Mexico, Guanajuato, Xichu, *Rzedowski 41342* [=*Constance 2371* (UC)]
Scandix pecten-veneris L.	O	Germany, Cult. Akademie der Wissenschaften (No. 2/77); *Downie 27* (ILL)
Smyrnium olusatrum L.	O	France, Cult. Jardin Botanique de Caen (No. 1492); *Downie 328* (ILL)
Torilis nodosa (L.) Gaertn.	O	France, Cult. Jardin Botanique de Caen (No. 1495); *Downie 322* (ILL)
Zizia aurea (L.) Koch	N	Canada, Cult. Jardin Botanique de Montréal (No. 60); *Downie 8* (ILL)
Apiaceae subfamily Hydrocotyloideae		
Centella erecta (L.f.) Fern.	O	Cuba, La Habana, Cayo La Rosa, Laguna Ariquenabo, *Stevens et al. 23626;* Cult. UC Botanical Garden, Berkeley [=*Constance 2336* (UC)]
Hydrocotyle bowlesioides Math. & Const.	N	Costa Rica, Cult. UC Botanical Garden, Berkeley (No. 61.1190) [=*Constance 222* (UC)]
Apiaceae subfamily Saniculoideae		
Eryngium planum L.	O	Ireland, Cult. National Botanic Gardens, Glasnevin (No. 93); *Downie 191* (ILL)
Araliaceae		
Aralia chinensis L.	O	China, Cult. Shanghai Botanic Garden (No. 28); *Downie 407* (ILL)
Pittosporaceae		
Pittosporum tobira (Thunb.) Aiton	O	China, Cult. Missouri Botanical Garden, St. Louis (No. 801425)

[a] New World (N) or Old World (O) geographic distribution.

levels of genetic divergence within the subfamily can be ascertained.

Outgroup taxa. Two representatives of Apiaceae subfamily Hydrocotyloideae (*Centella* and *Hydrocotyle*) and one representative each of Apiaceae subfamily Saniculoideae (*Eryngium*), Araliaceae (*Aralia*), and Pittosporaceae (*Pittosporum*) were chosen as outgroups (Table 1). Most authors, except Hutchinson (1973), have agreed that Apiaceae and Araliaceae are closely related (e.g., Thorne, 1973; Takhtajan, 1980; Cronquist, 1981), and many have suggested an affinity between these two taxa and Pittosporaceae (van Tieghem, 1884; Jay, 1969; Thorne, 1973, 1992; Dahlgren, 1980; Stuhlfauth *et al.*, 1985; Judd *et al.*, 1994). Recent phylogenetic analyses of cpDNA *rbc*L sequences (Plunkett *et al.*, in press) and morphologic and anatomic characters (Judd *et al.*, 1994) reveal that subfamilies Apioideae and Saniculoideae are each monophyletic and are sister-taxa.

DNA extraction. Total genomic DNAs were isolated from fresh leaf material of one or, rarely, more individual plants using the modified CTAB procedure of Doyle

and Doyle (1987) and further purified by centrifugation in cesium chloride–ethidium bromide gradients. Leaf material was either collected directly from the field, taken from flower- and fruit-bearing plants propagated from seed in the greenhouse, or obtained from accessioned plants cultivated at the University of California Botanical Garden, Berkeley (Table 1).

PCR amplification and sequencing strategy. The complete *rpo*C1 intron region and portions of the flanking exon regions in each genomic DNA were PCR-amplified using primers "5′*rpo*C1 exon" and "3′*rpo*C1 exon" (Fig. 1) in an equimolar ratio. Primers were designed by comparing the two *rpo*C1 exon sequences from tobacco, rice, and *Marchantia* and choosing regions highly conserved among the taxa. Details of the amplification reactions, and the DNA purification and sequencing strategies used, are the same as outlined previously (Downie and Katz-Downie, 1996). Each set of reactions was monitored by the inclusion of positive (tobacco DNA) and negative (no template) controls. Successful PCR amplifications resulted in a single DNA band corresponding to approximately 1100 bp. Forward primers "5′*rpo*C1 exon," "*rpo*C1 intron-1," "*rpo*C1 intron-2," and "*rpo*C1 intron-3," and reverse primer "*rpo*C1 exon2" (Figs. 1 and 2) were each used in the sequencing of each template DNA. In most instances, each of the template DNAs was sequenced twice with the same primer to ensure accuracy. All primers were synthesized by Operon Technologies, Inc. (Alameda, CA).

Sequence analysis. Group II introns, of which the intron in *rpo*C1 is classified, are characterized by six centrally radiating structural components (designated as domains I–VI; Michel *et al.*, 1989). These introns exhibit considerable evolutionary conservation of secondary structure, with each of these domains divided into highly conserved stem portions and, generally, less conserved loop portions (Michel and Dujon, 1983; Michel *et al.*, 1989). The alignment of *rpo*C1 intron sequences was done manually by comparing these sequences to conserved regions documented for *rpo*C1 and other group II introns (Michel *et al.*, 1989). Predictions of secondary structure by free-energy minimization were made using MULFOLD Version 2.0 (Zuker, 1989; Jaeger *et al.*, 1989) and used to guide the alignment when

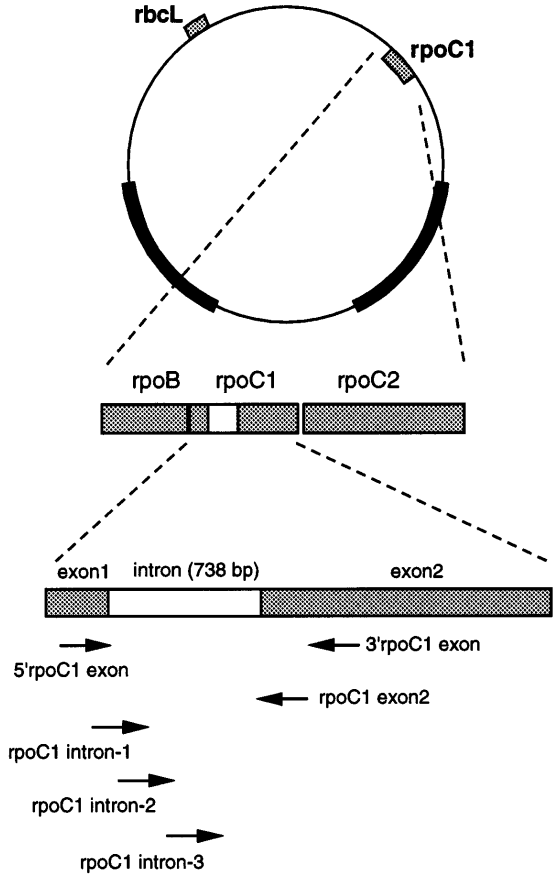
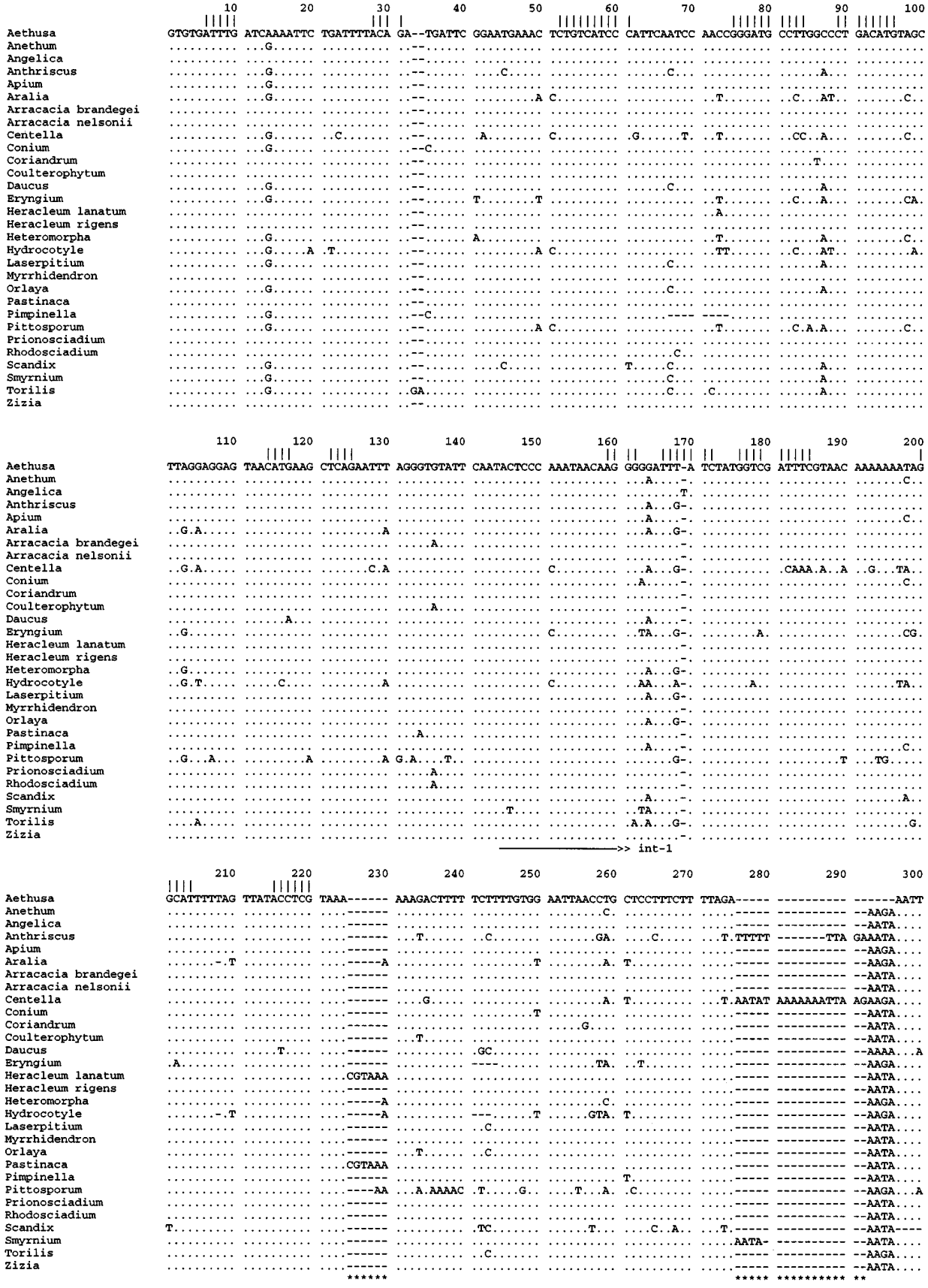


FIG. 1. Location of the *rpo*C1 gene and its intron in the tobacco chloroplast genome and its placement relative to the gene *rbc*L and cotranscribed ribosomal polymerase genes *rpo*B and *rpo*C2 (based on Shinozaki *et al.*, 1986). The gene *rpo*C1 is interrupted by an intron in tobacco and most other land plants; in tobacco, the intron is 738 bp in size (Shimada *et al.*, 1990). The thickened parts of the circle represent the 25.3-kb inverted repeat region. The arrows represent the approximate positions of the oligonucleotide primers used in PCR amplification (5′*rpo*C1 exon and 3′*rpo*C1 exon) and DNA sequencing (all but 3′*rpo*C1 exon). Primer sequences are as follows: 5′*rpo*C1 exon, GGTCTTCCTAG(C/T)TA(C/T)AT(C/T/A)GC; *rpo*C1 intron-1, ACTC CCAAATAA(A/C)AAGG; *rpo*C1 intron-2, TGGCATAACC(A/G)TCG AGG; *rpo*C1 intron-3, AGGAAGCCGTAAT(C/T)AA(A/G)G; 3′*rpo*C1 exon, AATAAGCA(C/T)AA(N)ACCATCCA; *rpo*C1 exon2, ATTTCAT ATTCGAA(C/T)AANCC. All sequences are written 5′ to 3′.

FIG. 2. Aligned *rpo*C1 intron nucleotide sequences from 30 representatives of Apiaceae, Araliaceae, and Pittosporaceae arranged alphabetically. Nucleotide sites are numbered 5′ to 3′ from the exon1–intron boundary to the intron–exon2 boundary. A, C, G, and T denote dATP, dCTP, dGTP, dTTP, respectively. Nucleotide identity with the first sequence is indicated by a dot; gaps required for alignment are indicated by hyphens; ambiguous regions excluded from the analysis are indicated by asterisks below the alignment. Complete taxon names are provided in Table 1. The size of each intron (in bp) is provided in the square brackets at the end of the alignment, and the locations of the three internal sequencing primers (Fig. 1) are indicated by the double-headed arrows. The approximate locations of the six major secondary-structural components characteristic of group II introns, designated as domains I through VI, are as follows (numbers refer to coordinates in alignment): domain I (7–441), domain II (447–525), domain III (533–679), domain IV (683–772), domain V (774–807), domain VI (816–842). The nucleotides likely participating in either secondary structure or three-dimensional pairings are indicated by vertical bars above the alignment.


Aethusa
Anethum
Angelica
Anthriscus
Apium
Aralia
Arracacia brandegei
Arracacia nelsonii
Centella
Conium
Coriandrum
Coulterophytum
Daucus
Eryngium
Heracleum lanatum
Heracleum rigens
Heteromorpha
Hydrocotyle
Laserpitium
Myrrhidendron
Orlaya
Pastinaca
Pimpinella
Pittosporum
Prionosciadium
Rhodosciadium
Scandix
Smyrnium
Torilis
Zizia
>> int-1

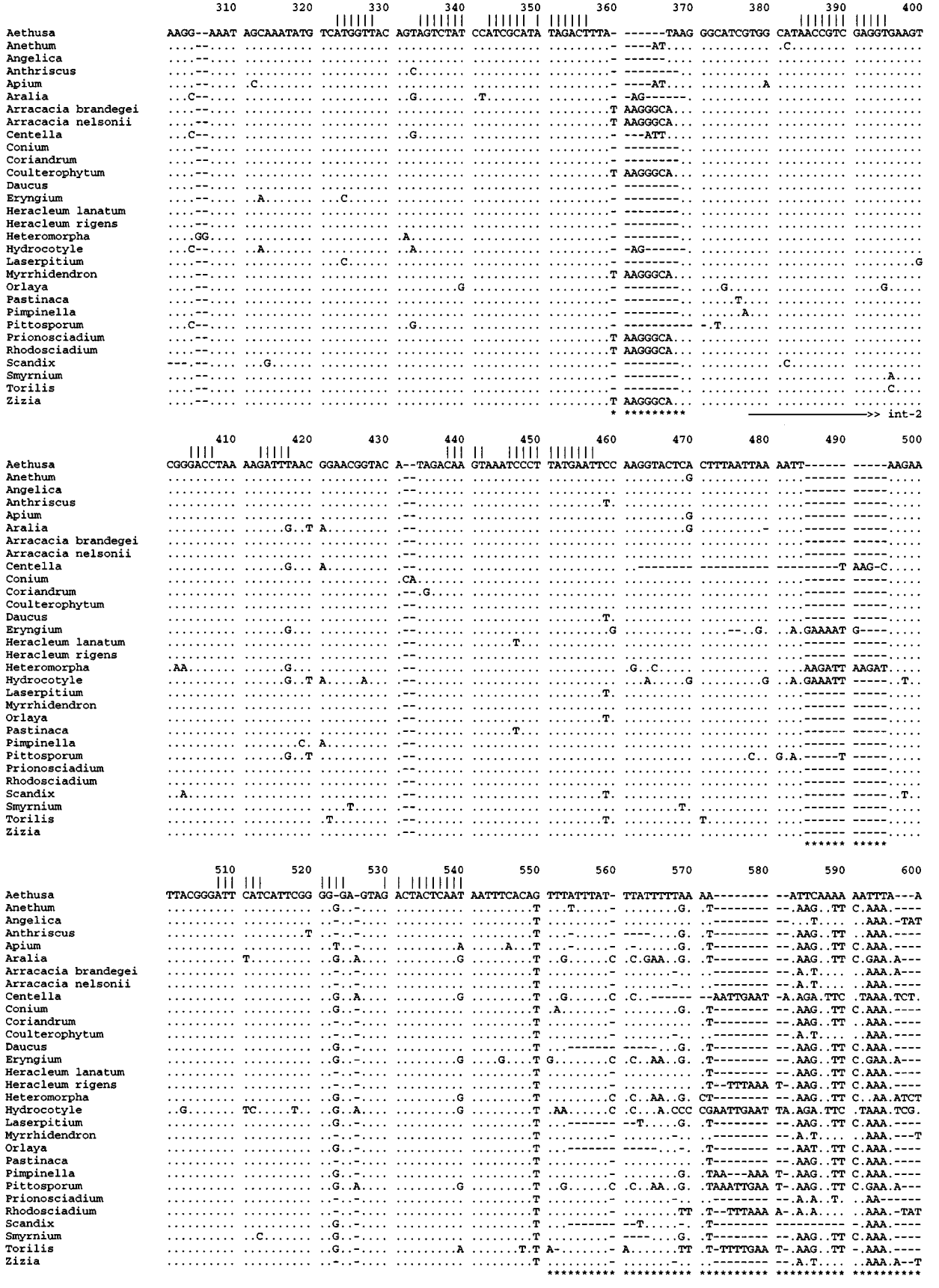


FIG. 2—*Continued*

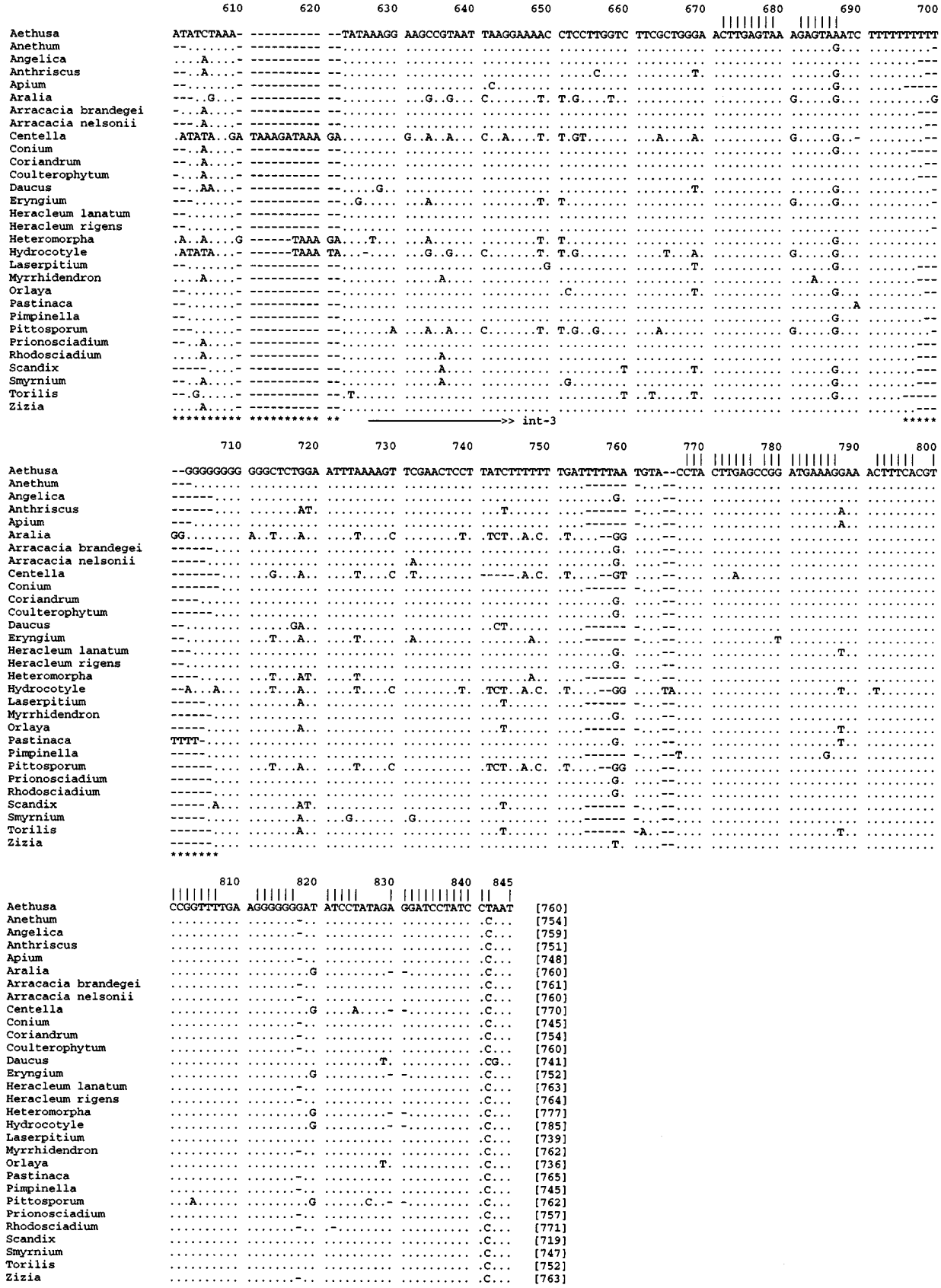


FIG. 2—*Continued*

appropriate. Boundaries of the exon and intron regions were determined by comparison of the DNA sequences to the corresponding boundaries in tobacco (Shinozaki *et al.*, 1986) and consensus splice sites in other plants for group II introns (Michel *et al.*, 1989).

Only the intron sequences were included in the analysis, as the sequencing strategy outlined above effectively ignored the flanking, and presumably more highly conserved, exon regions. Because the taxa being considered are closely related, pairwise nucleotide differences of unambiguously aligned positions were determined using the DISTANCE MATRIX option in PAUP Version 3.1.1 (Swofford, 1993). Thus, these divergence values were calculated simply as a proportion of divergent sites in each pairwise comparison with no provision made to account for superimposed events (multiple hits) which may have occurred at many positions. Only those positions which were in obvious alignment were used in the distance calculations and phylogenetic analyses. Gaps were treated as missing data. Transition/transversion ratios were calculated using MacClade Version 3 (Maddison and Maddison, 1992), and the A+T content of each intron for each species was calculated manually. The nucleotide sequence data reported in this study have been deposited with the EMBL/GenBank Data Libraries under Accession Nos. U36278–U36307.

Phylogenetic analysis. The resulting data matrix (excluding ambiguous positions) was analyzed by assuming unordered character states (i.e., Fitch parsimony) using PAUP run on either a Macintosh Quadra 700 or Power Macintosh 8100/100 AV computer. All HEURISTIC searches were replicated 500 times with RANDOM addition sequence and TREE BISECTION–RECONNECTION (TBR) branch swapping. The options MULPARS, STEEPEST DESCENT, COLLAPSE, and ACCTRAN optimization were selected. Initially, all searches were performed using equal character weighting. Bootstrap values (Felsenstein, 1985) were calculated from 100 replicate analyses using the HEURISTIC search strategy and SIMPLE addition sequence of the taxa. Owing to the large number of trees obtained upon the relaxation of parsimony and the memory capabilities of the computers, a "decay analysis" (Bremer, 1988) was not done. The number of additional steps required to force particular taxa into a monophyletic group was examined using the CONSTRAINTS option of PAUP. The amount of phylogenetic information in the parsimony analyses was estimated using the consistency index (CI, Kluge and Farris, 1969), retention index (RI, Farris, 1989), and g_1 statistic (Hillis and Huelsenbeck, 1992). The g_1 statistic was achieved by calculating the tree-length distribution of 10,000 random parsimony trees using PAUP's RANDOM TREES selection and was used to assess the amount of nonrandom structure in the data. Each indel event was mapped onto one of the resulting minimal-length cladograms in the most parsimonious way possible in order to test its congruence with the phylogeny constructed on the basis of nucleotide substitutions alone.

The pattern of *rpo*C1 intron evolution is clearly constrained by the presence of a high degree of secondary structure. Considerations of secondary structure in phylogenetic analyses, with regard to evaluating respective properties of paired (stem) and unpaired (loop) positions on tree reliability, are therefore important. However, because of the lack of consensus on how to deal with this issue [i.e., should paired positions be eliminated, weighted by one-half (or some other value) relative to unpaired positions to account for their non-independence (and observed levels of compensatory substitution), or be given equal weight (Smith, 1989; Wheeler and Honeycutt, 1988; Hillis and Dixon, 1991; Dixon and Hillis, 1993)], all three of these weighting schemes were considered here.

Character-state weighted parsimony analysis, in which transversions were weighted over transitions by factors of 1.1, 1.2, or 1.5 using PAUP's USERTYPE STEPMATRIX command, was also implemented. In this analysis, HEURISTIC searches were conducted using SIMPLE addition and TBR branch-swapping. These methods allow for the correction of multiple substitutions and differential transition/transversion probability based on empirical observation from the data. The ratio of 1.2 was selected based on the actually observed frequencies in the maximally parsimonious trees; the values of 1.1 and 1.5 were selected because they simply bracket the observed value of 1.2. Additionally, parsimony trees were constructed using transversions only. As the ratio of transitions to transversions approaches 1.0 (as is the case here), it is often assumed that transitions are becoming largely saturated and, thus, should be down-weighted or ignored altogether in phylogenetic analyses (Mindell and Honeycutt, 1990; Knight and Mindell, 1993; but see Reeder, 1995, for an exception).

Maximum likelihood phylogeny estimation was also explored utilizing the fastDNAml program (Version 1.0.6; Olsen *et al.*, 1994) based on the procedures of Felsenstein (1981). A maximum likelihood tree was inferred using a transition/transversion ratio of 1.2, randomizing the sequence addition order (JUMBLE), and by invoking the GLOBAL branch swapping search option. Empirical base frequencies were derived from the sequence data and used in the maximum likelihood calculations.

Tobacco and rice intron sequence comparisons. Pairwise comparisons of the 17 chloroplast introns shared between tobacco and rice (Shinozaki *et al.*, 1986; Hiratsuka *et al.*, 1989) were carried out in order to determine the extent of sequence conservation between

these regions and to explore the taxonomic level to which these additional intron sequences would be most effective in providing characters to resolve relationships. To date, tobacco and rice represent the only photosynthetic angiosperms for which comprehensive intron sequence data are available for the entire chloroplast genome. Unfortunately, the rice *rpo*C1 gene does not contain an intron (Hiratsuka *et al.*, 1989) so its rate of evolution relative to other rice chloroplast introns cannot be ascertained. Each pair of sequences was aligned using the PILEUP procedure of the University of Wisconsin GCG sequence analysis software package (Version 7, Genetics Computer Group, Madison, WI) on a VAX computer. Similarity between sequences was calculated as the proportion of aligned sites having identical nucleotides; gaps were excluded from the calculations. The default gap weight (penalty) of 5.00 and a gap length weight of 0.30 were selected, although it is realized that the position and lengths of gaps created in the alignment procedure may vary depending upon the gap weights used. Slight modifications of these gap penalties, however, did not drastically alter gap position or length and similarity values. No adjustment to the alignment was necessary when sequences forming secondary-structural features were considered.

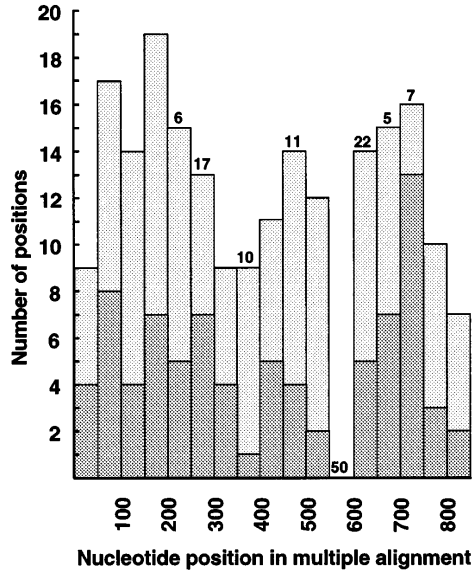


FIG. 3. Number of variable (light shaded bars) and potentially phylogenetically informative (dark shaded bars) positions distributed by 50-bp intervals in the multiple alignment of *rpo*C1 intron sequences for 30 taxa of Apiaceae, Araliaceae, and Pittosporaceae (Fig. 2). Intervals begin at the exon1–intron boundary and end at the intron–exon2 boundary; the last interval spans only 45 bp. The number of excluded nucleotide positions within each interval (owing to ambiguity in alignment) is indicated above each bar (e.g., all 50 nucleotides between positions 550 and 600 were omitted from the analysis).

RESULTS

Sequence analysis. Alignment of all 30 *rpo*C1 intron sequences resulted in a matrix of 845 positions (Fig. 2). Because of alignment ambiguities, however, it was necessary to delete 128 positions (15.1%)—these positions are identified by asterisks in Fig. 2. Of the remaining 717 unambiguously aligned nucleotide sites, 513 (71.5%) were unvarying, 123 (17.2%) were unique to individual taxa, and 81 (11.3%) had at least two nucleotide states in two or more sequences and were potentially informative phylogenetically. The distribution of the numbers of variable and informative nucleotide positions across the entire intron region is somewhat uniform, with the exception of a region exhibiting a peak of informative positions near its 3′ end (Fig. 3). Proper alignment of these unambiguously aligned sequences required the introduction of 25 gaps, 9 of which were 1 bp in length and 8 of which were 2 bp in length (Table 2). The remainder ranged from 3 to 22 bp in length, with the average size of all gaps being approximately 3 bp. Of these 25 gaps, 7 were potentially informative for phylogenetic analysis. These introns, like all noncoding DNA (Palmer, 1991), are AT-rich in base composition. Percentage A+T content ranged from 62.4 to 65.2%, and averaged 63.9%. These values are similar to those reported for other chloroplast introns in tobacco and rice (Shimada and Sugiura, 1991).

The nucleotides participating in either secondary structure or three-dimensional pairings, as inferred by a model based on the consensus of 68 group II intron sequences by Michel *et al.* (1989) and the results of the MULFOLD analysis, are indicated by vertical bars above the alignment in Fig. 2. The number of these positions is 232 (27% of the entire alignment, including ambiguous regions), of which only 8 (3%) were informative phylogenetically. Based on these data, the approximate locations of the 6 major structural domains in the intron were also realized (Fig. 2). Domain I is the largest, comprising almost 58% of the total length of the intron, whereas domains V and VI, each comprising about 4% of the length of the intron, are the smallest. The most variable domain, ascertained by dividing the number of variable (and unambiguous) positions in each region by its size, was domain II with 28.2% of its positions variable. This was followed by domains IV, I, III, VI, and V, with 26.7, 26.2, 20.4, 18.5, and 17.6% of positions variable, respectively. As expected, the vast majority of these variable positions, and all ambiguous positions, occurred in the single-stranded loop regions of the intron.

In direct pairwise comparisons of unambiguous positions among all accessions, sequence divergence ranged from identity (between *Arracacia brandegei* and *Prionosciadium*) to 11.3% of nucleotides (between *Scandix* and the outgroup *Pittosporum*) and averaged 3.8%. Comparison of only Apioideae sequence pairs yielded a maximum divergence value of 5.8% (between *Scandix*

TABLE 2

Characteristics of the 25 Length Mutations Inferred in the Multiple Alignment (Fig. 2) of Unambiguous *rpo*C1 Intron Sequences in 30 Taxa of Apiaceae, Araliaceae, and Pittosporaceae

No.	Position	Size (bp)	Informative?[a]	Type[b]
1	33–34	2	No	Insertion
2	67–74	8	No	Deletion
3	169	1	No	Insertion
4	208	1	Yes (2 taxa)	Deletion A
5	241–244	4	No	Deletion
6	241–243	3	No	Deletion
7	293–296	4	No	Deletion
8	297–303	7	No	Deletion
9	305–306	2	No	Insertion
10	370–371	2	No	Insertion
11	432–433	2	No	Insertion
12	463–484	22	No	Deletion
13	475–476	2	No	Deletion
14	480	1	No	Deletion
15	523	1	Yes (14 taxa)	Deletion B
16	526	1	Yes (26 taxa)	Deletion C
17	626	1	No	Deletion
18	690	1	No	Deletion
19	741–745	5	No	Deletion
20	755–761	7	Yes (13 taxa)	Deletion D
21	757–758	2	Yes (26 taxa)	Insertion E
22	765–766	2	No	Insertion
23	818	1	Yes (16 taxa)	Deletion F
24	822	1	No	Deletion
25	830–831	2	Yes (24 taxa)	Insertion G

[a] Length mutation potentially informative phylogenetically and, if yes, number of taxa possessing mutation.

[b] Type of length mutation relative to the outgroup *Pittosporum*. Potentially informative length mutations are identified as A through G.

and *Heteromorpha*). Excluding the basal *Heteromorpha* (discussed below), the maximum divergence value among all remaining apioid accessions was 4.6% (between *Scandix* and *Apium*). It should be noted that as a consequence of excluding a large region of low similarity from the distance calculations (positions 551–622, Fig. 2), these divergences are slightly lower than they would have been if this region was included.

Many indel events in noncoding regions involve short direct repeats that are likely to be the result of slipped-strand mispairing during replication (Takaiwa and Sugiura, 1982; Zurawski *et al.*, 1984). Such mutations were inferred here and include insertion events 1, 3, 9, 11, 21, 22, and 25 (Table 2), each involving either a 1- or 2-bp repeated sequence. Several larger repetitive insertions were identified but are in those regions of the alignment excluded from the analysis; these tandem repeats make it impossible to determine the homology of specific repeat units across all sequences in a multiple alignment. Perfect direct repeats of this type include the nucleotide sequence CGTAAA at positions 225–230 (in *Pastinaca* and one species of *Heracleum*), and the sequence TAAGGGCA at positions 360–367 [in *Arracacia* (both species), *Coulterophytum*, *Myrrhidendron*, *Prionosciadium*, *Rhodosciadium*, and *Zizia*]. Slightly imperfect repeat motifs include those in *Anthriscus* (positions 276–285), *Eryngium*, *Heteromorpha* and *Hydrocotyle* (positions 485–495), and *Centella*, *Heteromorpha* and *Hydrocotyle* (positions 611–622). A succession of T's and G's in all sequences between positions 691 and 713 of the alignment is also apparent.

Phylogenetic analysis. Parsimony analysis of equally weighted character states resulted in 30 maximally parsimonious topologies, whose strict consensus tree with accompanying bootstrap values is shown in Fig. 4. These trees have a length of 287 steps, CIs of 0.812 (all characters) and 0.656 (excluding uninformative characters), and a RI of 0.809. The CI value of 0.656 is considerably higher than that of the expected value of 0.432 for 30 taxa (Sanderson and Donoghue, 1989). The g_1 statistic for 10,000 random trees generated from these data was -1.125. This value is significantly more skewed than random data ($g_1 = -0.09$ for 250 variable positions and 25 or more taxa; $P < 0.01$), indicating that these data contain significant amounts of phylogenetic signal (Hillis and Huelsenbeck, 1992).

One of these 30 maximally parsimonious 287-step trees was arbitrarily chosen and is presented in Fig. 5 to indicate the number of nucleotide substitutions supporting each branch, as optimized by ACCTRAN in PAUP, and the distribution of the seven phylogenetically informative length mutations (Table 2). Four of these mutations (indels A, C, E, and G) were perfectly congruent (i.e., synapomorphic) with the phylogeny inferred by nucleotide substitutions; the remaining three (indels B, D, and F) were inferred to be homoplastic. Indels B and F are in regions of the alignment characterized by runs of G's and it is probable that a small but unobservable compression in each of these regions may have accounted for the independent derivation of deletion B in *Anthriscus* and the reversal of mutation F in *Aethusa*. Indel D, a 7-bp sequence inferred to be absent in the common ancestor of Apiaceae but present in a major clade of Apioideae, is also likely a reversal. An alternative yet nonparsimonious explanation to account for the distribution of this deletion is to suggest that it occurred four times during the evolution of the group (that is, independently in *Eryngium*, *Heteromorpha*, the clade consisting of *Anethum* through *Pimpinella*, and the clade consisting of *Anthriscus* through *Smyrnium*). The occurrence of homoplastic indels indicates, however, that such mutations can be misleading for studies of phylogenetic inference if used independently of nucleotide substitutions (Morton and Clegg, 1993; Golenberg *et al.*, 1993).

The elimination of all (232) inferred paired positions from a subsequent parsimony analysis, resulting in the removal of only eight informative nucleotide sites, did

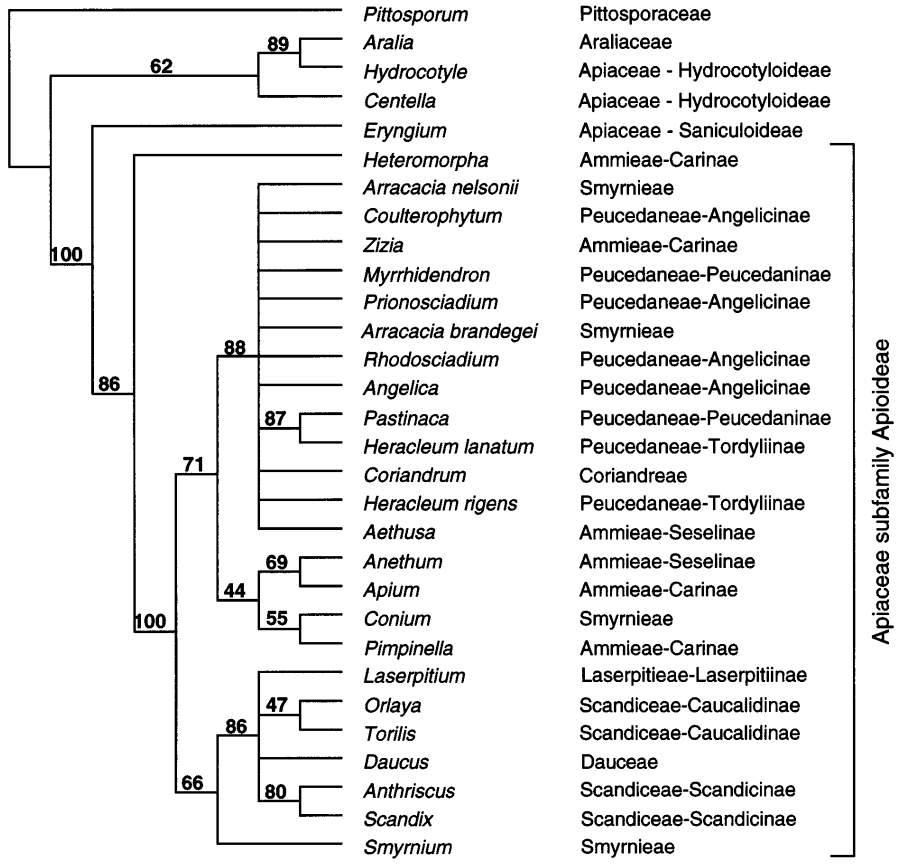


FIG. 4. Strict consensus of the 30 maximally parsimonious 287-step trees derived from unweighted parsimony analysis of chloroplast DNA *rpo*C1 intron sequences (CI excluding uninformative characters = 0.656, RI = 0.809). Numbers above the nodes indicate the number of times a monophyletic group occurred in 100 bootstrap replicates. Complete taxon names are provided in Table 1. Infrasubfamilial classification of Apiaceae subfamily Apioideae based on Drude (1898).

little to affect the topology of the phylogenetic estimate. The strict consensus of the resultant trees (n = 50, length = 246 steps, CI excluding uninformative characters = 0.650, RI = 796) was identical to that illustrated in Fig. 4, except for the collapse of the *Orlaya*+*Torilis* clade. Weighting paired positions by one-half relative to unpaired positions resulted in precisely the same topology and number of trees as when all positions were treated equally.

The average transition/transversion ratio in all *rpo*C1 intron sequences across all 30 minimal length trees, as determined by MacClade, was 1.17. When the average observed transition/transversion ratio of 1.2 was used in a weighted parsimony analysis, the same 30 trees resulted as in the equally weighted analysis. Identical results were obtained when transversions were weighted 1.1 or 1.5 times over transitions. When transversions were used alone the phylogenetic resolution obtained decreased substantially, largely due to the two-thirds reduction in the number of available informative characters. Although the basal relationships in the strict consensus tree derived from these data were very similar to those depicted in Fig. 4 and described below, much resolution was lost among the members of Apioideae. The low levels of DNA sequence divergence exhibited among these apioid taxa and their close evolutionary relatedness, however, suggest that transitions may not be wholly saturated and, therefore, should not be ignored.

The tree obtained using the maximum likelihood method and a transition/transversion ratio of 1.2 had a *ln* likelihood of −2703.5 (Fig. 6). This maximum likelihood tree possesses the same basal relationships as detected by the parsimony analyses. Regions of the topology that were weakly supported or unresolved in the parsimony analysis were similarly supported or unresolved when the maximum likelihood method was invoked.

Phylogenetic resolutions. Using either *Pittosporum* or *Aralia* as the outgroup, all phylogenetic analyses confirm the monophyly of subfamily Apioideae and its sister-group status to *Eryngium*, the only representative examined of subfamily Saniculoideae. Furthermore, the results reveal that, among all the members of Apioideae included in the analysis, *Heteromorpha* is

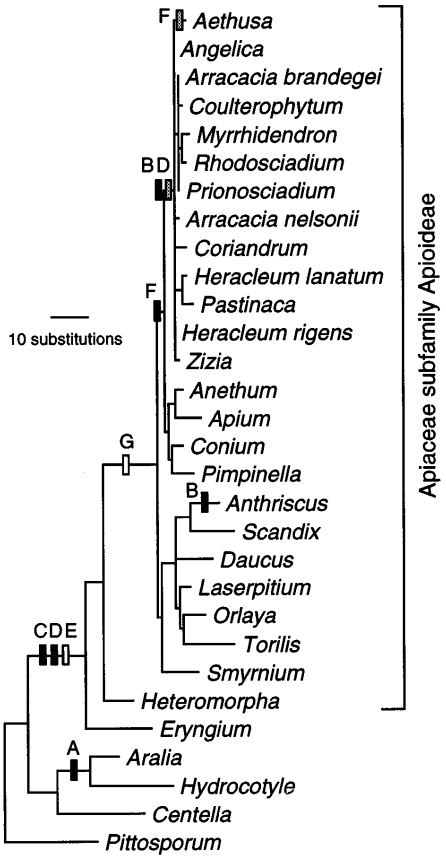


FIG. 5. One of 30 maximally parsimonious trees of 287 steps derived from parsimony analysis of *rpo*C1 intron sequence data using equally weighted character states. Lengths of branches are proportional to the number of inferred nucleotide substitutions occurring along them (note scale bar). The distribution of phylogenetically informative length mutations have been superimposed on the phylogram and are identified by letters corresponding to their locations in the multiple alignment and their designations in Table 2. Deletions, solid bars; insertions, open bars; reversals, shaded bars.

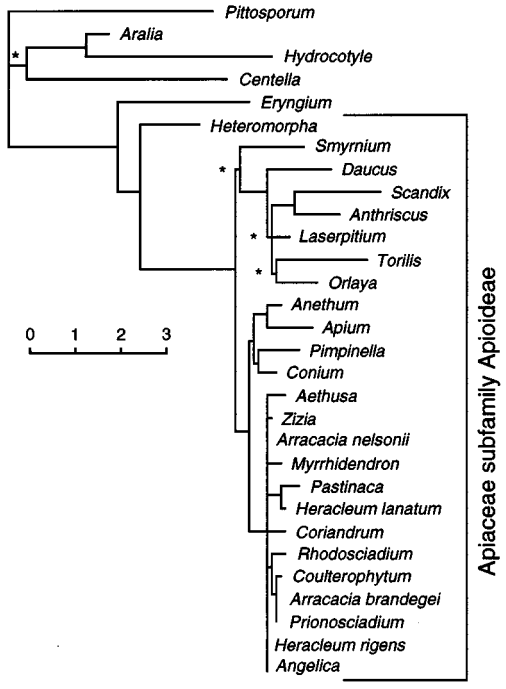


FIG. 6. Maximum likelihood tree constructed from unambiguous *rpo*C1 intron sequences using a transition/transversion ratio of 1.2. All branch lengths, unless otherwise indicated, are significantly positive at $P < 0.01$. Four branches are significantly positive at $P < 0.05$ and are indicated by asterisks. Branches of zero length have been collapsed. Complete taxon names are provided in Table 1. Scale distance is given as 100 × value.

the earliest diverging lineage. *Aralia* emerges from within a paraphyletic Apiaceae subfamily Hydrocotyloideae (represented by *Hydrocotyle* and *Centella*) and this clade is sister to Apioideae+Saniculoideae.

The phylogenetic analyses further reveal that, in the context of those species examined, Apioideae (less *Heteromorpha*) comprise two major clades. The first of these is represented by genus *Smyrnium* (Smyrnieae) and those taxa belonging to Drude's (1898) tribes Dauceae, Laserpitieae, and Scandiceae (Fig. 4). The two subtribes of Scandiceae, Caucalidinae (*Orlaya* and *Torilis*) and Scandicinae (*Anthriscus* and *Scandix*), are each monophyletic; the former clade, however, is supported weakly. The second major phylogenetic division within the subfamily comprises the two other generic representatives of the tribe Smyrnieae (*Arracacia* and *Conium*) and those taxa belonging to Drude's tribes Ammieae, Coriandreae, and Peucedaneae. Within this larger clade, *Anethum*+*Apium* and *Conium*+*Pimpinella* comprise a monophyletic group, although the relationships here are supported poorly in the parsimony analysis. Each of the two accessions examined of *Arracacia* (*A. nelsonii* and *A. brandegei*) and *Heracleum* (*H. lanatum* and *H. rigens*) do not arise together. Constraining the parsimony analysis so that the two *Arracacia* species are monophyletic produced minimal-length trees one step longer than those produced without the constraint. Forcing the two species of *Heracleum* together resulted in minimal-length trees of 289 steps, two steps longer than those without the constraint. *H. lanatum* and *Pastinaca* unite in all phylogenetic analyses to form a strongly supported clade. Relationships within each of the two major clades of Apioideae are largely unresolved owing to low levels of nucleotide sequence divergence among the taxa.

Tobacco and rice intron sequence comparisons. Seventeen of the introns occurring in tobacco cpDNA are also present in the chloroplast genome of rice; in rice, the chloroplast genes *rpo*C1 and *clp*P do not contain intervening sequences (Shinozaki *et al.*, 1986; Hiratsuka *et al.*, 1989). The results of the pairwise comparisons of these 17 shared intron sequences, with respect to their location in the chloroplast genome, are presented in Table 3. The five introns located within the inverted repeat regions of these genomes were highly

TABLE 3

Size and Percentage Similarity between Tobacco and Rice Chloroplast DNA Intron Sequences, Arranged in Descending Similarity with Respect to Genome Region

Interrupted gene	Intron size (bp)		% Similarity
	Tobacco	Rice	
Inverted repeat region			
3′-*rps*12	536	540	96.1
*trn*A-UGC	709	812	96.0
*ndh*B	679	712	94.7
*trn*I-GAU	707	947	94.0
*rpl*2	666	663	93.0
Large single-copy region			
*trn*V-UAC	571	597	78.3
*trn*G-UCC	691	678	76.6
*atp*F	695	828	74.7
ORF168/170 5′intron[a]	738	745	73.9
ORF168/170 3′intron[a]	783	729	73.5
*pet*D	742	745	70.7
*trn*L-UAA	503	540	70.3
*pet*B	753	811	69.2
*trn*K-UUU[b]	995	868	68.8
*rps*16	860	809	67.2
*rpl*16	1020	1059	64.5
*rpo*C1	738	Missing[c]	N/A
*clp*P 5′intron[d]	807	Missing[c]	N/A
*clp*P 3′intron[d]	642	Missing[c]	N/A
Small single-copy region			
*ndh*A	1148	987	67.1

[a] ORF168 in tobacco; ORF170 in rice.
[b] Includes both 3′ and 5′ noncoding portions; the gene *mat*K is excluded.
[c] This intron is missing from the rice chloroplast genome.
[d] ORF196.

conserved in nucleotide sequence (ranging between 93.0 and 96.1% similarity), whereas the 12 remaining introns, located within the (large and small) single-copy regions of these genomes, are much less conserved (ranging between 64.5 and 78.3% similarity).

DISCUSSION

Molecular evolution of rpoC1 intron sequences. The intron in *rpo*C1 can be further divided into subgroup IIB on the basis of the presence of the following three features recognized by Michel *et al.* (1989): (1) an internal loop with conserved motif AAGC in subdomain C1 of domain I (positions 118–121, Fig. 2); (2) a bulging A on the 3′ side of the stem region in domain VI (position 838); (3) a slightly asymmetrical terminal loop in subdomain D3 of domain I (position 341–349), which bears exon-binding site 1 (EBS1). Comparisons of *rpo*C1 exon–intron boundary sequences reveal that these regions are highly conserved in land plants. The 5′ (GTGTGATTTG) and 3′ (CTATCCCAAT) boundary sequences in most Apiaceae and relatives are identical to those comparable regions reported in tobacco (Shinozaki *et al.*, 1986) and spinach (Hudson *et al.*, 1988) *rpo*C1 intron sequences and only differ very slightly from those in *Marchantia* (Ohyama *et al.*, 1986), and *Aethusa* and *Daucus* (Fig. 2). Moreover, these boundary sequences are similar or identical to other splice sites for group II introns in other chloroplast protein-coding genes (Michel *et al.*, 1989).

Each of the six domains characteristic of group II introns can vary markedly in their acceptance of mutational change. For example, in a detailed investigation of cpDNA *trn*V(UAC) intron sequences, Learn *et al.* (1992) showed that domain II had the highest rate of sequence divergence, approaching the synonymous substitution rate reported for protein-coding genes. In contrast, stem-loop region D3 and domain V were determined to evolve much more slowly, at rates similar to those reported for nonsynonymous substitutions of several chloroplast genes. These disparate evolutionary rates appear to be tied to the functional importance of these regions in intron processing (Learn *et al.*, 1992; Clegg *et al.*, 1994). In our analysis, domain II was also deemed to be the most variable region, but only slightly more so than domains I and IV (the various subdomains of domain I were not identified or considered in this study). Conversely, domains V and VI exhibited the lowest amount of sequence divergence. The primary sequence and secondary structures of domains V and VI have previously been shown to be highly conserved in group II introns, as these regions are necessary for proper processing of the transcript (Michel and Dujon, 1983; Michel *et al.*, 1989).

Of the 30 *rpo*C1 intron sequences determined here, 22 fell between the narrow range of 745 and 765 bp; overall, their sizes varied from 719 to 785 bp (Fig. 2). These sizes lie within the range of those reported for two other angiosperms (spinach, 756 bp [Hudson *et al.*, 1988]; tobacco, 738 bp]).[2] Considering the size and type of gaps in the multiple alignment, single-nucleotide deletions were the most frequent (8/25 events) and the frequency of gaps decreased as their length increased (Table 2). As has been observed in other studies of noncoding DNA, a major source of variation is the generation of short (1–10 bp) deletions or duplications, most likely due to slipped-strand mispairing events (Levinson and Gutman, 1987). Overall, deletions of any size (17/25 events) were over twice as frequent as insertions. Again, these results support other studies where it has been demonstrated that deletions, particularly single-nucleotide deletions, usually outnumber inser-

[2] In tobacco, the intron in the gene *rpo*C1 was originally reported as being 739 bp in size (Shinozaki *et al.*, 1986). We have followed Shimada *et al.* (1990) in assuming that the intron is 738 bp due to the erroneous earlier report of a T (thymine) nucleotide after position 23,617 in the tobacco cpDNA sequence.

tions in noncoding regions (Golenberg *et al.*, 1993; Saitou and Ueda, 1994).

Phylogenetic relationships. Members of Apioideae, the typical "umbellifers," unequivocally form a monophyletic group, being distinguished from those in the other two subfamilies by the shared presence of compound umbels, a specialized fruit consisting of two one-seeded mericarps suspended from a common bifurcate carpophore, a soft endocarp that is sometimes hardened by woody subepidermal layers, a terminal style arising from the stylopodium, fruits without scales, an absence of stipules, and the widespread but not ubiquitous occurrence of flavones, methylated flavonoids, furanocoumarins, and phenylpropenes (Crowden *et al.*, 1969; Harborne, 1971; Hegnauer, 1971; Nielsen, 1971; Heywood, 1982). The molecular results presented here fully corroborate the monophyly of Apioideae.

Although the monophyly of Saniculoideae cannot be assessed as only one accession of this subfamily was available, Hydrocotyloideae (represented by *Centella* and *Hydrocotyle*) appear to be paraphyletic with the one member of Araliaceae (*Aralia*) arising from within. On the basis of *rbc*L sequence data, the hydrocotyloids appear to be an unnatural assemblage, representing at least four independent lineages, three of which are allied with Araliaceae (Plunkett *et al.*, in press). In light of these recent *rbc*L-based results, and considering the sparse sampling of Araliaceae and hydrocotyloids carried out in this study, the inferred paraphyly of Hydrocotyloideae may be incorrect.

Dawson (1971) has indicated that herbaceous Apioideae have likely evolved from montane tropical woody apioid ancestors. Although the majority of Apioideae are characterized by a herbaceous habit, several members, including *Heteromorpha* and *Myrrhidendron*, are woody. The wood anatomy of *Heteromorpha* (and *Pittosporum*) is much like that found in many Araliaceae. Most Apiaceae (including *Myrrhidendron*) have vessel elements with predominantly simple perforations, whereas *Heteromorpha*, *Pittosporum* and many Araliaceae possess double perforations (Rodriguez, 1971). On the basis of our results, the woody habit of *Myrrhidendron* is secondarily derived. The basal position of *Heteromorpha* in the subfamily, as ascertained here, is also indicated in a phylogeny based on *rbc*L data (Plunkett *et al.*, in press).

The two major subgroups of Apioideae identified in this study (excluding the basal *Heteromorpha*) were also revealed by cladistic analysis of nuclear ribosomal DNA ITS sequences (Downie and Katz-Downie, 1996). The close relationships among the allied and polyploid New World genera *Arracacia*, *Coulterophytum*, *Myrrhidendron*, *Prionosciadium*, and *Rhodosciadium* (Table 1), as inferred by these ITS comparisons, are not depicted in any of the trees presented here. Nucleotide divergence values of *rpo*C1 intron sequences among these accessions range from identity (between *Arracacia brandegei* and *Prionosciadium*) to 0.6% (between *Arracacia nelsonii* and *Rhodosciadium*) and are too low to be useful in providing sufficient phylogenetic information. Each of these taxa, however, along with the North American endemic *Zizia*, share an 8-bp insertion relative to *Pittosporum* (positions 360–367, Fig. 2), providing compelling evidence for their common ancestry. It is not unrealistic, however, to presume that apparently identical indels may have had multiple origins in unrelated taxa (Downie *et al.*, 1991; Golenberg *et al.*, 1993). The separation of *Anethum*, *Apium*, *Conium* and *Pimpinella* as a clade from the other examined members of their respective tribes or subtribes is also reflected in the ITS study where these taxa (and others not included in this analysis) lie basal within this major phylogenetic division of Apioideae. Of the four tribes and six subtribes recognized by Drude (1898) for which more than one generic representative was examined, only two subtribes (Caucalidinae and Scandicinae) are clearly monophyletic. Tribes Smyrnieae and Ammieae appear polyphyletic, and the putative monophyly of all remaining taxa remains unknown with the data at hand.

Utility of rpoC1 intron sequences in phylogeny estimation. Phylogenies derived from the *rpo*C1 intron are largely equivocal in their abilities to explain relationships among many of the members of Apioideae included in this study, owing to the extensive similarity exhibited among the sequences [513 of the 717 (71.5%) unambiguously aligned nucleotide sites were identical for all species in which they were present]. Much of this invariant sequence—189 of 513 sites (37%)—is integral to the formation of secondary structure for group II introns; many of these positions are also conserved in *Marchantia* plastid group II introns (Michel *et al.*, 1989). This, along with the small size of the region—a typical angiosperm *rbc*L sequence is almost twice its length, for example—reduces the number of phylogenetically informative characters available to the analysis. A sequence divergence value of between 5 and 15% among the taxa compared usually indicates an appropriate rate of divergence that will minimize multiple hits while still providing a sufficient number of characters to resolve relationships (Templeton, 1983; Ritland and Clegg, 1990; Olmstead and Palmer, 1994). Within Apioideae, divergence values in pairwise comparisons fell at the lower end of this range, approaching 5.8% of nucleotides. These low values indicate that *rpo*C1 intron sequences, unlike the introns in chloroplast genes *trn*L and *trn*V (Clegg *et al.*, 1986; Learn *et al.*, 1992; Fangan *et al.*, 1994; Gielly and Taberlet, 1994), are largely inappropriate to infer phylogeny among closely related taxa. Instead, the *rpo*C1 intron data appear well suited to comparisons among distantly related species, as sequence divergence values among members of

Pittosporaceae, Araliaceae, and some Apiaceae reached just over 11% of nucleotides.

The apparently limited resolution of these intron sequences among most Apioideae examined might also reflect a real biological phenomenon—the rapid and recent radiation of this lineage. To this end, the resolution obtained by the analysis of these intron sequences is not significantly worse than that provided by *rbc*L, as the latter also proved to be somewhat limited in its ability to resolve relationships among the higher apioids. However, resolution of such a putatively young group has been achieved, in part, using data from the more rapidly evolving chloroplast *mat*K gene (G. Plunkett, D. Soltis, and P. Soltis, unpublished data) and nuclear ITS regions (Downie and Katz-Downie, 1996). Therefore, the lack of variation exhibited by *rpo*C1 intron sequences in Apioideae should not preclude the use of these data in inferring relationships in other groups of plants of comparable and deeper taxonomic ranks, as greater phylogenetic resolution might be achieved in those groups having a longer evolutionary history.

The upper limits (i.e., phylogenetic depth) to which *rpo*C1 intron sequences can be applied have yet to be realized. Pairwise nucleotide differences of available *rpo*C1 intron sequences, excluding gaps, from *Apium* (this study), tobacco (Shinozaki *et al.*, 1986), and spinach (Hudson *et al.*, 1988) were 17.4% (for the *Apium*/tobacco comparison), 18.4% (for tobacco/spinach), and 21.7% (for *Apium*/spinach). Although extensive sequence conservation exists among these taxa (approximately 74% of all positions are invariant), numerous short length mutations, ranging from 1 to 12 bp, were also apparent, affecting 13% of all positions. This conservative mode of intron evolution plus the fact that introns are not subjected to codon bias, as is the case with protein-coding genes, suggest that these sequences should therefore be good choices for further analyses among distantly related species. Their small size and abundance of length mutations confounding alignment, however, largely preclude them from being useful at this level.

Phylogenetic utility of other cpDNA introns. Evolutionary rates vary among cpDNA introns, suggesting that not all introns are useful for phylogenetic analyses at the same taxonomic level. Pairwise comparisons of the 17 chloroplast introns shared between tobacco and rice indicate that five introns (i.e., *rpl*2, *ndh*B, 3′*rps*12, *trn*I and *trn*A) are extraordinarily conserved in nucleotide sequence (Table 3). This striking sequence conservation is likely due to the intron's position in the chloroplast genome's inverted repeat (Fig. 1), a region whose mutation rate is at least three times slower than that of single-copy sequences (Wolfe *et al.*, 1987). It may also reflect unusually strong functional constraints on the evolution of these particular introns. Obviously, these five introns will be most informative at the highest levels of evolutionary divergence. Sequence comparisons among the 11 large single-copy (LSC) region introns range from 64.5 to 78.3%, while the sole intron in the small single-copy (SSC) region has a sequence similarity between tobacco and rice of 67.1%. This heterogeneity in rate of cpDNA intron evolution, relative to genomic region, has been reported previously by Clegg *et al.* (1991). Thus, the 17–20 introns in the chloroplast genomes of most angiosperms encompass a wide range of evolutionary rates and are suitable for phylogenetic study over a diversity of taxonomic levels. The utility of the majority of these introns in phylogeny estimation and their molecular evolution have yet to be examined.

CONCLUSIONS

Partial concordance between the consensus trees derived from parsimony analyses of cpDNA *rpo*C1 intron (this study) and *rbc*L sequences (Plunkett *et al.*, in press) and nuclear ribosomal DNA internal transcribed spacers (Downie and Katz-Downie, 1996) suggests that *rpo*C1 intron nucleotide substitutions can prove useful for addressing phylogenetic relationships among distantly related members of Apiaceae subfamily Apioideae and between members of Apiaceae and Araliaceae/Pittosporaceae. Length mutations in these intron sequences also appear to be of potential value to phylogeny estimation, with one (an 8-bp insertion at positions 360–367) uniting *Zizia* with several polyploid Mexican and Central American endemics. In contrast, several homoplastic length mutations are also apparent and have the potential to confound cladistic analyses. Limitations imposed by the small size of the region can be overcome, in part, by combining these intron sequences with sequence data obtained from other chloroplast genes or noncoding regions.

Our analyses demonstrate that chloroplast *rpo*C1 intron sequence data have little power to resolve phylogenetic relationships at low taxonomic levels. It has too few informative sites to allow for analyses of close relatives, and it evolves too rapidly for reliable comparisons among distantly related angiosperms. Although length mutations within the intron or even the presence or absence of the intron itself can serve to delimit major taxonomic groupings at deep levels, these kinds of characters are sparse. At an appropriate level of divergence, however, *rpo*C1 intron data can provide phylogenetically useful information.

ACKNOWLEDGMENTS

The authors thank G. Plunkett, L. Constance, and the Botanical Garden of the University of California (Berkeley), for providing us with leaf material; the various botanical gardens cited in the text for providing seeds; L. Constance for confirming the plant identifications; W. Yang for laboratory assistance; and J. Palmer, G. Plunkett,

and two anonymous reviewers for helpful comments on an earlier draft of the manuscript. This work was supported by laboratory start-up funds from the University of Illinois, and by grants from the National Science Foundation (DEB-9407712) and the Campus Research Board of the University of Illinois.

REFERENCES

Arnold, M. L., Buckner, C. M., and Robinson, J. J. (1991). Pollen-mediated introgression and hybrid speciation in Louisiana irises. *Proc. Natl. Acad. Sci. USA* **88:** 1398–1402.

Bremer, K. (1988). The limits of amino acid sequence data in angiosperm phylogenetic reconstruction. *Evolution* **42:** 795–803.

Clegg, M. T., and Zurawski, G. (1992). Chloroplast DNA and the study of plant phylogeny: Present status and future prospects. *In* "Molecular Systematics of Plants" (P. S. Soltis, D. E. Soltis, and J. J. Doyle, Eds.), pp. 1–13. Chapman and Hall, New York.

Clegg, M. T., Ritland, K., and Zurawski, G. (1986). Processes of chloroplast DNA evolution. *In* "Evolutionary Processes and Theory" (S. Karlin and E. Nevo, Eds.), pp. 275–294. Academic Press, New York.

Clegg, M. T., Learn, G. H., and Golenberg, E. M. (1991). Molecular evolution of chloroplast DNA. *In* "Evolution at the Molecular Level" (R. K. Selander, A. G. Clark, and T. S. Whittam, Eds.), pp. 135–149. Sinauer Associates, Sunderland, MA.

Clegg, M. T., Gaut, B. S., Learn, G. H., Jr., and Morton, B. R. (1994). Rates and patterns of chloroplast DNA evolution. *Proc. Natl. Acad. Sci. USA* **91:** 6795–6801.

Cozens, A. L., and Walker, J. E. 1986. Pea chloroplast DNA encodes homologues of *Escherichia coli* ribosomal subunit S2 and the β′ subunit of RNA polymerase. *Biochem. J.* **236:** 453–460.

Cronquist, A. (1981). An integrated system of classification of flowering plants. Columbia Univ. Press, New York.

Crowden, R. K., Harborne, J. B., and Heywood, V. H. (1969). Chemosystematics of the Umbelliferae—A general survey. *Phytochemistry* **8:** 1963–1984.

Curtis, S. E., and Clegg, M. T. (1984). Molecular evolution of chloroplast DNA sequences. *Mol. Biol. Evol.* **1:** 291–301.

Dahlgren, R. T. (1980). A revised system of classification of the angiosperms. *Bot. J. Linn. Soc.* **80:** 91–124.

Dawson, J. W. (1971). Relationships of the New Zealand Umbelliferae. *In* "The Biology and Chemistry of the Umbelliferae" (V. H. Heywood, Ed.), pp. 43–61. Academic Press, New York.

Dixon, M. T., and Hillis, D. M. (1993). Ribosomal RNA secondary structure—Compensatory mutations and implications for phylogenetic analysis. *Mol. Biol. Evol.* **10:** 256–267.

Doebley, J., Durbin, M., Golenberg, E. M., Clegg, M. T., and Ma, D. P. (1990). Evolutionary analysis of the large subunit of carboxylase (*rbc*L) nucleotide sequence among the grasses (Gramineae). *Evolution* **44:** 1097–1108.

Downie, S. R., and Katz-Downie, D. S. (1996). A molecular phylogeny of Apiaceae subfamily Apioideae: Evidence from nuclear ribosomal DNA internal transcribed spacer sequences. *Am. J. Bot.* **83:** 234–251.

Downie, S. R., Olmstead, R. G., Zurawski, G., Soltis, D. E., Soltis, P. S., Watson, J. C., and Palmer, J. D. (1991). Six independent losses of the chloroplast DNA *rpl*2 intron in dicotyledons: Molecular and phylogenetic implications. *Evolution* **45:** 1245–1259.

Doyle, J. J. (1993). DNA, phylogeny, and the flowering of plant systematics. *Bioscience* **43:** 380–389.

Doyle, J. J., and Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* **19:** 11–15.

Drude, O. (1898). Umbelliferae. *In "Die natürlichen Pflanzenfamilien"* (A. Engler and K. Prantl, Eds.), Vol. 3, pp. 63–250. Wilhelm Engelmann, Leipzig, Germany.

Fangan, B. M., Stedje, B., Stabbetorp, O. E., Jensen, E. S., and Jakobsen, K. S. (1994). A general approach for PCR-amplification and sequencing of chloroplast DNA from crude vascular plant and algal tissue. *BioTechniques* **16:** 484–494.

Farris, J. S. (1989). The retention index and homoplasy excess. *Syst. Zool.* **38:** 406–407.

Felsenstein, J. (1981). Evolutionary trees from DNA sequences: A maximum likelihood approach. *J. Mol. Evol.* **17:** 368–376.

Felsenstein, J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* **39:** 783–791.

Gielly, L., and Taberlet, P. (1994). The use of chloroplast DNA to resolve plant phylogenies: Noncoding versus *rbc*L sequences. *Mol. Biol. Evol.* **11:** 769–777.

Golenberg, E. M., Clegg, M. T., Durbin, M. L., Doebley, J., and Ma, D. P. (1993). Evolution of a noncoding region of the chloroplast genome. *Mol. Phylogenet. Evol.* **2:** 52–64.

Harborne, J. B. (1971). Flavonoid and phenylpropanoid patterns in the Umbelliferae. *In* "The Biology and Chemistry of the Umbelliferae" (V. H. Heywood, Ed.), pp. 293–314. Academic Press, New York.

Hegnauer, R. (1971). Chemical patterns and relationships of Umbelliferae. *In* "The Biology and Chemistry of the Umbelliferae" (V. H. Heywood, Ed.), pp. 267–277. Academic Press, New York.

Heywood, V. H. (1982). General introduction to the taxonomy of the Umbelliferae. *In* Les Ombellifères. Actes du 2ème Symposium International sur les Ombellifères "Contributions Pluridisciplinaires à la Systématique" (A.-M. Cauwet-Marc and J. Carbonnier, Eds.), pp. 107–112. Monographs in Systematic Botany from the Missouri Botanical Garden, Vol. 6. Braun-Brumfield, Ann Arbor, MI.

Heywood, V. H. (1993). "Flowering Plants of the World," Oxford Univ. Press, New York.

Hillis, D. M., and Dixon, M. T. (1991). Ribosomal DNA: Molecular evolution and phylogenetic inference. *Q. Rev. Biol.* **66:** 411–453.

Hillis, D. M., and Huelsenbeck, J. P. (1992). Signal, noise, and reliability in molecular phylogenetic analyses. *J. Hered.* **83:** 189–195.

Hiratsuka, J., Shimada, H., Whittier, R., Ishibashi, T., Sakamoto, M., Mori, M., Kondo, C., Honji, Y., Sun, C.-R., Meng, B.-Y., Li, Y.-Q., Kanno, A., Nishizawa, Y., Hirai, A., Shinozaki, K., and Sugiura, M. (1989). The complete sequence of the rice (*Oryza sativa*) chloroplast genome: Intermolecular recombination between distant tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. *Mol. Gen. Genet.* **217:** 185–194.

Hudson, G. S., Holton, T. A., Whitfeld, P. R., and Bottomley, W. (1988). Spinach chloroplast *rpo*BC genes encode three subunits of the chloroplast RNA polymerase. *J. Mol. Biol.* **200:** 639–654.

Hutchinson, J. (1973). "The Families of Flowering Plants," Oxford Univ. Press, Oxford, England.

Jaeger, J. A., Turner, D. H., and Zuker, M. (1989). Improved predictions of secondary structures for RNA. *Proc. Natl. Acad. Sci. USA* **86:** 7706–7710.

Jay, M. (1969). Chemotaxonomic researches on vascular plants XIX. Flavonoid distribution in the Pittosporaceae. *Bot. J. Linn. Soc.* **62:** 423–429.

Johnson, L. A., and Soltis, D. E. (1994). *mat*K DNA sequences and phylogenetic reconstruction in Saxifragaceae s. str. *Syst. Bot.* **19:** 143–156.

Judd, W. S., Sanders, R. W., and Donoghue, M. J. (1994). Angiosperm family pairs: Preliminary phylogenetic analyses. *Harvard Pap. Bot.* **5:** 1–51.

Kim, K.-J., Jansen, R. K., Wallace, R. S., Michaels, H. J., and Palmer, J. D. (1992). Phylogenetic implications of *rbc*L sequence variation in the Asteraceae. *Ann. Missouri Bot. Gard.* **79:** 428–445.

Knight, A., and Mindell, D. P. (1993). Substitution bias, weighting of DNA sequence evolution, and the phylogenetic position of Fea's viper. *Syst. Biol.* **42:** 18–31.

Kluge, A. G., and Farris, J. S. (1969). Quantitative phyletics and the evolution of anurans. *Syst. Zool.* **18:** 1–32.

Learn, G. H., Jr., Shore, J. S., Furnier, G. R., Zurawski, G., and Clegg, M. T. (1992). Constraints on the evolution of plastid introns: The group II intron in the gene encoding tRNA-Val(UAC). *Mol. Biol. Evol.* **9:** 856–871.

Levinson, G., and Gutman, G. A. (1987). Slipped-strand mispairing: A major mechanism for DNA sequence evolution. *Mol. Biol. Evol.* **4:** 203–221.

Liston, A. (1992). Variation in the chloroplast genes *rpo*C1 and *rpo*C2 of the genus *Astragalus* (Fabaceae): Evidence from restriction site mapping of a PCR-amplified fragment. *Am. J. Bot.* **79:** 953–961.

Maddison, W. P., and Maddison, D. R. (1992). "MacClade, Version 3.0: Analysis of Phylogeny and Character Evolution," Sinauer Associates, Sunderland, MA.

Manen, J.-F., Natali, A., and Ehrendorfer, F. (1994). Phylogeny of Rubiaceae–Rubieae inferred from the sequence of a cpDNA intergene region. *Pl. Syst. Evol.* **190:** 195–211.

Michel, F., and Dujon, B. (1983). Conservation of RNA secondary structures in two intron families including mitochondrial-, chloroplast- and nuclear-encoded members. *EMBO J.* **2:** 33–38.

Michel, F., Umesono, K., and Ozeki, H. (1989). Comparative and functional anatomy of group II catalytic introns—A review. *Gene* **82:** 5–30.

Mindell, D. P., and Honeycutt, R. L. (1990). Ribosomal RNA in vertebrates: Evolution and phylogenetic applications. *Annu. Rev. Ecol. Syst.* **21:** 541–566.

Morton, B. R., and Clegg, M. T. (1993). A chloroplast DNA mutational hotspot and gene conversion in a noncoding region near *rbc*L in the grass family (Poaceae). *Curr. Genet.* **24:** 357–365.

Nielsen, B. E. (1971). Coumarin patterns in the Umbelliferae. *In* "The Biology and Chemistry of the Umbelliferae" (V. H. Heywood, Ed.), pp. 325–336. Academic Press, New York.

Ohyama, K., Fukuzawa, H., Kohchi, T., Shirai, H., Sano, T., Sano, S., Umesono, K., Shiki, Y., Takeuchi, M., Chang, Z., Aota, S., Inokuchi, H., and Ozeki, H. (1986). Complete nucleotide sequence of liverwort *Marchantia polymorpha* chloroplast DNA. *Pl. Mol. Biol. Rep.* **4:** 148–175.

Olmstead, R. G., and Palmer, J. D. (1994). Chloroplast DNA systematics: A review of methods and data analysis. *Am. J. Bot.* **81:** 1205–1224.

Olmstead, R. G., and Sweere, J. A. (1994). Combining data in phylogenetic systematics: An empirical approach using three molecular data sets in the Solanaceae. *Syst. Biol.* **43:** 467–481.

Olsen, G. J., Matsuda, H., Hagstrom, R., and Overbeek, R. (1994). fastDNAml: A tool for construction of phylogenetic trees of DNA sequences using maximum likelihood. *CABIOS* **10:** 41–48.

Palmer, J. D. (1991). Plastid chromosomes: Structure and evolution. *In* "Cell Culture and Somatic Cell Genetics of Plants" (L. Bogorad and I. K. Vasil, Eds.), Vol. 7A, pp. 5–53. Academic Press, San Diego, CA.

Plunkett, G. M., Soltis, D. E., and Soltis, P. S. Higher level relationships of Apiales (Apiaceae and Araliaceae) based on phylogenetic analysis of *rbc*L sequences. *Am. J. Bot.*, in press.

Reeder, T. W. (1995). Phylogenetic relationships among Phrynosomatid lizards as inferred from mitochondrial ribosomal DNA sequences: Substitutional bias and information content of transitions relative to transversions. *Mol. Phylogenet. Evol.* **4:** 203–222.

Rieseberg, L. H., Hanson, M. A., and Philbrick, C. T. (1992). Androdioecy is derived from dioecy in Datiscaceae: Evidence from restriction site mapping of PCR-amplified chloroplast DNA fragments. *Syst. Bot.* **17:** 324–336.

Ritland, K., and Clegg, M. T. (1990). Optimal DNA sequence divergence for testing phylogenetic hypotheses. *In* "Molecular Evolution, UCLA Symposia on Molecular and Cellular Biology" (M. T. Clegg and S. J. O'Brien, Eds.), Vol. 122, pp. 289–299. A. R. Liss, New York.

Rodriguez, R. L. C. (1971). The relationships of the Umbellales. *In* "The Biology and Chemistry of the Umbelliferae" (V. H. Heywood, Ed.), pp. 63–91. Academic Press, New York.

Saitou, N., and Ueda, S. (1994). Evolutionary rates of insertion and deletion in noncoding nucleotide sequences of primates. *Mol. Biol. Evol.* **11:** 504–512.

Sanderson, M. J., and Donoghue, M. J. (1989). Patterns of variation in levels of homoplasy. *Evolution* **43:** 1781–1795.

Savolainen, V., Manen, J. F., Douzery, E., and Spichiger, R. (1994). Molecular phylogeny of families related to *Cestrales* based on *rbc*L 5′ flanking sequences. *Mol. Phylogenet. Evol.* **3:** 27–37.

Shimada, H., and Sugiura, M. (1991). Fine structural features of the chloroplast genome: Comparison of the sequenced chloroplast genomes. *Nucleic Acids Res.* **19:** 983–995.

Shimada, H., Fukuta, M., Ishikawa, M., and Sugiura, M. (1990). Rice chloroplast RNA polymerase genes: The absence of an intron in *rpo*C1 and the presence of an extra sequence in *rpo*C2. *Mol. Gen. Genet.* **221:** 395–402.

Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hayashida, N., Matsubayashi, T., Zaita, N., Chunwongse, J., Obokata, J., Yamaguchi-Shinozaki, K., Ohto, C., Torazawa, K., Meng, B.-Y., Sugita, M., Deno, H., Kamagoshira, T., Yamada, K., Kusuda, J., Takaiwa, F., Kato, A., Tohdoh, N., Shimada, H., and Sugiura, M. (1986). The complete nucleotide sequence of the tobacco chloroplast genome: Its organization and expression. *EMBO J.* **5:** 2043–2049.

Smith, A. B. (1989). RNA sequence data in phylogenetic reconstruction: Testing the limits of its resolution. *Cladistics* **5:** 321–344.

Soltis, D. E., Morgan, D. R., Grable, A., Soltis, P. S., and Kuzoff, R. (1993). Molecular systematics of Saxifragaceae sensu stricto. *Am. J. Bot.* **80:** 1056–1081.

Steele, K. P., and Vilgalys, R. (1994). Phylogenetic analysis of Polemoniaceae using nucleotide sequences of the plastid gene *mat*K. *Syst. Bot.* **19:** 126–142.

Stuhlfauth, T., Fock, H., Huber, H., and Klug, K. (1985). The distribution of fatty acids including petroselinic and tariric acids in the fruit and seed oils of the Pittosporaceae, Araliaceae, Umbelliferae, Simarubaceae and Rutaceae. *Biochem. Syst. Ecol.* **13:** 447–453.

Swofford, D. L. (1993). "PAUP: Phylogenetic Analysis Using Parsimony, Version 3.1.1" Computer program distributed by the Illinois Natural History Survey, Champaign, IL.

Taberlet, P., Gielly, L., Pautou, G., and Bouvet, J. (1991). Universal primers for amplification of three non-coding regions of chloroplast DNA. *Pl. Mol. Biol.* **17:** 1105–1109.

Takaiwa, F., and Sugiura, M. (1982). Nucleotide sequences of the 16S–23S spacer region in an rRNA gene cluster from tobacco chloroplast DNA. *Nucleic Acids Res.* **8:** 2665–2676.

Takhtajan, A. (1980). Outline of the classification of flowering plants (Magnoliophyta). *Bot. Rev.* **46:** 225–359.

Templeton, A. R. (1983). Convergent evolution and nonparametric inferences from restriction data and DNA sequences. *In* "Statistical Analysis of DNA Sequence Data" (B. S. Weir, Ed.), pp. 151–179. Dekker, New York.

Thorne, R. F. (1973). Inclusion of the Apiaceae (Umbelliferae) in the Araliaceae. *Notes R. Bot. Garden, Edinburgh* **32:** 161–165.

Thorne, R. F. (1992). Classification and geography of the flowering plants. *Bot. Rev.* **58:** 225–348.

van Ham, R. C. H. J., Hart, H. 't, Mes, T. H. M., and Sandbrink, J. M. (1994). Molecular evolution of noncoding regions of the chloroplast genome in the Crassulaceae and related species. *Curr. Genet.* **25:** 558–566.

van Tieghem, P. (1884). Sur la structure et les affinités des Pittosporées. *Bull. Soc. Bot. Fr.* **31:** 384–385.

Wheeler, W. C., and Honeycutt, R. L. (1988). Paired sequence difference in ribosomal RNAs: Evolutionary and phylogenetic implications. *Mol. Biol. Evol.* **5:** 90–96.

Wolfe, K. H. (1991). Protein-coding genes in chloroplast DNA: Compilation of nucleotide sequences, data base entries, and rates of molecular evolution. *In* "Cell Culture and Somatic Cell Genetics of Plants" (L. Bogorad and I. K. Vasil, Eds.), Vol. 7B, pp. 467–482. Academic Press, San Diego, CA.

Wolfe, K. H., Li, W.-H., and Sharp, P. M. (1987). Rates of nucleotide substitution vary greatly among mitochondrial, chloroplast, and nuclear DNAs. *Proc. Natl. Acad. Sci. USA* **84:** 9054–9058.

Xiang, Q.-Y., Soltis, D. E., Morgan, D. R., and Soltis, P. S. (1993). Phylogenetic relationships of *Cornus* L. sensu lato and putative relatives inferred from *rbc*L sequence data. *Ann. Missouri Bot. Gard.* **80:** 723–734.

Zuker, M. (1989). On finding all suboptimal foldings of an RNA molecule. *Science* **244:** 48–52.

Zurawski, G., Clegg, M. T., and Brown, A. H. D. (1984). The nature of nucleotide sequence divergence between barley and maize chloroplast DNA. *Genetics* **106:** 735–749.