American Journal of Botany 84(2): 253–273. 1997.

RELATIONSHIPS IN THE CARYOPHYLLALES AS SUGGESTED BY PHYLOGENETIC ANALYSES OF PARTIAL CHLOROPLAST DNA ORF2280 HOMOLOG SEQUENCES[1]

STEPHEN R. DOWNIE,[2] DEBORAH S. KATZ-DOWNIE, AND KYUNG-JIN CHO

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

Phylogenetic relationships within the angiosperm order Caryophyllales were investigated by comparative sequencing of two portions of the highly conserved inverted repeat (totaling some 1100 base pairs) coinciding with the region occupied by ORF2280 in *Nicotiana,* the largest gene in the plastid genomes of most land plants. Data were obtained for 33 species in 11 families within the order and for one species each of Plumbaginaceae, Polygonaceae, and Nepenthaceae. These data, when analyzed along with previously published ORF (open reading frame) sequences from *Nicotiana, Spinacia, Epifagus,* and *Pelargonium* using parsimony, neighbor-joining, and maximum likelihood methods, reveal that: (1) *Amaranthus, Celosia,* and *Froelichia* (all Amaranthaceae) do not comprise a monophyletic group; (2) *Amaranthus* may be nested within a paraphyletic Chenopodiaceae; (3) *Sarcobatus* (Chenopodiaceae) is allied with Nyctaginaceae + Phytolaccaceae (the latter family excluding *Stegnosperma* but including *Petiveria*); and (4) Caryophyllaceae (with *Corrigiola* basal within the clade) are sister group to Chenopodiaceae + Amaranthaceae. Basal relations within the order remain obscure. Sequence divergence values in pairwise comparisons across all Caryophyllales taxa ranged from 0.1 to 5% of nucleotides. However, despite these low values, 23 insertion and deletion events were apparent, of which five were informative phylogenetically and bolstered several of the relationships listed above. A polymerase chain reaction (PCR) survey for ORF homolog length variants in representatives from 70 additional angiosperm families revealed major deletions, of ≈ 100 to 1400 base pairs, in 19 of these families. Although the ORF is located within the mutationally retarded inverted repeat region of most angiosperm chloroplast DNAs, this gene appears particularly prone to length mutation.

Key words: Caryophyllales; chloroplast DNA; deletion; inverted repeat; ORF2280 phylogeny.

Open reading frame 2280 (ORF2280), or its homolog, is the largest gene in the plastid genomes of *Nicotiana* and most other land plants (Ohyama et al., 1986; Zhou et al., 1988; Shimada and Sugiura, 1991; Wolfe, Morden, and Palmer, 1992; Downie et al., 1994; Wakasugi et al., 1994). In almost all angiosperm chloroplast DNAs (cpDNAs) examined, the gene is contained within a large inverted duplication, the so-called "inverted repeat (IR)" (Fig. 1), and is conserved in immediate location, being flanked by *trn*I-CAU and *trn*L-CAA located on the opposite strand (Fig. 2). ORF2280 encodes a protein of unknown function; however, it shares a few short amino-acid motifs, including parts of a nucleotide-binding site, with members of the CDC48 family of proteins and may possibly be a proteolytic ATPase (Wolfe, 1994).

Despite the presence of this ORF within the IR—the most evolutionarily conservative region of the chloroplast genome (Wolfe, Li, and Sharp, 1987; Palmer, 1991)—it appears remarkably tolerant of insertions and deletions (indels). Differences in length are apparent among distantly related species (Wolfe and Sharp, 1988; Zhou et al., 1988; Downie et al., 1994; Davis, 1995; Hahn, Givnish, and Sytsma, 1995), among closely related species in a genus (Sytsma and Gottlieb, 1986; Nimzyk, Schöndorf, and Hachtel, 1993), and even among isolates of a single species (Blasko et al., 1988). Furthermore, the ORF2280 locus is a pseudogene in at least four independent lineages of angiosperms owing to the kilobase-magnitude deletions apparent (Hiratsuka et al., 1989; Downie et al., 1994; Hahn, Givnish, and Sytsma, 1995).

Cladistic analyses of IR restriction site data from representatives of two subclasses of angiosperms reveal further that the numbers and frequencies of restriction site changes, as well as the number of detectable length variants, are greater in the region occupied by the ORF than anywhere else in the IR (Fig. 2). Comparative restriction site mapping of 99 Asteridae (sensu Cronquist, 1981) and outgroup cpDNAs with four restriction endonucleases identified a total of 77 restriction sites, 60 of which were variable (Downie and Palmer, 1992a). An analysis of 24 Caryophyllidae cpDNAs with ten restriction endonucleases yielded 161 restriction sites of which 101 were variable (Downie and Palmer, 1994). Expressed in terms of mutations per unit length (i.e., kilobases) of DNA, the areas within the IR with the highest frequencies of mutations are those corresponding to the regions occupied by ORF2280 or its homolog and the intergenic spacers between the ORF and *ndh*B (Fig. 2). In the Caryophyl-

[1] Manuscript received 8 September 1995; revision accepted 6 September 1996.

The authors thank the various botanical gardens cited in the text for providing us with leaf material; S. Sanderson, J. Clement, and J. Palmer for providing plant material or DNAs; L. Paulius for laboratory assistance; and H.-D. Behnke, J. Clement, J. Rodman, and two anonymous reviewers for their insightful criticisms and helpful comments on an earlier draft of the manuscript. This work was supported by laboratory start-up funds from the University of Illinois, and, in part, by grants from the National Science Foundation (DEB-9407712) and the Campus Research Board of the University of Illinois.

[2] Author for correspondence (Fax: 217 244-7246; e-mail: stephen_downie@qms1.life.uiuc.edu).

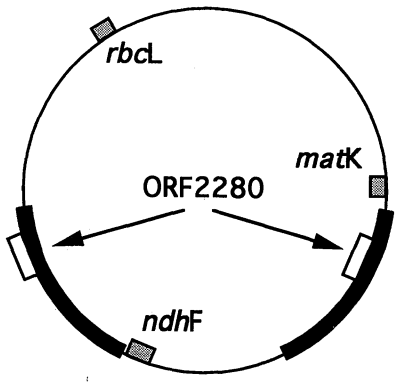


Fig. 1. Location of the gene ORF2280 in the *Nicotiana tabacum* chloroplast genome and its placement relative to single-copy genes *rbc*L, *mat*K, and *ndh*F (based on Shinozaki et al., 1986). The thickened parts of the circle represent the two 25.3-kb duplicated regions called the inverted repeat (IR). In *Nicotiana,* ORF2280 is ≈ 6.8 kb in size.

lidae study (Downie and Palmer, 1994), the highest frequency of mutation was in the extreme 3′ portion of the ORF, where a mutation density of 15.0 per kilobase (kb) of sequence was detected. In contrast, the lowest frequencies occurred in those regions occupied by two chloroplast ribosomal RNA genes (16S and 23S rDNA) and the 3′ portion of *ndh*B. This is not unexpected as the chloroplast ribosomal DNA genes have been previously shown to be very highly conserved evolutionarily (Rawson et al., 1981; Palmer, Singh, and Pillay, 1983). Although ORF2280 comprises some 6.8 kb of sequence (or ≈ 27% of the entire *Nicotiana tabacum* IR; Shinozaki et al., 1986), this region contributed approximately half of all the variable sites detected in each study. However, the similarly sized rRNA operon (Fig. 2), which covers some 29% of the IR region (or 7.3 kb of sequence), contributed only ≈ 20% of the detectable mutations. This pattern is also reflected in the distribution of restriction fragment length variants. All eight indel events (ranging in size between 200 and 600 base pairs [bp]) inferred from both studies, with the exception of the loss of the intron from gene *rpl*2, mapped into the region occupied by the ORF. The absence of any detectable length variation in intergenic spacer regions is surprising, as these comprise ≈ 23% of the *Nicotiana* IR (Shinozaki et al., 1986) and, presumably, could easily accommodate the disruptive effects of insertions and deletions.

The region encompassing the ORF and the spacers between the ORF and *ndh*B are the most mutationally dense within the IR; however, the characters they contribute to

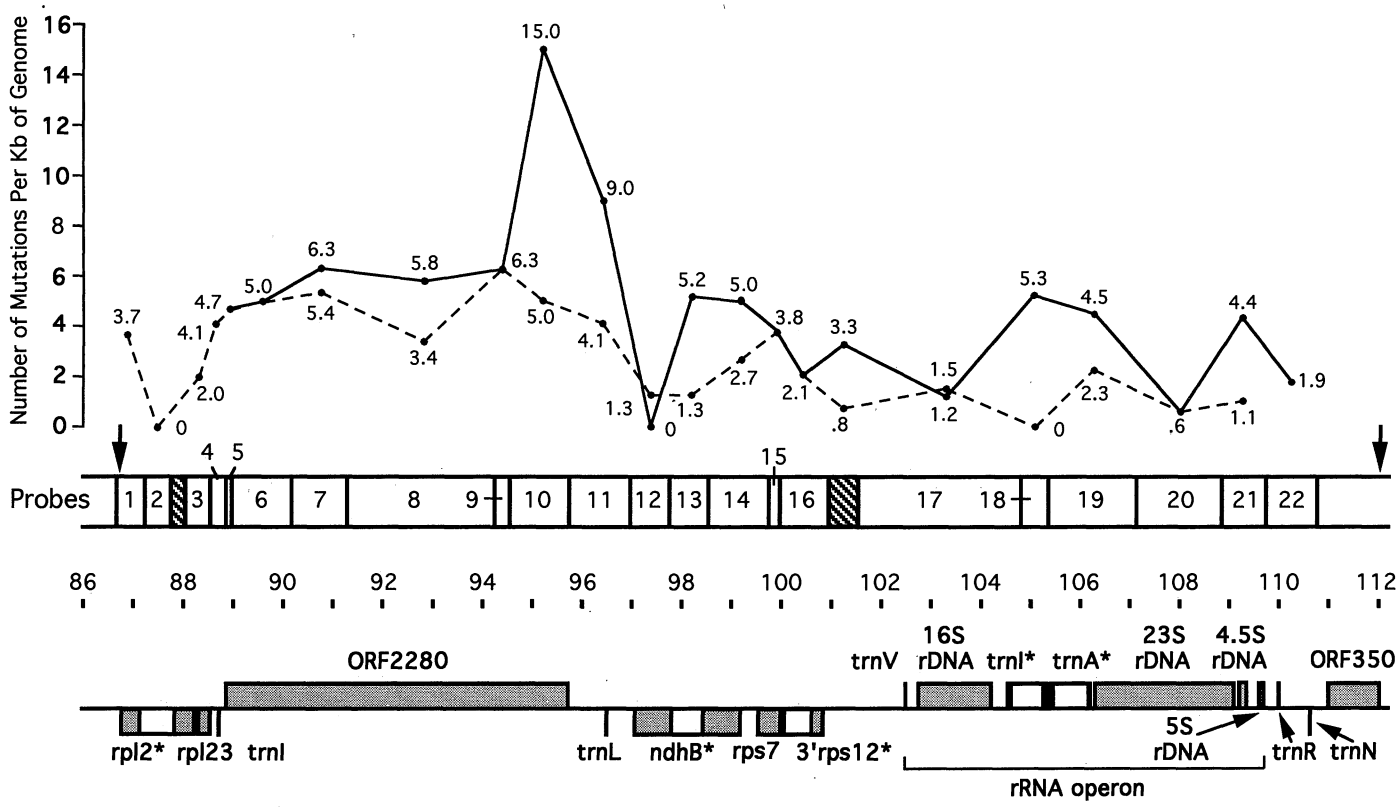


Fig. 2. Gene map of the *Nicotiana tabacum* cpDNA IR region showing the distribution of variable restriction sites per kilobase of sequence, as detected by the hybridization probes indicated, in representatives from angiosperm subclasses Asteridae and Caryophyllidae. Gene locations and sequence coordinates in kilobases (scale in middle) are from Shinozaki et al. (1986). Asterisks indicate genes containing introns; for these genes, filled boxes indicate exons and open boxes indicate introns. Genes on the top of the line are transcribed from left to right; those on the bottom are transcribed from right to left. The subclones used as hybridization probes in Downie and Palmer (1992a, 1994) are numbered from 1 to 22 (the regions between probes 2 and 3 and between probes 16 and 17 have not been subcloned) and range in size from 0.2 to 3.3 kb (averaging ≈ 1 kb). The boundaries of the IR in *Nicotiana* are indicated by the two downward pointing arrows. The numbers of mutations per kilobase of IR sequence over virtually the entire IR, as inferred by comparative restriction site mapping experiments of 99 Asteridae (hatched line; Downie and Palmer, 1992a) and 24 Caryophyllidae (solid line; Downie and Palmer, 1994) taxa, are indicated.

the phylogenetic analyses are not any more homoplastic than those characters obtained from the most conserved regions. Of the 101 variable restriction sites identified in the Caryophyllidae IR study, 62 of these mapped into these variable locales. The distribution of the number of inferred changes per character on the maximally parsimonious trees reveals that, for these 62 characters, the average number of steps per character was 1.6, with the range of possible steps extending from one to five. Only the slightly higher value of 1.7 was obtained for those characters scored from the rRNA operon.

This differential conservation of restriction sites and length variants within the IR region suggests that further comparative analyses of caryophyllalean cpDNAs might benefit by focusing exclusively upon the variable ORF region. The primary objective of this study is to assess the utility of cpDNA ORF2280 homolog sequences as a source of phylogenetic information. However, owing to the large size of the ORF in most angiosperm cpDNAs, this study will focus on only two internal portions of the gene. These two regions were chosen for this first comparative phylogenetic study of ORF homolog sequences because previous restriction site mapping studies had revealed several length polymorphisms therein (Downie and Palmer, 1992a, 1994). Thus, as a second and related objective, the availability of these sequence data affords an opportunity to further characterize these insertions and deletions and report on their usefulness as phylogenetic markers.

MATERIALS AND METHODS

Plant material—Fresh leaf material and/or DNAs from 49 species, representing 11 families of Caryophyllales (sensu Cronquist, 1981) and one species each of Plumbaginaceae (*Limonium*), Polygonaceae (*Rheum*), and Nepenthaceae (*Nepenthes*), were obtained from various sources (Table 1). Chloroplast or total cellular DNAs were isolated from fresh leaf material of one or, rarely, more individual plants using the methods of Palmer (1986) or Doyle and Doyle (1987), respectively, and purified on CsCl/ethidium bromide gradients. These taxa were selected for the following reasons. (1) They represent 11 of the 12 families of Caryophyllales recognized by Cronquist (1981) and, potentially, are maximally divergent evolutionarily within the order; only the small family Achatocarpaceae was not sampled. Thus, the utility of ORF2280 homolog sequences in providing characters to infer phylogeny at various stages of genetic divergence within the order can be ascertained. Many of the families accepted and circumscribed by Cronquist (and other authors) may not represent monophyletic entities (Hershkovitz, 1989; Rodman, 1990), but until more robust higher level analyses are available for the order, Cronquist's system seems to be a reasonable place to start. (2) Nineteen of these accessions represent precisely the same species that were used in a previous phylogenetic study of the order using cpDNA IR restriction site variation (Table 1; Downie and Palmer, 1994), enabling a comparison between the results obtained from each study.

Six outgroups were employed in the phylogenetic analyses. Among current classification systems and phylogenies based on *rbc*L sequences, a consensus favors an association between the Caryophyllales and the families Polygonaceae and Plumbaginaceae (Dahlgren, 1980; Takhtajan, 1980; Cronquist, 1981; Olmstead et al., 1992; Chase et al., 1993). Together, these three taxa comprise the subclass Caryophyllidae (Cronquist, 1981). CpDNA data have also suggested a close relationship of Nepenthaceae with the Caryophyllales (Albert, Williams, and Chase, 1992). Thus, *Rheum* (Polygonaceae), *Limonium* (Plumbaginaceae), and *Nepenthes* (Nepenthaceae) were chosen as outgroups. Because of the availability of previously published ORF sequences for *Nicotiana* (Shinozaki et al., 1986), *Spinacia* (Zhou et al., 1988), *Epifagus* (Wolfe, Morden, and Palmer, 1992), and *Pelargonium* (Downie et al., 1994), these taxa were also included.

PCR amplification and sequencing strategy—Double-stranded DNAs of two regions internal to ORF2280 in each genomic DNA were PCR (polymerase chain reaction)-amplified using primer pairs "ORF2" and "ORF3" or "ORF4" and "ORF5" in an equimolar ratio (Fig. 3). Primers were designed by comparing ORF2280 homolog sequences from *Nicotiana, Spinacia, Epifagus, Pelargonium,* and *Marchantia* (Ohyama et al., 1986) and choosing regions highly conserved among these taxa (see Downie et al., 1994, for a multiple alignment of the amino-acid sequences of these homologs). These 100-μL PCR amplifications contained (in order of addition) 65.6 μL of sterile water, 10.0 μL of 10 × *Taq* polymerase reaction buffer (Promega Corp., Madison, WI), 200 μmol/L of each deoxyribonucleotide triphosphate (dNTP; United States Biochemical Corp., Cleveland, OH), 1.5 mmol/L of $MgCl_2$, 2.0 Units of *Taq* DNA polymerase (Promega Corp., Madison, WI), 1.0 μmol/L of each primer pair, and a 0.5–1.0 μL aliquot of unquantified genomic (template) DNA. Each PCR cycle proceeded in the following manner: (1) 1 min at 94°C; (2) 1 min at 53°C; and (3) 1 min at 72°C. The first cycle was preceded by an initial denaturation step of 30 s at 94°C, and a 10-min 72°C extension period followed completion of the 35 thermal cycles. Each set of reactions was monitored by the inclusion of positive (*Nicotiana* and *Spinacia* cpDNAs) and negative (no template) controls.

Each amplified DNA fragment was electrophoresed in a 1% agarose gel (using 1 × TAE as the gel buffer), visualized with ethidium bromide, and then excised under low wavelength UV light with a scalpel. Successful PCR amplifications resulted in a single DNA band corresponding to 312 (*Spinacia*) or 636 (*Nicotiana*) bp for primer pair "ORF2" and "ORF3," and 456 (*Spinacia*) or 579 (*Nicotiana*) bp for primer pair "ORF4" and "ORF5." The gel slice containing the DNA fragment was transferred to a 1.5-mL microcentrifuge tube and the DNA was recovered using the Elu-Quik DNA Purification Kit (Schleicher & Schuell, Keene, NH). The purified DNA was resuspended in 20 μL of sterile water; this volume was sufficient for two to four sequencing reactions. Sequencing was done using the dideoxy chain termination method employing Sequenase (Version 2.0; United States Biochemical, Cleveland, OH) with α-^{35}S-dATP (Amersham Life Science, Arlington Heights, IL) as the labeling agent. Modifications to the sequencing protocol included denaturation of the DNA by boiling the DNA/primer/acetamide mix for 4 min, followed by snap-cooling the annealing mixture for 3 min in an ice water bath (Winship, 1989). Forward primers "ORF2," "ORF2a," "ORF4," and "ORF4a," and reverse primers "ORF3" and "ORF5" (Fig. 3) were each used in the sequencing of each template DNA. The complete sequencing of both DNA strands was not done; only some 20–40% of each of the two regions of the ORF was determined on both strands (i.e., where DNA sequences from primers "ORF2a" and "ORF3" and primers "ORF4a" and "ORF5" overlapped). In most instances, each of the DNAs was sequenced twice with the same primer, although it is acknowledged that multiple sequencing reactions using the same primer, as well as overlapping forward primers, can yield similar artifacts. All primers were synthesized by Operon Technologies (Alameda, CA).

Sequence ambiguities (i.e., base compressions or hard stops) were few but where they occurred they were resolved by resequencing the region using 7-deaza-dGTP or dITP in place of dGTP (United States Biochemical, Cleveland, OH). Up to ten sets of reactions were separated electrophoretically in 6% polyacrylamide gels in which the xylene cyanole dye marker was run 30 cm (for a short gel) or 60 cm (for a long gel). Gels were dried onto Whatman 3MM paper in a vacuum dryer and then exposed to X-ray film (Kodak XAR) for 2–4 d at room temperature.

TABLE 1. Caryophyllales and outgroup taxa examined for nucleotide and/or length variation in ORF2280 homolog sequences. Asterisks indicate those taxa sampled for major length variation only; double daggers indicate those taxa included in a previous study of cpDNA IR restriction site variation (Downie and Palmer, 1994). Available voucher specimens for Downie's collections will be housed at ILL; collection information for all other accessions is available upon request. Data for *Nicotiana, Epifagus, Pelargonium,* and *Spinacia* Publ. are based on published ORF homolog sequences and are not provided below (see text for references).

Taxon	Source and/or voucher
Caryophyllales	
Aizoaceae	
Tetragonia tetragonioides (Pallas) Kuntze‡	W. J. Beal Botanical Garden 89B423, Michigan State Univ.; *Downie 1070*
Amaranthaceae	
Amaranthus albus L.*	W. J. Beal Botanical Garden 3, Michigan State Univ.; *Downie 761*
Amaranthus spinosus L.*	W. J. Beal Botanical Garden 95, Michigan State Univ.; *Downie 758*
Amaranthus tricolor L.	J. Palmer Lab (Indiana Univ.) DNA Accession 47; no voucher
Amaranthus tricolor L.*	W. J. Beal Botanical Garden 96, Michigan State Univ.; *Downie 756*
Celosia argentea L. 'Century Red'‡	Brooklyn Botanical Garden; *Downie 1049*
Celosia argentea L. 'Pink Tassles'*	Brooklyn Botanical Garden; *Downie 1053*
Froelichia floridana (Nutt.) Moq.	*Clement 24* (TEX)
Basellaceae	
Anredera cordifolia (Ten.) Steenis‡	Matthaei Botanical Garden 840353, Univ. of Michigan; *Olmstead 51*
Cactaceae	
Pereskia grandiflora Haw.‡	Matthaei Botanical Garden, Univ. of Michigan; *Olmstead 46*
Caryophyllaceae	
Corrigiola littoralis L.‡	W. J. Beal Botanical Garden, Michigan State Univ.; *Downie 1035*
Dianthus sp.	Cult. Bloomington, Indiana; *Downie 1028*
Lychnis chalcedonica L.*	W. J. Beal Botanical Garden, Michigan State Univ.; *Downie 1032*
Petrorhagia saxifraga (L.) Link	W. J. Beal Botanical Garden, Michigan State Univ.; *Downie 1069*
Saponaria officinalis L.*	W. J. Beal Botanical Garden, Michigan State Univ.; *Downie 1030*
Silene latifolia Poiret	Indiana Univ. Greenhouse; *Downie 1019*
Silene schafta Gmel.‡	W. J. Beal Botanical Garden, Michigan State Univ.; *Downie 1033*
Chenopodiaceae	
Archiatriplex nanpinensis Chu*	China, Longkang, Nanping, Sichuan; *Downie 759*
Atriplex hastata L.	J. Palmer Lab (Indiana Univ.) Accession 45; no voucher
Beta vulgaris L.‡	J. Palmer Lab (Indiana Univ.) Accession 44; no voucher
Camphorosma monspeliaca L.	Stavropol Botanical Garden; *Downie 732*
Ceratoides lanata (Pursh) J. T. Howell*	Utah, Tooele Co., Rush Valley; 2 mi W of mi 11, UT 73; *Downie 728*
Ceratoides lanata (Pursh) J. T. Howell*	*Clement 3* (TEX)
Chenopodium berlandieri Moq.*	W. J. Beal Botanical Garden 121, Michigan State Univ.; *Downie 762*
Chenopodium capitatum (L.) Asch.*	W. J. Beal Botanical Garden 47, Michigan State Univ.; *Downie 757*
Chenopodium murale L.‡	J. Palmer Lab (Indiana Univ.) Accession 41; no voucher
Cycloloma atriplicifolium (Spreng.) J. M. Coulter	*Clement 12* (TEX)
Grayia spinosa (Hook.) Moq.	Utah, Tooele Co., 1 mi W of mi 11, UT 73; *Downie 729*
Hablitzia thamnoides Marsch.*	Finland, Univ. Turku Botanical Garden 154; *Downie 760*
Kochia sp.	J. Palmer Lab (Indiana Univ.) Accession 50; no voucher
Salicornia bigelovii Torr.	*Nesom 7500* (TEX)
Sarcobatus vermiculatus (Hook.) Torr.	Utah, Grantsville, mi 5, UT 138; *Downie 726*
Sarcobatus vermiculatus (Hook.) Torr.*	*Clement 4* (TEX)
Spinacia oleracea L.‡	J. Palmer Lab (Indiana Univ.) Accession 39; no voucher
Suaeda torreyana Wats.	Utah, Grantsville, mi 5, UT 138; *Downie 727*
Didiereaceae	
Alluaudia montagnacii Rauh var. *ascendens* Drake‡	Brooklyn Botanical Garden; *Downie 1055*
Didierea madagascariensis Baillon‡	Missouri Botanical Garden 821268; *Downie 1063*
Molluginaceae	
Mollugo verticillata L.‡	W. J. Beal Botanical Garden, Michigan State Univ.; *Downie 1068*
Nyctaginaceae	
Bougainvillea glabra Choisy‡	Indiana Univ. Greenhouse; *Downie 1020*
Mirabilis nyctaginea (Michx.) MacMillan‡	W. J. Beal Botanical Garden B87137, Michigan State Univ.; *Downie 1067*
Phytolaccaceae	
Petiveria alliacea L.	Univ. of Illinois Greenhouse, Urbana 85543; *Downie 714*
Phytolacca americana L.	Brooklyn Botanical Garden; *Downie 1054*
Stegnosperma halimifolium Benth.‡	Missouri Botanical Garden 720287; *Downie 1061*
Portulacaceae	
Calandrinia ciliata DC.	W. J. Beal Botanical Garden, Michigan State Univ.; *Downie 1066*
Claytonia perfoliata Donn‡	W. J. Beal Botanical Garden 90B1242W, Michigan State Univ.; *Downie 1031*
Portulaca oleracea L.‡	W. J. Beal Botanical Garden 109786, Michigan State Univ.; *Downie 1034*

TABLE 1. Continued.

Taxon	Source and/or voucher
Plumbaginaceae	
Limonium gmelinii Kuntze‡	W. J. Beal Botanical Garden, Michigan State Univ.; *Downie 465*
Polygonales	
Polygonaceae	
Rheum rhaponticum L.‡	J. Palmer Lab (Indiana Univ.) Accession 247; no voucher
Nepenthales	
Nepenthaceae	
Nepenthes alata Blanco 'Kosobe'	J. Palmer Lab (Indiana Univ.) Accession 918; M. Chase

Sequence analysis—Boundaries of the two sequenced ORF regions were determined by comparison of the DNA sequences to the previously published *Nicotiana, Spinacia,* and *Pelargonium* ORF homologs. Owing to their conservatism and lack of ambiguous indels, DNA sequences were aligned manually. Pairwise nucleotide differences of all aligned positions were determined using the DISTANCE MATRIX option in PAUP version 3.1 (insertion/deletion events were treated as missing data; Swofford, 1993). Thus, these divergence values were calculated simply as the proportion of divergent sites in each direct pairwise comparison with no provision made to account for superimposed events (multiple hits). The DNA sequences were translated to amino acid data using MacClade version 3.01 (Maddison and Maddison, 1992). MacClade was also used to examine, over all maximally parsimonious trees, patterns of transition (Ts)/transversion (Tv) bias, the number of changes per codon position, and the relative variability of base substitutions along both ORF sequences. G + C content was calculated manually. The sequences reported in this study are available from GenBank; their accession numbers are provided in Fig. 4.

Phylogenetic analysis—The resulting 40-taxon data matrix was analyzed initially 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 100 times with RANDOM addition sequence and TREE BISECTION-RECONNECTION (TBR) branch swapping. The options MULPARS, STEEPEST DESCENT, COLLAPSE and ACCTRAN optimization were selected. All characters were weighted equally with respect to codon position; similarly, nucleotide transformations within characters were equally weighted. Bootstrap values (Felsenstein, 1985) were

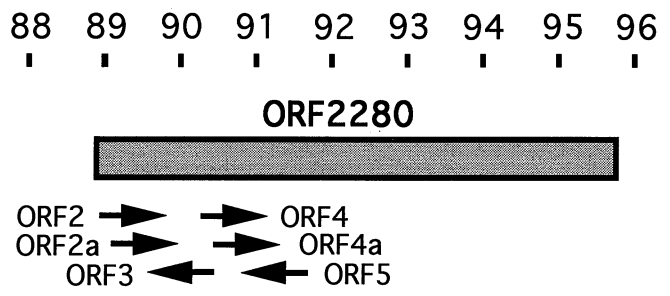


Fig. 3. Relative locations of the six oligonucleotide primers used in PCR amplification and DNA sequencing of ORF2280 homolog sequences. Location coordinates (in kb on scale and bp below) are from Shinozaki et al. (1986). Primer sequences, written 5′ to 3′, are as follows: ORF2: TCGCGGTGGTGGAGNAAYTGG (89414–89434); ORF2a: CAGCTTTTCGAAATCTTRGT (89642–89661); ORF3: CTTTCAATTGGCTAGARTCNG (90045–90065); ORF4: AGAAATGGCAGAYCTNTTYAC (90726–90746); ORF4a: TCTTTATTGGTTYTAYCTCC (90865–90884); ORF5: TTCCTGGAAATTYTTRCTNCC (91327–91347). Small aliquots of these primers for trial purposes are available from the senior author upon request.

first calculated from 100 replicate analyses using the HEURISTIC search strategy (with TBR branch swapping and SIMPLE addition sequence of taxa). Here a MAXTREE limit of 2500 trees per replicate was set, as the computer ran out of memory on a replication in an earlier analysis. This limit of 2500 saved trees was reached in 32 of the 100 bootstrap replicates, thus the values obtained may not supply an accurate measure of support. A subsequent analysis was done with 1000 bootstrap replicates and a MAXTREE limit of 100. Although this limit was achieved in 876 of the replicates and, thus, culminated in a less thorough analysis, it afforded an opportunity to examine a larger number of resamplings. 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) of two steps or more longer than the shortest trees could not be done. Because of possible base substitution rate heterogeneity between the two ORF regions, separate parsimony analyses of these regions were also performed and the results compared. The amount of phylogenetic information in the parsimony analyses was estimated using the consistency (CI, Kluge and Farris, 1969) and retention (RI, Farris, 1989) indices. The trees were rooted by positioning the root along the branch connecting *Nicotiana* to the rest of the network. The selection of either *Nicotiana* (Solanaceae), *Rheum* (Polygonaceae), or *Limonium* (Plumbaginaceae) as an outgroup, or all three simultaneously, did not affect the ingroup tree topology.

Gaps in the multiple alignment were incorporated into the parsimony analyses in one of two ways. First, each indel was scored and entered as a separate presence/absence (binary) character while treating gap positions as missing data (GAPMODE=MISSING; Swofford, 1993). Second, gap positions were retained as missing data but each indel was mapped a posteriori onto the resulting minimal-length cladograms in the most parsimonious way possible. A third method, in which sequence data for all taxa for any position are deleted when missing values occur in at least one of the taxa, was unrealistic owing to the large size of the gaps present in many of the taxa. For example, deleting those positions from the alignment where gaps occur in *Suaeda* (Chenopodiaceae) would entail losing almost half of the entire data matrix and much of the available phylogenetic information.

Distance trees were constructed using the neighbor-joining method (Saitou and Nei, 1987) implemented using the NEIGHBOR program in Felsenstein's (1993) phylogeny inference package (PHYLIP, version 3.5). Distance matrices were calculated using the DNADIST program of PHYLIP and the numbers of nucleotide substitutions (excluding gaps) were estimated using Kimura's (1980) two-parameter method. A range of transition/transversion (Ts/Tv) ratios was used (i.e., from 0.8 to 2.0) but did little to change the resultant tree topology. A bootstrap analysis of these data was done using 100 resampled data sets generated with the SEQBOOT program prior to calculating the distance matrices and neighbor-joining trees. PHYLIP's CONSENSE program was then used to construct a strict consensus tree.

The maximum likelihood method was also applied to these data using the program fastDNAml (version 1.0.6; Olsen et al., 1992, 1994), based on the procedures of Felsenstein (1981). Maximum likelihood trees

were inferred using a Ts/Tv ratio of either 1.0 or 2.0, randomizing the input order of sequences (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. Like the neighbor-joining algorithm, this method ignored gaps.

ORF homolog deletions in other angiosperms—In order to detect major length variants in ORF2280 homolog sequences in other groups of angiosperms, primers "ORF2" and "ORF5" (Fig. 3) were used in a PCR survey. In *Nicotiana,* the distance between these two primers is ≈ 1.9 kb (Shinozaki et al., 1986). DNAs from 70 families, representing most subclasses of angiosperms, were sampled (a complete list of taxa examined is available upon request). The methodology applied here was the same as described above, with the only exception being a reduction in the volume of each PCR reaction (25 μL instead of 100 μL). The ensuing PCR fragments were electrophoresed in 1% agarose gels, stained with ethidium bromide, and sized against *Eco*RI/*Hin*dIII-digested lambda DNA standards. Only those experiments where the PCR products yielded a single, bright DNA band were considered successful.

RESULTS

Sequence analysis—Aligned DNA sequences of both ORF portions (here designated as "ORF2-ORF3" and "ORF4-ORF5") for 40 caryophyllalean taxa and outgroups are presented in Fig. 4. This alignment required the introduction of many gaps, of which 23 were necessary to align the 36 Caryophyllidae representatives, *Nicotiana,* and *Nepenthes* (Table 2). Of these 23 indels, which ranged in length from three to 297 bp relative to *Nicotiana,* five were potentially informative and, of these five, only one represented an insertion. All indels could be aligned unambiguously, and none of them interrupted the reading frame of the gene. Five insertions (length mutations 6, 14, 20, 22, and 23, Table 2) represent direct repeats of 6–15 bp of adjacent sequence, and two mutations (5 and 11, Table 2) represent insertions, of 6 and 9 bp, that are not apparently related to adjacent nucleotide sequences. The 16 remaining indels represent deletions, relative to *Nicotiana.* For two of these deletions (4 and 9, Table 2), small direct repeats are found just prior to the deleted region and at the end of the deletion in related taxa. For example, the sequence TGGATTTG occurs both just prior to indel number 4 in *Dianthus* and *Petrorhagia* (position 218–225, Fig. 4) and at position 488–495 in other Caryophyllaceae. It has been suggested that slipped-strand mispairing events during replication may be a factor in creating these indels, particularly if they are associated with small direct repeats (Levinson and Gutman, 1987).

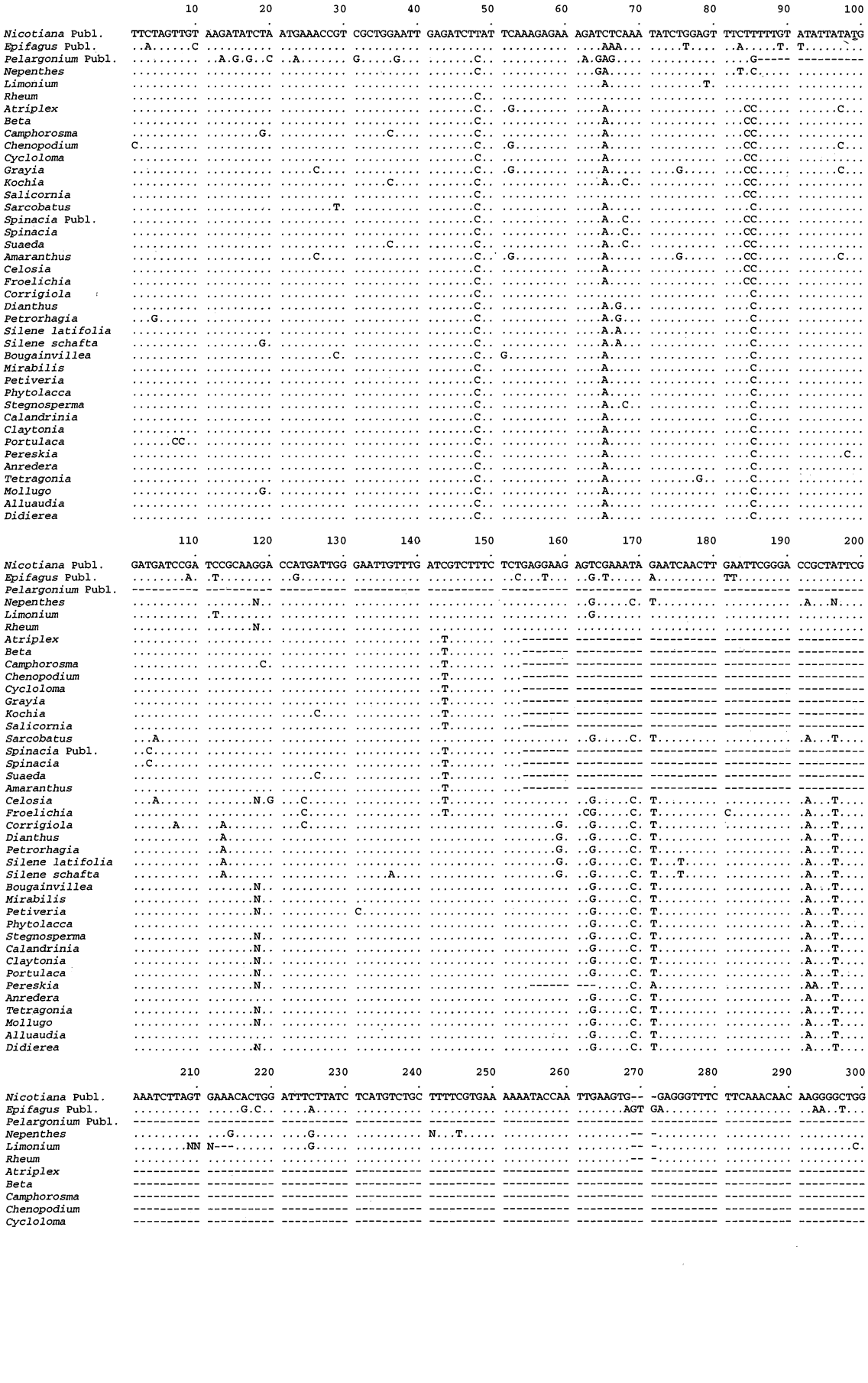
Alignment of all sequence positions resulted in a matrix of 1 156 characters × 40 taxa (46 240 entries). Of these characters, 116 (10%) had at least two nucleotide states in two or more sequences and were potentially informative phylogenetically, 845 (73%) were unvarying, and 195 (17%) were autapomorphic (Table 3). Third-codon position substitutions were only slightly higher than first-codon position substitutions (114 vs. 106, respectively) and both of these positions were higher than second-codon position substitutions (91; Table 3). Considering each portion of the ORF separately, the distribution of variable sites by codon position in "ORF4-ORF5" was approximately the same across all positions, whereas in "ORF2-ORF3" third-codon position substitutions were most numerous. The "ORF4-ORF5" region contained a slightly greater number of potentially informative sites than did "ORF2-ORF3." Both portions of the ORF are A + T rich, with a G + C content of 35–40%.

In direct pairwise comparisons of all positions among all accessions, sequence divergence values ranged from 0.1 to 20.4% of nucleotides (Table 3). Comparisons between the two accessions of *Spinacia* and between *Didierea* and *Alluaudia* both yielded a divergence value of 0.1%, as each pair of sequences varied by only a single mutation. The highest value occurred between *Epifagus* and *Pelargonium,* whose sequences varied at 76 sites. Among the Caryophyllales ORF homologs, the two highest sequence divergence values obtained were between *Silene schafta* (Caryophyllaceae) and *Spinacia* (Chenopodiaceae), and between *S. schafta* and *Grayia* (Chenopodiaceae), where each pair of sequences differed at 5% of all sites. The ranges of pairwise sequence divergence values within "ORF2-ORF3" and "ORF4-ORF5" were approximately the same (Table 3).

Phylogenetic analysis—Parsimony analysis of all data, including the five informative indels, resulted in 33 minimal length trees, whose strict consensus with accompanying bootstrap values is shown in Fig. 5. Each of these trees had a length of 447 steps, a CI (excluding uninformative substitutions) of 0.633, and a RI of 0.784. Bootstrap values calculated from either 100 or 1000 replicate analyses (and setting the maximum tree limit to 2 500 or 100, respectively) were comparable with one exception. Bootstrap support for the branch leading to *Alluaudia* and *Didierea* (both Didiereaceae) varied between 51 and 69% depending upon the approach used, even though their sequences differed by only one base substitution across both ORF regions. Results of the decay analysis revealed 5 693 trees at ≤ 448 steps and > 9 500 trees at ≤ 449 steps before the search was terminated. Major groups maintained upon the relaxation of parsimony include Amaranthaceae/Chenopodiaceae (ex-

→

Fig. 4. Aligned DNA sequences of two ORF2280 homolog regions from 40 representatives of Caryophyllales and outgroups. The first region ("ORF2-ORF3," Fig. 3) ranges from position 1 to 563 and corresponds to coordinates 89452 to 89999 in *Nicotiana* cpDNA (Shinozaki et al., 1986). The second region ("ORF4-ORF5," Fig. 3) ranges from position 564 to 1156 and corresponds to coordinates 90770 to 91317 in *Nicotiana* cpDNA. Positions 1157 to 1161, identified as characters A-E in this figure, refer to the five phylogenetically informative indels identified in Table 2. N = uncertain nucleotide state or gap; hyphens = gaps required for alignment; ? = indel obliterated by larger superimposed deletion. Complete taxon names are provided in Table 1. All previously unpublished DNA sequences have been deposited with GenBank under accession numbers U48509-U48544 (for "ORF2-ORF3") and U48545-U48580 (for "ORF4-ORF5") in their order presented herein. Published ORF homolog sequences for *Nicotiana, Epifagus, Pelargonium,* and *Spinacia* are available from GenBank under numbers Z00044, M81884, M83200, and X07908, respectively.


10 20 30 40 50 60 70 80 90 100
Nicotiana Publ. TTCTAGTTGT AAGATATCTA ATGAAACCGT CGCTGGAATT GAGATCTTAT TCAAAGAGAA AGATCTCAAA TATCTGGAGT TTCTTTTTGT ATATTATATG
Epifagus Publ.
Pelargonium Publ.
Nepenthes
Limonium
Rheum
Atriplex
Beta
Camphorosma
Chenopodium
Cycloloma
Grayia
Kochia
Salicornia
Sarcobatus
Spinacia Publ.
Spinacia
Suaeda
Amaranthus
Celosia
Froelichia
Corrigiola
Dianthus
Petrorhagia
Silene latifolia
Silene schafta
Bougainvillea
Mirabilis
Petiveria
Phytolacca
Stegnosperma
Calandrinia
Claytonia
Portulaca
Pereskia
Anredera
Tetragonia
Mollugo
Alluaudia
Didierea
110 120 130 140 150 160 170 180 190 200
Nicotiana Publ. GATGATCCGA TCCGCAAGGA CCATGATTGG GAATTGTTTG ATCGTCTTTC TCTGAGGAAG AGTCGAAATA GAATCAACTT GAATTCGGGA CCGCTATTCG
Epifagus Publ.
Pelargonium Publ.
Nepenthes
Limonium
Rheum
Atriplex
Beta
Camphorosma
Chenopodium
Cycloloma
Grayia
Kochia
Salicornia
Sarcobatus
Spinacia Publ.
Spinacia
Suaeda
Amaranthus
Celosia
Froelichia
Corrigiola
Dianthus
Petrorhagia
Silene latifolia
Silene schafta
Bougainvillea
Mirabilis
Petiveria
Phytolacca
Stegnosperma
Calandrinia
Claytonia
Portulaca
Pereskia
Anredera
Tetragonia
Mollugo
Alluaudia
Didierea
210 220 230 240 250 260 270 280 290 300
Nicotiana Publ. AAATCTTAGT GAAACACTGG ATTTCTTATC TCATGTCTGC TTTTCGTGAA AAAATACCAA TTGAAGTG-- -GAGGGTTTC TTCAAACAAC AAGGGGCTGG
Epifagus Publ.
Pelargonium Publ.
Nepenthes
Limonium
Rheum
Atriplex
Beta
Camphorosma
Chenopodium
Cycloloma

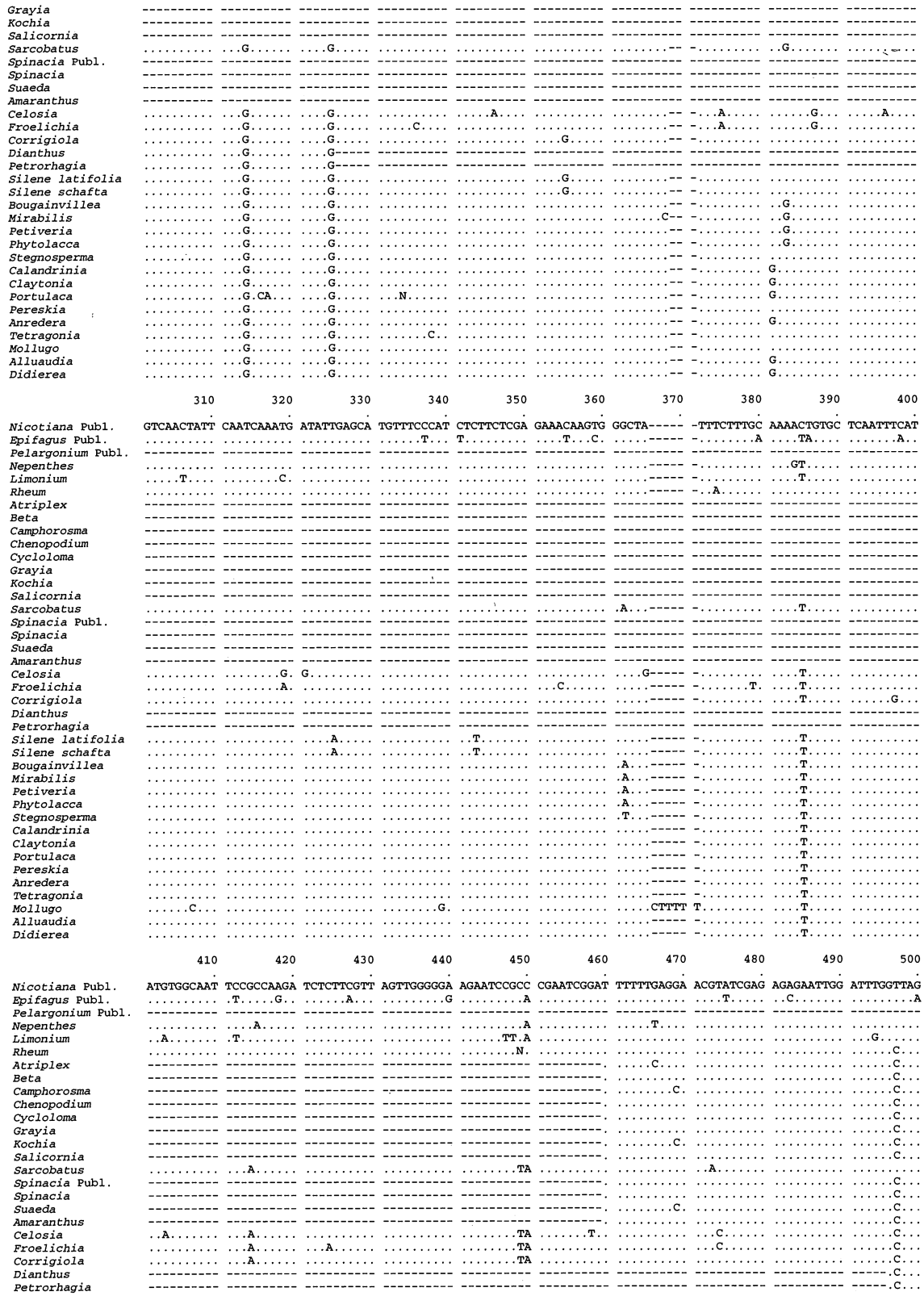


Fig. 4. *Continued.*

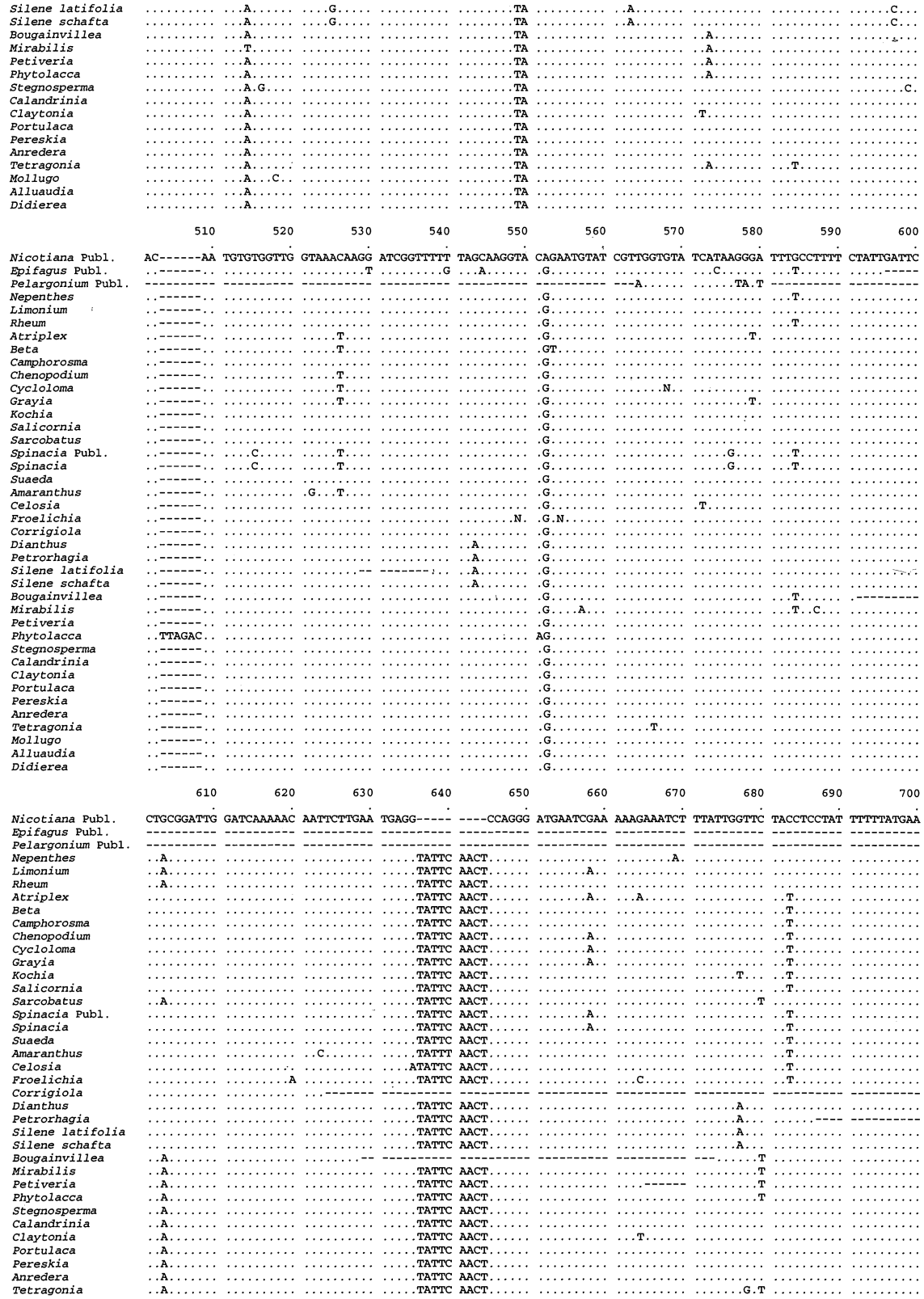


Fig. 4. *Continued.*

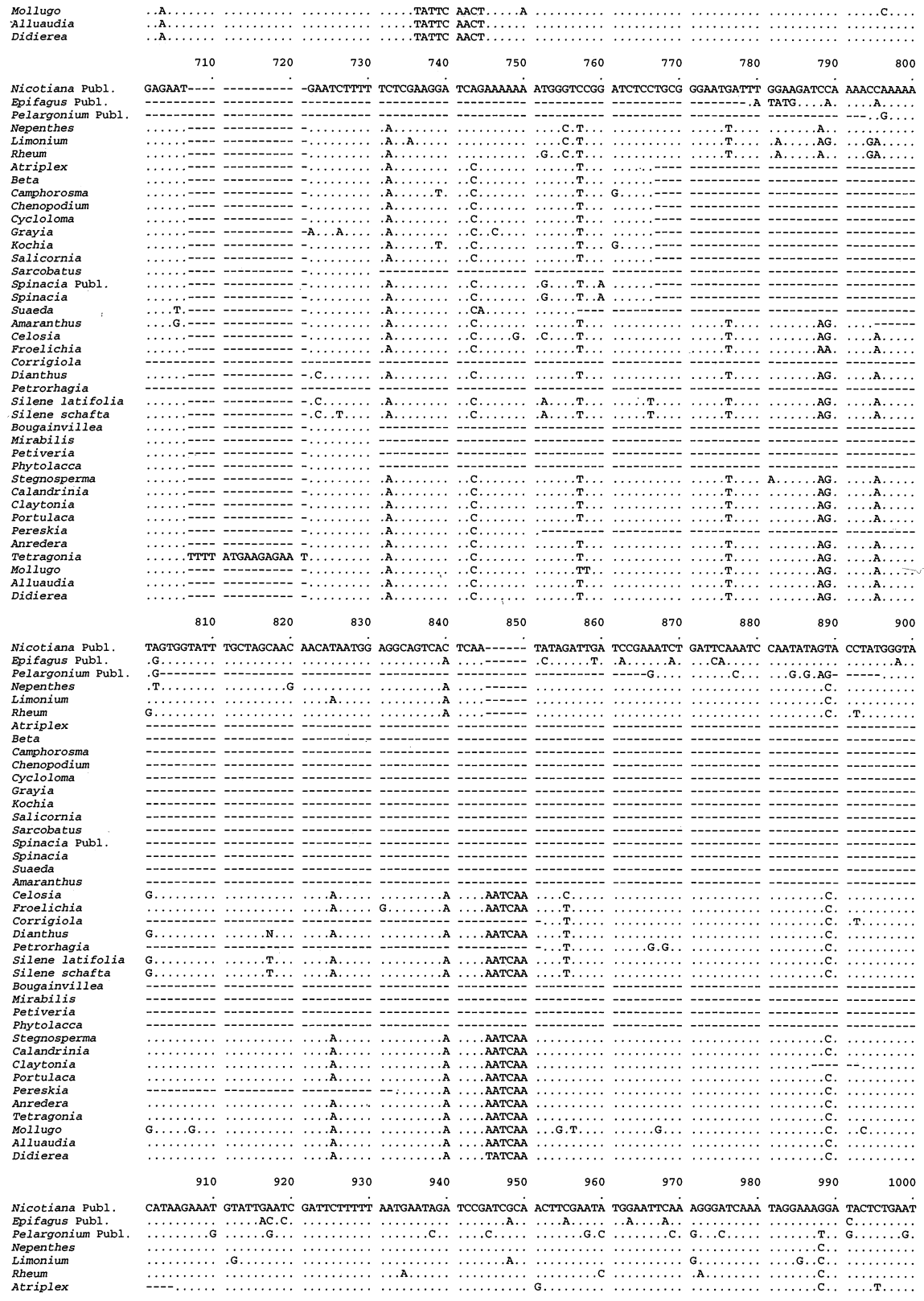


Fig. 4. *Continued.*

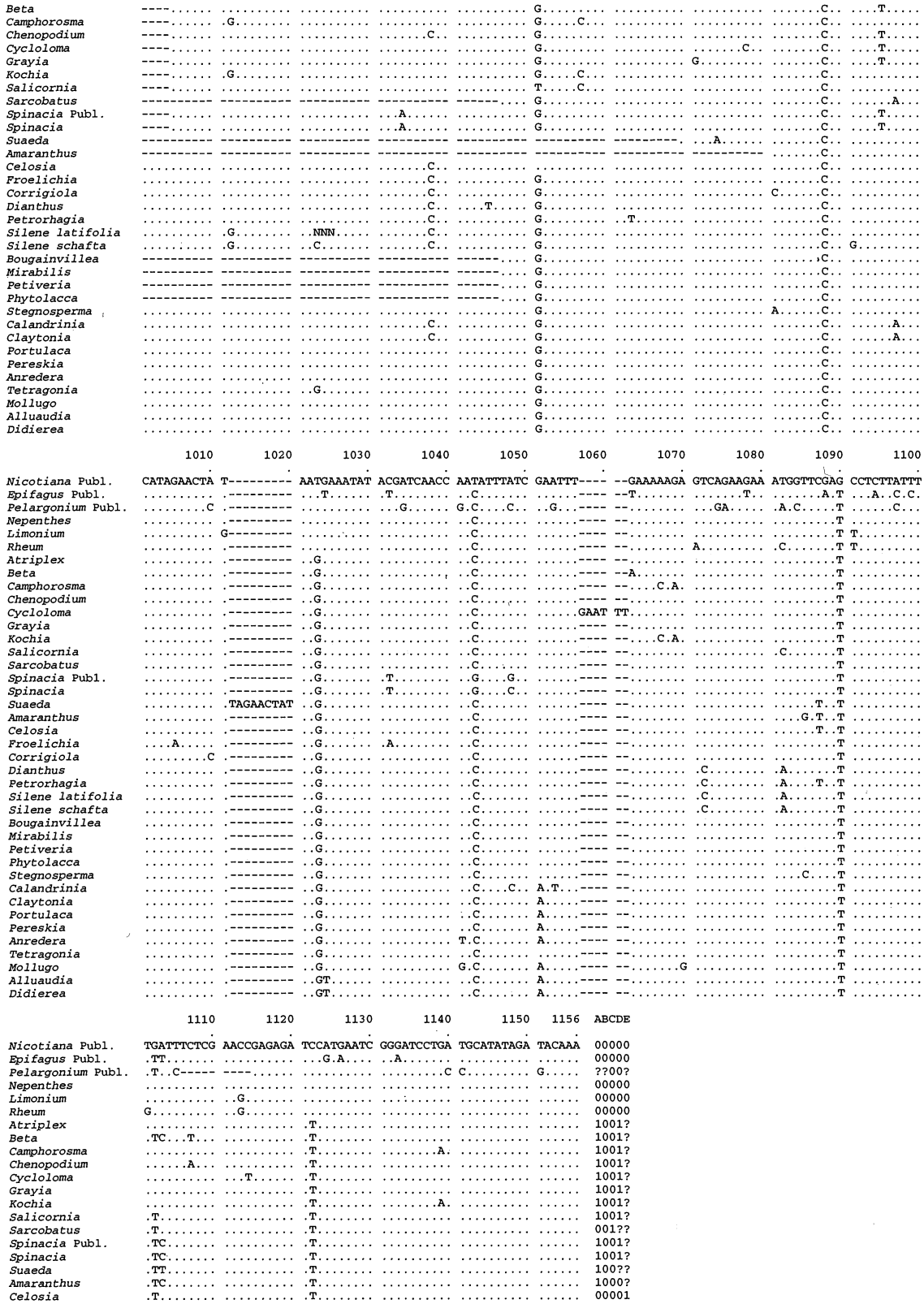


Fig. 4. *Continued.*

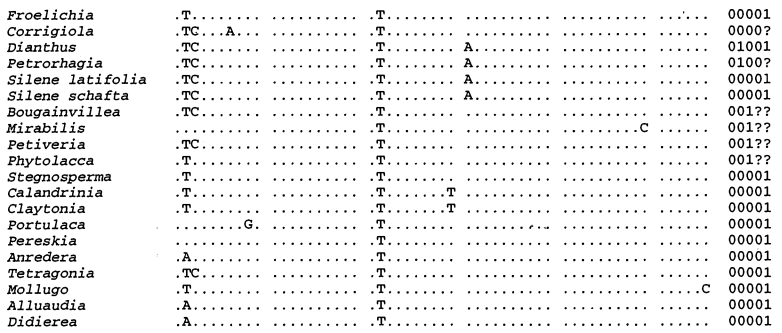


Fig. 4. *Continued.*

cluding *Sarcobatus*), Amaranthaceae/Chenopodiaceae+Caryophyllaceae, and *Sarcobatus*/Nyctaginaceae/Phytolaccaceae (excluding *Stegnosperma*)+*Tetragonia.*

One of these 33 trees was arbitrarily chosen and is presented in Fig. 6 to illustrate the number of characters supporting each clade, as optimized by ACCTRAN in PAUP. The distribution of the number of inferred changes per character on this single tree reveals that 27 characters change three times or more with the average number of steps per character being 0.38 (see Fig. 6, inset). Given the accelerated rate of evolution of the *Epifagus* and *Pelargonium* ORF homolog sequences (Downie et al., 1994), illustrated here by the long branches leading to these taxa, one cannot discount the possibility that the putative sister-group relationships between *Epifagus* and *Limonium,* and between *Pelargonium* and *Nepenthes,* are artifacts produced by the attraction of these branches during the analysis (Felsenstein, 1978). Excluding *Epifagus* and *Pelargonium* from the data set and rerunning the analysis resulted in the same 33 ingroup topologies as uncovered previously (tree length = 320 steps; CI excluding uninformative characters = 0.670; RI = 0.823). In these trees, *Nepenthes* was sister taxon to the Caryophyllales, with *Limonium* and then *Rheum* falling as successive sister taxa.

Reanalyzing the data without the five informative indels yielded 1 560 minimal length trees each of 441 steps (CI excluding uninformative characters = 0.628; RI = 0.769; Table 3). The topology of the strict consensus tree derived from these trees was not as resolved as that obtained when the gaps were included, owing to the collapse of the six nodes identified by arrows in Fig. 5. Only

TABLE 2. Characteristics of the 23 insertion/deletion events inferred in the multiple alignment of ORF2280 homolog sequences in 36 taxa of Caryophyllidae and *Nepenthes* relative to the outgroup *Nicotiana.*

No.	Position[a]	Size (bp)	Potentially informative phylogenetically?	Type	Taxa[b]
1	154–459	297	yes (A[c])	deletion	*Amaranthus, Atriplex, Beta, Camphorosma, Chenopodium, Cycloloma, Grayia, Kochia, Salicornia, Spinacia, Suaeda*
2	155–163	9	no	deletion	*Pereskia*
3	212–214	3	no	deletion	*Limonium*
4	226–495	261	yes (B[c])	deletion	*Dianthus, Petrorhagia*
5	366–371	6	no	insertion	*Mollugo*
6	503–508	6	no	insertion	*Phytolacca*
7	529–537	9	no	deletion	*Silene latifolia*
8	592–600	9	no	deletion	*Bougainvillea*
9	624–851	198	no	deletion	*Corrigiola*
10	629–673	36	no	deletion	*Bougainvillea*
11	636–644	9	no	insertion	All Caryophyllidae plus *Nepenthes*
12	665–670	6	no	deletion	*Petiveria*
13	687–851	144	no	deletion	*Petrorhagia*
14	707–721	15	no	insertion	*Tetragonia*
15	731–946	210	yes (C[c])	deletion	*Sarcobatus, Bougainvillea, Mirabilis, Petiveria, Phytolacca*
16	752–832	81	no	deletion	*Pereskia*
17	757–969	207	no	deletion	*Suaeda*
18	767–904	132	yes (D[c])	deletion	*Atriplex, Beta, Camphorosma, Chenopodium, Cycloloma, Grayia, Kochia, Salicornia, Spinacia*
19	795–980	180	no	deletion	*Amaranthus*
20	845–850	6	yes (E[c])	insertion	*Dianthus, Silene, Celosia, Froelichia, Stegnosperma, Calandrinia, Claytonia, Portulaca, Pereskia, Anredera, Tetragonia, Mollugo, Didierea, Alluaudia*
21	887–892	6	no	deletion	*Claytonia*
22	1012–1020	9	no	insertion	*Suaeda*
23	1057–1062	6	no	insertion	*Cycloloma*

[a] Coordinates refer to position in multiple alignment (Fig. 4).
[b] Taxa scored as "?" for indel are omitted.
[c] Letter code refers to character designation at end of alignment in Fig. 4.

TABLE 3. Characteristics and evolutionary features of the two ORF2280 homolog portions sequenced in 40 taxa of Caryophyllidae and outgroups, separately and combined. Values in brackets are for 34 Caryophyllales representatives only.

Characteristic	ORF2-ORF3[a]	ORF4-ORF5[b]	Combined
Length range (bp)	85–554 [251–554]	287–578 [293–578]	372–1 126 [610–1 126]
Length mean (bp)	434.1 [427.6]	467.6 [464.9]	901.7 [892.6]
Aligned length (bp)	563	593	1 156
Number of indels[c]	7	16	23
Number (and %) of constant nucleotide sites	416 (73.9)	429 (72.3)	845 (73.1)
Number (and %) of autapomorphic sites	95 (16.9)	100 (16.9)	195 (16.9)
Number (and %) of potentially informative sites	52 (9.2)	64 (10.8)	116 (10.0)
Number (and %) of variable sites by codon position			
1st	50 (34.0)	56 (34.1)	106 (34.1)
2nd	38 (25.9)	53 (32.3)	91 (29.3)
3rd	59 (40.1)	55 (33.5)	114 (36.7)
G and C content (%)	35.9–39.6 [36.7–39.6]	30.3–40.4 [32.6–35.0]	34.8–40.0 [34.8–40.0]
Uncorrected sequence divergence (%)	0–20.0 [0–5.6]	0.2–20.6 [0.2–5.2]	0.1–20.4 [0.1–5.0]
Parsimony analysis			
Number of shortest trees	78	>13 400	1 560
Length of shortest trees	176	247	441
Tree length by codon position			
1st	59	81	153
2nd	45	80	128
3rd	72	86	160
Consistency index (excluding uninformative characters)	0.817	0.612	0.628
Retention index	0.916	0.748	0.769
Ts/Tv[d]	0.88	0.86	0.83
Ts/Tv ratio by codon position			
1st	0.89	0.90	0.76
2nd	1.05	0.65	0.74
3rd	0.84	1.10	0.99

[a] This region, ranging from positions 1 to 563 in Fig. 4, corresponds to coordinate units 89452 to 89999 in *Nicotiana tabacum* cpDNA (Shinozaki et al., 1986).

[b] This region, ranging from positions 564 to 1156 in Fig. 4, corresponds to coordinate units 90770 to 91317 in *Nicotiana tabacum* cpDNA (Shinozaki et al., 1986).

[c] Determination of polarity based upon comparison to outgroup *Nicotiana* (*Epifagus* and *Pelargonium* excluded). See Table 2 for further characterization.

[d] Transition/transversion ratio average across all maximally parsimonious trees.

one of the five informative indels (indel C, Table 2) was perfectly congruent with this phylogeny. Indels A and B were inferred to be homoplastic, as each occurred along two branches of a trichotomy at the base of Chenopodiaceae/Amaranthaceae and within Caryophyllaceae, respectively. Indel D was detected in all Chenopodiaceae save two accessions; in these taxa (*Sarcobatus* and *Suaeda*), this indel was scored as "?" because the indel event was obliterated by larger deletions within this region. Furthermore, indel D is homoplastic when mapped onto the tree, as it was not detected in *Amaranthus*. Indel E, representing a 6-bp insertion relative to *Nicotiana*, was unambiguously detected in 15 taxa and scored as "?" in all other caryophyllalean taxa because it, too, was masked by larger deletions.

To examine relative variability of base substitutions within and between both ORF portions, the average number of character-state changes across all 1 560 shortest trees was mapped along the lengths of these regions using a 10-bp nonoverlapping window (Fig. 7). Within each ORF portion, this analysis revealed regions of high conservation and regions of considerable variability. The average number of base substitutions in "ORF2-ORF3" was 3.3 vs. 4.4 in "ORF4-ORF5," indicating a slightly higher frequency of character-state transformations in the latter. The average number of character-state changes over both ORF portions was 3.8. Steps calculated over all trees by codon position (for each ORF portion and both regions together) indicated that slightly more changes occurred in the third position followed by first and second positions (Table 3). Approximately 60% of these changes occurred in first- and second-codon positions, which are more likely to result in amino acid substitutions. The average Ts/Tv ratio across all 1 560 trees, as determined by MacClade, was 0.83, indicating slightly more transversions than transitions.

To investigate differences in tree topology and levels of homoplasy between the two regions of the ORF, parsimony analysis of each region was conducted. The strict consensus of the 78 maximally parsimonious trees derived from only "ORF2-ORF3" nucleotide substitutions (176 steps, CI excluding uninformative characters = 0.817, RI=0.916; Table 3) was slightly more resolved than the strict consensus tree derived from both ORF data sets (excluding gaps). The six taxa comprising Portulacaceae, Basellaceae, and Didiereaceae formed a clade, as did *Tetragonia* (Aizoaceae), *Sarcobatus* (Chenopodiaceae), both members of Nyctaginaceae, and *Petiveria* and *Phytolacca* of Phytolaccaceae. Caryophyllaceae was maintained as sister group to a clade comprised of amaranths and chenopods (except *Sarcobatus*). Separate analysis of "ORF4-ORF5" sequences generated a strict consensus tree that was consistent with, but considerably less resolved than, the strict consensus trees derived from both the combined analysis or the "ORF2-ORF3" analysis alone. In the "ORF4-ORF5" analysis, the number of shortest trees could not be ascertained, as the computer ran out of memory with 13 400 trees saved. This analysis was rerun by setting an arbitrary limit of 5 000 trees; each of these trees had a length of 247 steps, a CI excluding uninformative characters of 0.612, and a RI of 0.748 (Table 3). Maintained here was a monophyletic group consisting of all Caryophyllaceae, Amaranthaceae, and Chenopodiaceae (less *Sarcobatus*); all other taxa fell as a large polytomy to this clade. Although both ORF regions contained approximately the same number of phylogenetically informative substitutions, the "ORF4-ORF5" region contained more homoplastic characters, as determined by the lower consistency and retention indices, than did "ORF2-ORF3." The Ts/Tv ratio with respect to codon position also differed markedly between the two regions of the ORF (Table 3). Most chloroplast genes show a transversion bias at the first- and second-codon positions, whereas transitions predominate at the third position (Clegg, 1993). In "ORF2-ORF3," more transitions

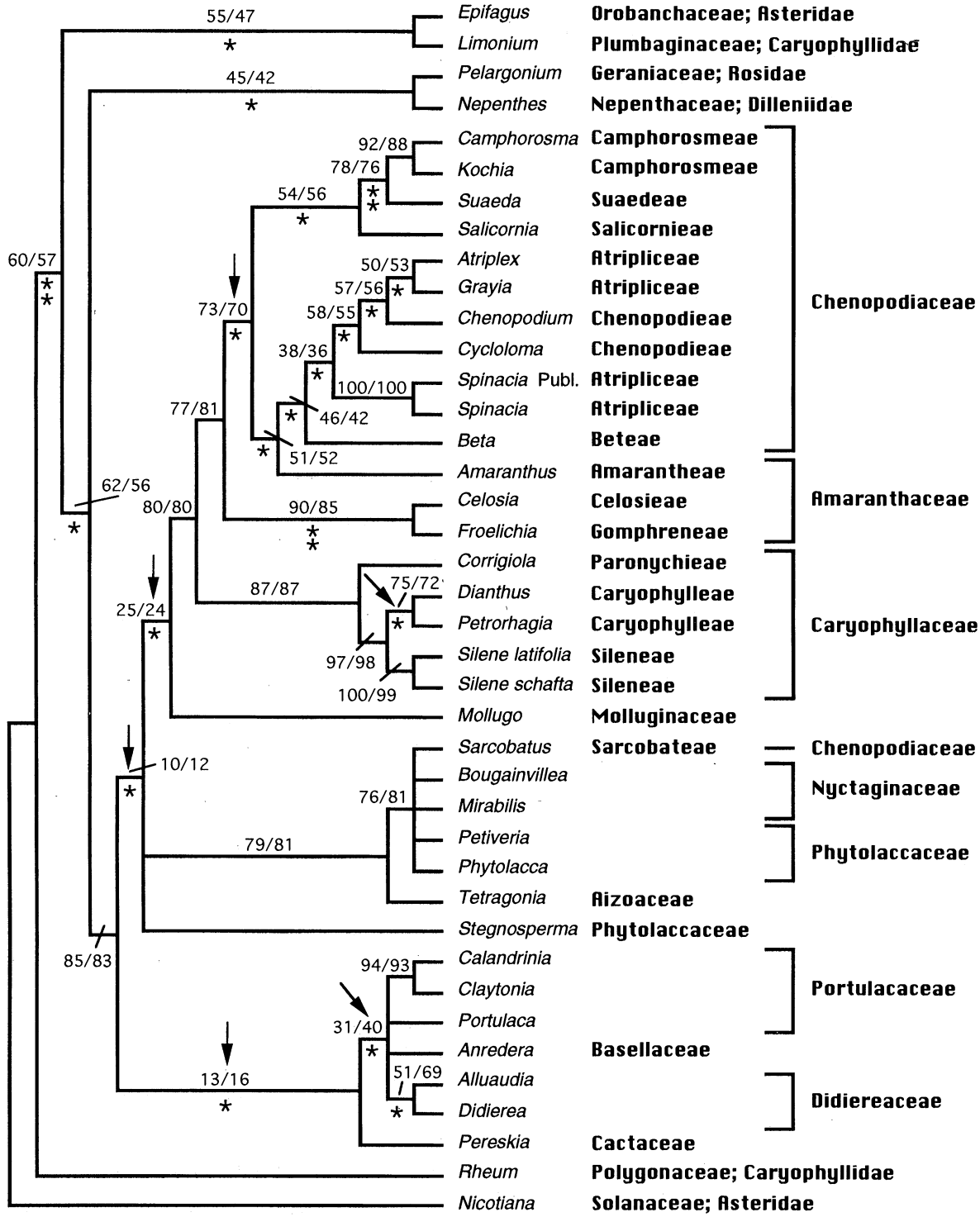


Fig. 5. Strict consensus of 33 maximally parsimonious 447-step trees derived from unweighted parsimony analysis of partial cpDNA ORF2280 homolog nucleotide sequences and five phylogenetically informative length mutations (CI excluding uninformative characters = 0.633, RI = 0.784) from 40 taxa of Caryophyllales and outgroups. Paired numbers above the nodes indicate the number of times the clade occurred in 100 or 1000 bootstrap replicates when the number of MAXTREES was set at 2500 or 100, respectively. Single or double asterisks below the nodes indicate that the branch collapses at 448 or 449 steps, respectively. Arrows indicate nodes that collapse when the five length mutations are excluded from the analysis. Complete taxon names are provided in Table 1. Intrafamilial classifications of Chenopodiaceae, Amaranthaceae, and Caryophyllaceae based on Ulbrich (1934), Schinz (1934), and Pax and Hoffmann (1934), respectively. Familial and subclass designations based on Cronquist (1981).

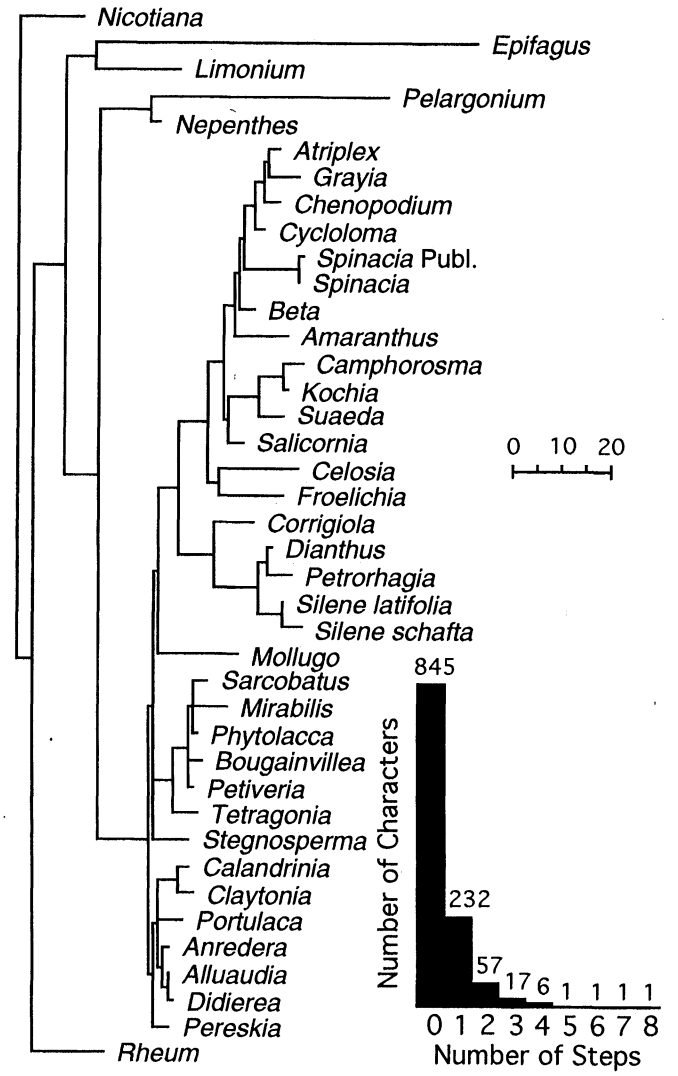


Fig. 6. One of 33 maximally parsimonious trees of 447 steps derived from unweighted parsimony analysis of partial cpDNA ORF2280 homolog nucleotide sequences and five phylogenetically informative length mutations (Table 2). CI excluding uninformative characters = 0.633; RI = 0.784. Lengths of branches are proportional to the number of inferred mutations (note scale bar). The histogram (inset) summarizes the number of inferred changes (steps) per character on this tree. Complete taxon names are provided in Table 1.

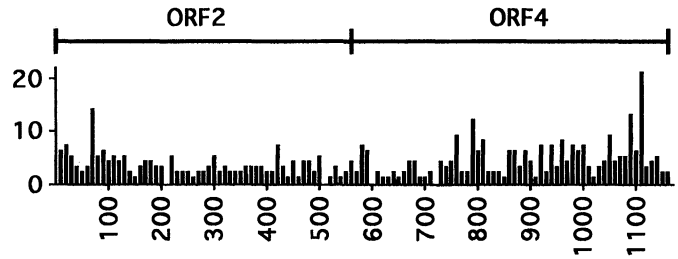


Fig. 7. Distribution of base substitutions across both ORF2280 homolog portions ("ORF2-ORF3" and "ORF4-ORF5"), as estimated by the average number of character-state changes over all 1560 shortest trees. Each bar represents the number of base substitutions using a 10-bp nonoverlapping window.

were apparent in the second position than in the third position.

The tree obtained from the neighbor-joining analysis (and using a Ts/Tv ratio of 1.0; Fig. 8) was similar to that obtained using parsimony when both ORF regions are included but informative gaps are not (i.e., the "collapsed" Fig. 5). Major differences between them include (1) the union of *Limonium* with *Rheum* (and not *Epifagus*) at the base of the tree, (2) the position of *Amaranthus* within the chenopods, (3) the association of *Pereskia* with *Mollugo* and not with Portulacaceae/Basellaceae/Didiereaceae, and (4) the failure of *Froelichia* and *Celosia* (both Amaranthaceae) to remain monophyletic. Within the Caryophyllales, however, the lengths of many of the internal branches are small, and these branches are supported weakly in the bootstrap analysis (<80% of 100 replicates). Thus, many of the differences between the parsimony and neighbor-joining analyses occur as a result of the branching order of weakly supported lineages.

The tree produced from the maximum likelihood method, using a Ts/Tv ratio of 2.0, had a log likelihood of −4325.10 (Fig. 9). This maximum likelihood tree was very similar to the trees constructed using parsimony and the entire data matrix (but, again, excluding informative indels). Here *Amaranthus* arises as sister taxon to a clade consisting of *Beta, Spinacia, Cycloloma, Chenopodium, Grayia,* and *Atriplex,* and *Froelichia* and *Celosia* are maintained as monophyletic, however their relationship to *Amaranthus* and the chenopods is not clear. When a Ts/Tv ratio of 1.0 was used, the resultant maximum-likelihood tree (log likelihood of −4289.23) was similar to that observed in Fig. 9, with the exception that *Amaranthus, Beta,* and *Spinacia* now formed a trichotomy. This clade was sister group to a monophyletic *Cycloloma, Chenopodium, Grayia,* and *Atriplex.*

Phylogenies estimated using neighbor-joining and maximum likelihood methods and parsimony analysis of nucleotide substitutions only reveal the following similar and noteworthy relationships. (1) *Amaranthus, Celosia,* and *Froelichia* (all Amaranthaceae) do not comprise a monophyletic group. In the maximum likelihood and parsimony analyses, *Amaranthus* arises from within a paraphyletic Chenopodiaceae. (2) *Sarcobatus* (Chenopodiaceae) is allied with Nyctaginaceae + Phytolaccaceae (the latter excluding *Stegnosperma* but including *Petiveria*). (3) Caryophyllaceae (with *Corrigiola* basal within the clade) are sister group to Chenopodiaceae + Amaranthaceae. (4) *Tetragonia* (Aizoaceae) is allied with Nyctaginaceae + Phytolaccaceae + *Sarcobatus.* (5) *Nepenthes* exhibits a closer relationship to the order than either *Rheum* (Polygonaceae) or *Limonium* (Plumbaginaceae). The unexpected and apparently novel positions of *Amaranthus* and *Sarcobatus* in the cladograms are unlikely to represent misidentifications or other errors as PCR-amplifications of additional representatives of these same genera (Table 1) produced the same unique-sized fragments as the taxon sequenced. The ORF data, however, do not provide strong support for many of the clades within the Caryophyllales, particularly among the basal nodes, as indicated by the low bootstrap percentages and many small internal branch lengths. Thus, these data do not elucidate basal relationships within the order.

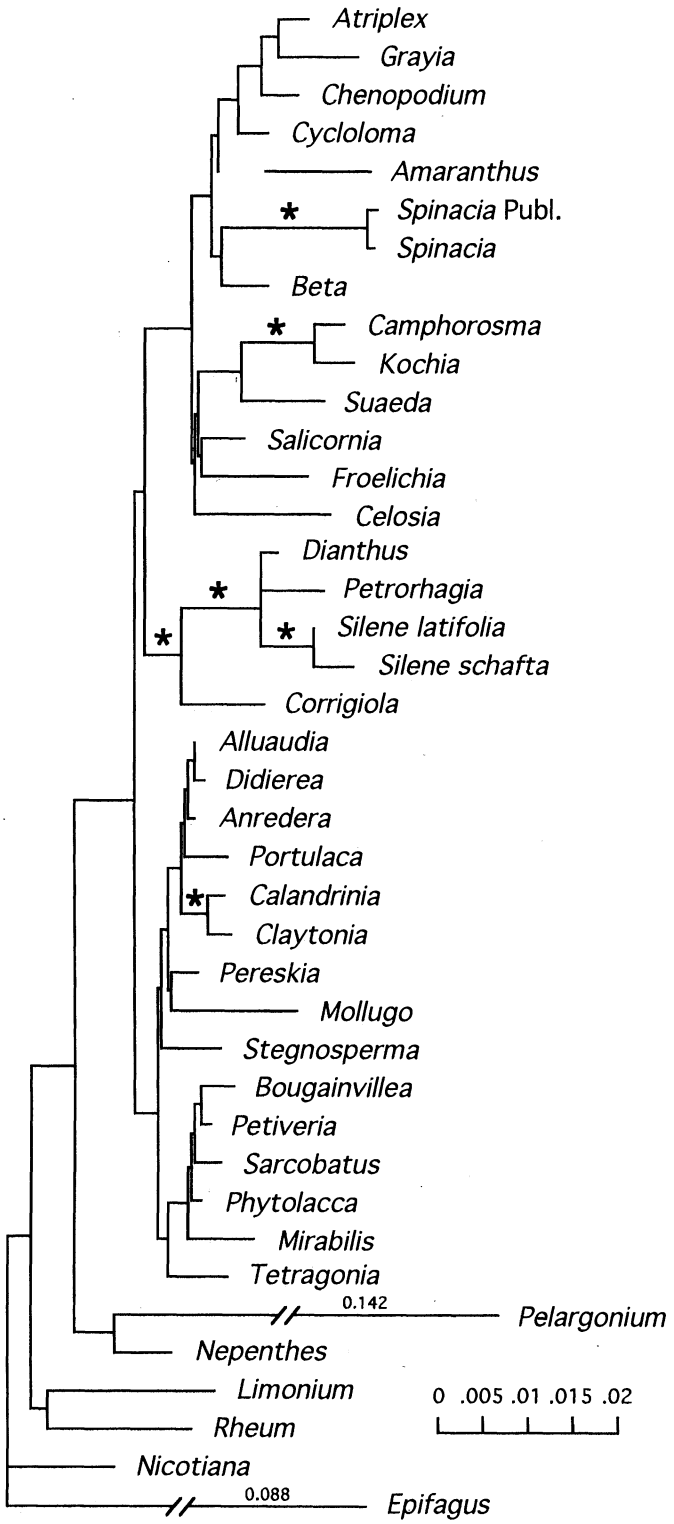


Fig. 8. Neighbor-joining tree inferred from the analysis of substitution rates estimated from the two-parameter method of Kimura (1980) for 40 combined "ORF2-ORF3" and "ORF4-ORF5" sequences. Length of branches are proportional to distances (note scale bar) and branch lengths less than 0.00060 have been collapsed; branches leading to *Pelargonium* and *Epifagus* are extraordinarily long and have been truncated (their lengths are indicated, however). Branches supported in ≥ 80 of 100 bootstrap replicate analyses are indicated by asterisks. Complete taxon names are provided in Table 1.

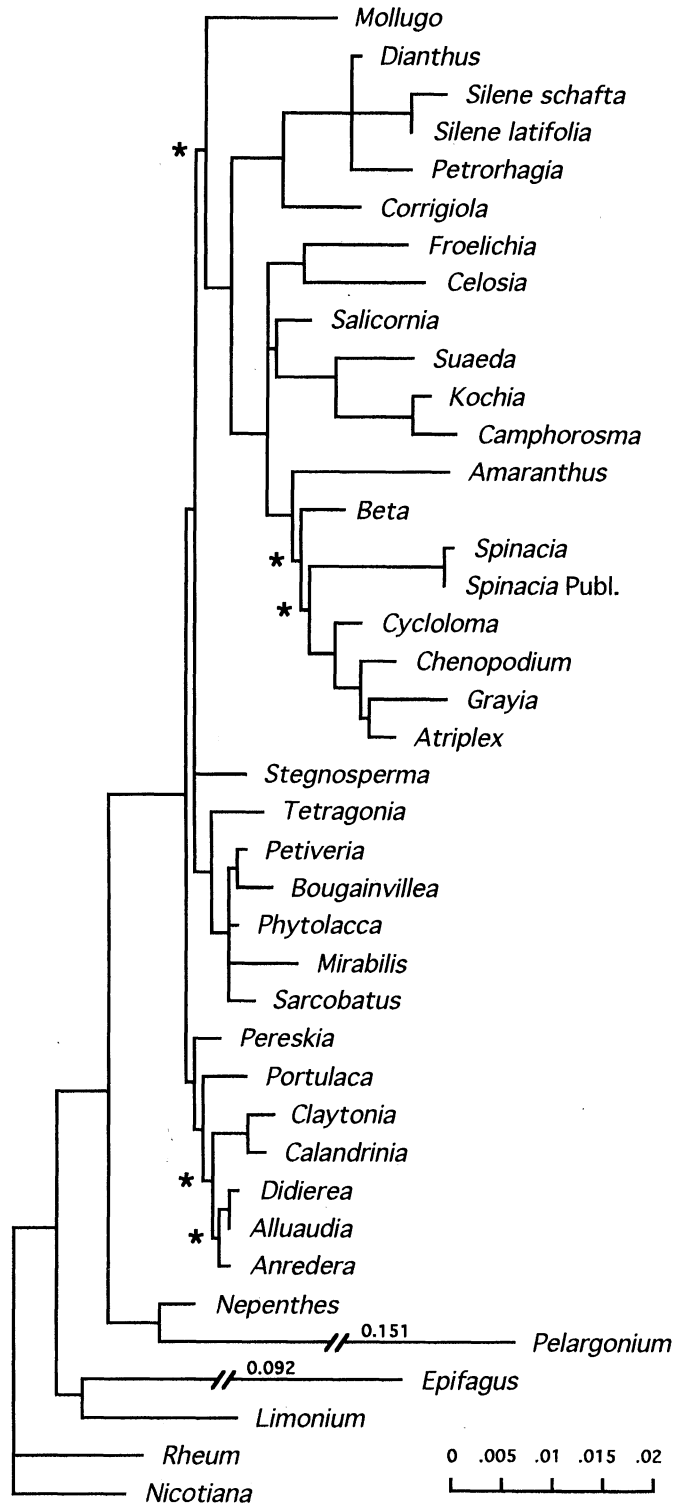


Fig. 9. Maximum likelihood tree constructed from partial ORF2280 homolog sequences using a Ts/Tv ratio of 1.0. The log likelihood of this tree is −4 325.10. Branch lengths are proportional to distances (note scale bar). All branch lengths, unless otherwise indicated, have values significantly positive at $P<0.01$. Branches identified by asterisks are supported by values significantly positive at $P<0.05$.

TABLE 4. Angiosperms possessing major deletions, relative to *Nicotiana tabacum,* in ORF2280 homolog sequences. In all, representatives from 70 families (list available upon request) were surveyed for length variation using PCR and primer pair "ORF2" and "ORF5" (Fig. 3).

Family	Species	Deletion (bp)
Apiaceae	*Coriandrum sativum*	500
	Daucus carota	500
Araliaceae	*Aralia chinensis*	500
	Hedera helix	500
Aristolochiaceae	*Aristolochia durior*	100
Asclepiadaceae	*Asclepias* (three spp.)	500
	Periploca sepium	500
Bignoniaceae	*Clytostoma callistigioides*	500
Campanulaceae	*Campanula garganica*	1400
Caprifoliaceae	*Kolkwitzia amabilis*	500
	Symphoricarpos alba	500
	Triosteum sp.	500
Convolvulaceae	*Cuscuta* sp.	600
Dipsacaceae	*Dipsacus sativus*	500
Geraniaceae	*Pelargonium* × *hortorum*	900
	Erodium ciconium	400
	Sarcocaulon vanderietiae	700
Iridaceae	*Crocus vernus*	100
Liliaceae	*Allium cernuum*	100
	Erythronium albidum	200
	Tricyrtis latifolia	100
	Tulipa sp.	200
Lobeliaceae	*Lobelia mildbraedii*	150
Onagraceae	*Oenothera missouriensis*	600
Paeoniaceae	*Paeonia lactiflora*	350
Pittosporaceae	*Billardiera scandens*	500
Ranunculaceae	*Aquilegia vulgaris*	450
Scrophulariaceae	*Striga asiatica*	550
Verbenaceae	*Clerodendrum ugandense*	500

With the incorporation of the five informative indels in the parsimony analysis (Fig. 5), greater resolution is achieved than by using the nucleotide substitutions alone. This analysis reveals that *Celosia* + *Froelichia* (both Amaranthaceae) are sister group to all chenopods (excluding *Sarcobatus*); however, Chenopodiaceae are paraphyletic with *Amaranthus* arising from within. Indel C strengthens the union among *Sarcobatus, Petiveria, Phytolacca,* and both Nyctaginaceae members as all possess this deletion. Similarly, *Dianthus* and *Petrorhagia,* both of tribe Caryophylleae, form a clade on the basis of synapomorphy indel B. The inclusion of these indels also results in the formation of a (very weakly supported) group containing representatives of Portulacaceae, Basellaceae, Didiereaceae, and Cactaceae.

Major structural variation in other angiosperms—Results of the PCR survey to detect length mutations in ORF2280 homolog sequences using primers "ORF2" and "ORF5" revealed that the vast majority of the angiosperm cpDNAs examined yielded a single 1.9-kb fragment, the size expected in *Nicotiana* cpDNA. Major deletions, ranging in size from 100 bp to 1.4 kb, were observed from 31 DNA accessions representing 19 families (Table 4). No insertions (> 50 bp) were detected.

DISCUSSION

Phylogenetic utility of ORF2280 homolog sequences—Several factors constrain the usefulness of ORF2280 homolog nucleotide sequences in phylogenetic studies. First, the evolutionarily conservative nature of these sequences precludes robust hypotheses of relationships among allied taxa. For both ORF regions and across 40 species (representing 18 families) some 10% of nucleotides were potentially informative and, within the Caryophyllales, sequence divergence values among pairwise comparisons approached only 5% of nucleotides. Congeners had identical or nearly identical sequences, although structural variation was apparent, as seen in a 9-bp gap in one of two species of *Silene* (indel 7, Table 2). Length variation in ORF sequences among congeners, or even among isolates of a single species, is not novel and has been reported elsewhere (Sytsma and Gottlieb, 1986; Blasko et al., 1988; Nimzyk, Schöndorf, and Hachtel, 1993).

Second, the ubiquitous presence of large, informative indels in these ORF sequences creates problems on how to best include these data in a cladistic analysis (i.e., should they be weighted greater than, lesser than, or equal to, a nucleotide substitution, or should they even be included at all?). To date, there is no consensus on how to best treat these data. In this study, all nucleotide substitutions and the five indels were weighted equally although the latter are obviously a much less frequent occurrence. Excluding potentially informative indels from an analysis may result in trees that are less resolved, as seen in Fig. 5.

Third, plants with greatly rearranged cpDNA IRs, including those possessing major deletions within their ORF homologs (such as those observed in some Campanulaceae, Geraniaceae, Cuscutaceae, Orobanchaceae, Passifloraceae, and several monocot families including the grasses; Hiratsuka et al., 1989; Downie and Palmer, 1992a, b; Downie et al., 1994; Davis, 1995; Hahn, Givnish, and Sytsma, 1995), can pose problems in the sense that the remaining sequence, if any, may not provide enough nucleotide bases to analyze. These rearrangements, however, can serve as additional characters for phylogenetic purposes (reviewed in Downie and Palmer, 1992b; Olmstead and Palmer, 1994).

Fourth, in those plastid genomes where the ORF lies outside of the evolutionarily conservative IR region, the rate of evolution for that portion of the genome may be 4–5 times greater than if it were still contained within the IR (Wolfe, Li, and Sharp, 1987; Downie et al., 1994). Although the molecular mechanisms responsible for maintaining the low frequency of mutations in IR sequences are not understood, the presence of two identical copies of this region in most angiosperm cpDNAs could imply the operation of some sort of gene conversion or copy-correction process (Palmer, 1991). Chloroplast genomes lacking an IR, such as those of conifers (Raubeson and Jansen, 1992) and some legumes (Lavin, Doyle, and Palmer, 1990), or those chloroplast genomes with IRs of substantially different sizes, such as those of some Apiaceae (S. Downie, unpublished data), may also be problematic, particularly if a study includes plants both with and without structural alterations to their IRs, as substantial differences in substitution rates among lineages may lead to artifactual results in a parsimony analysis (Felsenstein, 1978).

Lastly, despite the generally conservative nature of

ORF homolog sequences, accelerated rates of ORF sequence evolution evidently exist in some taxa (e.g., *Pelargonium* and *Epifagus*; Downie et al., 1994), even though these genes are contained within an IR. As indicated above, the presence of these taxa in an analysis based solely upon parsimony may confound interpretation of relationships.

Large indels appear to be a common occurrence within the ORF, affirmed in the sequences obtained here (Table 2), the results of PCR amplifications of a diversity of angiosperm cpDNAs (Table 4), and by way of reports available in literature (e.g., Sytsma and Gottlieb, 1986; Jansen and Palmer, 1988; Kellogg, 1992; Downie and Palmer, 1992a, 1994; Manos, Nixon, and Doyle, 1993; Davis, 1995; Hahn, Givnish, and Sytsma, 1995). Indels can be considered phylogenetically highly informative characters for cladistic study and have been used to bolster monophyly for those taxa which share identical mutations (e.g., Lloyd and Calder, 1991; Mes and Hart, 1994), although it is not unrealistic to presume that apparently identical indels may have had multiple origins in unrelated taxa (Downie et al., 1991, 1994; Golenberg et al., 1993; Davis, 1995). Only one of the five informative length mutations detected in this study (indel C, Table 2) unambiguously defined a monophyletic group, as it mapped without homoplasy onto the strict consensus tree derived from the analysis of base substitutions alone. On this same tree, indels A and B were inferred to be homoplastic as each fell along two of three branches of a trichotomy. However, when these characters are included alongside nucleotide substitutions and analyzed using parsimony (Fig. 5), these regions of the cladograms become perfectly resolved, and indels A and B each occur without homoplasy. The nature of indel D is not altogether clear. Indel D, absent in *Amaranthus* but present in all but two chenopods, occurs in approximately the same position as length mutation 19 (Table 2), a 180-bp deletion unique to *Amaranthus*. It is possible that this 180-bp deletion occurred prior to the mutation event giving rise to indel event D, thus prohibiting the latter from occurring in this lineage by removing one of its endpoints. Other scenarios are also plausible: indel D has occurred twice independently in the Caryophyllales (in the lineages leading to *Camphorosma-Salicornia* and *Atriplex-Beta*), or a reversal (i.e., the unlikely gain and precise insertion of an identical or near identical sequence) has occurred on the branch leading to *Amaranthus* after the deletion originated in the common ancestor of the chenopods. Like indel D, the status of indel E, is also not clear. Indel E is clearly absent in the outgroups *Nicotiana, Epifagus, Nepenthes, Limonium,* and *Rheum* (Fig. 4) and is seen in those 14 caryophyllalean genera identified in Table 2. However, the superimposition of several larger indel events in Chenopodiaceae, *Amaranthus, Corrigiola, Petrorhagia,* Phytolaccaceae (sensu stricto), and Nyctaginaceae (Fig. 4) within this region obliterates any previous event that may have occurred in these taxa.

Of the noncaryophyllid families for which more than one species was examined that exhibited a detectable deletion (e.g., Apiaceae, Araliaceae, Asclepiadaceae, and Caprifoliaceae; Table 4), the occurrence of similarly sized fragments might reinforce monophyly of each of these groups. This 500-bp deletion may also serve as a marker uniting all families, as representatives from Bignoniaceae, Dipsacaceae, Pittosporaceae, and Verbenaceae all possess the same sized mutation. On the basis of cpDNA *rbc*L sequences, these families can all be thought of to belong to an expanded Asteridae (Olmstead et al., 1993). It is stressed, however, that in the absence of the underlying sequence data, the homology of similar-sized deletions, even in putatively related taxa, is not assured.

In summary, while these ORF data are indeed useful in establishing some hypotheses of relationship within the Caryophyllales, particularly when indels are considered alongside nucleotide substitutions, the characters they contribute to the phylogenetic analyses are few. These data fail to elucidate some long-standing problems, such as the position(s) of the two anthocyanin-producing taxa (Caryophyllaceae and Molluginaceae) and the proper placement of *Stegnosperma.* Furthermore, these data do not provide the resolution necessary to ascertain the deeper level relationships within the order. These data do, however, provide precise information on the location and size of indels, of which there are many. As stated above, these indels can be used to bolster support for several clades recognized using nucleotide substitutions alone.

Phylogenetic resolutions—Owing to the lack of resolution among many of the clades depicted in the phylogenetic trees, as ascertained by the low bootstrap values and/or short branch lengths, this discussion will deal with only those few best supported (and most intriguing) relationships suggested by the study. Included here is the polyphyly of Amaranthaceae, the placement of *Sarcobatus* alongside Phytolaccaceae and Nyctaginaceae, and the grouping of Chenopodiaceae/Amaranthaceae with Caryophyllaceae.

Amaranthaceae/Chenopodiaceae—The amaranths and chenopods have long been recognized as closely related so it is not surprising that they arise together here. On the basis of *rbc*L sequence comparisons (Rettig, Wilson, and Manhart, 1992; Chase et al., 1993; Manhart and Rettig, 1994), only three genera within this alliance have been considered to date: *Atriplex* and *Spinacia* (Chenopodiaceae) and *Amaranthus* (Amaranthaceae). With the exception of the Rettig, Wilson, and Manhart (1992) analysis, where the Chenopodiaceae form a monophyletic group (albeit weakly supported), *Spinacia* groups with *Amaranthus* rather than with *Atriplex,* suggestive of a paraphyletic Chenopodiaceae. The possibility that a monophyletic Amaranthaceae may be nested within Chenopodiaceae has been diagrammed previously by Carolin (1983) on the basis of leaf trichome data. The difficulty in accepting such a hypothesis, as indicated by Rodman (1994), is that such an arrangement requires a reversal to multiovulate gynoecia (characteristic of tribe Celosieae of Amaranthaceae), as most other amaranths (including *Froelichia*) and all chenopods possess a single basal ovule (Cronquist, 1981). The results presented here are unique in suggesting that Amaranthaceae are polyphyletic, with *Celosia* (Celosieae, Amaranthoideae) and *Froelichia* (Gomphreneae, Gomphrenoideae) forming a clade that is sister group to Chenopodiaceae + *Amaranthus.* The presence of uniovulate *Amaranthus* (Amarantheae,

Amaranthoideae) but not multiovulate *Celosia* (barring exceptions; Townsend, 1993) within Chenopodiaceae makes the above hypothesis more palatable.

Downie and Palmer (1994), on the basis of sampling very few taxa, considered Amaranthaceae and Chenopodiaceae sister groups, with the cpDNA IRs of the latter characterized structurally by the loss of an ≈ 300-bp fragment. *Amaranthus,* however, was not included in that study. The sequence data obtained herein confirm this (297-bp) deletion in all chenopod (except *Sarcobatus*; see below) and *Amaranthus* cpDNAs; this deletion was not observed in *Celosia* and *Froelichia.* Surveying for the occurrence of this deletion in other Amarantheae could serve as a molecular marker to identify other taxa arising within Chenopodiaceae. Within the chenopods, a 6-kb inversion has been detected in the chloroplast genomes of *Atriplex* and *Chenopodium* but not in *Beta, Kochia,* or *Spinacia* (Downie and Palmer, 1994). In the phylogenies inferred here, *Atriplex* and *Chenopodium,* along with *Grayia,* form a clade. Surveying for this inversion in *Grayia,* and in other members of tribes Atripliceae and Chenopodieae, might also prove profitable.

Phytolaccaceae, Nyctaginaceae, and Sarcobatus–Phytolacca and *Petiveria* (Phytolaccaceae) and *Mirabilis* and *Bougainvillea* (Nyctaginaceae) are supported as a distinct clade on the basis of ORF nucleotide substitutions and by the common possession of a 210-bp deletion. Again, their union is not unexpected, as the close association between these taxa has been known for some time. What is intriguing, however, is the juxtaposition of *Sarcobatus* (Chenopodiaceae) with these taxa. Pairwise nucleotide divergence values between ORF sequences from *Sarcobatus, Phytolacca, Petiveria, Mirabilis,* and *Bougainvillea* ranged from 0.4 to 1.1% of nucleotides, with the number of pairwise mutation differences varying from 4 to 10. In contrast, sequence divergence values between *Sarcobatus* and any other chenopod were substantially higher, ranging between 2.3 and 3.7% of nucleotides. *Sarcobatus* also possesses the same 210-bp deletion as detected in *Phytolacca, Petiveria, Mirabilis,* and *Bougainvillea*; this deletion was not apparent in any other examined caryophyllid taxon. The genus *Sarcobatus,* endemic to North America, is the only member of Ulbrich's (1934) Chenopodiaceae subfamily Sarcobatoideae. *Sarcobatus* is distinctive within the Chenopodiaceae in having form-P3cf sieve-element plastids with a central globular crystal (Behnke, 1993, 1994). Although this plastid form is not found in any other examined member of Chenopodiaceae, it does resemble those plastid characters present in some genera of Nyctaginaceae (Behnke, 1994). Pollen morphology also supports, in part, the distinction of *Sarcobatus* from other Chenopodiaceae (Nowicke, 1994).

Caryophyllaceae—The five examined species of Caryophyllaceae constitute a monophyletic group in this analysis. The groupings of *Dianthus* with *Petrorhagia,* both of tribe Caryophylleae, and tribes Caryophylleae with Sileneae (*Silene*), both of subfamily Caryophylloideae, are consistent with traditional taxonomic treatment (Pax and Hoffmann, 1934). *Corrigiola,* included in either Caryophyllaceae (subfamily Paronychioideae), Molluginaceae, or Illecebraceae (discussed in Gilbert, 1987), is allied with the four other representatives of Caryophyllaceae examined. *Dianthus* and *Petrorhagia* cpDNA IRs share a 261-bp deletion in their ORF homologs; this mutation is not apparent in *Lychnis* or *Saponaria* cpDNAs (both also Caryophylloideae; Table 1). *Corrigiola, Petrorhagia,* and *Silene* also each possess unique deletions of 198, 144, and 9 bp, respectively (Table 2). These indels may prove useful in circumscribing monophyletic groups within Caryophyllaceae as additional taxa are examined.

The proposed sister-group relationship between Caryophyllaceae and Amaranthaceae/Chenopodiaceae is also supported, in part, by similarities in their pantoporate pollen and, considering only the paronychioid Caryophyllaceae within the former, their floral morphology (Nowicke and Skvarla, 1980; Kühn et al., 1993; Nowicke, 1994). It has been presumed that pantoporate pollen may have arisen independently in these two groups (Nowicke and Skvarla, 1980; Cronquist, 1988). If, however, the palynological similarities are not convergence, then the multiovulate condition found in *Celosia* (discussed above) has indeed occurred independently. Phylogenetic analysis of nucleotide sequence data from the plastid *rbc*L gene also suggests an affinity among these three families (Manhart and Rettig, 1994).

Conclusions—Although the comparative analysis of *rbc*L sequence variation has clearly overshadowed the use of other organellar DNA sequences in phylogenetic studies, data from several other coding and noncoding regions are rapidly becoming available and exhibit great potential for inferring evolutionary relationships at various hierarchical levels. However, the IR region of the chloroplast genome, a region encompassing almost a third of the entire molecule in most angiosperms, has been largely ignored (at least from the framework of comparative sequencing) primarily because of its highly conservative nature. We have demonstrated that by focusing on one of the most variable regions of the IR, specifically the region occupied by ORF2280 in *Nicotiana,* a new set of data is obtained with which to assess evolutionary relationships. ORF2280 homolog sequences appear to be particularly prone to length mutations relative to most other regions within the IR, and can be used alongside nucleotide substitutions to infer evolutionary history. Because the number of variable nucleotide sites provided by the ORF sequences is low, greater resolution of relationships within the Caryophyllales using this region will likely require additional sequence data from the more variable 3′ portion of the ORF and the intergenic spacers between the gene and *ndh*B. These regions could provide the additional characters necessary to more fully resolve relationships within the order. Sequence data from these regions are currently being obtained.

LITERATURE CITED

ALBERT, V. A, S. E. WILLIAMS, AND M. W. CHASE. 1992. Carnivorous plants: phylogeny and structural evolution. *Science* 257: 1491–1495.

BEHNKE, H.-D. 1993. Further studies of the sieve-element plastids of the Caryophyllales including *Barbeuia, Corrigiola, Lyallia, Microtea, Sarcobatus,* and *Telephium. Plant Systematics and Evolution* 186: 231–243.

———. 1994. Sieve-element plastids: their significance for the evolution and systematics of the order. *In* H.-D. Behnke and T.J. Mabry [eds.], Caryophyllales: evolution and systematics, 87–121. Springer-Verlag, Berlin.

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. *Current Genetics* 14: 287–292.

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

CAROLIN, R. C. 1983. The trichomes of the Chenopodiaceae and Amaranthaceae. *Botanische Jahrbücher für Systematik* 103: 451–466.

CHASE, M. W., ET AL. 1993. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbc*L. *Annals of the Missouri Botanical Garden* 80: 528–580.

CLEGG, M. T. 1993. Chloroplast DNA sequences and the study of plant evolution. *Proceedings of the National Academy of Sciences, USA* 90: 363–367.

CRONQUIST, A. 1981. An integrated system of classification of flowering plants. Columbia University Press, New York, NY.

———. 1988. The evolution and classification of flowering plants. New York Botanical Garden, Bronx, NY.

DAHLGREN, R. 1980. A revised system of classification of angiosperms. *Botanical Journal of the Linnean Society* 80: 91–124.

DAVIS, J. I. 1995. A phylogenetic structure for the monocotyledons, as inferred from chloroplast DNA restriction site variation, and a comparison of measures of clade support. *Systematic Botany* 20: 503–527.

DOWNIE, S. R., AND J. D. PALMER. 1992a. Restriction site mapping of the chloroplast DNA inverted repeat: a molecular phylogeny of the Asteridae. *Annals of the Missouri Botanical Garden* 79: 266–283.

———, AND ———. 1992b. Use of chloroplast DNA rearrangements in reconstructing plant phylogeny. *In* P. Soltis, D. Soltis, and J. Doyle [eds.], Molecular systematics of plants, 14–35. Chapman and Hall, New York, NY.

———, AND ———. 1994. A chloroplast DNA phylogeny of the Caryophyllales based on structural and inverted repeat restriction site variation. *Systematic Botany* 19: 236–252.

———, R. G. OLMSTEAD, G. ZURAWSKI, D. E. SOLTIS, P. S. SOLTIS, J. C. WATSON, AND J. D. PALMER. 1991. Six independent losses of the chloroplast DNA *rpl*2 intron in dicotyledons: molecular and phylogenetic implications. *Evolution* 45: 1245–1259.

———, D. S. KATZ-DOWNIE, K. H. WOLFE, P. J. CALIE, AND J. D. PALMER. 1994. Structure and evolution of the largest chloroplast gene (ORF2280): internal plasticity and multiple gene loss during angiosperm evolution. *Current Genetics* 25: 367–378.

DOYLE, J. J., AND J. L. DOYLE. 1987. A rapid DNA isolation for small quantities of fresh tissue. *Phytochemical Bulletin* 19: 11–15.

FARRIS, J. S. 1989. The retention index and homoplasy excess. *Systematic Zoology* 38: 406–407.

FELSENSTEIN, J. 1978. Cases in which parsimony and compatibility methods will be positively misleading. *Systematic Zoology* 27: 401–410.

———. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of Molecular Evolution* 17: 368–376.

———. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39: 783–791.

———. 1993. PHYLIP (Phylogeny Inference Package) version 3.5s. Distributed by the author. Department of Genetics, University of Washington, Seattle.

GILBERT, M. G. 1987. The taxonomic position of the genera *Telephium* and *Corrigiola*. *Taxon* 36: 47–49.

GOLENBERG, E. M., M. T. CLEGG, M. L. DURBIN, J. DOEBLEY, AND D. P. MA. 1993. Evolution of a noncoding region of the chloroplast genome. *Molecular Phylogenetics and Evolution* 2: 52–64.

HAHN, W. J., T. J. GIVNISH, AND K. J. SYTSMA. 1995. Evolution of the monocot chloroplast inverted repeat. I. Evolution and phylogenetic implications of the ORF 2280 deletion. *In* P. J. Rudall, P. J. Cribb, D. F. Cutler, and C. J. Humphries [eds.], Monocotyledons: systematics and evolution, 579–587. Royal Botanic Gardens, Kew.

HERSHKOVITZ, M. A. 1989. Phylogenetic studies in Centrospermae: a brief appraisal. *Taxon* 38: 602–610.

HIRATSUKA, J., H. SHIMADA, R. WHITTIER, T. ISHIBASHI, M. SAKAMOTO, M. MORI, C. KONDO, Y. HONJI, C.-R, SUN, B.-Y. MENG, Y.-Q. LI, A. KANNO, Y. NISHIZAWA, A. HIRAI, K. SHINOZAKI, AND M. SUGIURA. 1989. The complete sequence of the rice (*Oryza sativa*) chloroplast genome: intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. *Molecular and General Genetics* 217: 185–194.

JANSEN, R. K., AND J. D. PALMER. 1988. Phylogenetic implications of chloroplast DNA restriction site variation in the Mutisieae (Asteraceae). *American Journal of Botany* 75: 753–766.

KELLOGG, E. A. 1992. Tools for studying the chloroplast genome in the Triticeae (Gramineae): an *Eco*RI map, a diagnostic deletion, and support for *Bromus* as an outgroup. *American Journal of Botany* 79: 186–197.

KIMURA, M. 1980. A simple method for estimating evolutionary rates of base substitution through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* 16: 111–120.

KLUGE, A. G., AND J. S. FARRIS. 1969. Quantitative phyletics and the evolution of anurans. *Systematic Zoology* 18: 1–32.

KÜHN, U., V. BITTRICH, R. CAROLIN, H. FREITAG, I. C. HEDGE, P. UOTILA, AND P. G. WILSON. 1993. Chenopodiaceae. *In* K. Kubitzki, J. G. Rohwer, and V. Bittrich [eds.], The families and genera of flowering plants, 13–19. Springer-Verlag, Berlin.

LAVIN, M., J. J. DOYLE, AND J. D. PALMER. 1990. Evolutionary significance of the loss of the chloroplast DNA inverted repeat in the Leguminosae subfamily Papilionoideae. *Evolution* 44: 390–402.

LEVINSON, G., AND G. A. GUTMAN. 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Molecular Biology and Evolution* 4: 203–221.

LLOYD, D. G., AND V. L. CALDER. 1991. Multi-residue gaps, a class of molecular characters with exceptional reliability for phylogenetic analyses. *Journal of Evolutionary Biology* 4: 9–21.

MADDISON, W. P., AND D. R. MADDISON. 1992. MacClade, version 3.0. Sinauer, Sunderland, MA.

MANHART, J. R., AND J. H. RETTIG. 1994. Gene sequence data. *In* H.-D. Behnke and T.J. Mabry [eds.], Caryophyllales: evolution and systematics, 235–246. Springer-Verlag, Berlin.

MANOS, P. S., K. C. NIXON, AND J. J. DOYLE. 1993. Cladistic analysis of restriction site variation within the chloroplast DNA inverted repeat region of selected Hamamelididae. *Systematic Botany* 18: 551–562.

MES, T. H. M., AND H.'T HART. 1994. *Sedum surculosum* and *S. jaccardianum* (Crassulaceae) share a unique 70 bp deletion in the chloroplast DNA *trn*L (UAA)-*trn*F (GAA) intergenic region. *Plant Systematics and Evolution* 193: 213–221.

NIMZYK, R., T. SCHÖNDORF, AND W. HACHTEL. 1993. In-frame length mutations associated with short tandem repeats are located in unassigned open reading frames of *Oenothera* chloroplast DNA. *Current Genetics* 23: 265–270.

NOWICKE, J. W. 1994. Pollen morphology and exine ultrastructure. *In* H.-D. Behnke and T.J. Mabry [eds.], Caryophyllales: evolution and systematics, 167–221. Springer-Verlag, Berlin.

———, AND J. J. SKVARLA. 1980. Pollen morphology: the potential influence in higher order systematics. *Annals of the Missouri Botanical Garden* 66: 633–700.

OHYAMA, K., H. FUKUZAWA, T. KOHCHI, H. SHIRAI, T. SANO, S. SANO, K. UMESONO, Y. SHIKI, M. TAKEUCHI, Z. CHANG, S.-I. AOTA, H. INOKUCHI, AND H. OZEKI. 1986. Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast genome. I. Cloning and gene identification. *Nature* 322: 572–574.

OLMSTEAD, R. G., B. BREMER, K. M. SCOTT, AND J. D. PALMER. 1993. A parsimony analysis of the Asteridae sensu lato based on *rbc*L sequences. *Annals of the Missouri Botanical Garden* 80: 700–722.

———, H. J. MICHAELS, K. M. SCOTT, AND J. D. PALMER. 1992. Monophyly of the Asteridae and identification of their major lineages inferred from DNA sequences of *rbc*L. *Annals of the Missouri Botanical Garden* 79: 249–265.

———, AND J. D. PALMER. 1994. Chloroplast DNA systematics: a review of methods and data analysis. *American Journal of Botany* 81: 1205–1224.

OLSEN, G. J., R. OVERBEEK, N. LARSEN, T. L. MARSH, M. J. MCCAUGHEY, M. A. MACIUKENAS, W.-M. KUAN, T. J. MACKE, Y. XING, AND C.

R. WOESE. 1992. The ribosomal database project. *Nucleic Acids Research* 20: 2199–2200.

———, H. MATSUDA, R. HAGSTROM, AND R. OVERBEEK. 1994. fastDNAml: a tool for construction of phylogenetic trees of DNA sequences using maximum likelihood. *Computer Applications in the Biosciences* 10: 41–48.

PALMER, J. D. 1986. Isolation and structural analysis of chloroplast DNA. *Methods in Enzymology* 118: 167–186.

———. 1991. Plastid chromosomes: structure and evolution. *In* L. Bogorad and I.K. Vasil [eds.], The molecular biology of plastids, vol. 7A, 5–53, Cell culture and somatic cell genetics of plants. Academic Press, San Diego, CA.

———, G. P. SINGH, AND D. T. N. PILLAY. 1983. Structure and sequence evolution of three legume chloroplast DNAs. *Molecular and General Genetics* 190: 13–19.

PAX, F, AND K. HOFFMANN. 1934. Caryophyllaceae. *In* A. Engler and H. Harms [eds.], Die natürlichen Pflanzenfamilien, 2d. ed., vol. 16C: 275–364. Wilhelm Engelmann, Leipzig.

RAUBESON, L. A., AND R. K. JANSEN. 1992. A rare chloroplast-DNA mutation is shared by all conifers. *Biochemical Systematics and Ecology* 20: 17–24.

RAWSON, J. Y. R., M. T. CLEGG, K. THOMAS, C. RINEHART, AND B. WOOD. 1981. A restriction endonuclease map of the ribosomal RNA genes and short single copy DNA sequence of the pearl millet chloroplast genome. *Gene* 16: 11–19.

RETTIG, J. H., H. D. WILSON, AND J. R. MANHART. 1992. Phylogeny of the Caryophyllales—gene sequence data. *Taxon* 41: 201–209.

RODMAN, J. E. 1990. Centrospermae revisited, part 1. *Taxon* 39: 383–393.

———. 1994. Cladistic and phenetic studies. *In* H.-D. Behnke and T. J. Mabry [eds.], Caryophyllales: evolution and systematics, 279–301. Springer-Verlag, Berlin.

SAITOU, N., AND M. NEI. 1987. The neighbor-joining method: a new method for reconstructing evolutionary trees. *Molecular Biology and Evolution* 4: 406–425.

SCHINZ, H. 1934. Amaranthaceae. *In* A. Engler and H. Harms [eds.], Die natürlichen Pflanzenfamilien, 2d. ed., vol. 16C: 7–85. Wilhelm Engelmann, Leipzig.

SHIMADA, H., AND M. SUGIURA. 1991. Fine structural features of the chloroplast genome: comparison of the sequenced chloroplast genomes. *Nucleic Acids Research* 19: 983–995.

SHINOZAKI, K., ET AL. 1986. The complete nucleotide sequence of the tobacco chloroplast genome. *Plant Molecular Biology Reporter* 3: 111–147.

SWOFFORD, D. L. 1993. PAUP: Phylogenetic analysis using parsimony, version 3.1. Computer program distributed by the Illinois Natural History Survey, Champaign, IL.

SYTSMA, K. J., AND L. D. GOTTLIEB. 1986. Chloroplast DNA evolution and phylogenetic relationships in *Clarkia* sect. *Peripetasma* (Onagraceae). *Evolution* 40: 1248–1261.

TAKHTAJAN, A. 1980. Outline of the classification of flowering plants (Magnoliophyta). *Botanical Review (Lancaster)* 46: 225–359.

TOWNSEND, C. C. 1993. Amaranthaceae. *In* K. Kubitzki, J. G. Rohwer, and V. Bittrich [eds.], The families and genera of flowering plants, 70–91. Springer-Verlag, Berlin.

ULBRICH, E. 1934. Chenopodiaceae. *In* A. Engler and H. Harms [eds.], Die natürlichen Pflanzenfamilien, 2d. ed., vol. 16C: 375–584. Wilhelm Engelmann, Leipzig.

WAKASUGI, J. TSUDZUKI, S. ITO, M. SHIBATA, AND M. SUGIURA. 1994. A physical map and clone bank of the black pine (*Pinus thunbergii*) chloroplast genome. *Plant Molecular Biology Reporter* 12: 227–241.

WINSHIP, P. R. 1989. An improved method for directly sequencing PCR amplified material using dimethyl sulphoxide. *Nucleic Acids Research* 17: 1266.

WOLFE, K. H. 1994. Similarity of plastid ORF2280 to the CDC48 family: a proteolytic ATPase? *Current Genetics* 25: 379–383.

———, W.-H. LI, AND P. M. SHARP. 1987. Rates of nucleotide substitution vary greatly among mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences, USA* 84: 9054–9058.

———, C. W. MORDEN, AND J. D. PALMER. 1992. Function and evolution of a minimal plastid genome from a nonphotosynthetic parasitic plant. *Proceedings of the National Academy of Sciences, USA* 89: 10648–10652.

———, AND P. M. SHARP. 1988. Identification of functional open reading frames in chloroplast DNA. *Gene* 66: 215–222.

ZHOU, D.-X., O. MASSENET, F. QUIGLEY, M. J. MARION, F. MONÉGER, P. HUBER, AND R. MACHE. 1988. Characterization of a large inversion in the spinach chloroplast genome relative to *Marchantia*: a possible transposon-mediated origin. *Current Genetics* 13: 433–439.