INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.
Phylogenetic and molecular evolutionary studies of chloroplast DNA variation in the Campanulaceae

Cosner, Mary Elizabeth, Ph.D.
The Ohio State University, 1993
PHYLOGENETIC AND MOLECULAR EVOLUTIONARY
STUDIES OF CHLOROPLAST DNA VARIATION
IN THE CAMPANULACEAE

DISSERVATION

Presented in Partial Fulfillment of the Requirements
for the Degree Doctor of Philosophy in the
Graduate School of The Ohio State University

By

Mary Elizabeth Cosner, B.G.S., B.S., M.S.

The Ohio State University
1993

Dissertation Committee:  
D.J. Crawford
M.L. Evans
R.K. Jansen
A.A. Snow

Approved by

Adviser
Department of Plant Biology
ACKNOWLEDGMENTS

I thank Bob Jansen for encouragement, support, guidance, patience, inspiration, and friendship. He encouraged me to persevere and keep a positive attitude when the project seemed insurmountable. He was always available to help with problems and was willing to spend as much time as necessary with me and others; his interest in and enthusiasm for my research was an inspiration. I especially appreciate his fairness, sense of humor, and straightforward approach. I was fortunate to have Bob for my adviser, but I am even more fortunate to have him for a friend. I thank Loni Walker and Cathy Cosner for continued patience, love, and support. Thanks also to Daniel Crawford.
VITA

Mary Elizabeth Cosner
Department of Plant Biology
The Ohio State University
1735 Neil Avenue, Columbus, Ohio 43210

1981......................B.S., General Studies,
              Capital University,
              Columbus, Ohio Summa Cum Laude.

1983......................B.S., Horticulture,
              The Ohio State University,
              Columbus, Ohio Summa Cum Laude.

1986-1988................Graduate Teaching Associate,
              The Ohio State University,
              Columbus, Ohio

1988......................M.S., Botany,
              The Ohio State University,
              Columbus, Ohio

1988-1989................Graduate Teaching Associate,
              University of Connecticut,
              Storrs, Connecticut

1989-1992................Graduate Research Associate,
              The Ohio State University,
              Columbus, Ohio

1992......................Graduate Teaching Associate,
              The Ohio State University,
              Columbus, Ohio

1993......................Graduate Fellow,
              The Ohio State University,
              Columbus, Ohio
PUBLICATIONS


FIELDS OF STUDY

Botany: Plant Molecular Systematics and Evolution
<table>
<thead>
<tr>
<th>TABLE OF CONTENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGMENTS ................................................................. ii</td>
</tr>
<tr>
<td>VITA ...................................................................................... iii</td>
</tr>
<tr>
<td>LIST OF TABLES ........................................................................ viii</td>
</tr>
<tr>
<td>LIST OF FIGURES ......................................................................... x</td>
</tr>
<tr>
<td>CHAPTER PAGE</td>
</tr>
<tr>
<td>I. PHYLOGENETIC RELATIONSHIPS IN THE CAMANULALES</td>
</tr>
<tr>
<td>BASED ON rbcL SEQUENCES ......................................................... 1</td>
</tr>
<tr>
<td>Introduction ............................................................................. 1</td>
</tr>
<tr>
<td>Materials and Methods ................................................................. 4</td>
</tr>
<tr>
<td>Results ..................................................................................... 8</td>
</tr>
<tr>
<td>Discussion ............................................................................... 11</td>
</tr>
<tr>
<td>Literature Cited ......................................................................... 26</td>
</tr>
<tr>
<td>II. THE HIGHLY REARRANGED CHLOROPLAST GENOME OF</td>
</tr>
<tr>
<td>TRACHELIUM CAERULEUM: MULTIPLE INVERSIONS, INVERTED</td>
</tr>
<tr>
<td>REPEAT EXPANSION AND CONTRACTION, TRANSPOSITION,</td>
</tr>
<tr>
<td>AND SEVERAL REPEAT FAMILIES .................................................. 40</td>
</tr>
<tr>
<td>Introduction ............................................................................. 40</td>
</tr>
<tr>
<td>Materials and Methods ................................................................. 42</td>
</tr>
<tr>
<td>Results ..................................................................................... 44</td>
</tr>
<tr>
<td>Discussion ............................................................................... 51</td>
</tr>
<tr>
<td>Literature Cited ......................................................................... 70</td>
</tr>
<tr>
<td>III. CHLOROPLAST DNA STRUCTURAL EVOLUTION IN THE</td>
</tr>
<tr>
<td>CAMANULACEAE ........................................................................... 93</td>
</tr>
<tr>
<td>Introduction ............................................................................. 93</td>
</tr>
<tr>
<td>Materials and Methods ................................................................. 97</td>
</tr>
<tr>
<td>Results ..................................................................................... 100</td>
</tr>
<tr>
<td>Discussion ............................................................................... 109</td>
</tr>
<tr>
<td>Literature Cited ......................................................................... 127</td>
</tr>
</tbody>
</table>
IV. PHYLOGENETIC IMPLICATIONS OF COMPLEX CHLOROPLAST
GENOME STRUCTURE IN THE CAMPAULACEAE..............157

Introduction...........................................157
Materials and Methods............................163
Results................................................166
Discussion..........................................168
Literature Cited....................................185

LITERATURE CITED...........................................218
# LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLES</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Treatments of families in the Campanulales/Asterales complex according to four major systems of classification</td>
<td>32</td>
</tr>
<tr>
<td>2. Taxa for which \textit{rbcL} sequences were obtained in this study. Species that were sequenced from herbarium material are denoted by an asterisk (*)</td>
<td>33</td>
</tr>
<tr>
<td>3. Sizes of HindIII restriction fragments of \textit{Trachelium caeruleum} cpDNA larger than 0.5 kb cloned into pUC19. Fragment numbers are the same as in Fig. 4 and approximate sizes are in kb. Chloroplast DNA HindIII fragments of 12.5, 8.3, 4.4, 3.8, and 1.5 kb were not cloned</td>
<td>78</td>
</tr>
<tr>
<td>4. Species mapped for chloroplast DNA structural rearrangements</td>
<td>133</td>
</tr>
<tr>
<td>5. Size variation in Campanulaceae chloroplast DNA including genome sizes (in kb), size and extent of IR, and insertions larger than about 5 kb. Sizes of insertions are followed by their general locations in the genome</td>
<td>134</td>
</tr>
<tr>
<td>6. Select classification systems of Campanulaceae. Major intrafamilial subdivisions indicated but only genera (Table 7) sampled for this study included. Genera in Table 7 that are sunk in other taxa in these systems are not listed</td>
<td>192</td>
</tr>
<tr>
<td>7. Species in Campanulaceae mapped for chloroplast DNA structural rearrangements</td>
<td>197</td>
</tr>
<tr>
<td>8. Chloroplast DNA rearrangement characters in the Campanulaceae. Numbers refer to tobacco cpDNA hybridization probes. T=transposition (T'=second transposition (Chapter 3, Fig. 12) of most of</td>
<td></td>
</tr>
</tbody>
</table>
53-56), I=inversion, i=insertion (followed by approximate size in kb), D=deletion/divergence, IRc and IRe=IR contraction or expansion, respectively (followed by single copy region at which event occurred). All other characters are inversion endpoints.

9. Data matrix of 84 cpDNA rearrangement characters*.
0/1 indicates endpoints scored 0 (absent) in analyses one and two, and scored 1 (present) in analyses three and four (see text for details).
?=missing data.

ix
LIST OF FIGURES

FIGURES PAGE

1. Strict consensus tree of 32 equally parsimonious trees of 4003 steps (consistency index=0.28) generated using rbcL sequence data from 117 taxa. The parsimony analysis employed the TBR, ACCTRAN, and MULPARS options of PAUP version 3.1.1 (Swofford 1993). Boxed genera indicate sequences obtained for this study. Labelling of groups follows Olmstead & al. (1993) ...............................................34

2. One of four equally parsimonious trees of 1354 steps (consistency index=0.49) based on rbcL sequences of 44 taxa. Numbers indicate changes at each node and along each branch. Brackets indicate familial/ordinal/group circumscription..36

3. Detail of Campanulales/Asterales clade based on rbcL analysis of 44 taxa (topology of Campanulales/Asterales was identical in all four equally parsimonious trees). Numbers indicate changes at each node and along each branch, numbers in parentheses are bootstrap values, and d refers to decay indices.................................38

4. Clone bank of Trachelium caeruleum cpDNA. HindIII clones were digested with HindIII and separated on a 1.0% agarose gel. The arrows indicate the migration of the pUC19 plasmid vector. Clone numbers correspond to those in Table 3. The larger of two fragments below the vector band in lane 21 is a non-cpDNA fragment cloned together with the cpDNA fragment.........................79

5. Physical and gene map of Trachelium caeruleum cpDNA. Complete circles from outside to inside show restriction sites and approximate restriction fragment sizes for HindIII, EcoRV, XbaI, BamHI, and SstI in that order. Enzymes mapped but not
pictured include BglII and EcoRI; double digests using HindIII and the remaining 6 enzymes not shown. The two black bars represent the minimum extent of the inverted repeat and the open extensions of these bars its maximum possible extent. Locations, orientations and approximate lengths of genes are shown by arrows labelled with gene names; sizes of ORFs are those of tobacco. Outermost arrows labelled A-O indicate blocks of genes and their orientations relative to tobacco. Numbered lines indicate the positions of elements belonging to 5 families of dispersed repeats (r). Repeat 5 is a small segment of the 23S rRNA gene located in both the 23S rDNA and ORF1901. Brackets indicate presumed insertions not homologous to tobacco cpDNA; numbered brackets designate regions that may contain pseudogenes of clpP and 5'-rps12 (1), accD (2), and part of ORF1901 (3). The small arrow shows the location of a 213 bp tobacco cpDNA probe that does not hybridize to Trachelium DNA.

6. Model for the evolution of Trachelium caeruleum cpDNA from a tobacco-like ancestor. The model includes 7 inversions, 1 transposition, expansion of the IR into the SSC region, and contraction of the IR at the LSC end. The top gene map and first lettered map show tobacco, and the bottom gene map and last lettered map show Trachelium. Only genes at rearrangement endpoints are given. Arrows on top gene map indicate rearrangement endpoints; those with asterisks designate the transposition. Letters correspond to blocks of genes and are the same as in Fig. 5; arrows under letters indicate orientations of gene blocks. The bars indicate location and extent of the IR. Numbered arrows designate inversion endpoints and the following map gives the resulting inverted genome structure. "T" stands for transposition. Insertions/deletions in Trachelium relative to tobacco are not shown. The ordering of events shown is arbitrary except as explained in text.

7. Model for the evolution of Trachelium caeruleum cpDNA from a tobacco-like ancestor. The model includes 10 inversions and loss and regrowth of the IR. Designations are the same as Fig. 6. Inversion numbers are identical to those in Fig. 6.
with the additions of inversions a, b, and c. Rearrangement endpoints are the same as Fig. 6, but asterisked arrows in that figure indicate inversion endpoints for this model. The sequence of events shown is arbitrary except as explained in text

8. Model for the evolution of *Trachelium caeruleum* cpDNA from a tobacco-like ancestor. The model includes 9 inversions, expansion of the IR into the SSC region, and contraction of the IR at the LSC end. Designations are the same as Fig. 6. Inversion numbers are identical to those in Fig. 6 with the additions of inversions d and e. Inversions d and e must occur in the IR copy shown. Open bars on IR indicate maximum possible extent of IR following inversion d. Rearrangement endpoints are the same as Fig. 6, but asterisked arrows in that figure indicate inversion endpoints for this model. The sequence of events shown is arbitrary except as explained in text

9. Model for RNA-mediated transposition of 5'-*rps12* and *clpP* from the LSC region to the IR, and duplicative transposition of a small segment of 23S rDNA, in *Trachelium caeruleum* cpDNA. Open boxes indicate presumably functional genes, shaded boxes designate putative pseudogenes, and the black box represents a partial copy of the 23S rRNA gene. Dashed line indicates the DNA segment separating the regions pictured. Triangles represent the transposition events (in the case of 5'-*rps12* and *clpP* this involves insertion of a cDNA made by reverse transcription of a chloroplast transcript of these genes). The model can be modified to represent a DNA-mediated transposition by replacing the 5'-*rps12* and *clpP* pseudogenes with an insertion of unknown content and origin

10. Short dispersed repeats in the chloroplast genome of *Trachelium caeruleum*. Filters containing *Trachelium* total DNA digested with HindIII and electrophoresed on a 1.2% agarose gel were hybridized with each of five *Trachelium* cpDNA HindIII clones. Sizes in kb of the HindIII inserts are given above each lane. The outside lanes contain lambda DNA molecular weight standards with fragment sizes in kb indicated. The 1.2 kb and 4.5
kb fragments contain the 23S rDNA and its duplicated segment, respectively

11. Chloroplast DNA gene map of Nicotiana tabacum (tobacco). The circular chromosome has been linearized and separated onto four lines (A-D). Genes above the line are transcribed from left to right and those below the line are transcribed from right to left. Numbers in blocks below genes represent 106 tobacco cpDNA hybridization probes used in this study. Numbers above probes designate tobacco coordinates in kb.

12. Chloroplast DNA restriction site maps of five species (Table 4) in the Campanulaceae for three enzymes (S=SstI, H=HindIII, EV=EcoRV) illustrating different extents of the IR and the results of three transpositions involving clpP and 5'-rps12 (tobacco probes 53-56). Numbered blocks indicate approximate sizes (in kb) and locations of restriction fragments; the sizes of unnumbered fragments could not be directly determined due to insertions. Maps for BamHI, BglII, EcoRI and double digests available but not shown. Dark bars indicate approximate location and extent of the IR. Numbers below map show hybridization of tobacco cpDNA probes (see Fig. 11); approximate extents of contiguous blocks of tobacco fragments are bounded by parentheses. Gene names indicate the approximate location of the gene's midpoint; genes above are transcribed from left to right and those below are transcribed from right to left.

13. Linear cpDNA maps for 18 species (Table 4) in Campanulaceae showing order of hybridized tobacco probes. Lines under maps represent approximate location and extent of IR. Asterisks indicate 23S rDNA duplicative transposition; (*) is partial deletion/divergence of 23S rDNA transposition. Size and location of large insertions designated by "i" followed by size in kb (insertions less than 5 kb not shown). Underlined numbers indicate sequence absent from an otherwise intact block of probes. Ambiguous hybridization of tobacco probes not shown (see text).

14. Chloroplast DNA restriction site maps of portions of Asyneuma and Musschia showing large insertions.
Numbers above maps indicate tobacco cpDNA probes. Enzymes: S=SstI, EV=EcoRV, H=HindIII, EI=EcoRI, Bg=BglII, Ba=BamHI.

15. Strict consensus tree of two equally parsimonious trees based on 84 cpDNA structural characters. This tree has 88 steps and a consistency index of 0.90 (excluding autapomorphies) I=inversion, i=insertion >5 kb, T=transposition, D=deletion/divergence. Insertions in the region of tobacco's large ORF1901 not shown.

16. Model for the evolution of Trachelium caeruleum cpDNA from a tobacco-like ancestor (Chapter 2). The model includes 7 inversions, 1 transposition, expansion of the IR into the SSC region, and contraction of the IR at the LSC end. The top gene map and first lettered map show tobacco, and the bottom gene map, last lettered map, and ordering of tobacco probe numbers in parentheses (not to scale) show Trachelium. Only genes at rearrangement endpoints are given. Arrows on top gene map indicate rearrangement endpoints; those with asterisks designate the transposition. Letters correspond to blocks of genes; arrows under letters indicate orientations of gene blocks. The bars indicate location and extent of the IR. Numbered arrows designate inversion endpoints and the following map gives the resulting inverted genome structure. "T" stands for transposition. Insertions/deletions in Trachelium relative to tobacco are not shown.

17. Model for evolution of cpDNA structure in Wahlenbergia, Merciera, Prismatocarpus, and Roella from a Trachelium-like ancestor, showing only inversions. Numbers in parentheses show order of hybridized tobacco cpDNA probes. Inversion endpoints are given as numbered arrows. The final map shows only features shared by the three species; autapomorphies of individual taxa not pictured.

18. Model for evolution of cpDNA structure in Asyneuma, Legousia, Petromarula, and Triodanus from a Trachelium-like ancestor. Numbers in parentheses show order of hybridized tobacco cpDNA probes. Arrows indicate transfer of a 25 kb region from the
LSC region to the SSC region (mechanism unknown). Final map shows only features shared by the four species; autapomorphies of individual taxa not pictured..........................151

19. Models for evolution of cpDNA structure in A. Edraianthus and Symphyandra, B. Campanula and Adenophora, C. Jasione, from a Trachelium-like ancestor, showing only inversions. Numbers in parentheses show order of hybridized tobacco cpDNA probes. Inversion endpoints are given as numbered arrows. Inversions not shown for Jasione.......153

20. Model for evolution of cpDNA structure in Cyananthus, Codonopsis, and Platycodon from a tobacco-like ancestor, showing only inversions. Numbers in parentheses show order of hybridized tobacco cpDNA probes. Inversion endpoints are given as numbered arrows. First two inversions shared by all three species, first five shared by Codonopsis and Cyananthus. Remaining five inversions in Platycodon not pictured..........155

21. Linear cpDNA maps for 18 species (Table 7) in Campanulaceae showing order of hybridized tobacco probes. Lines under maps indicate location and extent of IR. Asterisks indicate 23S rDNA duplicative transposition; (*) is partial deletion/divergence of 23S rDNA transposition. Size and location of large insertions designated by "i" followed by size in kb (insertions less than 5 kb not shown). Underlined numbers indicate sequence absent from an otherwise intact block of probes...............................20?

22. Strict consensus tree of 241 equally parsimonious trees of 94 steps generated using 84 cpDNA rearrangement characters from 18 taxa in the Campanulaceae. CI=0.76 (excluding autapomorphies). Numbers indicate changes at each node and along each branch; numbers below are bootstrap values........................................206

23. Strict consensus tree of 12 equally parsimonious trees of 125 steps (CI=0.78) based on 84 cpDNA rearrangement characters from 18 taxa in the Campanulaceae. The analysis gave a weight of two to non-inversion characters. Numbers indicate changes

xv
at each node and along each branch; numbers below are bootstrap values.

24. Two equally parsimonious trees based on 84 cpDNA rearrangement characters in the Campanulaceae. The identical two trees were found in two separate cladistic parsimony analyses. The first involved recoding certain inversion endpoint characters, resulting in trees of 88 steps (CI=0.90). The second used both recoding of endpoints and weighting non-inversion characters, resulting in trees of 119 steps (CI=0.91). See text for details. Numbers indicate changes at each node and along each branch (for analysis that recoded but did not weight certain characters); numbers below are bootstrap values (identical in both analyses).

25. Two models of seven steps each for the evolution of cpDNA structure in Platycodon from a tobacco-like ancestor, showing only inversions. Numbers in parentheses show order of hybridized tobacco cpDNA probes. Inversion endpoints are given as numbered arrows. Locations of probes 6-7, 8, and 9 believed to be the result of transposition (Chapter 3).

26. Strict consensus tree of two equally parsimonious trees (Fig. 24) showing evolution of structural rearrangements in the Campanulaceae. I=inversion, i=insertion >5 kb, T=transposition, D=deletion/divergence. Insertions in the region of tobacco's large ORF1901 not shown. Letters following taxon names indicate Campanulinae (C), Platycodinae (P), and Wahlenberginae (W) of Schönland (1889). The genera Triodanus and Leqousia were treated as Specularia, and Asyneuma and Petromarula as Phyteuma by Schönland.

27. Model for evolution of cpDNA structure in Wahlenbergia, Merciera, Prismatocarpus, and Roella from a Trachelium-like ancestor, showing only inversions. Numbers in parentheses show order of hybridized tobacco cpDNA probes. Inversion endpoints are given as numbered arrows. The final map shows only features shared by the three species; autapomorphies of individual taxa not pictured.

xvi
CHAPTER I

PHYLOGENETIC RELATIONSHIPS IN THE CAMPANULALES
BASED ON rbcL SEQUENCES

Introduction

The Campanulales are an angiosperm order of varying and controversial circumscription (Lammers 1992). The most inclusive delimitation of the order would include 12 taxa: (1) a core group of five taxa (Campanulaceae, Cyphiaceae, Lobeliaceae, Pentaphragmataceae, and Sphenocleaceae) that have been included by essentially all authors; (2) six taxa (Asteraceae, Brunoniaceae, Calyceraceae, Donatiaceae, Goodeniaceae, and Stylidiaceae) that have been included in the order by some authors but removed by others; and (3) Menyanthaceae, traditionally part of Gentianales but recently assigned to Campanulales on the basis of chloroplast DNA (cpDNA) data (see below).

The major classification systems currently in use differ considerably in the classification of these taxa (Table 1); see Lammers (1992) for a more complete review. Although different in the ranking of taxa, both the
systems of Takhtajan (1987) and Thorne (1992, and unpublished update) recognize an association among all taxa by uniting them as members of Asteridae and Asteranae, respectively. In contrast, Cronquist (1988) discounts any relationship between Campanulales and either Calyceraceae or Asteraceae. Calycerales and Asterales are allied with Dipsacales and Rubiales, respectively, whereas Campanulales are thought to be derived from or near Solanales. Likewise, Dahlgren (1980, 1983; Dahlgren & al. 1981; updated in G. Dahlgren, 1989a,b) dissociates many of the orders, placing Goodeniales and Dipsacales (Calyceraceae) in the Gentiananae, Campanulales and Asterales in the Asteranae, and Stylidiales in the Ericanae. It is clear that there is considerable disagreement surrounding the Campanulales/Asterales complex regarding placement of taxa, assignment of taxonomic rank, and affinities with other orders.

Recent molecular studies have supplied a new set of data with which to assess evolutionary relationships. Chloroplast DNA studies, including restriction site comparisons, rearrangement analysis, and sequencing of the *rbcL* gene have all contributed significantly to our understanding of relationships within the Asteridae and other plant groups. *rbcL* data have indicated that Calyceraceae and Goodeniaceae collectively comprise a
clade that is the sister group to the Asteraceae, providing new evidence for resolving the debate surrounding the question of the nearest extant relative of this large family (Michaels & al. 1993; Olmstead & al. 1992, 1993). A surprising result of rbcL analyses is that Menyanthaceae, traditionally assigned to Gentianales, are the sister group to the clade consisting of Asteraceae, Calyceraceae, and Goodeniaceae (Michaels & al. 1993; Olmstead & al. 1992, 1993), a position also supported by restriction site comparisons of the cpDNA inverted repeat (Downie & Palmer 1992). Campanulaceae and Lobeliaceae, termed the "campanulad" clade by Lammers (1992), form the sister group to the "asterads" (Asteraceae, Calyceraceae, Menyanthaceae, and Goodeniaceae). The sister group of this Campanulales/Asterales complex is a clade consisting of the "dipsacads" and "apiads" (see Lammers 1992). These overall relationships are also supported by restriction site analysis of the cpDNA inverted repeat (Downie & Palmer 1992). In addition, rbcL data place Corokia (Cornaceae) with the asterads near Menyanthaceae (Chase & al. 1993; Michaels & al. 1993; Olmstead & al. 1993). Although this relationship was unexpected, Corokia's placement in the Cornaceae has been widely debated. For example, Eyde (1966, 1988) suggested the genus is misplaced in Cornaceae, Takhtajan (1987) placed Corokia
and Argophyllum in the Argophyllaceae, and Patel (1973) observed that Corokia is similar to Escalloniaceae in wood anatomy. As indicated by Michaels & al. (1993), further studies of Corokia, such as floral development, pollen ultrastructure, and embryology, are needed to more completely understand this relationship.

Lammers (1992) reexamined circumscription and classification of the Campanulales/Asterales complex through the integration of a diverse array of data. As he pointed out in his synthesis, severe gaps exist in several of the data sets, with certain taxa being widely studied, while others are poorly understood. The purpose of the present study was to sequence the chloroplast gene rbcL from members of those families of the Campanulales/Asterales complex that had not been previously examined. These data, combined with rbcL sequences from other laboratories, were then used to assess phylogeny of the group.

Materials and Methods

The rbcL gene was sequenced from ten species in six families of the order Campanulales (Table 2). Only the monogeneric families Donatiaceae and Brunoniaceae were not represented; efforts to employ the polymerase chain
reaction (PCR) to amplify the \textit{rbcL} gene from several different herbarium specimens of \textit{Donatia} and \textit{Brunonia} failed. Sources of DNA included fresh material, herbarium specimens, or plants grown from seeds (Table 2). Total DNA was extracted from fresh material according to the CTAB method of Doyle & Doyle (1987), and purified on CsCl/ethidium bromide gradients. A small-scale CTAB extraction procedure was employed to obtain total DNA from herbarium material.

Two synthetic oligonucleotides were used as PCR amplification primers to generate a double-stranded DNA fragment of approximately 1550 base pairs (bp) containing the entire coding sequence for the chloroplast gene \textit{rbcL}. The 5' primer is based on the first 26 nucleotides of the gene and is degenerate for A/G at position 18 to account for the single sequence difference between tobacco and maize in this region. The 3' primer is based on a 24 bp sequence approximately 100 bp downstream from the gene's termination codon.

One hundred ul PCR reactions consisted of 0.2 mM dNTPs, 1 unit of Tfl polymerase (Epicentre Technologies), 1x Tfl polymerase buffer (includes 1.5 mM MgCl), and a one ul aliquot of unquantified DNA. The thermocycler was programmed for one initial denaturation cycle consisting of three minutes denaturation at 95C, one minute primer
annealing at 55-60°C, and one minute extension at 72°C, followed by 30 cycles of one minute denaturation at 95°C, one minute annealing at 55-60°C, and one minute extension at 72°C. A final extension period of five minutes at 72°C terminated the PCR reactions. "Hot-start" PCR was utilized for certain DNAs isolated from herbarium specimens, for which the initial denaturation cycle consisted of three minutes denaturation at 97°C, one minute annealing at 55-60°C, and one minute extension at 72°C. The Tfl polymerase was added after the completion of the three minute 97°C denaturation.

The PCR products were electrophoresed in standard or low-melting temperature agarose, excised from the gel, and purified using GeneClean (Biol01). Double-stranded PCR products were directly sequenced using the Sequenase Version 2.0 (USB) dideoxy chain-termination method, employing seven forward and 4 reverse rbcL oligonucleotide primers. Two of these were identical to the 5' and 3' PCR primers; the remaining internal primers were kindly supplied by G. Zurawski. Modifications to the Sequenase protocol include denaturation of the double-stranded DNA by boiling the DNA/primer mix for three minutes, followed by snap-chilling the annealing mixture for 5-10 minutes in an ice water bath (Winship 1989). In addition, one ul DMSO was added to both the labelling and termination
reactions to reduce the effects of DNA secondary structure. Sequencing products were electrophoresed in 6% polyacrylamide gels; gels were fixed in 10% acetic acid, dried, and exposed to x-ray film.

A total of 1428 bp was compared comprising the entire coding region of most species sequenced for this study; sequences were aligned manually. *rbcL* sequences obtained from the Campanulales were compared with previously published *rbcL* sequences from members of the Asteridae and outgroups (Olmstead & al. 1992, 1993) for a total of 117 taxa analyzed. Parsimony analyses were performed using PAUP version 3.1.1 (Swofford 1993) on a Macintosh Quadra 700 with the Tree Bisection Reconnection (TBR), ACCTRAN, and MULPARS options. Trees were rooted using the outgroup taxa indicated in Fig. 1. In addition, a topological constraints analysis was performed in which the monophyly of *Sphenoclea* and Campanulaceae, Lobeliaceae, Cyphiaceae, and Stylidiaceae was forced.

A subset of the 117 taxa was then chosen in order to more rigorously test relationships in the Campanulales/Asterales complex. The subset consisted of 44 taxa, including all members of the Dipsacales, Apiales, Asterales, and Campanulales analyzed within the larger data set, as well as three species from the Gentianales and the Ilex clade, and seven rosids used as outgroup
taxa. The same heuristic search methods were employed except that one hundred random entries of data were performed in an attempt to find all equally parsimonious trees (Maddison 1991). The amount of support for monophyletic groups was evaluated using 100 bootstrap replicates (Felsenstein 1985) and a decay analysis (Donoghue & al. 1992; Hillis & Dixon 1989). Trees more than three steps longer than the most parsimonious trees could not be examined because the large number of trees generated exceeded the memory capacity of our computer. Four separate analyses were performed in which topological constraints were imposed that forced the monophyly of 1) Cyphiaceae; 2) Campanulaceae and Lobeliaceae; 3) Campanulaceae, Lobeliaceae, and Pentaphragmataceae; and 4) Cyphiaceae, Campanulaceae, Lobeliaceae, Pentaphragmataceae, and Stylidiaceae. These were done in order to determine how many additional steps are necessary to create these monophyletic groups.

Results

The coding region of rbcL is 1428 bp for all taxa sequenced for this study except for that of Sphenoclezeyslanica, which is three codons longer.
The parsimony analysis of the larger data set consisting of 117 taxa produced 32 equally parsimonious trees of 4003 steps; the strict consensus tree is shown in Fig. 1. The most surprising result is the placement of Sphenoclea in the Solanales with Hydrolea and Montinia, in sharp contrast to its typical classification within the Campanulales. The topological constraints analysis forcing the monophyly of Sphenoclea and the Campanulaceae, Cyphiaceae, Lobeliaceae, and Stylidiaceae resulted in trees with 4042 steps, 39 steps longer than the shortest trees. This topological constraints analysis was not run to completion because the large number of trees generated exceeded the memory of the computer.

More rigorous analyses (i.e., 100 random entries of data, bootstrap and decay analyses) of the smaller data set of 44 taxa produced four equally parsimonious trees of 1354 steps with a consistency index of 0.49. Of 1428 positions, 488 were variable, and 303 were phylogenetically informative. The topology of the Campanulales/Asterales complex is identical in the four trees; all branch collapsing in the strict consensus tree occurs outside of the complex. Therefore, only one of the four equally parsimonious trees is shown in Fig. 2. For convenience, the two major asterad/campanulad clades in the rbcL phylogeny will be referred to here as Asterales
(Asteraceae, Calyceraceae, Goodeniaceae, Menyanthaceae, Pentaphragmataceae, Corokia) and Campanulales (Campanulaceae, Cyphiaceae, Lobeliaceae, Stylidiaceae).

The reduced rbcL tree (Fig. 3) has several phylogenetic implications. (1) The Campanulaceae/Cyphiaceae/Lobeliaceae clade is very strongly supported as indicated by the high bootstrap (100%) and decay index (d>3) values. (2) The Cyphiaceae are paraphyletic. Ten additional steps, for a total of 1364, were needed for the Cyphiaceae to be monophyletic. (3) The campanulads and lobeliads are not sister taxa. It required three additional steps to force the monophyly of Campanulaceae and Lobeliaceae. (4) There is moderate support for the inclusion of Stylidiaceae within the Campanulales (bootstrap value=72; decay index=3). (5) Pentaphragma is more closely allied with the Asterales than the Campanulales. Only two additional steps (total 1356) were required to place Pentaphragma within the Campanulales, but 36 more steps were necessary to force the monophyly of Campanulaceae, Lobeliaceae, and Pentaphragmataceae.
Discussion

The *rbcL* phylogeny provides new insights into relationships within the Campanulales/Asterales complex. The implications of the data on the classification and phylogeny of the six families sequenced here are discussed below.

**Campanulaceae/Lobeliaceae.** The campanulads and lobeliads have long been recognized as closely related taxa that consistently appear together in classification systems. Most disagreement concerns the assignment of taxonomic rank, in particular whether to recognize the taxa as two subfamilies of Campanulaceae, or as the separate families Campanulaceae and Lobeliaceae. Among those preferring to recognize two families include Dahlgren (1980, 1983), deCandolle (1830), Federov (1972), Kovanda (1978), Lammers (1992), and Takhtajan (1987), whereas Cronquist (1988), Schönland (1889), Thorne (1992), and Wagenitz (1964) relegated the two taxa to subfamilies Campanuloideae and Lobeloideae. Cronquist (1988) admitted that treating the two as families vs. subfamilies is "purely a matter of taste". For a detailed review of taxonomic and phylogenetic considerations regarding Campanulaceae and Lobeliaceae, see Lammers (1992).
The *rbcL* phylogeny (Fig. 2) indicates that Campanulaceae and Lobeliaceae are not a monophyletic group. Two groups can be distinguished within the Campanulaceae, the *Codonopsis/Cyananthus* clade and the *Trachelium/Campanula* clade, each strongly supported in 100 bootstrap replicates (Fig. 3). Several authors have suggested that *Codonopsis* and *Cyananthus* are primitive members in Campanulaceae (Dunbar 1975a,b; Hong & Ma 1991; Lammers, 1992; Takhtajan 1987; Thulin 1975). Although only four campanulad genera are represented, the *rbcL* phylogeny is in agreement that *Codonopsis/Cyananthus* and *Trachelium/Campanula* are separate groups. *Nemacladus* (Cyphiaceae) is the sister taxon to Campanulaceae, and the *Nemacladus/Campanulaceae* clade forms the sister group to the Lobeliaceae.

The monophyly of the Campanulaceae is not well supported in the bootstrap or decay analyses (Fig. 3), apparently due to the family's close relationship with *Nemacladus*. In addition, the *Nemacladus/Campanulaceae/Lobeliaceae* clade is not particularly well supported, occurring in only 55 of 100 bootstrap replicates and having a decay index of 2 (Fig. 3).
Cyphiaceae. The Cyphiaceae are a heterogeneous family consisting of five genera. Nemacladus, Parishella, and Pseudonemacladus occur in western North America, Cyphocarpus is a small Chilean genus, and Cyphia consists of about 60 species in southern and tropical Africa. Although there has never been any doubt that the family belongs to the Campanulales, generic affinities have been debated. The taxa are frequently given an intermediate position between Campanulaceae and Lobeliaceae (Cronquist 1981; Dunbar 1975b; Dunbar & Wallentinus 1976; Kovanda 1978). The cyphiad genera share several features with both families, including the zygomorphic corolla of Lobeliaceae and the free anthers of Campanulaceae. Many authors have suggested that the family as circumscribed does not form a natural group (Bentham 1875; Dunbar 1975b, 1984; Dunbar & Wallentinus 1976; Lammers 1992; Schönland 1889; Takhtajan 1987).

The genera have been variously classified by different authors. For example, Thorne (1992) placed Cyphia in Cyphioideae and Cyphocarpus in Lobelioideae, two different subfamilies within Campanulaceae. Hutchinson (1973) placed Cyphocarpus in Campanulaceae and the remaining four genera in Lobeliaceae. Schönland (1889), Wimmer (1968) and others placed all five genera in Campanulaceae subfamily Cyphioideae, while several authors
(eg. deCandolle 1830; Federov 1972; Kovanda 1978; Lammers 1992) have elected to recognize the group as the Cyphiaceae. Takhtajan (1987) divided the taxa into the three families Cyphiaceae (Cyphia), Cyphocarpaceae (Cyphocarpus), and Nemacladaceae (Nemacladus, Parishella, and Pseudonemacladus). Lammers (1992) essentially agreed with Takhtajan that, on the basis of morphology, palynology, and geography, the Cyphiaceae fall into three apparently natural groups (North American, Chilean, and African) that may best be delimited as separate families. He chose, however, to recognize a single family because very little is known about the group as a whole, particularly regarding chemistry and embryology.

The rbcL phylogeny supports the notion that the Cyphiaceae consist of three different but closely related clades (Fig. 2). The family is clearly not monophyletic in the rbcL tree, but the monophyly of Campanulaceae, Cyphiaceae, and Lobeliaceae is strongly indicated, with this group appearing in all 100 bootstrap replicates and having a decay index greater than 3 (Fig. 3). The most obvious non-molecular synapomorphy uniting these three families is the presence of a network of latex-producing articulated laticifers (Lammers 1992).
Stylidiaceae. The Stylidiaceae are a small, primarily Australian family consisting of five genera. The family, whose taxonomic placement has been the subject of considerable disagreement, is the basal member of the Campanulales clade in the rbcL phylogeny (Fig. 2). This relationship is well supported by a bootstrap value of 72 and a decay index of 3 (Fig. 3). The varied classification of the family has included placement not only in or near the Campanulales, but in the Goodeniales (Hutchinson 1969), near the Saxifragaceae (Thorne 1976), in the Cornales (Dahlgren 1980, 1983; Dahlgren & al. 1981), and near the Ericales (G. Dahlgren 1989a).

Although the Stylidiaceae have unusual features that appear aberrant in the Campanulales (for example, anomalous secondary growth; stamens two and adnate to style, forming an irritable gynandrium; anther thecae divergent and apically confluent), they also share with certain taxa in the Campanulales/Asterales complex three characteristics that are of limited distribution among dicots. These include the storage of starch as inulin [ubiquitous in Campanulales/Asterales (Pollard & Amuti 1981)], development of terminal endosperm haustoria [found in Campanulaceae, Lobeliaceae, Pentaphragmataceae (Kapil & Vijayaraghavan 1965; Lammers 1992)], and production of iridoids [produced in Calyceraceae, Goodeniaceae,
Menyanthaceae (Jensen & al. 1975; Kaplan & Gottlieb 1982). All taxa examined in Stylidiaceae produce carbocyclic iridoids, whereas taxa in the Asterales produce simple seco-iridoids (Jensen & al. 1975; Kaplan & Gottlieb 1982). The difference in classes of iridoid compounds between Stylidiaceae and Asterales was one of several factors leading Lammers (1992) to suggest a closer affinity of Stylidiaceae to Ericales, the only dicot order in which terminal endosperm haustoria and inulin are found together with carbocyclic iridoids (Lammers 1992).

In a recent study of floral development, Erbar (1992) argues for an association of Stylidiaceae with the Campanulales. The Stylidiaceae are characterized by early sympetaly, in which a corolla ring primordium is initiated before the petal primordia on its rim, a condition that characterizes the entire Campanulales/Asterales complex. On the other hand, petal primordia are initiated separately and only later become fused (late sympetaly) in the Ericales (Leins 1964). In addition, Erbar (1992) suggests that no morphological, palynological, or embryological feature of the Stylidiaceae contradicts placement of the family in or near Campanulales. The rbcL data support her conclusions.
**Pentaphragmataceae.** The Pentaphragmataceae are a monogeneric family of about 30 species native to eastern Asia, Malaysia, and New Guinea. Most authors have suggested a close relationship between Pentaphragmataceae and the Campanulaceae, and at various times *Pentaphragma* has been included within the Campanulaceae (eg., Hutchinson 1973; Schönland 1889; Takhtajan 1980). Airy Shaw (1942, 1954, 1973) was the only author to suggest other affinities. He allied *Pentaphragma* with the Begoniaceae, based mainly on its asymmetrical leaf base.

In the rbcL phylogeny, *Pentaphragma* occurs as the basal member of the Asterales. This relationship, however, is not strongly supported, occurring in only 35 bootstrap replicates and having a decay index of 2 (Fig. 3). Although its relationship with the Asterales is questionable, there is little doubt that *Pentaphragma* is allied with Campanulales/Asterales but does not belong in the family Campanulaceae. In the topological constraints analysis, 36 additional steps were necessary to force the monophyly of Pentaphragmataceae and the campanulad/lobeliad complex.

*Pentaphragma* is characterized by several features that are absent from the remainder of the Campanulales/Asterales complex, including strongly asymmetrical leaf bases, inflorescence consisting of an
axillary sympodial helicoid cyme, trilobate pollen (Dunbar 1978), and the hypanthium adnate to the ovary by only five narrow longitudinal septa that form intervening nectariferous lacunae.

Pentaphragma appears to be intermediate between the two orders in one embryological feature. Those members examined in Campanulales have terminal endosperm haustoria, taxa in Asterales lack them, and Pentaphragma is characterized by endosperm haustoria at the micropylar end only (Kapil & Vijayaraghavan 1965). Most other features (excluding those unique features listed above) of Pentaphragmataceae are found in certain families in both the Asterales and Campanulales, such as various morphological, palynological, and embryological characteristics.

Sphenocleaceae. Perhaps the most striking feature of the rbcL phylogeny is the inclusion of Sphenoclea in a clade with Hydrolea and Montinia as basal members of the Solanales (Fig. 1). This result was particularly surprising given the consistency with which Sphenoclea has been classified in or near the Campanulaceae (Table 1); only Airy Shaw (1948, 1973) felt the genus was misplaced there, and he allied it with the Phytolaccaceae. Kovanda (1978) mentioned that certain features suggested
affinities with the Lythraceae, although he maintained the classification of *Sphenoclea* within Campanulales.

The Sphenocleaceae are a monogeneric family of one or two species, easily recognized by a densely spicate inflorescence with each sessile flower subtended by a bract and two bracteoles (Rosatti 1986). They are different from Campanulales/Asterales in having imbricate corolla lobes, tetracytic stomates, and a circumscissile capsule (Lammers 1992). Dunbar (1975b) indicated that the smooth pollen of *Sphenoclea* has no relationship with any campanulad genera. The placement of both *Hydrolea* and *Montinia* as basal in the Solanales was also unexpected (see Morgan & Soltis 1993; Olmstead & al. 1993), since *Montinia* is typically classified in the Saxifragaceae (or segregated from Saxifragaceae in the Escalloniaceae, Grossulariaceae, or Montiniaceae), and *Hydrolea* is generally assigned to the Hydrophyllaceae, although a somewhat aberrant member (Constance 1963). The completely bilocular ovary and axile placentation of *Hydrolea* differentiate it from the unilocular ovary and parietal placentation of the remainder of the Hydrophyllaceae, and accounts for its segregation by some authors (Brown 1818; Choisy 1833). Based on wood anatomy, Carlquist (1989) felt *Montinia* is more similar to Myrtales than Rosales (Saxifragales), and also indicated that the presence of
iridoid compounds (Dahlgren & al. 1977) complicates its classification.

Conclusions. Although it must be emphasized that the rbcL phylogeny is a gene tree (Doyle 1992), we believe it is reasonable to assume the phylogeny accurately represents the species tree, particularly at these higher taxonomic levels. Several conclusions may be put forward regarding phylogeny and classification within the Campanulales/Asterales complex based on the rbcL trees.

The rbcL phylogeny indicates that there are two major clades within the Campanulales/Asterales complex that roughly correspond to the traditional circumscription of these two orders (Fig. 2). The Campanulales and Asterales have generally been treated as separate orders, but at times have been merged into a single order (see Lammers 1992; Wagenitz 1964). Although Campanulales/Asterales appears in only 54 bootstrap replicates, it is supported by 12 mutations, compared to its sister group, Apiales/Dipsacales, which is supported by half the number of changes (Fig. 2) and has no support in the bootstrap analysis. Within the Campanulales/Asterales complex, the Campanulales is considerably more strongly supported than is the Asterales (Fig. 3). All of these observations
taken together indicate the delimitation of either one order or two closely related orders.

The rbcL phylogeny also indicates that both Pentaphragma and Sphenoclea, often recognized as members of the family Campanulaceae, should be treated as separate families. The Sphenocleaceae do not belong in either the Campanulales or Asterales, but are more closely allied to Solanales (Fig. 1). Although clearly a member of the Campanulales/Asterales complex, the precise position of the Pentaphragmataceae is uncertain. The rbcL phylogeny only weakly supports a basal placement in Asterales (Fig. 3).

The taxonomic position of Stylidiaceae in Campanulales is supported by rbcL data. It is moderately well-supported as the basal lineage within the order (Fig. 3), and thus it seems unnecessary to place the family in its own order as done by several authors (see above).

The rbcL phylogeny indicates that the Cyphiaceae are paraphyletic (Fig. 2). It appears that the three cyphiad groups may best be treated as families (Cyphiaceae, Cyphocarpaceae, Nemacladaceae) as Takhtajan (1987) suggested. However, additional data regarding the embryology and chemistry of these taxa are sorely needed to further elucidate their positions. Nemacladus is the sister taxon to Campanulaceae in the rbcL tree, but it
seems presumptive to include it within the Campanulaceae, particularly with such substantial gaps in our knowledge regarding this taxon. With the inclusion of *Nemacladus* in the analysis, Campanulaceae and Lobeliaceae do not form a monophyletic assemblage, supporting the recognition of these two clades as separate families.

If the *rbcL* phylogeny is an accurate representation of evolutionary relationships, parallel evolution of many character states has occurred in the Campanulales/Asterales complex, and reliance on a limited number of distinctive morphological features to resolve relationships would be extremely difficult. The characteristic pollen presentation mechanisms of certain groups will serve to illustrate this point (see Lammers (1992) for more detail). In some taxa (Asteraceae, Calyceraceae, Campanulaceae, Lobeliaceae), this syndrome is characterized by protandry accompanied by a close association of anthers around the style and introrse pollen discharge onto the style for presentation to pollinators (with variations according to taxon). A second type of mechanism exists in Goodeniaceae/Brunoniaceae, in which pollen is shed into a cup-like indusium. Certain taxa that apparently lack a specialized pollination presentation syndrome seem to have characteristics approaching the condition. For example,
stamen filaments are connate in some cyphiad genera, and species of *Cyphia* possess a unique uniaperturate fluid-filled stigmatic cavity. Other taxa in the Campanulales/Asterales complex lack any specialization altogether. The manner in which these various pollination mechanisms are related has frustrated taxonomists for many years; the molecular phylogeny suggests the multiple origin of these features. For a discussion of other features (morphological or otherwise) and evolutionary trends in Campanulales/Asterales, see Lammers (1992).

Lammers (1992) enumerated three non-molecular characteristics (two embryological and one chemical) that could be used to separate Campanulales and Asterales. Multinucleate tapetal cells, the absence of endosperm haustoria, and the mevalonate pathway characterize Asterales, while binucleate tapetal cells, terminal endosperm haustoria, and no mevalonate pathway are features of the Campanulales. The inclusion of Pentaphragmataceae in the Asterales (rather than Campanulales) in the rbcL phylogeny necessitates the following refinement of evolutionary hypotheses regarding these three characters. (1) Binucleate tapetal cells are plesiomorphic in the Campanulales/Asterales complex, and multinucleate cells were gained preceding the evolution of Menyanthaceae, Goodeniaceae, Calyceraceae, and
Asteraceae. (2) Endosperm haustoria at the micropylar end constitute the ancestral condition. Micropylar haustoria were lost in the ancestor of the Menyanthaceae, Goodeniaceae, Calyceraceae, and Asteraceae. In the Campanulales, weakly aggressive chalazal haustoria were gained in the basal family (Stylidiaceae), becoming equally aggressive in the remaining families. (3) The plesiomorphic state is a functional mevalonic acid pathway that possibly produced iridoids. Seco-iridoids are found in Asterales, with the loss of iridoids and a shift to sesquiterpene lactones occurring in the Asteraceae. Carbocyclic iridoids are produced in the Stylidiaceae, while the mevalonate pathway is apparently non-functional in the remainder of the Campanulales. With regard to these trends, it is crucial to note that no embryological or chemical data are available for the Cyphiaceae, Corokia has not been studied embryologically, and it is not known if Pentaphragma produces iridoids. These missing data are critical for a more complete understanding of evolution of these characters in the Campanulales/Asterales complex.

Phylogenetic analysis of rbcL sequences has provided new insights into evolutionary relationships of the Campanulales/Asterales complex. Complete resolution of relationships of the group will require additional sampling of rbcL sequences from Brunoniaceae and
Donatiaceae and more critical phylogenetic analysis of non-molecular data.
LITERATURE CITED


Table 1. Treatment of families in the Campanulales/Asterales complex according to four major systems of classification.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Campanulaceae</td>
<td>Campanulales:</td>
<td>Campanulales:</td>
<td>Campanulales:</td>
<td>Campanulales:</td>
</tr>
<tr>
<td>Campanulaceae</td>
<td>Campanulales:</td>
<td>Campanulales:</td>
<td>Campanulaceae</td>
<td>Campanulaceae</td>
</tr>
<tr>
<td>Cyphiaceae</td>
<td>Campanulales:</td>
<td>Campanulales:</td>
<td>Campanulales:</td>
<td>not treated</td>
</tr>
<tr>
<td>Cyphocarpaceae</td>
<td>Campanulales:</td>
<td>Campanulales:</td>
<td>Campanulaceae</td>
<td></td>
</tr>
<tr>
<td>Nemacladaceae</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lobeliaceae</td>
<td>Campanulales:</td>
<td>Lobeliaceae</td>
<td>Campanulales:</td>
<td>Lobeliaceae</td>
</tr>
<tr>
<td>Lobeliaceae</td>
<td>Campanulales:</td>
<td>Lobeliaceae</td>
<td>Campanulaceae</td>
<td></td>
</tr>
<tr>
<td>Pentaphragmataceae</td>
<td>Campanulales:</td>
<td>Pentaphragmatae</td>
<td>Campanulales:</td>
<td>Campanulales:</td>
</tr>
<tr>
<td>Pentaphragmataceae</td>
<td>Campanulales:</td>
<td>Pentaphragmatae</td>
<td>Campanulaceae</td>
<td>Campanulaceae</td>
</tr>
<tr>
<td>Sphenocleaceae</td>
<td>Campanulales:</td>
<td>Sphenocleaceae</td>
<td>Campanulales:</td>
<td>Campanulaceae</td>
</tr>
<tr>
<td>Styliidae</td>
<td>Styliidae:</td>
<td>Styliidae:</td>
<td>Campanulales:</td>
<td>Styliidae:</td>
</tr>
<tr>
<td>Styliidae</td>
<td>Styliidae:</td>
<td>Styliidae:</td>
<td>Campanulaceae</td>
<td>Styliidae:</td>
</tr>
<tr>
<td>Donatiaceae</td>
<td>Styliidae:</td>
<td>Campanulales:</td>
<td>Campanulales:</td>
<td>Styliidae:</td>
</tr>
<tr>
<td>Donatiaceae</td>
<td>Donatiaceae:</td>
<td>Donatiaceae:</td>
<td></td>
<td>Styliidae:</td>
</tr>
<tr>
<td>Asteraceae</td>
<td>Asterales:</td>
<td>Asterales:</td>
<td>Asterales:</td>
<td>Asterales:</td>
</tr>
<tr>
<td>Asteraceae</td>
<td>Asteraceae:</td>
<td>Asteraceae:</td>
<td>Asteraceae:</td>
<td>Asteraceae:</td>
</tr>
<tr>
<td>Calyceraceae</td>
<td>Calycerales:</td>
<td>Calyceraceae:</td>
<td>Asterales:</td>
<td>Dipsacales:</td>
</tr>
<tr>
<td>Calyceraceae</td>
<td>Calyceraceae:</td>
<td>Calyceraceae:</td>
<td></td>
<td>Calyceraceae:</td>
</tr>
<tr>
<td>Goodeniaceae</td>
<td>Goodeniiales:</td>
<td>Campanulales:</td>
<td>Campanulales:</td>
<td>Goodeniiales:</td>
</tr>
<tr>
<td>Goodeniaceae</td>
<td>Goodeniiales:</td>
<td>Goodeniaceae:</td>
<td>Goodeniaceae:</td>
<td>Goodeniaceae:</td>
</tr>
<tr>
<td>Brunoniaceae</td>
<td>Goodeniiales:</td>
<td>Campanulales:</td>
<td>Campanulales:</td>
<td>Goodeniaceae:</td>
</tr>
<tr>
<td>Brunoniaceae</td>
<td>Goodeniiales:</td>
<td>Brunoniaceae:</td>
<td>Goodeniaceae:</td>
<td>Goodeniaceae:</td>
</tr>
<tr>
<td>Menyanthaceae</td>
<td>Gentianales:</td>
<td>Solanales:</td>
<td>Campanulales:</td>
<td>Cornales:</td>
</tr>
<tr>
<td>Menyanthaceae</td>
<td>Menyanthaceae</td>
<td>Menyanthaceae</td>
<td>Menyanthaceae</td>
<td>Menyanthaceae</td>
</tr>
</tbody>
</table>

ᵃ updated in unpublished report
ᵇ updated posthumously by G. Dahloren (1989a,b)
Table 2. Taxa for which rbcL sequences were obtained in this study. Species that were sequenced from herbarium material are denoted by an asterisk (*).

<table>
<thead>
<tr>
<th>Species</th>
<th>GENBANK accession #</th>
<th>Source/voucher</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Campanulaceae:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Codonopsis ovata</em> Benth.</td>
<td>L18797</td>
<td>Cosner 175</td>
</tr>
<tr>
<td><em>Cyananthus microphyllus</em> Benth.</td>
<td>L18795</td>
<td>Cosner 176</td>
</tr>
<tr>
<td><em>Trachelium caeruleum</em> L.</td>
<td>L18793</td>
<td>Cosner 173</td>
</tr>
<tr>
<td><strong>Cyphiaceae:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Cyphia elata</em> Harv.</td>
<td>L18796</td>
<td>Dahlstrand 1360 (F)*</td>
</tr>
<tr>
<td><em>Cyphocarpus rigescens</em> Miers</td>
<td>L18792</td>
<td>Zollner 8333 (MO)*</td>
</tr>
<tr>
<td><em>Nemacladus ramosissimus</em> Nutt.</td>
<td>L18791</td>
<td>Boyd et al. 4566 (F)*</td>
</tr>
<tr>
<td><strong>Lobeliaceae:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Clermontia kakeana</em> Meyen</td>
<td>L18789</td>
<td>Lammers 8059 (F)</td>
</tr>
<tr>
<td><strong>Pentaphragmataceae:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Pentaphragma ellipticum</em> Poulsen</td>
<td>L18794</td>
<td>Hugh Tan, Natl. Univ. of Singapore</td>
</tr>
<tr>
<td><strong>Sphenocleaceae:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Sphenoclea zeylanica</em> Gaertn.</td>
<td>L18798</td>
<td>Hill et al. 13195 (F)*</td>
</tr>
<tr>
<td><strong>Stylidiaceae:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Stylidium graminifolium</em> Sw.</td>
<td>L18790</td>
<td>Cosner 174</td>
</tr>
</tbody>
</table>
Fig. 1. Strict consensus tree of 32 equally parsimonious trees of 4003 steps (consistency index=0.28) generated using \textit{rbcL} sequence data from 117 taxa. The parsimony analysis employed the TBR, ACCTRAN, and MULPARS options of PAUP version 3.1.1 (Swofford 1993). Boxed genera indicate sequences obtained for this study. Labelling of groups follows Olmstead & al. (1993).
Fig. 2. One of four equally parsimonious trees of 1354 steps (consistency index=0.49) based on rbcL sequences of 44 taxa. Numbers indicate changes at each node and along each branch. Brackets indicate familial/ordinal/group circumscription.
Fig. 3. Detail of Campanulales/Asterales clade based on rbcL analysis of 44 taxa (topology of Campanulales/Asterales was identical in all four equally parsimonious trees). Numbers indicate changes at each node and along each branch, numbers in parentheses are bootstrap values, and d refers to decay indices.
FIGURE 3
CHAPTER II

THE HIGHLY REARRANGED CHLOROPLAST GENOME OF TRACHELIELUM CAERULEUM: MULTIPLE INVERSIONS, INVERTED REPEAT EXPANSION AND CONTRACTION, TRANSPOSITION AND SEVERAL REPEAT FAMILIES

Introduction

The chloroplast genome of photosynthetic land plants consists of a circular chromosome ranging in size from about 120 to 217 kilobases (kb). Chloroplast DNA (cpDNA) consists of a large inverted repeat (IR) separated by large and small single copy regions. Exceptions include conifers (Lidholm et al. 1988; Raubeson and Jansen 1992; Strauss et al. 1988) and certain taxa in a few angiosperm families, the most notable of which are six tribes of the Fabaceae (legumes) (Lavin et al. 1990; Palmer, 1991; Palmer and Thompson 1981), which have lost one segment of the IR. The IR is generally 20-30 kb in length, but ranges in size from about 10-76 kb (Palmer, 1991). Expansion or contraction of the IR occurs infrequently; the most drastic case of expansion is found in geranium, in which the IR has grown to 76 kb (Palmer et al. 1987).
Gene content, gene order, and nucleotide sequence are highly conserved in cpDNA (Palmer 1991), particularly in the IR (Downie and Palmer 1992a; Wolfe et al. 1987). Few cases of chloroplast gene loss have been reported in land plants (Palmer 1991) and rearrangement of genes is relatively rare. Most cases in which regions of cpDNA have been inverted involve one or only a few inversions. For example, a 22 kb derived inversion separates most members of the angiosperm family Asteraceae from subfamily Barnadesioideae and other angiosperms (typified by tobacco) (Jansen and Palmer 1987a,b). Three inversions characterize all examined members of the grass family (Poaceae) (Doyle et al. 1992; Hiratsuka et al. 1989; Howe 1985; Quigley and Weil 1985). More complicated genomes exhibiting several to many inversions (for example, conifers and some legumes) have also lost one segment of the IR, thus destabilizing the genome (Palmer 1985).

The existence of multiple cpDNA rearrangements in two closely related angiosperm families, Campanulaceae and Lobeliceae, was noted several years ago following a survey of restriction site mutations (Jansen and Palmer 1987b; Palmer 1985; Downie and Palmer 1992b). We have found that the chloroplast genome of Trachelium caeruleum (Campanulaceae) is highly rearranged relative to that of the majority of land plants. At least seven inversions,
IR contraction/expansion, two major presumed deletions, several insertions, and a putative transposition have transformed the chloroplast genome of Trachelium from a tobacco-like ancestor.

Materials and Methods

Chloroplast-enriched DNA was isolated from Trachelium caeruleum by the sucrose gradient method (Palmer 1986). The cpDNA was digested with the restriction endonuclease HindIII, resulting in 26 fragments larger than 0.5 kb, ranging in size from about 1.1 to approximately 18 kb. The HindIII-digested cpDNA was shotgun-cloned into the plasmid vector pUC19. The ligation mixture was then used to transform E. coli strain DH5alpha. Plasmid DNA was recovered using the alkaline lysis method of Birnboim and Doly (1979), and sizes of plasmid inserts determined by HindIII restriction digestion followed by visualization of fragments in agarose gels. Fragments that were not successfully shotgun-cloned were cut out of a 1% SeaPlaque (FMC) agarose gel and ligated into pUC19 in agarose.

Total DNA from Trachelium caeruleum was prepared according to the CTAB isolation method of Doyle and Doyle (1987). Agarose gel electrophoresis, bidirectional
transfer of DNA onto Zetabind (AMF CUNO) nylon membranes, labelling of recombinant plasmids by nick translation, filter hybridizations, and autoradiography were carried out essentially as described by Palmer (1986). Restriction enzyme digests of Trachelium total DNA were electrophoresed in 1.2% agarose gels. Hybridization probes consisted of cloned HindIII cpDNA fragments from T. caeruleum. Fragments that were not cloned were cut out of a 1% SeaPlaque (FMC) agarose gel and nick translated directly in the agarose. In addition, 109 cloned tobacco cpDNA fragments were used as probes in hybridization experiments. These probes were subcloned from larger cloned cpDNA fragments (Olmstead and Palmer 1992; Sugiura et al. 1986) and average about 1.2 kb in size.

The polymerase chain reaction (PCR) was employed to determine if introns are present in the gene clpP in Trachelium. One forward and one reverse primer each (for a total of eight 14- to 21-mer primers) for exon 3, intron 2, exon 2, and intron 1 were constructed using available sequences from tobacco (Shinozaki et al. 1986), Epifagus virginiana (Wolfe et al. 1992), or a consensus. Fifty ul PCR reactions consisted of 0.2 mM dNTPs, 0.5 unit of Tfl polymerase (Epicentre Technologies), 1x Tfl polymerase buffer (includes 1.5 mM MgCl), approximately 1 uM primers, and a 0.5 ul aliquot of unquantified DNA. The
thermocycler was programmed for one initial denaturation cycle consisting of three minutes denaturation at 95°C, one minute primer annealing at 55-60°C, and one minute extension at 72°C, followed by 30 cycles of one minute denaturation at 95°C, one minute annealing at 55-60°C, and one minute extension at 72°C. A final extension period of five minutes at 72°C terminated the PCR reactions.

Results

Clone bank of *Trachelium caeruleum*

The shotgun cloning strategy together with the cloning of gel-isolated fragments of approximately 18, 11.5, 9.6, and 7.4 kb resulted in a clone bank consisting of 21 of 26 *HindIII* cpDNA fragments from *Trachelium* representing 82% of the genome (Fig. 4; Table 3). Uncloned fragments include those of about 12.5, 8.3, 4.4, 3.8, and 1.5 kb in size.

Physical and gene mapping

Total DNA from *Trachelium* was digested with the restriction endonucleases *BamHI, BglII, EcoRI, EcoRV,*
HindIII, SstI, and XbaI, and double digests were carried out using HindIII and the remaining six enzymes. The cloned HindIII fragments from Trachelium and gel isolations of uncloned fragments were used as homologous probes in order to construct restriction site maps for the seven enzymes by the identification of overlapping fragments (Fig. 5). Summation of restriction fragment sizes of the seven enzymes indicates that Trachelium has a chloroplast genome of about 162 kb, consisting of an IR of about 27 kb, and large and small single copy regions of approximately 100 and 8 kb, respectively.

Hybridizations using 109 small cloned tobacco cpDNA fragments (average size 1.2 kb) were used to map the genome organization of Trachelium relative to tobacco and to determine the locations of chloroplast genes (Fig. 5). It was determined in earlier experiments that hybridizations with 43 larger tobacco clones (average size 3.1 kb) (Olmstead and Palmer 1992) that are frequently used in mapping experiments were not sufficient to resolve structural maps of Trachelium. The small tobacco probes consist of parts or all of one to a few genes and associated non-coding sequence, providing differential hybridization of 5' and 3' ends of many chloroplast genes, thus making it possible to determine direction of transcription for these genes (Fig. 5). The direction of
transcription of other genes was inferred either from their co-transcription in tobacco with genes for which 5' and 3' differential hybridization was possible, or by comparisons of their known locations in polycistronic transcription units in tobacco compared with their similar locations in Trachelium (Fig. 5). Because many of the tobacco cpDNA probes contain more than one gene, it was not always possible to determine which genes define inversion endpoints. A conservative approach was taken that kept operons known from other plants intact, but it should be noted that locations of some genes were inferred.

Most of the 109 small tobacco chloroplast DNA probes hybridized to Trachelium cpDNA, but lack of hybridization of four tobacco probes suggests possible deletions in Trachelium. Two probes containing the 5' and 3' segments of tobacco's ORF512 (identified as accD by Li and Cronan 1992), respectively, showed no hybridization to Trachelium. Only two other tobacco chloroplast probes did not hybridize to Trachelium. One is a small probe of 213 base pairs (bp), about half of which consists of the 5' end of tobacco's large ORF2280. The second probe is located in another large open reading frame of tobacco, ORF1901. No hybridization was detected for a 1415 bp segment of this gene.
Structural organization of the *Trachelium* chloroplast genome

Results from restriction site and gene mapping show that the chloroplast genome of *Trachelium* is radically rearranged compared to that of other land plants (Fig. 5). Genome modifications include changes in gene order and orientation, extent of the IR at both large single copy (LSC) and small single copy (SSC) ends, and gene content. Figures 6, 7, and 8 diagram three possible series of evolutionary events involving the fewest numbers of rearrangements necessary to produce the *Trachelium* genome structure from a tobacco-like ancestor. Rearrangements in *Trachelium* consist of no fewer than seven inversions, IR expansion/contraction, one putative gene loss, partial deletions of two large ORFs, several insertions, and a possible transposition (see below). However, complex hybridization patterns indicate that a more complicated evolutionary scenario consisting of a greater number of rearrangements may actually have occurred. In many instances, individual tobacco probes hybridized in several locations in the cpDNA. Many of these are readily explained by the obvious splitting of the sequence by a rearrangement endpoint. In some cases, the complex
hybridization patterns were not as easily interpreted, and were presumably due either to "endpoint remnants" left by previous inversions, or to small dispersed repeats within the genome. Although Figs. 6, 7, and 8 portray step-wise series, it must be emphasized that the order of events is not known. Hybridization experiments show that at least two of the mutations have disrupted what have been reported in other land plants to be transcription units. Genes within both the atpA and clpP operons have been separated by different rearrangements.

Mapping experiments show that, although IR size in Trachelium is similar to that of the majority of other angiosperms, its gene content is significantly different. Several genes that are single copy in tobacco are duplicated (i.e., are located in the IR) in Trachelium and vice versa (Fig. 5). The difference in extent of Trachelium's IR has resulted from spreading of the repeat into the SSC region and shrinkage at the LSC end. Genes between and including ndhE and rps15 in the IR of Trachelium are located in the SSC region in tobacco (Fig. 5), whereas genes between and including 3'-rps12 and ORF2280 in the LSC of Trachelium are in tobacco's IR (Fig. 5). In addition, 5'-rps12 and clpP that reside in the IR of Trachelium are in the LSC of tobacco and most other land plants. The rRNA operon, a conserved feature of the
IR in all IR-containing plants, remains in *Trachelium*'s IR despite the disruptive changes in IR structure; however, the operon now comprises the set of IR genes nearest to the LSC region in *Trachelium*, whereas the operon is closer to the SSC end of the IR in other angiosperms.

Transposition in *Trachelium* cpDNA

Evidence suggests that transposition has played a role in the evolution of the chloroplast genome of *Trachelium*. Figure 6 depicts a model in which part of the clpP operon consisting of the genes clpP and 5'-rps12 has been relocated from the LSC region to the IR. The two genes in *Trachelium* are now located within a large open reading frame (ORF1901 of tobacco) near its 5' end (Fig. 9). The third gene of the operon, rpl20, has remained in the LSC region. A gap of about 3.0 kb exists between rpl20 and psbB where clpP and 5'-rps12 normally reside; this gap lies within a series of genes of about 20 kb that is otherwise unrearranged relative to tobacco. PCR of clpP using primers for exons 2 and 3, and introns 1 and 2, indicates that both introns exist in clpP in *Trachelium*. The 9.6 kb HindIII *Trachelium* clone (the IR clone that contains clpP and 5'-rps12) was used as the template in PCR reactions.
It also appears that a duplicative transposition of a segment of the 23S rRNA gene has taken place in *Trachelium*. A 1680 bp tobacco chloroplast DNA probe consisting entirely of coding sequence (the 5' and 3' ends of the gene are located in two separate probes) from within this gene hybridizes both to the presumably functional 23S rRNA locus and to a region approximately 14 kb downstream in *Trachelium* (Fig. 9). The exact size of the duplication is unknown, but the smallest restriction fragment containing the duplicated DNA is an XbaI fragment of approximately 1.0 kb. The partially duplicated 23S rDNA is located within the same large open reading frame (ORF1901) to which part of the clpP operon was also putatively transposed (Fig. 9).

Dispersed repeated sequences in *Trachelium* cpDNA

Certain cloned *Trachelium* HindIII cpDNA fragments hybridize only to themselves (e.g., fragments of 2.4 and 4.2 kb; Fig. 10) when hybridized to HindIII digests of total *Trachelium* DNA, but others hybridize to one or more additional HindIII fragments. For example, a 4.5 kb HindIII clone hybridizes to itself and to a 1.2 kb fragment (Fig. 10). The reciprocal hybridization of the 1.2 kb clone to the 4.5 kb fragment indicates the
existence of a region of homologous sequence between the two fragments (Fig. 10). In another case, a 9.6 kb clone hybridizes to itself and to fragments of 4.9 and 3.8 kb (Fig. 10). Reciprocal hybridization indicates that the 4.9 and 3.8 kb fragments do not hybridize to each other, and thus share different sequence homologies with the 9.6 kb fragment. There are at least five clear sets of dispersed repeats within Trachelium's chloroplast genome (Fig. 5). Many other Trachelium HindIII clones hybridize very faintly to additional HindIII fragments, but these data are more difficult to interpret.

Discussion

Chloroplast genome evolution in Trachelium is notable in five major respects. First, as the result of at least seven inversions, gene order in Trachelium is highly rearranged relative to most other land plants. Second, the IR has spread into the adjacent SSC region but has been reduced in size at its LSC end, resulting in larger and smaller LSC and SSC regions, respectively. Third, substantial evidence suggests that transposition, a rarely reported event in cpDNA evolution (Palmer 1991), has contributed to the scrambled gene order of Trachelium.
Fourth, at least five families of dispersed repeats have been identified in Trachelium, whereas most chloroplast genomes lack small repeated sequences. Fifth, at least two rearrangements (either two inversions or one inversion and one transposition) have resulted in the disruption of what have been reported as two different transcriptional units in other land plants.

Evolution of chloroplast genome structure in Trachelium

Evolutionary models. The chloroplast genome structure of Trachelium is highly rearranged relative to that of most other land plants. The complexity of the genome makes it difficult to determine with certainty the number and types of rearrangement events that have occurred. However, we have constructed three possible models (Figs. 6, 7 and 8) to describe the evolution of Trachelium cpDNA from a tobacco-like ancestor. Although we present models that attempt to minimize the numbers of evolutionary steps, it must be emphasized that a more complicated series of events may have occurred. The most complicating factor in postulating evolutionary scenarios for Trachelium's chloroplast genome is the relocation of clpP and 5'-rps12 (sequence block I in Figs. 6, 7, and 8) from the LSC region to the IR.
The first model (Fig. 6) proposes a series of steps involving seven inversions, one transposition, one IR contraction, and one IR expansion. The only constraint on ordering of these events is that inversion 1 (Fig. 6) must occur following the IR contraction, since one inversion endpoint lies within what is typically the IR in most angiosperms. This proposed constraint is based on the assumption that inversions between the IR and either single copy region would be highly disruptive and unlikely to occur. Two other inversions (2 and 4) involving sequences found in the LSC region of *Trachelium* but in the IR of most angiosperms could have occurred either before or after IR shrinkage but for simplicity are diagrammed in Fig. 6 as happening after the IR contraction. The other four inversions lie in what is typically the LSC region in most land plants. Besides the single constraint mentioned above, the rearrangements could have occurred in any sequence; the ordering of events in Fig. 6 is arbitrary.

The second evolutionary model (Fig. 7) consists of ten inversions, one IR loss, and one IR regrowth. This scenario is constrained in that loss of the IR segment must occur before inversion 1, inversion 1 must occur before inversions a and b, and inversions a and b must occur before the IR regrowth. A fourth constraint involves the relative orientations of single copy regions
in the ancestral genome. Chloroplast DNA exists in two orientations in equimolar quantities due to intramolecular recombination between the paired segments of the IR (Kolodner and Tewari 1979; Palmer 1983; Stein et al. 1986). This "flip-flop" recombination results in two molecules whose only difference is the relative orientations of their single copy regions. For the second model to operate, the single copy segments must exist in the orientations shown (Fig. 7); for comparison, the opposite orientation is diagrammed in Fig. 6. All other rearrangements could have occurred in any order.

The evolutionary scenarios presented in Figs. 6 and 7 were constructed based on the assumption that the transfer of clpP and 5'-rps12 from the LSC region to the IR would not have been mediated by inversion(s) between the IR and LSC region. Presumably, the stability of the IR would render such an inversion highly unlikely and difficult to envision mechanistically since a rearrangement of this type would be extremely disruptive to the IR. However, given the complexity of Trachelium's chloroplast genome, it is conceivable that inversion between the IR and the LSC region may have occurred, and is proposed as a mechanism in Fig. 8. This model (Fig. 8) invokes expansion/contraction of the IR and nine inversions, at least one of which must occur between the IR and LSC
region, accompanied by temporary growth and/or shrinkage of the IR. If inversions d and e of Fig. 8 occurred with the IR intact, both the transposition of Fig. 6 and the IR loss/regrowth and one inversion of Fig. 7 would be eliminated. The effect of an inversion between the IR and LSC region on the IR is unknown, but it seems likely that inversion d would be accompanied by either severe shrinkage or extreme growth of the IR; Fig. 8 shows the presumed minimum and maximum extent of the IR following the inversion. Inversion e reverses most of inversion d, and it is possible that these two events occurred in concert rather than in an ordered fashion, thus reducing disruption to the IR. The only constraint on ordering of events is that inversion 1 must occur before inversions d and e. In addition, the IR contraction is diagrammed (Fig. 8) as occurring before inversion 1 to eliminate an additional inversion between the IR and LSC region, since one endpoint of inversion 1 lies in what is typically the IR in angiosperms. For proper orientation of sequence block I, inversions d and e must occur between the LSC region and IR A (Fig. 8), adding a third constraint.

We favor the model that involves transposition (Fig. 6) for three major reasons. First, this model postulates two fewer steps than Fig. 7, and three fewer than Fig. 8, by eliminating three and two inversions (and associated
growth/shrinkage of the IR in Fig. 8), respectively, but adding one transposition. Second, it invokes considerably less radical and complex change to the IR. In particular, although this model proposes extensive contraction of the IR, the rDNA core, a feature of virtually all cpDNA IRs, remains intact. The rDNA containing IR is temporarily lost in the model of Fig. 7, and although the effect of inversion d of Fig. 8 on the IR is unknown, the temporary exclusion of the rDNA from the IR is possible. Third, this model is less constrained, having only one constraint compared to four and three in Figs. 7 and 8, respectively. All three models propose events that do not typically occur in chloroplast genomes. Loss and subsequent regrowth of the IR has never been suggested as an evolutionary mechanism in the chloroplast genome, transposition has only rarely been postulated (see below), and inversion between the IR and a single copy region is extremely unlikely.

Chloroplast DNA mapping data from other genera in the Campanulaceae supply strong evidence against the IR loss/regrowth model (Fig. 7). Essentially all taxa examined in this family share the same IR/LSC junction, including three putatively basal genera in which clpP and 5'-rps12 have not been transferred from the LSC region to the IR (M. Cosner unpublished). If the IR was lost in the
ancestor of Trachelium and most other members of the Campanulaceae as a means to effect the inversional transfer of these two genes to the IR, it would be extremely unlikely for the IR to regrow with the same IR/LSC junction as the basal genera that retain clpP and 5'-rps12 in their LSC regions. A modification of the IR loss/regrowth model compatible with this evidence would include shrinkage of the IR at the LSC end severe enough that the inversion of clpP and 5'-rps12 could occur entirely within the LSC region, thus avoiding inversion between the IR and LSC region.

It is believed that the IR stabilizes cpDNA by decreasing intrachromosomal recombination (Palmer 1985). Conifers (Lidholm et al. 1988; Raubeson and Jansen 1992; Strauss et al. 1988) and some legumes (Lavin et al. 1990; Palmer 1991; Palmer and Thompson 1981), in which chloroplast genomes have become highly rearranged, have also lost one copy of the inverted repeat, whereas Trachelium retains its IR. It is clear that factors other than the loss of one copy of the IR have served to destabilize cpDNA and predispose Trachelium's chloroplast genome to rearrangement.

Endpoint reuse. Regardless of which evolutionary model is accepted, reuse of inversion endpoints has occurred in Trachelium. According to Fig. 7, there are
ten inversions and 13 endpoints. If all inversions used
different endpoints, the maximum number of endpoints would
be 20. In Fig. 6, seven inversions use ten endpoints;
maximally 14 endpoints could exist. The nine inversions
of Fig. 8 use 13 endpoints rather than the maximum 18.
Without knowing the precise order of events, it is not
possible to determine the exact nature of the endpoint
reuse. The cpDNA of the closely related family
Lobeliaceae is also rearranged, but not as radically as in
the Campanulaceae. More extensive endpoint reuse has
apparently occurred in Lobeliaceae. Among 18 taxa from
three genera examined, 11 inversions sharing 10 endpoints
were found (Knox, et al., 1993).

The reuse of endpoints led Knox et al. (1993) to
postulate the existence of recombinational "hot spots"
(Kung et al. 1982; Ogihara et al. 1988, 1991) in
Lobeliaceae. tRNA genes have been associated with cpDNA
inversion endpoints in several families of angiosperms
including Asteraceae (Jansen and Palmer 1987b), Fabaceae
(Michalowski et al. 1987; Palmer et al. 1988), Onagraceae
(Herrmann et al. 1983), and Poaceae (Hiratsuka et al.
Rodermel et al. 1987) as well as in some conifers (Lidholm
and Gustafsson 1991; Strauss et al. 1988; Tsudzuki et al.
1992), a fern (Hasebe and Iwatsuki 1992), a moss (Calie
and Hughes 1987), and a liverwort (Ohyama et al. 1986). The inversions may have been mediated through recombination across short repeated sequences found within or near the tRNA genes (Bonnard et al. 1985; Howe 1985; Howe et al. 1988; Hiratsuka et al. 1989; Rodermel 1992; Shimada and Sugiura 1989), but the exact mechanism by which the repeats may facilitate recombination is unclear. Most of the endpoints in Lobeliaceae are associated with tRNA genes (Knox et al. 1993). In Trachelium, six or seven endpoints (depending on which model is accepted) occur near one or more tRNA genes (Fig. 6). Although tRNA genes and associated repeats may play a role in cpDNA inversion, not all inversion endpoints lie near these genes, indicating other mechanistic factors are also involved.

Disruption of cpDNA operons in Trachelium

At least two operons have been disrupted by rearrangements in Trachelium. One endpoint of inversion 1 (Fig. 6) lies between two genes for subunits of the H⁺-ATPase complex, atpH and atpI, reported to be transcriptionally linked in spinach (Hennig and Herrmann 1986; Hudson et al. 1987; Westhoff et al. 1985), pea (Cozens et al. 1986; Hudson et al. 1987), rice (Kanno and
Hirai 1993), and wheat (Högglund et al. 1990). Another rearrangement has placed clpP and 5'-rps12 in Trachelium's IR, while rpl20 has remained in the LSC region. These three genes were shown to be co-transcribed in liverwort (Kohchi et al. 1988), rice (Kanno and Hirai 1993), and maize (Weglöhner and Subramanian 1992). The clpP operon was disrupted either by inversion b or d (Figs. 7 and 8, respectively) or by the transposition postulated in Fig. 6. Only two other studies report disruption of cpDNA transcription units by rearrangements. The mung bean rpl23 operon is split by an inversion (Palmer et al. 1988), whereas a transposition has been implicated in the disruption of subclover's rpoB operon (Milligan et al. 1989). A third possible case involves the formation of a new transcription unit following gene rearrangement. Nagano et al. (1991) described an operon containing trnQ, zfpA (accD or ORF371), psaI, ORF231, and petA in pea. trnQ was presumably moved adjacent to the other genes as the result of inversion in pea. It is not clear, however, if the relocation of trnQ disrupted a previously existing operon, because trnQ is reported to be transcribed singly in mustard (Neuhaus 1989; Nagano et al. 1991), whereas in rice it is cotranscribed with rps16 (Kanno and Hirai 1993), a gene that is absent from a few plant groups including some legumes such as pea (Nagano et al. 1991;
Downie and Palmer 1992b). Nevertheless, trnQ in pea is currently part of an operon different from that in rice, which contains ORF185, ORF230, and petA (Kanno and Hirai 1993). It is possible that new transcription linkages have been formed in Trachelium as well.

Chloroplast DNA insertions and deletions in Trachelium

For the simplicity of modelling, cpDNA insertions and deletions in Trachelium are not shown in Figs. 6, 7, and 8. As mentioned above, two tobacco chloroplast DNA probes containing the 5' and 3' ends of the gene accD do not hybridize to Trachelium, suggesting that the gene has been deleted from the Trachelium chloroplast genome. It cannot be ruled out that the sequence has simply diverged significantly from that of tobacco, because mapping in the region of the putative deletion indicates sufficient space (approximately 2.0 kb) to contain a gene of this size (Fig. 5). It is possible that accD is a pseudogene in Trachelium. It is also conceivable that the unidentified 2.0 kb segment was inserted before or after the deletion of accD. Regardless of whether deletion, insertion, or gene divergence has occurred, an inversion endpoint is associated with the location (Fig. 5).
A 213 bp probe containing the 5' end of tobacco's large ORF2280 does not hybridize to *Trachelium* (Fig. 5). Although this open reading frame is noted for sustaining fairly sizable deletions while remaining an intact ORF (Blasko et al. 1988; Zhou et al. 1988), it is possible that this gene is not functional in *Trachelium*. The putative deletion may be fairly recent, because the 213 bp tobacco probe hybridizes to cpDNA in some other members of Campanulaceae (M. Cosner unpublished). In addition, no hybridization was detected for a 1415 bp segment of tobacco's ORF1901, although, as seen with accD, mapping indicates a space sufficiently large to accommodate the segment, suggesting divergence rather than deletion (Fig. 5).

Several apparent insertions (as many as seven) are present in *Trachelium*'s cpDNA (Fig. 5). These appear to consist of DNA unique to *Trachelium*, as no hybridization to tobacco in these regions was detected. Three of these may not be insertions, but possibly consist of nonfunctional genes (putative pseudogenes) that have sufficiently diverged and no longer hybridize to tobacco cpDNA. One of these is the approximately 2.0 kb segment where accD normally resides, and the second is located where the 1415 bp tobacco ORF1901 probe should hybridize (both discussed above). The third is a 3.0 kb segment
located between the genes psbB and rpl20, within an otherwise unrearranged stretch of DNA (Fig. 5), the location from which clpP and 5'-rpsl2 were either transposed or inverted (see below) (Fig. 9). It is possible, however, that all three of these DNA segments are insertions of unknown content and origin, rather than pseudogenes. Four other insertions are present, three of which average about 2.0 kb in size and occur at inversion endpoints (Fig. 5). The fifth insertion is about 1.0 kb and is not located at an inversion endpoint (Fig. 5).

Transposition in Trachelium cpDNA

Variations in gene order in land plant chloroplast genomes are most often the result of inversions or gene deletions. Only recently has evidence pointed to a possible role for transposition events in cpDNA evolution. Milligan et al. (1989) reported evidence for the occurrence of at least two transpositions in subclover cpDNA. One of these is a duplicative transposition of part of the ribosomal protein gene rpl23, and the second, an apparent transposition of rpoB and rpoC1, is similar to the action of transposable elements. In addition, small segments of certain ribosomal protein genes have shifted positions in the chloroplast genomes of some wheat
species, presumably by transposition (Bowman et al., 1988; Ogihara et al., 1988).

The genes \textit{clpP} and 5'-\textit{rps12}, found in the LSC region in most land plants, have been transferred to the IR in \textit{Trachelium}, presumably either by transposition (Fig. 6) or inversion (Figs. 7 and 8). The new location of the genes in the IR is more simply explained by transposition (Fig. 6), particularly since the removal of the two genes left an otherwise unrearranged segment of DNA. Three inversions and several genomic constraints must be invoked in order to relocate the two genes by inversion in Fig. 7, and two inversions and one constraint are necessary in the model of Fig. 8; at least one of the inversions (Fig. 8) must occur between the IR and LSC region. If transposition was indeed the mechanism, it is unclear whether the event was DNA or RNA mediated.

Some evidence suggests that retrotransposition (insertion of a cDNA made by reverse transcription of a chloroplast transcript) may have been responsible for the relocation. A gap of approximately 3.0 kb is present in the location (between \textit{rpl20} and \textit{psbB}) where \textit{clpP} and 5'-\textit{rps12} are found in most plants, and this gap is sufficient in size to house the two genes. It may be that the 3.0 kb region represents pseudogenes of \textit{clpP} and 5'-\textit{rps12} that resulted from sequence divergence of the original genes.
following insertion of the cDNA in the IR. However, the gap could also be explained as an insertion of unknown origin, as there are a few other such regions in Trachelium's chloroplast genome that appear unique to Trachelium, that is, they do not hybridize to tobacco chloroplast DNA probes (Fig. 5).

In order for the genes to function in their new location, it would be necessary to acquire proper promoter sequences. 5'-rps12 and clpP are located within a large open reading frame (ORF1901 of tobacco) very near its 5' end (Fig. 9). It is possible that the genes have remained functional through the cannibalization of the presumably nonfunctional ORF's promoter.

PCR amplification of introns and coding regions of the current location of clpP shows that both clpP introns are present. If retrotransposition played a role in the insertion of the two genes in the IR, the presence of introns in clpP indicates that the cDNA was made from a partially processed transcript from which rpl20 was removed, but from which clpP introns were not yet spliced. Absence of introns would provide evidence for the involvement of an RNA intermediate, but intron presence is merely inconclusive. Kohchi et al. (1988) found multiple mRNAs are transcribed from the clpP operon (clpP, 5'-rps12, rpl20), and they also described the occurrence of
ordered processing and splicing of the transcripts that
involved processing prior to splicing of the clpP introns.
This is consistent with a retrotransposition hypothesis
involving an RNA species containing both 5'-rps12 and clpP
(including introns) but not rpl20.

Although retrotransposition has largely been
attributed to eukaryotic genomes (see reviews in Boeke and
Corces 1989; Coffin 1993; Gabriel and Boeke 1993; McClure
1991; Weiner et al. 1986), reverse transcriptase-like
genes have been discovered in mitochondrial introns and
plasmids (e.g. Kuiper and Lambowitz 1988; Michel and Lang
1985; Xiong and Eickbush 1988; reviewed in Boeke and
Corces 1989; McClure 1991; Weiner et al. 1986), in a plant
mitochondrial genome as an ORF not associated with an
intron (Schuster and Brennicke 1987), and in an intron of
a plastid gene from a green alga (Kück 1989). In
addition, transposition has played an apparent role in the
evolution of certain other extensively rearranged
chloroplast genomes (see above).

Although certain features of the Trachelium
chloroplast genome suggest an RNA mediated transposition
of the two genes, none of the data contradict a DNA
mediated transposition of 5'-rps12 and clpP, and there is
some evidence that supports such an event. In another
species in the Campanulaceae, the putatively transposed
clpP and 5'-rps12 genes have been transferred from the IR to a unique location in the LSC region within an otherwise unrearranged stretch of DNA (Cosner, unpublished). This suggests a DNA mediated transposition event similar to a transposable element. In a DNA mediated transposition, the necessity of acquiring proper promoter sequences is avoided, as it is conceivable that the promoter could be transferred as well. The region of putative transposition as well as the site housing the suspected pseudogenes in *Trachelium* will be sequenced in an attempt to resolve whether DNA or RNA mediated transposition or inversion was involved.

Dispersed repeats and inversion mechanisms

Comparisons among chloroplast genomes of highly divergent plant lineages show a remarkable conservation of chromosome structure, gene content, and gene sequence. In particular, the large IR is especially conserved, showing an even slower rate of sequence evolution than either single copy region (Downie and Palmer 1992a; Wolfe et al. 1987). The IR is believed to confer an inherent structural stability on cpDNA by limiting recombination between other segments of the chromosome (Palmer 1985). With the exception of the large inverted repeat,
chloroplast genomes seldom contain repeated sequences, possibly related to a constraint on size of the chloroplast DNA molecule (Palmer 1985). Exceptions include cpDNA of geranium (Palmer et al. 1987), wheat (Bowman et al., 1988; Bowman and Dyer 1986; Howe 1985; Ogihara et al. 1988; Quigley and Weil 1985), rice (Shimada and Suguira 1989), subclover (Milligan et al., 1989), broad bean (Bonnard et al. 1985) and Douglas fir (Tsai and Strauss 1989), whose chloroplast genomes all contain a series of short dispersed repeats. In these species, dispersed repeats have been associated with inversion endpoints and are thought to have facilitated inversions through recombination across homologous repeats.

Trachelium's chloroplast genome contains at least five families of dispersed repeats (Figs. 5 and 10). All of these are associated with various rearrangements, including inversions, insertions, and possible transpositions, thereby suggesting a mechanism for at least some of the rearrangements. One of the repeated sequences consists of part of the 23S rRNA gene, presumably the result of a duplicative transposition (see above) (Fig. 9), whereas the remainder are of unidentified content. It is not known how these repeats may have originated and for what reasons they are present in the chloroplast genomes of so few plant species. Although
dispersed repeats may play a role in the destabilization of the chloroplast genome of Trachelium, not all of Trachelium's inversion endpoints are associated with dispersed repeats. It is possible that in some instances the small repeats were lost as the result of inversion, but it seems likely that additional inversion mechanisms were involved in Trachelium as well.

In addition to small dispersed repeats, insertions and deletions are frequently associated with inversion endpoints in Trachelium (Fig. 5), and may be part of the inversion process. However, it is not known if the insertions and deletions are part of the cause of inversion, or are the result of inversion. Since such radically rearranged chloroplast genomes are typically associated with loss of one copy of the IR, it is probable that other unknown factors have contributed to the rearrangement of cpDNA in Trachelium.
LITERATURE CITED


Table 3. Sizes of HindIII restriction fragments of *Trachelium caeruleum* cpDNA larger than 0.5 kb cloned into pUC19. Fragment numbers are the same as in Fig. 4 and approximate sizes are in kb. Chloroplast DNA HindIII fragments of 12.5, 8.3, 4.4, 3.8, and 1.5 kb were not cloned.

<table>
<thead>
<tr>
<th>Fragment number</th>
<th>Fragment size</th>
<th>Fragment number</th>
<th>Fragment size</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>18.0</td>
<td>12</td>
<td>4.2</td>
</tr>
<tr>
<td>2</td>
<td>11.5</td>
<td>13</td>
<td>3.7</td>
</tr>
<tr>
<td>3</td>
<td>9.6</td>
<td>14</td>
<td>2.5</td>
</tr>
<tr>
<td>4</td>
<td>7.4</td>
<td>15</td>
<td>2.5</td>
</tr>
<tr>
<td>5</td>
<td>7.2</td>
<td>16</td>
<td>2.4</td>
</tr>
<tr>
<td>6</td>
<td>7.2</td>
<td>17</td>
<td>2.4</td>
</tr>
<tr>
<td>7</td>
<td>6.2</td>
<td>18</td>
<td>2.3</td>
</tr>
<tr>
<td>8</td>
<td>5.4</td>
<td>19</td>
<td>1.2</td>
</tr>
<tr>
<td>9</td>
<td>4.9</td>
<td>20</td>
<td>1.2</td>
</tr>
<tr>
<td>10</td>
<td>4.5</td>
<td>21</td>
<td>1.1</td>
</tr>
<tr>
<td>11</td>
<td>4.4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Fig. 4. Clone bank of *Trachelium caeruleum* cpDNA. HindIII clones were digested with HindIII and separated on a 1.0% agarose gel. The arrows indicate the migration of the pUC19 plasmid vector. Clone numbers correspond to those in Table 3. The larger of two fragments below the vector band in lane 21 is a non-cpDNA fragment cloned together with the cpDNA fragment.
Fig. 5. Physical and gene map of *Trachelium caeruleum* cpDNA. Complete circles from outside to inside show restriction sites and approximate restriction fragment sizes for HindIII, EcoRV, XbaI, BamHI, and SstI in that order. Enzymes mapped but not pictured include BglII and EcoRI; double digests using HindIII and the remaining 6 enzymes not shown. The two black bars represent the minimum extent of the inverted repeat and the open extensions of these bars its maximum possible extent. Locations, orientations and approximate lengths of genes are shown by arrows labelled with gene names; sizes of ORFs are those of tobacco. Outermost arrows labelled A-0 indicate blocks of genes and their orientations relative to tobacco. Numbered lines indicate the positions of elements belonging to 5 families of dispersed repeats (r). Repeat 5 is a small segment of the 23S rRNA gene located in both the 23S rDNA and ORF1901. Brackets indicate presumed insertions not homologous to tobacco cpDNA; numbered brackets designate regions that may contain pseudogenes of clpP and 5'-rps12 (1), accD (2), and part of ORF1901 (3). The small arrow shows the location of a 213 bp tobacco cpDNA probe that does not hybridize to *Trachelium* DNA.
Fig. 6. Model for the evolution of *Trachelium caeruleum* cpDNA from a tobacco-like ancestor. The model includes 7 inversions, 1 transposition, expansion of the IR into the SSC region, and contraction of the IR at the LSC end. The top gene map and first lettered map show tobacco, and the bottom gene map and last lettered map show *Trachelium*. Only genes at rearrangement endpoints are given. Arrows on top gene map indicate rearrangement endpoints; those with asterisks designate the transposition. Letters correspond to blocks of genes and are the same as in Fig. 5; arrows under letters indicate orientations of gene blocks. The bars indicate location and extent of the IR. Numbered arrows designate inversion endpoints and the following map gives the resulting inverted genome structure. "T" stands for transposition. Insertions/deletions in *Trachelium* relative to tobacco are not shown. The ordering of events shown is arbitrary except as explained in text.
FIGURE 6
Fig. 7. Model for the evolution of *Trachelium caeruleum* cpDNA from a tobacco-like ancestor. The model includes 10 inversions and loss and regrowth of the IR. Designations are the same as Fig. 6. Inversion numbers are identical to those in Fig. 6 with the additions of inversions a, b, and c. Rearrangement endpoints are the same as Fig. 6, but asterisked arrows in that figure indicate inversion endpoints for this model. The sequence of events shown is arbitrary except as explained in text.
Loss of IR B

Inversion 1

Inversion a

Inversions b, c

IR regrowth

Inversions 2, 3, 4, 5, 6, 7

FIGURE 7
Fig. 8. Model for the evolution of *Trachelium caeruleum* cpDNA from a tobacco-like ancestor. The model includes 9 inversions, expansion of the IR into the SSC region, and contraction of the IR at the LSC end. Designations are the same as Fig. 6. Inversion numbers are identical to those in Fig. 6 with the additions of inversions d and e. Inversions d and e must occur in the IR copy shown. Open bars on IR indicate maximum possible extent of IR following inversion d. Rearrangement endpoints are the same as Fig. 6, but asterisked arrows in that figure indicate inversion endpoints for this model. The sequence of events shown is arbitrary except as explained in text.
IR contraction/expansion

Inversion 1

Inversion d

Inversion e

Inversions 2, 3, 4, 5, 6, 7

FIGURE 8
Fig. 9. Model for RNA-mediated transposition of 5'-rps12 and clpP from the LSC region to the IR, and duplicative transposition of a small segment of 23S rDNA, in Trachelium caeruleum cpDNA. Open boxes indicate presumably functional genes, shaded boxes designate putative pseudogenes, and the black box represents a partial copy of the 23S rRNA gene. Dashed line indicates the DNA segment separating the regions pictured. Triangles represent the transposition events (in the case of 5'-rps12 and clpP this involves insertion of a cDNA made by reverse transcription of a chloroplast transcript of these genes). The model can be modified to represent a DNA-mediated transposition by replacing the 5'-rps12 and clpP pseudogenes with an insertion of unknown content and origin.
FIGURE 9
Fig. 10. Short dispersed repeats in the chloroplast genome of *Trachelium caeruleum*. Filters containing *Trachelium* total DNA digested with *HindIII* and electrophoresed on a 1.2% agarose gel were hybridized with each of five *Trachelium* cpDNA *HindIII* clones. Sizes in kb of the *HindIII* inserts are given above each lane. The outside lanes contain lambda DNA molecular weight standards with fragment sizes in kb indicated. The 1.2 kb and 4.5 kb fragments contain the 23S rDNA and its duplicated segment, respectively.
FIGURE 10
Introduction

Chloroplast DNA (cpDNA) of land plants is highly conserved in size and structure, gene content (about 120 genes), and nucleotide sequence (Palmer 1985, 1991; Zurawski and Clegg 1987), contributing to its utility in molecular evolutionary research. The circular chromosome of photosynthetic land plants ranges in size from about 120 to 217 kilobase pairs (kb), but the average size for most angiosperms is approximately 160 kb (Palmer 1991; Sugiura 1992). A prominent feature of cpDNA is a large repeated sequence, the two copies of which exist in opposite orientations, separating the remainder of the molecule into large and small single copy regions. This inverted repeat (IR) is generally about 25 kb in angiosperms, but ranges in size from approximately 10-76 kb (Palmer 1991).
Certain land plants have lost one copy of the IR, including six tribes of legumes (Lavin et al. 1990; Palmer et al. 1987b), conifers (Lidholm et al. 1988; Raubeson and Jansen 1992; Strauss et al. 1988), two genera in the Geraniaceae (Palmer 1991), one genus in the Scrophulariaceae (Palmer 1991), and the non-photosynthetic angiosperm Conopholis (Palmer 1991). Although gene content and extent of the IR is rather fixed in angiosperms, expansion/contraction of the IR has occurred in some taxa. Extreme examples include geranium, whose IR has become about three times the size of that of most other angiosperms (Palmer et al. 1987a), and Coriandrum, which has an IR less than half the typical size (Downie and Palmer 1992).

Although cpDNA structure is highly conserved, various types of genome rearrangements have been observed. In addition to the IR mutations mentioned above, other types of cpDNA structural rearrangements include inversions, insertions/deletions, and possible transpositions (Palmer 1991; Palmer et al. 1988a). In land plants, few cases of chloroplast gene or intron loss have been reported, and rearrangement of genes is fairly rare (Downie and Palmer 1992; Palmer 1991). Examples of cpDNA gene loss (for a more complete review see Downie and Palmer 1992) include the transfer of rpl22 to the nucleus in members of the
Fabaceae (Downie and Palmer 1992; Palmer 1991; Spielman et al. 1988), and the independent losses of rps16 and accD (ORF512) in several unrelated angiosperm genera or families (Downie and Palmer 1992; Palmer 1991).

In contrast with gene losses, most length mutations in cpDNA are small 1-10 base pair (bp) insertions or deletions that occur in non-coding regions (Golenberg et al. 1993, Palmer 1985, 1991), mainly resulting from replication errors (Golenberg et al. 1993, Palmer 1991; Zurawski et al. 1984). Larger insertions/deletions of non-coding regions occur less frequently and are likely due to recombinational processes (Ogihara et al. 1988; Palmer 1991). Very rarely can coding sequences tolerate length mutation. Insertions/deletions sustained in protein genes must generally preserve the reading frame, and length mutations in rDNA are primarily found in single-stranded loop segments of rRNA (Palmer 1991).

The arrangement of genes in cpDNA is rather conserved in land plants, with atypical genomes deviating by only one to a few inversions. For example, three inversions characterize all examined members of the Poaceae (grasses) (Doyle et al. 1992; Hiratsuka et al. 1989; Howe 1985; Quigley and Weil 1985), and a 22 kb derived inversion distinguishes most taxa in the Asteraceae from subfamily Barnadesioideae (Jansen and Palmer 1987a,b). All species
of the Fabaceae (legumes) share a 50 kb inversion (Palmer and Thompson 1982), and members of Phaseoleae subtribe Phaseolinae (subfamily Papilionoideae) are characterized by a 78 kb inversion (Bruneau et al. 1990; Palmer et al. 1988b). Two other lineages of legumes have extensive rearrangements (Milligan et al. 1989; Palmer et al. 1987b; Palmer et al. 1988b) that occurred subsequent to the loss of one copy of the IR, as is the independent case in conifers (Lidholm et al. 1988; Raubeson and Jansen 1992; Strauss et al. 1988). Although most cases of extreme genome scrambling are found in taxa that have lost one copy of the IR, certain exceptions exist. Taxa with five or more inversions that also have an intact IR include two species of Lobelia (Knox et al. 1993), geranium (Palmer et al. 1987a), and two species of Clematis (Hoot and Palmer in press).

Alterations of gene order in land plant cpDNA usually result from inversions or gene deletions, but some evidence suggests that transposition may be responsible in some instances. Milligan et al. (1989) reported data suggestive of at least two transpositions in subclover cpDNA. One of these is a duplicative transposition of a partial copy of the ribosomal protein gene rpl23, and the second appears to be the transposition of the genes rpoB and rpoC1 similar to a transposable element. Another case
involves the putative transposition of small segments of certain ribosomal protein genes to new positions in the chloroplast genomes of some wheat species (Bowman et al. 1988; Ogihara et al. 1988).

The presence of several uncharacterized cpDNA rearrangements in two closely related angiosperm families, Campanulaceae and Lobelieaceae, was reported several years ago following a survey of restriction site mutations (Jansen and Palmer 1987b; Palmer 1985; Downie and Palmer 1992). We have found that the chloroplast genomes in members of the Campanulaceae are drastically rearranged relative to those of most land plants. Over 40 inversions, at least eight putative transpositions, expansion and contraction of the IR, one presumed gene loss, deletions in two large open reading frames (ORFS), about 18 large insertions, and numerous smaller insertions have been involved in cpDNA evolution in the Campanulaceae.

Materials and Methods

Total DNA was isolated from one species in each of 18 genera in the Campanulaceae (Table 4) according to the
CTAB isolation method of Doyle and Doyle (1987). Agarose gel electrophoresis, bidirectional transfer of DNA onto Zetabind (AMF CUNO) nylon membranes, labelling of recombinant plasmids by nick translation, filter hybridizations, and autoradiography were carried out essentially as described by Palmer (1986). DNAs were digested with the restriction endonucleases BamHI, BglII, EcoRI, EcoRV, HindIII, and SstI, and double digests were carried out using HindIII and the remaining five enzymes. Digestion products were electrophoresed in 1.2% agarose gels until the bromphenol blue tracking dye had migrated 15 cm. Single and double digests were produced in triplicate for all taxa, yielding six identical filters for each species following bidirectional southern blotting of gels. Hybridization probes consisted of 106 small tobacco cpDNA probes averaging 1.2 kb in size (Fig. 11) that were subcloned (provided by J. Palmer) from larger cloned cpDNA fragments (Olmstead and Palmer, 1992; Sugiura et al. 1986). In addition, 21 cloned HindIII cpDNA fragments from Trachelium caeruleum (Campanulaceae) (Chapter 2) were used in hybridization experiments. Contiguous single and double digest fragments were mapped using the overlap hybridization technique described in Palmer (1986).
Phylogenetic methods are described in detail in Chapter 4. In brief, these analyses were complicated by the fact that it was not possible to determine which inversions were shared among taxa in many instances. Initially all inversion characters were scored as endpoints (hence, some inversions may have been scored twice), whereas other structural changes were defined as a single event. This approach resulted in 84 characters (Chapter 4), which were examined using heuristic parsimony analyses. Three additional parsimony analyses were performed by character weighting or recoding. The first modification involved giving a weight of two to any character that was not an inversion endpoint, which would compensate for coding inversions as two characters. The second modification recoded inversion characters by utilizing knowledge of the order of inversions in some groups. The third modification combined both character weighting and recoding. The parsimony analyses were conducted using PAUP version 3.1.1 (Swofford 1993) on a Macintosh IIci computer with the Tree Bisection Reconnection (TBR), ACCTRAN, and MULPARS options, using tobacco as the outgroup.
Results

Physical and gene mapping

Hybridizations using 106 small cloned tobacco cpDNA fragments (Fig. 11) were used to map cpDNA relative to tobacco for the 18 taxa listed in Table 4. The tobacco probes contain parts or all of one to a few genes and associated non-coding sequence, providing differential hybridization of the 5' and 3' ends of many chloroplast genes, permitting the direction of transcription to be directly determined for these genes (Fig. 12). The orientation of other genes was inferred either from their co-transcription in tobacco with genes for which 5' and 3' differential hybridization was available, or by comparisons of their known locations in operons in tobacco or other plants compared to their similar placement in Campanulaceae. Because many tobacco probes contain more than one gene, it was not always possible to determine which genes are located at inversion endpoints. The locations of some genes were inferred by taking a conservative approach that preserved polycistronic transcription units known from other plants.
Complete cpDNA restriction site maps for the enzymes BamHI, BglII, EcoRI, EcoRV, HindIII, and SstI and double digests with HindIII and the other five enzymes were constructed for 16 of the 18 taxa in Table 4. Nearly complete restriction site maps were constructed for Roella and Jasione. Inadequate quantities of DNA bound to nylon membranes prohibited mapping the small single copy (SSC) region of Roella, because signals on filters became increasingly weak in later rounds of hybridization. In Jasione, restriction site maps for the SSC/IR junction and part of the SSC region could not be resolved with the data gathered.

Most tobacco fragments hybridized to DNAs of the Campanulaceae, except for four that failed to hybridize to nearly all taxa. Tobacco probes 45 and 46 (references to numbers will refer to tobacco cpDNA of Fig. 11) contain the 5' and 3' ends, respectively, of the gene accD. Neither of these clones hybridized to any DNAs in the Campanulaceae, with two exceptions; 46 hybridized relatively strongly to Codonopsis cpDNA, but only very faintly to Platycodon. This suggests that accD may have been partially to totally lost in members of the Campanulaceae, but it is also possible that the sequence has diverged significantly from that of tobacco. This was suggested for Trachelium caeruleum (Chapter 2) because
there is sufficient space in the presumed location of the gene to retain part or all of a pseudogene. This is also true for nearly all other genera examined in the family. However, given the unusually high number of large cpDNA insertions in the family (see below), it is conceivable that these regions of unidentified DNA arose independently of accD, or are even associated with its deletion. accD also appears to be absent from six genera in the closely related family Lobeliaceae (M. Cosner unpublished; Downie and Palmer 1992; Knox et al. 1993).

A 1415 bp segment of tobacco's large ORF1901 (probe 106) failed to hybridize to all genera except Codonopsis (moderately weak signal) and Platycodon (very weak signal). As described for accD, all genera have large insertions in this region, making it impossible to rule out divergence. However, sizes of the insertions in many genera (up to 11 kb in Petromarula) far exceed the space requirement for a partial pseudogene corresponding to this portion of the gene. Legousia, Triodanum, Asyneuma, and Petromarula may share an additional deletion/divergence within the ORF (Figs. 12 and 13) as indicated by negative hybridization of probe 105. However, most genera gave only weak hybridization signals for this probe, and because probe 105 was one of the last probes hybridized (as filters were wearing out), these data should be
treated as inconclusive. It may be significant, however, that these four genera are grouped together by several other features as well.

Strong hybridization of a 213 bp tobacco probe (78) to Codonopsis, Platycodon, and Cyananthus and no hybridization in the rest of the taxa suggest relatively recent loss of this sequence, corresponding to the 5' end of tobacco's large ORF2280. Weak hybridization in many genera of certain other fragments homologous to this gene indicate the gene may be nonfunctional in some taxa in Campanulaceae. However, ORF2280 is noted for its capacity for sustaining relatively large insertions/deletions while remaining an intact open reading frame (Blasko et al. 1988; Zhou et al. 1988).

Size variation

Chloroplast genomes in the Campanulaceae range in size from about 162-221 kb (Table 5). Differences in size are primarily due to variation in extent of the inverted repeat and the presence in many taxa of large stretches of DNA that do not hybridize to tobacco or Trachelium cpDNA. The exact sizes of two taxa, Roella and Jasione, could not be determined due to the mapping difficulties mentioned above.
Size and extent of the IR vary significantly among taxa, and in only two instances are the junctions between the IR and a single copy region the same as tobacco. All of the genera except Jasione share the same IR/SSC junction (Fig. 12). The IR has expanded about 10 kb into the SSC region at its end nearest IR A (see tobacco Fig. 11); the junction now lies between ndhE and psaC (within tobacco fragment 101). The exact SSC/IR junction could not be determined for Jasione (see above), but hybridization data indicate that its IR has expanded even further into the SSC region.

Four different large single copy (LSC) region/IR junctions were mapped in the Campanulaceae (Table 5, Fig. 12). Codonopsis and Cyananthis share the same IR/LSC junction with tobacco, spanning probe 74 (Figs. 11 and 12). Platycodon's IR has expanded into probe 61 (Fig. 12) in the LSC region nearest IR B (see tobacco Fig. 11). The IRs of all other taxa have radically contracted at the LSC end nearest IR B. All but one of these genera share the same IR/LSC junction (about a 15 kb IR contraction) that excludes tobacco's ORF2280 from the IR, but includes trnL (Fig. 12). This junction coincides with an inversion endpoint at probes 84/90 (Fig. 12). Merciera's IR/LSC junction is found at or near another inversion endpoint (probes 85/79), presumably a derived condition resulting
from regrowth of part of the IR containing tobacco's ORF2280.

Also contributing to overall genome size variation among taxa are large insertions of unidentified DNA that do not hybridize to tobacco or Trachelium cpDNA. There are a total of 18 insertions of 5 kb or larger (Table 5, Figs. 13 and 14) found in eight genera. These insertions are variously distributed in the LSC and SSC regions and the IR of different taxa (Table 5). Numerous additional insertions less than 5 kb are found in all taxa, and it is presumed that large insertions contributed to the inability to map the SSC region in Jasione.

Transposition

Evidence from cpDNA restriction site mapping suggests that at least eight transpositions have occurred during the evolution of cpDNA in the Campanulaceae. One example involves the genes clpP and 5'-rps12 (tobacco 53-56), which have ostensibly been transposed from the LSC region to the IR (Fig. 13 and Edraianthus of Fig. 12), resulting in the disruption of the clpP operon (for more details see Chapter 2). This putative transposition is found in 15 of the 18 taxa mapped in the Campanulaceae (Fig. 13). Two additional transpositions involving this region have
subsequently occurred in select genera in the family (Fig. 12); the two genes have been moved to a unique and otherwise unrearranged location in the LSC region of *Asyneuma* (Fig. 12). The transposition of these genes left behind presumably non-coding (as determined by faint hybridization) segments of probes 53 and 56 (Fig. 12), allowing the sequence of events to be readily interpreted. In a separate event shared by the genera *Wahlenbergia*, *Merciera*, *Prismatocarpus*, and *Roella*, part of the transposed region corresponding to probe 53 and part of 54 was relocated to a unique site in the LSC region (Figs. 12 and 13). Although *clpP* (probes 54-56) most certainly remained in the IR, it is not known if 5'-*rps12* (probe 54) was moved to the LSC region. This seems likely, however, since strong hybridization occurred for probe 54 in both the IR (*clpP*) and in the LSC region (5'-*rps12*).

Other putative transpositions include the relocation of DNA sequences homologous to probe 16 and part of 17 in *Codonopsis*, fragment 28 in *Cyananthus*, and 5-8 in *Legousia* (Figs. 12 and 13). The putative transposition in *Legousia* has apparently disrupted both the genes *matK* and *trnK*, since portions of these genes are found in tobacco 3-5. In *Platycodon*, the disruption and scattering throughout the genome of DNA hybridizing to clones 6-9 may also be
the result of transposition (Figs. 12 and 13). This has presumably resulted in the disruption of the gene rps16.

What appears to be a duplicative transposition of a small segment of the 23S rRNA gene is shared by the taxa Trachelium, Campanula, Symphyandra, Jasione, and Adenophora. A tobacco probe (93) internal to the gene hybridizes to both the 23S rDNA and to a location over 10 kb downstream from the gene in the IR. The data indicate that this duplicative transposition is also shared by Edraianthus, but comparatively faint hybridization in this genus suggests subsequent partial loss or divergence of the duplicated DNA.

Inversions

At least 42 inversions were found in the 18 chloroplast genomes mapped in the Campanulaceae. Linear maps of the 18 taxa showing only the order of the tobacco probes are diagrammed in Fig. 13. In many instances, individual tobacco fragments hybridized in several locations in the chloroplast genome. Many of these were easily explained due to the obvious splitting of the region by a rearrangement endpoint. In some cases, the complex patterns of hybridization were not as readily interpreted, and were presumably due to "endpoint
"remnants" leftover from other inversions, transposition of small pieces of DNA, or small dispersed repeats within the genomes. At least five such families of repeats have been documented in *Trachelium caeruleum* (Chapter 2). It is important to note that such extraneous hybridization patterns due to repeats, small transpositions, inversion leftovers, or other undetermined phenomena are not shown in the cpDNA maps of Figs. 12 or 13.

The estimate of at least 42 inversions assumes that many of these are shared among taxa; however, a greater number may have actually occurred. Because of the unusually high number of inversions, it was impossible to unambiguously order all of the events. Within small groups of taxa it was possible to determine a probable sequence of inversions. A phylogenetic tree depicting possible evolutionary relationships among the taxa based on inversions and other rearrangements is given in Fig. 15. The tree was constructed using a combination of cladistic parsimony analyses using inversion endpoints and other rearrangements as characters, combined with observations suggesting a logical ordering of events in certain instances (see discussion and Chapter 4). This method produced two equally parsimonious trees, and the strict consensus tree is shown in Fig. 15. Evolutionary scenarios involving series of rearrangements (Figs. 16-20)
could then be constructed that are compatible with both the cladogram and cpDNA mapping data, and are discussed in detail below.

Discussion

The chloroplast genomes of members of the Campanulaceae are the most rearranged of any land plants examined. A minimum of 42 inversions, five separate IR expansion/contraction events, more than eight putative transpositions, 18 large insertions greater than 5 kb, numerous large insertions less than 5 kb, one presumed gene loss, and two deletions within large ORFs have occurred during cpDNA evolution in the family. The number of rearrangements far exceeds that reported in any other group, making it difficult to determine the exact numbers and evolutionary sequence of rearrangement events. When inversions have been reported in other families, they are generally few in number and easily characterized; the sequence of events is readily determined in most cases. The number of rearrangements in the Campanulaceae has been estimated conservatively, and it must be emphasized that a far greater number of events may have occurred. It was
not possible to determine the evolutionary sequence of events in some cases, although certain inversions could be ordered with confidence.

Evolutionary relationships among chloroplast genomes

The high number of cpDNA rearrangements in the Campanulaceae makes it difficult to determine the evolutionary sequence of inversions. In some cases, the order of inversions is clear, providing useful information regarding comparative chloroplast genome evolution. In other instances, chloroplast genomes bear little resemblance to each other, making it nearly impossible to envision a series of evolutionary steps between one genus and another. To assist in developing an evolutionary scenario, we performed parsimony-based cladistic analyses of different codings of rearrangement characters. The fact that cladograms from all four analyses are congruent and exhibit low levels of homoplasy (Chapter 4) gives us confidence that these trees are a robust estimate of phylogenetic relationships among chloroplast genomes of the Campanulaceae. This phylogenetic hypothesis permits the construction of an evolutionary scenario (Figs. 16-20) that is consistent with both the tree topology and cpDNA mapping data.
The 18 examined genera of the Campanulaceae share three derived cpDNA mutations relative to tobacco (Fig. 15). The first is an expansion of the IR into the SSC region (Figs. 12 and 13), a condition also shared by the closely related family Lobeliaceae (M. Cosner unpublished; E. Knox pers. comm.). All taxa also share the deletion or divergence of all or part of the gene accD, and the family is united by one inversion (Fig. 15).

Fifteen of the 18 genera share a drastic contraction of the IR at its LSC end, six inversions, and the putative transposition of clpP and 5'-rps12 (probes 53-56) into the IR (Fig. 15). The absence of these features in Platycodon, Codonopsis, and Cyananthus strongly indicates their basal position in the Campanulaceae. Four groups are delimited among the 15 taxa with more derived chloroplast genomes.

Figure 16 depicts the most parsimonious evolutionary scenario constructed (Chapter 2) to explain the evolution of cpDNA in Trachelium caeruleum from a tobacco-like ancestor. The series of events includes seven inversions, one transposition (probes 53-56), and expansion/contraction of the IR. In the simplest scheme, all 15 advanced taxa share each of these rearrangements (Figs. 13 and 15). Each of the advanced groups in Fig. 15
is characterized by additional rearrangements beyond those shown in Fig. 16.

When using Trachelium's chloroplast genome as a starting point (Fig. 16), a logical progression of inversions can be hypothesized for each of the other groups (Fig. 15) using the following rationale. An inversion is considered derived if it disrupts a series of tobacco probes that is contiguous in Trachelium (and tobacco), that is, if one or both endpoints disrupts an unrearranged sequence block. Wahlenbergia has evolved three additional inversions relative to Trachelium (Fig. 17); three inversion endpoints break up sequence blocks that are intact in Trachelium (Fig. 17). This group is also delimited by the transposition of 5'-rps12 from the IR to the LSC region (Figs. 12 and 13). Three additional inversions unite Prismatocarpus, Merciera, and Roella (Fig. 17).

The unique IR configuration of Merciera is presumed to be derived within this clade. Although the extent of Merciera's IR at its LSC end is intermediate in size between those of Cvananthus and Codonopsis, and the derived genera, its genome structure indicates the species is not evolutionarily intermediate. Merciera's IR expanded subsequent to the origin of structural changes it shares with Roella and Prismatocarpus (Fig. 15).
The genera *Asyneuma*, *Petromarula*, *Triodanu*s, and *Legousia* are united (Fig. 15) by a mutation that placed an approximately 25 kb segment of what is normally the LSC region into the SSC region (Fig. 18). This rearrangement may have resulted from either a transposition or two inversions (Fig. 15 arbitrarily indicates two inversions). In an inversion scenario, one inversion would relocate the sequence to the SSC region, followed by a second to reinvert the remainder of the genome. This is very difficult to envision mechanistically, since it is unclear what effect such inversions would have on the IR. It is also difficult to imagine such a large region moving by transposition, and it is possible that whatever events were necessary to create such a genome happened in concert, rather than as a series of discrete steps.

The group containing *Trachelium* is characterized by several additional rearrangements beyond those depicted in Fig. 16. Although *Musschia* appears as a separate clade, it shares the same basic chloroplast genome structure as *Trachelium*. However, it does not have the 23S rDNA duplicative transposition that unites all taxa in the *Trachelium* group (Fig. 15). Figure 19 outlines a possible series of events indicating relationships within the group. *Edraianthus* and *Symphyandra* share an additional inversion that disrupts an intact sequence block in
Trachelium (Fig. 19A). Campanula and Adenophora are united by a single inversion (different than the one joining Edraianthus/Symphvandra) that also separates contiguous probes (Fig. 19B). Two additional derived inversions resulted in Adenophora's chloroplast genome structure (Fig. 19B).

The most rearranged chloroplast genome in the family is undoubtedly that of Jasione (Fig. 13). It is unrecognizable as a member of the Trachelium group except for the presence of the 23S rDNA duplicative transposition. At least eight more inversions (beyond the seven shared by all 15 advanced taxa) and the clearly derived IR expansion into the SSC region are needed to modify Jasione's cpDNA from a Trachelium-like ancestor (Fig. 19C, inversions not shown). It may share at least one of these inversions with Adenophora, since both taxa share a derived endpoint (37/28) relative to the rest of the Trachelium group (Fig. 13).

Relationships of chloroplast genomes among Platycodon, Codonopsis, and Cyananthus, and how they relate to the 15 advanced taxa are more difficult to determine. Most inversions in these three taxa appear unrelated to those in the remainder of the family, although this is difficult to say with certainty. The genomes of Cyananthus and Codonopsis are more similar to
each other than either is to *Platycodon*, and none of the three bear much resemblance to the advanced genera (Fig. 13). Only one sequence block in the LSC region of *Platycodon* is found in *Cyananthus* and *Codonopsis* (Fig. 13). This consists of tobacco probes 41-44, a sequence that remains intact in all genera in the family (Fig. 13). The cladogram suggests that an inversion creating this block of genes was shared by all members of the Campanulaceae (inversion 1 in Fig. 20); this inversion has one endpoint at the site of deletion/divergence of the gene accD (probes 45-46). Inversion endpoints are also found in this location in the Lobeliaceae, but the actual inversions apparently are different (M. Cosner unpublished; Knox et al. 1993).

A series of events depicting evolutionary relationships among *Platycodon*, *Codonopsis*, and *Cyananthus* compatible with the cladogram is given in Fig. 20. In addition to the single family-wide inversion, *Platycodon*, *Codonopsis*, and *Cyananthus* share one inversion, and *Platycodon* has five unique inversions, for a total of seven (Fig. 20). *Codonopsis* and *Cyananthus* share three additional inversions (a total of five shared), *Codonopsis* has two more inversions, and *Cyananthus* three (Fig. 20).

It must be emphasized that we have constructed several other possible scenarios, but none as compatible
with the cladogram (Fig. 15) as these (Figs. 16-20). The actual number of inversions and their order are unknown, although our interpretation indicates 42 inversions occurred. Although the cladogram indicates that only one inversion is shared by all taxa in the Campanulaceae, it is possible that other inversions unite the family, but that endpoints of ancient inversions have been obliterated by more recent rearrangements, rendering older inversions undetectable in some or all taxa. It is also possible that more inversions are shared between basal and derived taxa. For example, *Platycodon* shares a sequence block (probes 25-16) that is ubiquitous in the 15 more advanced taxa (Fig. 13). However, it is impossible to say if shared or separate inversions were responsible for this similarity.

The proposed evolutionary histories presented in Figs. 16-20 do not depict events, other than inversions, that are unique to individual taxa. However, most taxa have features not found in other genera. Three genomes have IR boundaries not shared by other taxa, some have putative transpositions, and all genera have large insertions of unknown origin. Only the locations of insertions 5 kb or larger are shown in Fig. 13. These are also given in Fig. 15, but insertions in the region of tobacco's ORF1901 are not shown on the tree because most
taxa have at least small insertions in this location, and it cannot be determined if any of the events are shared (see below).

Length mutation

Chloroplast genomes in the Campanulaceae range in size from approximately 162-221 kb (Table 5). Many of these are considerably larger than is typical for angiosperms, and although some of the size variation can be attributed to differences in IR length, much of the variation is due to large insertions in most genera. Although these are referred to here as insertions, their mode of origin is unknown. Land plant cpDNA is noted for its compact nature and lack of major size variation; there have been no reports of nongenic insertions larger than 1 kb or insertions of foreign DNA (Palmer 1991). It has been suggested that selection may operate on land plant cpDNA to maintain small genome size and to eliminate unnecessary DNA sequences (Palmer 1985, 1991).

All genomes in the Campanulaceae have insertions of 1 kb or larger, and these range up to nearly 20 kb (Table 5). The insertions appear to be taxon-specific, that is, they do not hybridize to tobacco or Trachelium cpDNA probes. In some instances, proposed lengths (Table 5) of
insertions are the minimum possible. Restriction site maps for regions bordering the insertions could not always be linked together, since insertions are taxon-specific. However, comparative mapping within the family and the use of double digests greatly increased mapping resolution and allowed for greater confidence in interpreting the extents of the insertions (Fig. 14).

In no instance have such sizable cpDNA insertions of unknown content been discovered in land plants. Several small insertions unique to subclover cpDNA were found by Milligan et al. (1989), but none were bigger than 1 kb. The large insertions in Campanulaceae are comparable in size to those found in cpDNA of certain green algae such as *Chlamydomonas* (Turmel et al. 1987); algal chloroplast genomes are hypothesized to evolve under more relaxed selective constraints (Palmer 1985, 1991). It seems likely that recombination involving short dispersed repeats may contribute to the formation of length mutations (reviewed in Palmer 1985, 1991). Other than the large IR, repeated sequences are rare in land plant cpDNA, but are more commonly found in algal chloroplast genomes, along with an increase in variously sized insertions (Palmer 1985, 1991). At least five families of dispersed repeats exist in *Trachelium caeruleum* (Campanulaceae) cpDNA (Chapter 2). It is possible that recombination
involving repeated sequences may be responsible for the large insertions in the Campanulaceae.

As mentioned above, mapping unique DNA sequences using heterologous probes is problematic since linear portions of maps bordering the insertions can not always be linked together. However, mapping even the largest insertions in genera such as *Adenophora* (Fig. 14) was possible because insertions consisted of large regions containing no restriction sites for the enzymes surveyed. This is suggestive of the amplification of non-coding, possibly repeated sequences resulting from recombinational events (Palmer et al. 1985, Turmel et al. 1987). However, there is no evidence of the length heterogeneity that is characteristic of unequal crossing-over between tandem repeats, as is found in *Euglena* cpDNA (eg. Hallick and Buetow 1989; Schlunegger and Stutz 1984; Siemeister and Hachtel 1989). At present, the insertion of foreign DNA can not be ruled out.

Not all of the insertions consist of large regions devoid of restriction sites. In some cases, complete maps within insertions are not available because the locations of restriction sites for more frequent cutting enzymes could not be determined within the borders of the insertions (Fig. 14). However, the overall size of an insertion could still be determined if a large enough
restriction fragment of a less frequently cutting endonuclease spanned the region (see Musschia, Fig. 14). In some instances, it was possible to estimate only the minimum size of insertions.

All of the large insertions appear to be taxon-specific, that is, they are not shared by two or more genera. One region of the chloroplast genome seems to be an insertional "hot-spot". Most genera have additional DNA within tobacco's large ORF1901 near its 5' end, seemingly associated with the loss or divergence of a large part of this ORF. Although the general locations are the same, the sizes of the insertions vary greatly among genera, and it is therefore impossible to say if any of the additions share a common origin.

In addition to large insertions, variation in genome size can also be attributed to differences in IR size. Inverted repeats in the Campanulaceae range in size from 20 to 42 kb (Table 5). This size variation is due to 1) expansion or contraction of the IR, 2) changes in gene content due to transpositions or inversions, or 3) large insertions.

Although the IR/single copy junctions are relatively fixed in angiosperms, variation does occur. For example, the IR of Coriandrum sativum is less than half the size of those of typical angiosperms, having contracted about 10
kb at its LSC end (Downie and Palmer 1992). The IR of *Oryza sativa* has expanded into single copy regions in three of the four possible directions, but several deletions within the IR have resulted in an IR smaller than tobacco's (Hiratsuka et al. 1989). The drastic expansion of the IR of *Pelargonium* into large and small single copy regions increased IR size to approximately three-fold that of other angiosperms (Palmer et al. 1987a). However, in no case has the IR expanded to duplicate the gene *psbA* in angiosperms (Palmer 1985; Stein et al. 1992); the IR A/LSC junction appears the most evolutionarily conserved of all four IR/single copy junctions.

In contrast, the IR in several genera of ferns has expanded to include *psbA*, followed by at least two inversions within the IR (Stein et al. 1992). The *psbA* gene is also located in the IR of certain IR-containing algae such as some species of *Chlamydomonas* (Lemieux and Lemieux 1985; Turmel et al. 1987; Woessner et al. 1987), certain diatoms (Bourne et al. 1992; Kowallik 1989), *Chlorella ellipsoidea* (Yamada 1991), and the marine chromophyte *Olisthodiscus luteus* (Reith and Cattolico 1986).

Inverted repeat expansion or contraction has occurred at two IR/single copy junctions in the Campanulaceae (Fig.
12). A total of five separate expansion/contraction events have occurred, three of which are unique to different genera (Jasione, Merciera, and Platycodon). None of these events occurred at the IR A/LSC junction, and psbA remains single copy in the Campanulaceae. The expansion of IR A into the small single copy region is shared by the closely related family Lobeliaceae (M. Cosner unpubl.; E. Knox pers. comm.)

Transposition

Changes in gene order in angiosperm cpDNA are usually the result of inversions or gene loss. Some evidence indicates that transposition may also be a factor in chloroplast genome evolution. We propose that at least eight cases of transposition have occurred during cpDNA evolution in the Campanulaceae. This estimate of the frequency of transposition events is very conservative, and it is likely that a higher number actually occurred. Two of these are diagrammed and discussed in greater detail in Chapter 2. One is a putative transposition of the genes clpP and 5'-rps12 from the LSC region to the IR (found in 15 genera), and the second involves the duplicative transposition of a small portion of the 23S rRNA gene (shared by six taxa).
An attempt was made to attribute genome rearrangements in Campanulaceae to inversions whenever reasonable. Several incidents of cpDNA inversion have been well-documented in various families (see Downie and Palmer 1992 for a review), but transposition has rarely been suggested as an evolutionary mechanism in chloroplast genomes, and most cases are not well-documented. However, in the Campanulaceae, some genome arrangements are more consistent with transposition than inversion. For example, there are at least two cases in which inversions between the IR and LSC region would be necessary to produce the genome order found. Although it may be possible, we feel such an inversion would be very unlikely since it would be extremely disruptive to the IR (Chapter 2). One case, the movement of clpP and 5'-rps12 (tobacco 53-56) from the IR to the LSC region is discussed in detail in Chapter 2. A second case involves the relocation of the sequence homologous to tobacco probes 5-8 from the LSC region to the IR in Legousia (Fig. 13).

Transposition may also have resulted in genome arrangements in which a small stretch of DNA is missing from an otherwise unrearranged stretch of DNA and placed in a unique location (frequently another unrearranged region). For example, in Codonopsis, probes 16-17 are absent from an otherwise intact sequence (9-36), but are
present within another unrearranged sequence consisting of probes 46-56 (Fig. 12). In order for this to occur by inversion, one inversion would place 16-17 in its current location, followed by a second to reinvert the remainder of the sequence. Although this is possible, it seems less likely, and requires an additional step.

In addition to those discussed above, at least five other cases of presumed transposition have occurred (see results and Figs. 12 and 13). It is not known if any of the transpositions were DNA or RNA mediated; evidence in Trachelium (Chapter 2) indicates that either is possible in the transposition of *clpP* and 5'-*rps12* (53-56). However, two further transpositions of parts of this sequence suggest a mechanism similar to a transposon may have been involved.

Conclusions

Chloroplast genome structure is typically highly conserved among land plants. When cpDNA rearrangements have been discovered, they are generally few in number, easily characterized, and devoid of homoplasy. In most of the few cases in which chloroplast genomes are highly rearranged, one copy of the IR has been lost, presumably resulting in destabilization of the molecule (Palmer 1985,
1991). The Campanulaceae have the most rearranged chloroplast genomes discovered to date, yet all taxa examined retain an IR. It is clear that other factors must be involved in destabilizing the genomes in the Campanulaceae. Small dispersed repeats and the putative loss of accD were implicated as possible mechanisms for rearrangement in Trachelium caeruleum (Chapter 2).

Due to the unprecedented number of inversions in the Campanulaceae, determining the evolutionary order of events is extremely difficult. The evolutionary hypotheses presented here are only one set of possible interpretations of the data. Although we have attempted to produce the shortest evolutionary schemes and to avoid homoplasy, it is very possible (and perhaps likely) that longer, more complicated scenarios actually occurred, in which some inversions occurred in parallel, or in which different series of inversions produced similar genome structures in different taxa. Despite the difficulties in interpreting such a complex set of rearrangements, the utility of chloroplast DNA in the Campanulaceae as an evolutionary tool is evident. The data presented here have produced the most highly resolved intrafamilial phylogeny based entirely on cpDNA structural features. The Campanulaceae also provide a model system for further
studies of the mechanisms of evolution of chloroplast DNA structural changes.
LITERATURE CITED


### Table 4. Species mapped for chloroplast DNA structural rearrangements.

<table>
<thead>
<tr>
<th>Species</th>
<th>Source/Voucher</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Adenophora confusa</em> Nannf.</td>
<td>M. Cosner 177</td>
</tr>
<tr>
<td><em>Asyneuma virgatum</em> (Labill.) Bornm.</td>
<td>Berlin-Dahlem 0104*</td>
</tr>
<tr>
<td><em>Campanula elatines</em> L.</td>
<td>T. Ayers/BH88-287</td>
</tr>
<tr>
<td><em>Cyananthes viridis</em> Wall. in Roxb.</td>
<td>T. Ayers/BH88-229</td>
</tr>
<tr>
<td><em>Cyananthes lobatus</em> Wall.</td>
<td>M. Cosner 179</td>
</tr>
<tr>
<td><em>Edraianthus graminifolius</em> (L.) A.DC.</td>
<td>T. Ayers/BH88-195</td>
</tr>
<tr>
<td><em>Jasione heldreichii</em> Boiss. &amp; Orph.</td>
<td>T. Ayers/BH88-208</td>
</tr>
<tr>
<td><em>Legousia falcata</em> (Ten.) Fritsch</td>
<td>Berlin-Dahlem 0143*</td>
</tr>
<tr>
<td><em>Merciera tenuifolia</em> DC.</td>
<td>K. Steiner 2445</td>
</tr>
<tr>
<td><em>Musschia aurea</em> Dum.</td>
<td>T. Ayers/BH88-274</td>
</tr>
<tr>
<td><em>Petromarula pinnata</em> (L.) DC.</td>
<td>T. Ayers</td>
</tr>
<tr>
<td><em>Platycodon grandiflorum</em> DC.</td>
<td>T. Ayers/BH88-216</td>
</tr>
<tr>
<td><em>Prismatocarpus diffusus</em> DC.</td>
<td>K. Steiner 2448</td>
</tr>
<tr>
<td><em>Roella ciliata</em> L.</td>
<td>T. Ayers</td>
</tr>
<tr>
<td><em>Symphyandra hofmannii</em> Pant.</td>
<td>T. Ayers/BH88-225</td>
</tr>
<tr>
<td><em>Trachelium caeruleum</em> L.</td>
<td>M. Cosner 173</td>
</tr>
<tr>
<td><em>Triodanus perfoliata</em> (L.) Nieuwl.</td>
<td>M. Cosner 178</td>
</tr>
<tr>
<td><em>Wahlenbergia gloriosa</em> Loth.</td>
<td>T. Ayers/BH88-217</td>
</tr>
</tbody>
</table>

*Botanischer Garten and Botanisches Museum, Berlin-Dahlem*
Table 5. Size variation in Campanulaceae chloroplast DNA including genome sizes (in kb), size and extent of IR, and insertions larger than about 5 kb. Sizes of insertions are followed by their general locations in the genome.

<table>
<thead>
<tr>
<th>Genus</th>
<th>Size</th>
<th>IR&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Insertions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adenophora</td>
<td>221</td>
<td>30 (3)</td>
<td>5 (IR), 6 (LSC), 9 (LSC), 18 (LSC), 19 (SSC)</td>
</tr>
<tr>
<td>Asyneuma</td>
<td>167</td>
<td>20 (3)</td>
<td>5 (LSC)</td>
</tr>
<tr>
<td>Campanula</td>
<td>191</td>
<td>37 (3)</td>
<td>7 (IR)</td>
</tr>
<tr>
<td>Codonopsis</td>
<td>180</td>
<td>40 (1)</td>
<td></td>
</tr>
<tr>
<td>Cyananthus</td>
<td>168</td>
<td>38 (1)</td>
<td></td>
</tr>
<tr>
<td>Edraianthus</td>
<td>163</td>
<td>27 (3)</td>
<td></td>
</tr>
<tr>
<td>Jasione</td>
<td>&gt;169</td>
<td>&gt;30 (5)</td>
<td></td>
</tr>
<tr>
<td>Legousia</td>
<td>186</td>
<td>35 (3)</td>
<td>5 (IR)</td>
</tr>
<tr>
<td>Merciera</td>
<td>175</td>
<td>32 (4)</td>
<td></td>
</tr>
<tr>
<td>Musschcia</td>
<td>213</td>
<td>42 (3)</td>
<td>9 (LSC), 15 (LSC), 16 (IR)</td>
</tr>
<tr>
<td>Petromarula</td>
<td>195</td>
<td>32 (3)</td>
<td>8 (SSC), 9 (SSC), 11 (IR)</td>
</tr>
<tr>
<td>Platycodon</td>
<td>172</td>
<td>41 (2)</td>
<td></td>
</tr>
<tr>
<td>Prismatocarpus</td>
<td>202</td>
<td>30 (3)</td>
<td>5 (LSC), 5 (LSC), 15 (LSC)</td>
</tr>
<tr>
<td>Roella</td>
<td>&gt;145</td>
<td>&gt;21 (3)</td>
<td></td>
</tr>
<tr>
<td>Symphyandra</td>
<td>167</td>
<td>30 (3)</td>
<td></td>
</tr>
<tr>
<td>Trachelium</td>
<td>162</td>
<td>27 (3)</td>
<td></td>
</tr>
<tr>
<td>Triodanmus</td>
<td>183</td>
<td>32 (3)</td>
<td>7 (IR)</td>
</tr>
<tr>
<td>Wahlenbergia</td>
<td>165</td>
<td>26 (3)</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>IR types (see text for details) follow IR size in parentheses: Types 1-4: SSC/IR junction tobacco probe 101; 1) LSC/IR junction probe 74; 2) LSC/IR junction probe 61; 3) LSC/IR junction probes 84/90; 4) LSC/IR junction probes 85/79; 5) SSC/IR junction unknown (but see text), LSC/IR same as type 3.

<sup>b</sup>See text for explanation.
Fig. 11. Chloroplast DNA gene map of *Nicotiana tabacum* (tobacco). The circular chromosome has been linearized and separated onto four lines (A-D). Genes above the line are transcribed from left to right and those below the line are transcribed from right to left. Numbers in blocks below genes represent 106 tobacco cpDNA hybridization probes used in this study. Numbers above probes designate tobacco coordinates in kb.
FIGURE 11
Fig. 12. Chloroplast DNA restriction site maps of five species (Table 4) in the Campanulaceae for three enzymes (S=SstI, H=HindIII, EV=EcoRV) illustrating different extents of the IR and the results of three transpositions involving clpP and 5'-rps12 (tobacco probes 53-56). Numbered blocks indicate approximate sizes (in kb) and locations of restriction fragments; the sizes of unnumbered fragments could not be directly determined due to insertions. Maps for BamHI, BglII, EcoRI and double digests available but not shown. Dark bars indicate approximate location and extent of the IR. Numbers below map show hybridization of tobacco cpDNA probes (see Fig. 11); approximate extents of contiguous blocks of tobacco fragments are bounded by parentheses. Gene names indicate the approximate location of the gene's midpoint; genes above are transcribed from left to right and those below are transcribed from right to left.
<table>
<thead>
<tr>
<th></th>
<th>0.6</th>
<th>0.8</th>
<th>1.0</th>
<th>1.2</th>
<th>1.4</th>
<th>1.6</th>
<th>1.8</th>
<th>2.0</th>
<th>2.2</th>
<th>2.4</th>
<th>2.6</th>
<th>2.8</th>
<th>3.0</th>
<th>3.2</th>
<th>3.4</th>
<th>3.6</th>
</tr>
</thead>
<tbody>
<tr>
<td>s</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>15</td>
<td>16</td>
</tr>
<tr>
<td>EV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Arynamina**

<table>
<thead>
<tr>
<th></th>
<th>0.6</th>
<th>0.8</th>
<th>1.0</th>
<th>1.2</th>
<th>1.4</th>
<th>1.6</th>
<th>1.8</th>
<th>2.0</th>
<th>2.2</th>
<th>2.4</th>
<th>2.6</th>
<th>2.8</th>
<th>3.0</th>
<th>3.2</th>
<th>3.4</th>
<th>3.6</th>
</tr>
</thead>
<tbody>
<tr>
<td>s</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>15</td>
<td>16</td>
</tr>
<tr>
<td>EV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Maccari**

<table>
<thead>
<tr>
<th></th>
<th>0.6</th>
<th>0.8</th>
<th>1.0</th>
<th>1.2</th>
<th>1.4</th>
<th>1.6</th>
<th>1.8</th>
<th>2.0</th>
<th>2.2</th>
<th>2.4</th>
<th>2.6</th>
<th>2.8</th>
<th>3.0</th>
<th>3.2</th>
<th>3.4</th>
<th>3.6</th>
</tr>
</thead>
<tbody>
<tr>
<td>s</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>15</td>
<td>16</td>
</tr>
<tr>
<td>EV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Platyodon**

<table>
<thead>
<tr>
<th></th>
<th>0.6</th>
<th>0.8</th>
<th>1.0</th>
<th>1.2</th>
<th>1.4</th>
<th>1.6</th>
<th>1.8</th>
<th>2.0</th>
<th>2.2</th>
<th>2.4</th>
<th>2.6</th>
<th>2.8</th>
<th>3.0</th>
<th>3.2</th>
<th>3.4</th>
<th>3.6</th>
</tr>
</thead>
<tbody>
<tr>
<td>s</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>15</td>
<td>16</td>
</tr>
<tr>
<td>EV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**FIGURE 12**

138
Figure 12 (continued)
Fig. 13. Linear cpDNA maps for 18 species (Table 4) in Campanulaceae showing order of hybridized tobacco probes. Lines under maps represent approximate location and extent of IR. Asterisks indicate 23S rDNA duplicative transposition; (*) is partial deletion/divergence of 23S rDNA transposition. Size and location of large insertions designated by "i" followed by size in kb (insertions less than 5 kb not shown). Underlined numbers indicate sequence absent from an otherwise intact block of probes. Ambiguous hybridization of tobacco probes not shown (see text).
Figure 13 (continued)

**Triodanus**


**Wahlenbergia**

(1-11)\(60-56\)\(53-49\)\(37-40\)\(35-28\)\(11-15\)\(76-60\)\(27-26\)\(44-41\)\(47-48\)\(36-35\)\(53,54\)\(25-16\)\(90-84\)\(79-84\)\(90-96\)\(56-54\)\(105-101\)\(100-97\)\(101-105\)\(54-56\)\(96-90\)

**Merciera**

(1-10)\(49-53\)\(28-35\)\(40-37\)\(60-56\)\(11-15\)\(76-60\)\(27-26\)\(44-41\)\(47-48\)\(36-35\)\(53,54\)\(25-16\)\(17-16\)\(90-85\)\(79-84\)\(90-96\)\(56-54\)\(105-101\)\(100-97\)\(101-105\)\(54-56\)\(96-90\)

**Prismalocarpus**

(1-10)\(49-53\)\(28-35\)\(40-37\)\(60-56\)\(11-15\)\(76-60\)\(27-26\)\(44-41\)\(47-48\)\(36-35\)\(53,54\)\(25-16\)\(17-16\)\(90-85\)\(79-84\)\(90-96\)\(56-54\)\(105-101\)\(100-97\)\(101-105\)\(54-56\)\(96-90\)

**Roelina**

(1-10)\(49-53\)\(28-35\)\(40-37\)\(60-56\)\(11-15\)\(76-60\)\(27-26\)\(44-41\)\(47-48\)\(36-35\)\(53,54\)\(25-16\)\(17-16\)\(90-85\)\(79-84\)\(90-96\)\(56-54\)

**Jasione**


**Codonopsis**

(1-8)\(36-9\)\(40\)\(56-60\)\(37-39\)\(44-41\)\(46-56\)\(61-76\)\(96-77\)\(106-101\)\(100-97\)\(101-106\)\(77-95\)\(76-74\)

**Cyananthus**

(1-8)\(28\)\(36-26\)\(40\)\(56-60\)\(37-39\)\(25-9\)\(37\)\(44-41\)\(47-48\)\(56-49\)\(61-96\)\(105-101\)\(100-97\)\(101-105\)\(96-74\)

**Platycodon**

(1-5)\(29-36\)\(56-50\)\(28-26\)\(9\)\(50-46\)\(41-44\)\(37-40\)\(16-25\)\(10-15\)\(56-96\)\(106-101\)\(100-97\)\(101-106\)\(96-61\)
Fig. 14. Chloroplast DNA restriction site maps of portions of Asyneuma and Musschia showing large insertions. Numbers above maps indicate tobacco cpDNA probes. Enzymes: S=SstI, EV=EcoRV, H=HindIII, EI=EcoRI, Bg=BglII, Ba=BamHI.
### Musschla

<table>
<thead>
<tr>
<th></th>
<th>S</th>
<th>S/H</th>
<th>EV</th>
<th>EV/H</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>16.5</td>
<td>3.7</td>
<td>9.0</td>
<td>3.7</td>
<td>5.4</td>
</tr>
<tr>
<td></td>
<td>9.8</td>
<td>1.9</td>
<td>3.7</td>
<td>0.4</td>
<td>10.0</td>
</tr>
<tr>
<td></td>
<td>9.2</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
</tr>
</tbody>
</table>

### Adenophora

<table>
<thead>
<tr>
<th></th>
<th>S</th>
<th>S/H</th>
<th>EV</th>
<th>EV/H</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>6.5</td>
<td>4.9</td>
<td>10.5</td>
<td>4.9</td>
<td>5.7</td>
</tr>
<tr>
<td></td>
<td>27</td>
<td>20</td>
<td>5.0</td>
<td>2.1</td>
<td>1.9</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>4.6</td>
<td>19</td>
<td>10.0</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>1.8</td>
<td>0.8</td>
<td>25</td>
<td>4.6</td>
<td>1.0</td>
</tr>
</tbody>
</table>

---

**FIGURE 14**
Fig. 15. Strict consensus tree of two equally parsimonious trees based on 84 cpDNA structural characters. This tree has 88 steps and a consistency index of 0.90 (excluding autapomorphies) I=inversion, i=insertion >5 kb, T=transposition, D=deletion/divergence. Insertions in the region of tobacco's large ORF1901 not shown.
Fig. 16. Model for the evolution of *Trachelium caeruleum* cpDNA from a tobacco-like ancestor (Chapter 2). The model includes seven inversions, 1 transposition, expansion of the IR into the SSC region, and contraction of the IR at the LSC end. The top gene map and first lettered map show tobacco, and the bottom gene map, last lettered map, and ordering of tobacco probe numbers in parentheses (not to scale) show *Trachelium*. Only genes at rearrangement endpoints are given. Arrows on top gene map indicate rearrangement endpoints; those with asterisks designate the transposition. Letters correspond to blocks of genes; arrows under letters indicate orientations of gene blocks. The bars indicate location and extent of the IR. Numbered arrows designate inversion endpoints and the following map gives the resulting inverted genome structure. "T" stands for transposition. Insertions/deletions in *Trachelium* relative to tobacco are not shown.
FIGURE 16
Fig. 17. Model for evolution of cpDNA structure in Wahlengbergia, Merciera, Prismatocarpus, and Roella from a Trachelium-like ancestor, showing only inversions. Numbers in parentheses show order of hybridized tobacco cpDNA probes. Inversion endpoints are given as numbered arrows. The final map shows only features shared by the three species; autapomorphies of individual taxa not pictured.
**Trachelium**


**Inversion 1**


**Inversion 2**


**Wahlenbergia**


**Inversion 3**


**Inversion 4**


**Merciera, Prismatocarpus, Roella**


**FIGURE 17**
Fig. 18. Model for evolution of cpDNA structure in Asyneuma, Legousia, Petromarula, and Triodanus from a Trachelium-like ancestor. Numbers in parentheses show order of hybridized tobacco cpDNA probes. Arrows indicate transfer of a 25 kb region from the LSC region to the SSC region (mechanism unknown). Final map shows only features shared by the four species; autapomorphies of individual taxa not pictured.
**Trachelium**


---

**Asyneuma, Legousia, Petromarula, Triodanus**

Fig. 19. Models for evolution of cpDNA structure in A. Edraianthus and Symphyandra, B. Campanula and Adenophora, C. Jasione, from a Trachelium-like ancestor, showing only inversions. Numbers in parentheses show order of hybridized tobacco cpDNA probes. Inversion endpoints are given as numbered arrows. Inversions not shown for Jasione.
**FIGURE 19**

<table>
<thead>
<tr>
<th>Inversion 1</th>
<th>Inversion 2</th>
<th>Inversion 3</th>
<th>Inversion 4</th>
<th>Inversion 5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Trachelium</strong></td>
<td><strong>Campanula</strong></td>
<td><strong>Trachelium</strong></td>
<td><strong>Adenophora</strong></td>
<td><strong>Jasione</strong></td>
</tr>
</tbody>
</table>

---
Fig. 20. Model for evolution of cpDNA structure in *Cyananthus*, *Codonopsis*, and *Platycodon* from a tobacco-like ancestor, showing only inversions. Numbers in parentheses show order of hybridized tobacco cpDNA probes. Inversion endpoints are given as numbered arrows. First two inversions shared by all three species, first five shared by *Codonopsis* and *Cyananthus*. Remaining five inversions in *Platycodon* not pictured.
FIGURE 20

   \[ \text{Inversion 1} \]

   \[ \text{Inversion 2} \]

   \[ \text{Inversions 3,4} \]

   \[ \text{Inversion 5} \]

   \[ \text{Inversions 6,7} \]

   \[ \text{Inversion 8} \]

   **Cyananthus**

   \[ \text{Inversions 9,10} \]

   **Codonopsis**

   \[ \text{Inversions 9,10} \]

   **Platyodon**

   \[ \text{Inversions 11-15} \]
CHAPTER IV

PHYLOGENETIC IMPLICATIONS OF COMPLEX CHLOROPLAST GENOME STRUCTURE IN THE CAMPA NULACEAE

Introduction

The Campanulaceae sensu stricto are a nearly cosmopolitan angiosperm family consisting of latex-bearing, primarily perennial herbs or occasional subshrubs that typically have alternate leaves, sympetalous corollas, inferior ovaries, and capsular fruits. The family is traditionally considered to be closely related to the Lobeliaceae, Cyphiaceae, Pentaphragmataceae, and Sphenocleaceae; all of these taxa have been included in the Campanulaceae at varying taxonomic rank by different authors. The Campanulaceae in its strict sense are comprised of 600 (Kovanda 1978) to 950 (Takhtajan 1987) species distributed among 35 (Kovanda 1978) to 55 (Takhtajan 1987) genera. As these disparate accounts suggest, generic circumscription and intrafamilial classification vary widely according to author. As few as two (A. de Candolle 1830) and as many as 18 (Kolakovsky
1987) tribes have been recognized in the family (Table 6). The Campanuleae and Wahlenbergieae (at whatever rank) are typically the largest, most inclusive tribes in all treatments, with segregate tribes consisting of only one to a few genera.

The most comprehensive treatment of the Campanulaceae remains the monograph of A. deCandolle (1830), who recognized two groups corresponding to the Wahlenbergieae and Campanuleae (Table 6). In this broad sense, the Campanuleae are characterized by simple basal leaves and simple, alternate, cauline leaves that may be quite different in shape than the basal leaves. Flowers are solitary or borne in cymes or racemes, and have five petals that are proximally fused. The inferior ovary has 3-5 carpels and develops into a capsule that usually dehisces by lateral pores (rarely a berry). The Wahlenbergieae are mostly perennials characterized by simple, alternate, cauline leaves. Flowers are solitary or borne in cymes or heads, and petals may be free, proximally fused, or distally fused. The ovary is inferior, semi-inferior, or superior, and consists of two, three, or five carpels. The fruit is generally a capsule dehiscing by apical pores or valves (rarely a berry). Both groups have five stamens with filaments that are often proximally dilated and anthers with introrse
dehiscence; nectaries are generally present, and many ovules are attached to axile placentae. The entire family is characterized by secondary pollen presentation in which protandry is combined with a close association of anthers around the style and introrse pollen discharge onto the style for presentation to pollinators. This syndrome is similar to that found in Lobeliaceae and Asteraceae, but invaginating stylar hairs are unique to the Campanulaceae.

Federov's (1972) more recent work recognized eight tribes (Table 6), but only included taxa present in the former Soviet Union. Although Kolakovsky's (1987) is the only modern attempt to produce an intrafamilial classification of the Campanulaceae nearly worldwide (Table 6), the scope of the work is limited compared to that of either A. deCandolle (1830) or Federov (1972) and is likely an unnatural system.

Floral characteristics are more or less homogeneous family-wide, whereas capsule characters vary considerably, providing the basis for most intrafamilial classification schemes. Campanuleae typically include taxa with capsules dehiscing by lateral pores, whereas Wahlenbergieae have taxa with capsules dehiscing by apical valves. Ovary characters such as carpel number and position have also been important in traditional classifications. The tribe Platycodoneae (Yeo 1993) or subtribe Platycodinae
6) is sometimes segregated. It is defined by carpels that are equal in number to and alternate with the calyx lobes, whereas in Campanulaceae and Wahlenbegieae the carpels are often fewer than calyx lobes, or if the same in number then opposite them (Gadella 1964; Kovanda 1978; Thulin 1975). Little correlation appears to exist among diagnostic features, however, resulting in classifications that are undoubtedly unnatural (Kovanda 1978). In certain instances it is difficult to discern the rationale behind tribal placement of individual genera.

The high level of disagreement among both inter- and intrafamilial classifications of the Campanulaceae indicates a need for a modern assessment of the controversy. The chloroplast genome has proven to be a useful tool for phylogenetic reconstruction. Chloroplast DNA (cpDNA) of land plants is highly conserved in nucleotide sequence, gene order, and genome size and structure; its relatively slow rate of evolution makes it an excellent molecule for phylogenetic and evolutionary studies (Palmer 1987; Palmer et al. 1988a; Zurawski and Clegg 1987). Chloroplast genomes of photosynthetic angiosperms average about 160 kilobase pairs (kb) in size; the circular chromosome is divided by a large inverted repeat (average size 25 kb) into large and small single copy regions (Palmer 1985, 1991; Sugiura 1992). Three
major approaches to the study of cpDNA variation including restriction site mapping, gene sequencing, and analysis of major structural rearrangements have been used for phylogenetic investigations (Palmer et al. 1988a). This chapter focuses on the phylogenetic distribution of structural changes of the chloroplast genomes of the Campanulaceae. Four major categories of cpDNA structural rearrangements occur: 1) inversions, 2) insertions/deletions, 3) inverted repeat (IR) expansion/contraction or loss, and 4) transpositions (Palmer 1991; Palmer et al. 1988a).

Major structural rearrangements are relatively rare, therefore they are extremely useful as phylogenetic markers because they are readily polarized and typically lack homoplasy (Downie and Palmer 1992; Palmer et al. 1988a). When cpDNA inversions have been discovered, they are generally few in number and easily characterized. In groups in which more than one inversion has been documented, the order of events is usually readily determined. Inversions have been useful in phylogenetic reconstruction, particularly in the Fabaceae (legumes). All members of the family share a 50 kb inversion (Palmer and Thompson 1982), and members of Phaseoleae subtribe Phaseolinae (subfamily Papilionoideae) have a 78 kb inversion (Bruneau et al. 1990; Palmer et al. 1988b). Two
other lineages of legumes are characterized by extensive rearrangements (Milligan et al. 1989; Palmer et al. 1987b; Palmer et al. 1988b) that occurred following the loss of one copy of the IR. Although legumes constitute but one example of the phylogenetic utility of cpDNA rearrangements, inversions have also been useful in several other groups of land plants including Asteraceae (Jansen and Palmer 1987a,b), Cactaceae (Wallace 1993), Chenopodiaceae (Downie and Palmer 1992), Geraniaceae (Downie and Palmer 1992; Palmer et al. 1987a; Price and Palmer 1993), Lobeliaceae (Knox et al. 1993), Onagraceae (Hachtel et al. 1991; Herrmann et al. 1983; Sytsma et al. 1993), Poaceae (Doyle et al. 1992; Hiratsuka et al. 1989; Howe 1985; Quigley and Weil 1985), Ranunculaceae (Hoot and Palmer in press; Johansson and Jansen 1991), conifers (Raubeson and Jansen 1992a; Strauss et al. 1988), ferns (Hasebe and Iwatsuki 1990, 1992; Stein et al. 1986, 1992), lycopods (Raubeson and Jansen, 1992b) and bryophytes (Calie and Hughes 1987; Ohyama et al. 1988).

Although cpDNA gene losses and IR mutations have also provided phylogenetically useful information (reviewed by Downie and Palmer 1992), this paper will focus on cpDNA inversions and their phylogenetic interpretation and utility. We have previously reported that the chloroplast genomes of taxa in the Campanulaceae are the most highly
rearranged of any land plants so far examined (Chapter 3). At least 42 inversions, five separate IR expansion/contraction events, more than eight putative transpositions, one presumed gene loss, two deletions within large open reading frames (ORFs), 18 large insertions greater than 5 kb in size, and numerous smaller insertions between 1 and 5 kb, have contributed to chloroplast genome scrambling in the Campanulaceae. Due to the unprecedented number of inversions, it is not possible to unambiguously determine the evolutionary order of most inversions or in some cases to even define the inversion events with certainty. This complex situation poses special problems for using these structural rearrangements to estimate phylogenetic relationships. In this chapter we develop alternative character codings for the numerous inversions and compare the results of parsimony analyses of the different data sets. Furthermore, the systematic implications of the cpDNA phylogeny of the Campanulaceae are discussed.

Materials and Methods

Total DNA was isolated from one species in each of 18 genera in the Campanulaceae (Table 7) according to the CTAB method of Doyle and Doyle (1987). DNAs were digested
with the restriction endonucleases BamHI, BglII, EcoRI, EcoRV, HindIII, and SstI, and double digests were carried out using HindIII and the remaining five enzymes. Hybridization probes consisted of 106 small tobacco cpDNA probes (average size 1.2 kb) provided by J. Palmer (see Chapter 3). Twenty-one cloned HindIII cpDNA fragments from Trachelium caeruleum (Campanulaceae) (Chapter 2) were also used as hybridization probes. Single and double digest restriction site maps were constructed for all 18 species. For further detail regarding laboratory techniques and mapping procedures see Chapter 3.

A data matrix of 84 cpDNA rearrangement characters (Table 8) prepared for cladistic analysis is shown in Table 9. Because it was not possible to unambiguously identify many inversion events, nearly all inversionsal characters were initially defined as inversion endpoints; 53 of the 84 characters comprise endpoints. Other characters were more easily defined and consisted of eight transpositions, five IR expansions or contractions, one gene loss, 14 large insertions (>5 kb), one partial deletion/divergence of a duplicative transposition, and two inversions. For simplicity, unique insertions less than 5 kb were arbitrarily excluded from the analysis. Some taxa had insertions larger than 5 kb (sizes varied greatly among taxa) near the deletion/divergence of a
region corresponding to part of tobacco's large ORF1901 (Chapter 3). These were not included because it is unknown if any of the insertions have a common origin among taxa.

Cladistic parsimony analyses were performed on four variations of the 84 character data set (Table 9). The first analysis used the data set with no modifications. The second involved giving all non-endpoint characters a weight of two. The third analysis relied on the identification of inversions that are shared among some taxa, as well as interpreting the order of certain inversions. In such cases, endpoints of those inversions presumed to be shared were scored as present even if they are currently absent in certain taxa due to their disruption by additional inversions. The fourth analysis combined the weighting of analysis two with the character recoding of analysis three. The rationale behind defining and coding characters and data analysis are discussed in detail in the Discussion section.

The parsimony analyses were conducted using PAUP version 3.1.1 (Swofford 1993) on a Macintosh IIci computer with the Tree Bisection Reconnection (TBR), ACCTRAN, and MULPARS options. Tobacco was used as the outgroup since it has the ancestral chloroplast genome structure present in the majority of angiosperms (Downie and Palmer 1992).
One hundred random entries of data were also performed for each data set in an attempt to find all shortest trees (Maddison 1991), and the amount of support for monophyletic groups was evaluated using 100 bootstrap replicates for each data set (Felsenstein 1985).

Results

Complete restriction site maps were constructed for 16 of the 18 taxa, and nearly complete maps were constructed for Jasione and Roella. It was not possible to map the small single copy (SSC) region of Roella because hybridization signals became increasingly weak in later rounds of hybridization (Chapter 3). Rearrangements involving the IR/SSC junction and SSC region of Jasione prohibited totally resolving restriction site maps (Chapter 3).

Linear cpDNA maps showing the hybridization patterns of 106 consecutively numbered tobacco cpDNA probes for the 18 taxa are diagrammed in Fig. 21. For sizes, locations, and gene content of the tobacco fragments see Chapter 3; any reference to clone numbers will refer to tobacco cpDNA probes unless specified otherwise. Mapping data show that chloroplast genomes in the Campanulaceae are drastically rearranged relative to those of other land plants.
Several types of rearrangements contributed to cpDNA evolution in the family, and these are described in Chapter 3. To summarize, at least 42 inversions, five separate IR expansion/contraction events, more than eight putative transpositions, one presumed gene loss, two deletions within large ORFs, 18 large insertions greater than 5 kb in size, and numerous insertions between 1 and 5 kb were found in chloroplast genomes in the Campanulaceae.

The cladistic parsimony analysis of the unmodified 84 character data set produced 241 equally parsimonious trees of 94 steps with a consistency index (CI) of 0.76 (excluding autapomorphies); the strict consensus tree is given in Fig. 22. The analysis that gave a weight of two to non-endpoint characters resulted in 12 equally parsimonious trees of 125 steps (CI=0.78), the strict consensus of which is shown in Fig. 23. The third analysis using the modified coding of certain endpoint characters (see above and Discussion) resulted in two equally parsimonious trees (Fig. 24A,B) of 88 steps (CI=0.90). Both the strict and 50% majority rule consensus trees are identical in topology to tree A of Fig. 24. The fourth analysis, which combined character weighting and the character recoding described in analysis three, produced two equally parsimonious trees of 119
steps (CI=0.91) that were identical to those found in the third analysis (Fig. 24A,B).

Discussion

The chloroplast genomes in the Campanulaceae are radically rearranged relative to those of most other land plants. In other families in which inversions occur, there are usually only one to a few easily characterized inversions. Structural rearrangements are very powerful phylogenetic characters because of their relative rarity and lack of homoplasy (Downie and Palmer 1992; Palmer et al. 1988a). For example, both the loss of one copy of the IR and inversions are extremely useful characters in legume phylogeny (Lavin et al. 1990; Palmer et al. 1987b, 1988b). Although cpDNA inversions have been discovered in a wide range of land plant families (reviewed in Downie and Palmer 1992), only two possible cases of homoplasy involving inversions have been reported. These include two parallel inversions in the Ranunculaceae (Hoot and Palmer in press) and one within the Caryophyllales (Downie and Palmer submitted). Several independent cases of parallel gene loss have also been reported among land plants (Downie and Palmer 1992).
Phylogenetic analysis of cpDNA rearrangements

The number of inversions in the Campanulaceae presents several difficulties for interpreting their phylogenetic and evolutionary significance. There is no precedent for analyzing such a complex set of rearrangements within a family, and our approach is discussed below.

The first major problem was simply defining the inversion events. Beginning with a tobacco-like ancestor, an attempt was made to find the fewest number of inversions that would result in a particular gene order. In some cases, a given genome structure could be attained through different inversion events that used the same number of steps. For example, the chloroplast genome of *Platycodon* could have evolved from a tobacco-like ancestor by two different models of seven inversions each (Fig. 25); none of the inversions is the same in the two scenarios. Thus it is not possible to determine which of these two scenarios (if either) may have occurred.

The phylogenetic trees in Figs. 22-24 indicate that *Codonopsis*, *Platycodon*, and *Cyananthus* are basal within the family. The chloroplast genomes of *Codonopsis* and *Cyananthus* are more similar to each other than either is to *Platycodon* (as reflected by strong bootstrap (96%)}
support (Fig. 24) for this clade), and none of the three closely resembles the rest of the family (Fig. 21). The 15 derived taxa share several features (Fig. 26) that suggest they may have had a common ancestor (Chapter 3). Trachelium and Musschia have the least rearranged chloroplast genomes, and it is possible to derive the other 13 genome structures from a Trachelium-like ancestor (Chapter 3), provided the following assumption is accepted. An inversion is considered advanced if it disrupts an unrearranged sequence block in Trachelium (and therefore tobacco), that is, if one or both endpoints of an inversion separates a series of contiguous tobacco probes. When employing this rationale, it is fairly simple to determine that Wahlenbergia has evolved three additional inversions relative to Trachelium (Fig. 27), in which three inversion endpoints break up sequence blocks that are intact in Trachelium. Three additional inversions unite Merciera, Prismatocarpus, and Roella (Fig. 27). If using Trachelium as a starting point, it is also possible to derive the remainder of the taxa using this logic (see Chapter 3).

In addition to determining what inversions occurred, a second major problem involves interpreting which inversions are shared between basal and advanced genera. Although a clear series of evolutionary events can be
envisioned among the 15 derived taxa without the aid of computer-based analysis, their genomes are quite distinct from the three basal taxa. A high degree of structural divergence of cpDNAs between these two groups complicates interpretation of evolutionary relationships. Some sequence blocks are shared between certain primitive and advanced genera. For example, all taxa share the sequence block 41-44, *Platyodon* shares 16-25 with all advanced taxa, and 47-48 is common to *Cyananthurus* and all advanced genera (Fig. 21). However, it is not known if these sequence blocks were derived in different taxa by shared or unique inversions, particularly when most other aspects of genome structure are so dissimilar between basal and advanced groups.

Although the ideal way to analyze rearrangement data is to determine presence or absence of an event, this is clearly not possible in some cases in the Campanulaceae. Even though the order of events can be determined among the derived taxa with some confidence, the levels of resolution among the three basal genera, and between basal and derived taxa, are low when interpreting rearrangements by inspection of gene maps only. Because of this lack of resolution, and because of the inherent complexity of the data, we felt a new method of character analysis of the rearrangement data for phylogenetic purposes was
warranted. Our approach involved coding endpoints, along with more easily defined rearrangements, as characters in a cladistic parsimony analysis. Endpoints were defined as two non-contiguous tobacco probes that are now adjacent in one or more species (Table 8).

Using endpoints as characters has several drawbacks. The first is that only endpoints currently present in genomes are counted. Any endpoints that were obliterated by subsequent inversions would not be counted; shared inversions might then be overlooked, possibly resulting in misinterpretation of relationships. The second problem is the inadvertant weighting of certain events over others. If both endpoints of an inversion are still intact in a genome, the inversion is scored twice, if only a single endpoint remains the inversion is counted only once, and if both endpoints have been lost the inversion will not be included at all. This is a problem in the Campanulaceae because there appears to be at least some endpoint reuse (Chapter 2). The third problem is that non-inversion characters will be unintentionally downweighted because they are only scored once.

The advantage of using endpoints as characters is that it allows comparisons between basal and advanced genera that would not be possible by simply inspecting gene maps. In addition, it allows an independent
assessment of the evolutionary hypotheses formulated by interpretation of restriction site maps. We have employed four versions of the 84 character data set (Table 8) in an attempt to compensate for the shortcomings of the approach. The first analysis used the data set with no modifications. The second analysis gave a weight of two to non-endpoint characters in an attempt to compensate for the unintentional weight given to inversions in which both endpoints are present. This, of course, results in downweighting inversions in which only one endpoint remains and still fails to include inversions whose endpoints are both gone.

The third analysis used information gathered from interpretations of genome arrangements, particularly those of the 15 derived taxa. In Chapter 3, evolutionary models were constructed for the three major groups delimited among these 15 taxa utilizing the rationale outlined above for determining the order of inversions. When it was determined that inversions were shared between two or more genera, the taxa were coded as having an endpoint even if it was subsequently lost due to the disruption of the endpoint by subsequent inversions. We believe that the progression of inversions we have proposed within and between these groups of taxa are very accurate.
The fourth analysis combines the above modifications with weighting of non-endpoint characters to again attempt to assign weight evenly among characters. As expected, the strict consensus of 241 trees in analysis one is the least resolved of all trees (Fig. 22). The trees found in analyses 2, 3, and 4 are nearly identical, the differences occurring in the relative placements of Jasione, Campanula, and Adenophora (Figs. 23 and 24). Analysis two (Fig. 23) suggests that Jasione and Adenophora form a clade, with Campanula their sister taxon. Tree B of two equally parsimonious trees of both analyses 3 and 4 indicates that Campanula and Adenophora are sister taxa, forming a clade with Jasione (Fig. 24); this relationship collapses in the strict consensus tree, which is identical to tree A (Fig. 24).

Assessing both the cladograms and cpDNA maps allowed us to favor one of the two equally parsimonious trees of Fig. 24. Tree B is not compatible with our interpretation of cpDNA mapping data because it requires a reversal of an inversion in Campanula. In our evolutionary model for this group (Chapter 3), Campanula and Adenophora share a derived inversion relative to Trachelium, and Adenophora is characterized by two additional inversions (Fig. 26). In tree B (Fig. 24) a derived endpoint (character 31, Table 8) is the only character uniting Jasione, Campanula,
and Adenophora, followed by reversal in Campanula. If Adenophora and Jasione do share the inversion associated with this endpoint (37/28), the relationship is quite compatible with that shown in Fig. 22, in which Campanula is sister to a clade consisting of Jasione and Adenophora. There is no mapping evidence to indicate that Jasione, Adenophora, and Campanula share this inversion, followed by the reinversion of the region to its original orientation in Campanula (bootstrap support (Fig. 24) is also low). Interpretation of the situation is complicated by the fact that Jasione is quite divergent from all other taxa; approximately 15 inversions, an IR mutation, and the duplicative transposition are necessary to convert its genome from a tobacco-like ancestor (Chapter 3). It is therefore difficult to interpret its affinities with any confidence.

The most important results of all of the analyses are those related to the three basal taxa, since a reasonable interpretation of inversions among the advanced groups could be attained by inspection of genome maps (and is supported by the parsimony analyses). The cladistic analyses suggest that one inversion is shared by the entire family, along with two non-inversion characters (Fig. 26). Six additional inversions and two non-inversion characters unite the 15 derived genera (Fig.
This provides strong evidence that *Codonopsis*, *Platycodon*, and *Cyananthus* are basal within the family. *Platycodon*, *Cyananthus*, and *Codonopsis* share one inversion, and *Cyananthus* and *Codonopsis* share three more (Fig. 26).

It is possible that more complex evolutionary scenarios occurred, in which some inversions evolved in parallel, or in which similar genome structures resulted from a different set of inversions. Although we have attempted to produce the shortest evolutionary schemes and to avoid homoplasy, it is very possible that longer, more complicated scenarios actually occurred, especially because the Campanulaceae seem predisposed to cpDNA rearrangements. However, given the general lack of homoplasy in rearrangement data, we feel our phylogeny is a reasonable estimate of relationships within the family.

The parsimony analyses may underestimate the number of inversions shared between primitive and advanced genera, because evidence of shared inversions may have been lost. In addition, the methods cannot address whether similarities such as sequence blocks shared between primitive and advanced taxa (see above) have any real meaning. However, they provide an independent assessment of relationships among inversions (and hence, among genera) that could only be guessed at without
rigorous parsimony analyses. This is particularly important for assessing relationships among primitive and advanced taxa because their chloroplast genomes have diverged significantly.

Despite the complexity of the cpDNA rearrangements in the Campanulaceae, the resulting phylogeny (Fig. 26) is a robust estimate of intrafamilial relationships. Very little homoplasy exists in the rearrangement data sets, particularly in analyses three and four (Fig. 24); all of the homoplastic characters are inversion endpoints. There is no homoplasy in the resulting evolutionary hypothesis we have constructed based on the cladistic analyses and study of genome structures. Furthermore, the homoplasy in endpoint characters appears to be the result of different inversions producing similar endpoints. There was complete congruence among trees (Figs. 22-24) produced using the four separate analyses employed here. High consistency indices, high bootstrap support for most monophyletic groups (Figs. 22-24), and agreement with other types of data (see below) indicate that the cpDNA phylogeny is robust. Although the rearrangements in Campanulaceae are complex, their phylogenetic utility is evident. These analyses indicate that rearrangements are very powerful systematic characters in the Campanulaceae.
Phylogenetic implications of cpDNA data

Chloroplast DNA structural rearrangements are very powerful characters for reconstructing phylogeny. In most cases only very broad groups can be circumscribed by rearrangements since events such as inversions are fairly rare (Palmer et al. 1988a). Because there are so many rearrangements in the Campanulaceae, smaller groups can be identified, resulting in the most highly resolved phylogeny based entirely on cpDNA rearrangements produced in any land plant family.

Most traditional classifications of the Campanulaceae are based mainly on capsule dehiscence and ovary position and arrangement. A. deCandolle (1830) divided the family into two groups corresponding to Wahlenbergieae and Campanuleae (Table 6). Other authors followed this scheme, with the additional recognition of either the tribe Merciereae (A. deCandolle 1839) or the subtribe Platycodinae (Kovanda 1978; Schönland 1889) (Table 6). Bentham and Hooker (1876) divided genera among five groups based on fruit features, but did not formally designate subtribes. More recent workers recognize seven (Takhtajan 1987), eight (Federov 1972), or as many as 18 (Kolakovsky 1987) tribes (Table 6).
As Kovanda (1978) and Thulin (1975) recognized, classification of the Campanulaceae based on capsule characters is not a natural system. Neither the Campanuleae nor Wahlenbergieae (at whatever taxonomic rank) are monophyletic based on cpDNA rearrangements (Fig. 26). Likewise, no traditional classification suggests that Codonopsis, Platycodon, and Cyananthus are basal in the family as suggested by cpDNA data, with the exception of Takhtajan's (1987) system; however, he suggested only Cyananthus (in its own tribe Cyanantheae) as the most primitive member of the family, placing Platycodon and Codonopsis in other tribes (Table 6).

Studies of pollen ultrastructure indicate that Platycodon, Codonopsis, and Cyananthus are basal members of Campanulaceae (Dunbar 1975a,b). These taxa have colpate to colporate apertures, whereas the rest of the family (as surveyed here) have porate grains (Avetisjan 1967; Chapman 1966; Dunbar 1975a,b; Erdtman 1952). The evolutionary scheme based on pollen morphology presented by Dunbar (1975b) suggests that Cyananthus (colpate) and Codonopsis (colpate) are more closely related to each other than either is to Platycodon (colporate), which is also supported by the cpDNA tree (Fig. 26). Thulin (1975) suggested the heterogeneity in fruit characters among taxa with elongate apertures may warrant further subdivision.
Thulin (1975) believed that pollen morphology should constitute a key part of any modern reassessment of relationships in the Campanulaceae. He suggested that all taxa with elongated apertures should be removed from Campanuleae and Wahlenbergieae, and those with porate grains removed from Schönland's Platycodinae. All genera with elongated apertures are Asian, with the exception of Canarina, which is African (Thulin 1975). Following the removal of colpate and colporate taxa, Campanuleae sensu Schönland are comprised of northern hemisphere genera, whereas Wahlenbergieae contain southern hemisphere taxa, with the exceptions of Edraianthus, Githopsis, and Jasione (although Jasione occurs in north Africa as well as Europe). Chloroplast DNA data indicate that Jasione's and Edraianthus' affinities lie with northern hemisphere species rather than with Wahlenbergieae. Githopsis was not available for this study, but both Thulin (1975) and Morin (1983) felt the genus could be allied with Campanuleae.

The groups delimited by cpDNA rearrangements have geographical as well as palynological integrity. Wahlenbergia is primarily a southern hemisphere old world genus (Smith 1992); W. gloriosa, mapped for this study, is Australian (Smith 1992). Roellia, Merciera, and Prismatocarpus are all endemic to South Africa (eg.
Adamson 1950; Dyer 1975; Phillips 1926). The 10 genera in the *Trachelium* and *Legousia* clades are primarily European to Eurasian (although *Triodanum* and *Campanula* are also North American) (eg. Federov 1972; Rosatti 1986; Schönland 1889; Tsoong 1935; Tutin 1976), and *Musschia* is endemic to the island of Madeira (Hansen and Sunding 1993; Lowe 1857). The three basal genera are all Asian (eg. Federov 1972; Tsoong 1935), and thought by Thulin (1975) to be basal in the Campanulaceae based on pollen morphology and geography.

The cpDNA data are compatible with an *rbcL* study of the Campanulales (Cosner et al. in press) and are also largely congruent with a serological study of the Campanulaceae (Gudkova 1991). Although only four genera from the Campanulaceae were included in the *rbcL* study, results indicate a similar split between basal and derived taxa. Although the cpDNA and serological studies differed somewhat in the taxa sampled, both included a group containing *Trachelium*, *Adenophora*, and *Campanula*. They also agreed in the grouping of *Asyneuma* and *Petromarula* in a clade. The only discrepancy was in the placement of *Legousia*; the serological study placed this genus basal to all others surveyed (Gudkova 1991).

Chloroplast DNA data suggest affinities of several controversial genera (Fig. 26). Schönland (1889) united
Musschia and Platycodon as Platycodinae, clearly incompatible with both cpDNA and pollen evidence. A. deCandolle (1830) was unsure of Merciera's taxonomic position because its four basal ovules and single-seeded (by abortion) unilocular capsule (Phillips 1926) is unique in the Campanulaceae (Lammers 1992). This genus was later placed in Merciereae (A.P. deCandolle 1839), but is allied with other southern African genera in the cpDNA analysis (Fig. 26).

Adenophora and Symphyandra are segregates from Campanula, based on the presence of a conspicuous tubular nectariferous disc and connate anthers, respectively. Adenophora's chloroplast genome is derived relative to Campanula elatines (Fig. 26), but this position should be considered preliminary pending further sampling within Adenophora and the large genus Campanula. Surprisingly, Symphyandra is more closely related to Edraianthus than Campanula (Fig. 26). Edraianthus has traditionally been considered close to Wahlenbergia (A. deCandolle 1830).

Much controversy surrounds the taxonomy of the genera Triodanuss and Legousia. In some treatments, both genera were included under the name Specularia (eg. A.P. deCandolle 1830; Schönland 1889), although the correct name for Specularia is Legousia (Rosatti 1986). McVaugh (1945, 1948) and Fernald (1946) disagreed regarding the
circumscription of the genera; Fernald felt that *Triodanus* as a genus is very weak and should be remerged with *Legousia*. McVaugh (1945) argued that the two genera should either remain separate or both be sunk in *Campanula*. In his system, both species studied here (*T. perfoliata* and *L. falcata*) belong to *Triodanus*. As expected, *Triodanus* and *Legousia* belong to the same cpDNA clade, united by an unusual mutation that transferred a large segment of the large single copy (LSC) region to the SSC region (Chapter 3). However, *Legousia* has a putative transposition not found in *Triodanus*, which has a unique large insertion (Chapter 3).

The most characteristic morphological feature of *Jasione* is its inflorescence, a dense globose capitulum subtended by loose involucral bracts. The unique nature of *Jasione* in Campanulaceae led Dumortier (1829) to place it in its own family. Given its unusual morphology, it is interesting to note that *Jasione* has the most rearranged chloroplast genome examined in the Campanulaceae, having undergone around 15 inversions and a rare IR mutation (Chapter 3).

Despite the difficulties in interpreting such a complex set of rearrangements, the systematic utility of chloroplast DNA in the Campanulaceae is evident. The data indicate that traditional classifications based on fruit
and ovary characters are unnatural, and suggest affinities of several difficult genera. Further sampling within large genera such as *Campanula* and *Wahlenbergia*, as well as in genera such as *Adenophora* that are often considered unnatural, will be necessary to further elucidate relationships among chloroplast genomes. It is likely that intrafamilial relationships can be further resolved by including other genera in rearrangement analyses. Of particular interest are other putatively primitive taxa that have pollen with elongated apertures (Avetisian 1967; Chapman 1966; Dunbar 1975a,b; Erdtman 1952) such as *Campanumoea, Canarina, Leptocodon*, and *Ostrowskia*. 
LITERATURE CITED


Table 6. Select classification systems of Campanulaceae. Major intrafamilial subdivisions indicated but only genera (Table 7) sampled for this study included. Genera in Table 7 that are sunk in other taxa in these systems are not listed.

<table>
<thead>
<tr>
<th>A. deCandolle (1830)</th>
<th>A.P deCandolle (1839)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Campanulaceae</strong></td>
<td><strong>Campanulaceae</strong></td>
</tr>
<tr>
<td><strong>Subtribus I</strong></td>
<td><strong>Wahlenbergieae</strong></td>
</tr>
<tr>
<td>Jasione</td>
<td>Jasione</td>
</tr>
<tr>
<td>Codonopsis</td>
<td>Platyodon</td>
</tr>
<tr>
<td>Platyodon</td>
<td>Codonopsis</td>
</tr>
<tr>
<td>Wahlenbergia</td>
<td>Wahlenbergia</td>
</tr>
<tr>
<td>Prismatocarpus</td>
<td>Prismatocarpus</td>
</tr>
<tr>
<td>Roella</td>
<td>Roella</td>
</tr>
<tr>
<td><strong>Subtribus II</strong></td>
<td><strong>Campanuleae</strong></td>
</tr>
<tr>
<td>Petromarula</td>
<td>Petromarula</td>
</tr>
<tr>
<td>Campanula</td>
<td>Campanula</td>
</tr>
<tr>
<td>Trachelium</td>
<td>Trachelium</td>
</tr>
<tr>
<td>Adenophora</td>
<td>Adenophora</td>
</tr>
<tr>
<td>Symphyandra</td>
<td>Symphyandra</td>
</tr>
<tr>
<td>Musschia</td>
<td>Musschia</td>
</tr>
<tr>
<td><strong>Genus incertae sedis</strong></td>
<td><strong>Merciereae</strong></td>
</tr>
<tr>
<td>Merciera</td>
<td>Merciera</td>
</tr>
<tr>
<td>Federov (1972)</td>
<td>Takhtajan (1987)</td>
</tr>
<tr>
<td>----------------------------------</td>
<td>-------------------------------------</td>
</tr>
<tr>
<td>Campanulaceae</td>
<td>Campanulaceae</td>
</tr>
<tr>
<td>Sphenocleoidae</td>
<td>Cyanantheae</td>
</tr>
<tr>
<td>Campanuloideae</td>
<td>Cyananthus</td>
</tr>
<tr>
<td>Campanuleae</td>
<td>Wahlenbergieae</td>
</tr>
<tr>
<td>Campanula</td>
<td>Wahlenbergia</td>
</tr>
<tr>
<td>Symphyandra</td>
<td>Edraianthus</td>
</tr>
<tr>
<td>Adenophora</td>
<td>Jasione</td>
</tr>
<tr>
<td>Peracarpeae</td>
<td>Codonopsis</td>
</tr>
<tr>
<td>Ostrowskieae</td>
<td>Merciera</td>
</tr>
<tr>
<td></td>
<td>Roella</td>
</tr>
<tr>
<td></td>
<td>Prismatocarpus</td>
</tr>
<tr>
<td>Michauxieae</td>
<td>Platycodoneae</td>
</tr>
<tr>
<td>Phyteumateae</td>
<td>Platycodon</td>
</tr>
<tr>
<td>Asyneuma</td>
<td>Musschia</td>
</tr>
<tr>
<td>Legousia</td>
<td>Campanuleae</td>
</tr>
<tr>
<td>Wahlenbergieae</td>
<td>Adenophora</td>
</tr>
<tr>
<td>Codonopsis</td>
<td>Campanula</td>
</tr>
<tr>
<td>Platycodon</td>
<td>Legousia</td>
</tr>
<tr>
<td>Edraiantheae</td>
<td>Triodanlus</td>
</tr>
<tr>
<td>Edraianthus</td>
<td>Michauxieae</td>
</tr>
<tr>
<td>Jasioneae</td>
<td>Phyteumateae</td>
</tr>
<tr>
<td>Jasione</td>
<td>Asyneuma</td>
</tr>
<tr>
<td></td>
<td>Trachelium</td>
</tr>
<tr>
<td></td>
<td>Petromarula</td>
</tr>
<tr>
<td></td>
<td>Peracarpeae</td>
</tr>
<tr>
<td>Schönländ (1889)</td>
<td>Kovanda (1978)</td>
</tr>
<tr>
<td>------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>Campanulaceae</td>
<td>Campanulaceae</td>
</tr>
<tr>
<td>Lobelioideae</td>
<td>Campanulinae</td>
</tr>
<tr>
<td>Cyphioideae</td>
<td>Campanula</td>
</tr>
<tr>
<td>Symphyandra</td>
<td>Adenophora</td>
</tr>
<tr>
<td>Campanuloideae</td>
<td>Legousia</td>
</tr>
<tr>
<td>Pentaphragmeae</td>
<td></td>
</tr>
<tr>
<td>Sphenocleae</td>
<td>Wahlenberginae</td>
</tr>
<tr>
<td>Campanuleae</td>
<td>Wahlenbergia</td>
</tr>
<tr>
<td>Campanulinae</td>
<td>Codonopsis</td>
</tr>
<tr>
<td>Adenophora</td>
<td>Edraianthus</td>
</tr>
<tr>
<td>Symphyandra</td>
<td>Jasione</td>
</tr>
<tr>
<td>Trachelium</td>
<td>Platycodinae</td>
</tr>
<tr>
<td>Campanula</td>
<td>Platycodon</td>
</tr>
<tr>
<td>Wahlenberginae</td>
<td>Platycodon</td>
</tr>
<tr>
<td>Cyrananthus</td>
<td>Musschia</td>
</tr>
<tr>
<td>Jasione</td>
<td></td>
</tr>
<tr>
<td>Prismatocarpus</td>
<td></td>
</tr>
<tr>
<td>Merciera</td>
<td></td>
</tr>
<tr>
<td>Edraianthus</td>
<td></td>
</tr>
<tr>
<td>Wahlenbergia</td>
<td></td>
</tr>
<tr>
<td>Codonopsis</td>
<td></td>
</tr>
<tr>
<td>Roella</td>
<td></td>
</tr>
<tr>
<td>Roella</td>
<td>Platycodon</td>
</tr>
<tr>
<td>Edraianthus</td>
<td>Musschia</td>
</tr>
<tr>
<td>Jasione</td>
<td></td>
</tr>
<tr>
<td>Prismatocarpus</td>
<td></td>
</tr>
<tr>
<td>Merciera</td>
<td></td>
</tr>
<tr>
<td>Edraianthus</td>
<td></td>
</tr>
<tr>
<td>Wahlenbergia</td>
<td>Platycodon</td>
</tr>
<tr>
<td>Codonopsis</td>
<td>Musschia</td>
</tr>
<tr>
<td>Roella</td>
<td></td>
</tr>
<tr>
<td>Platycodon</td>
<td></td>
</tr>
<tr>
<td>Musschia</td>
<td></td>
</tr>
</tbody>
</table>
Table 6 (continued)

Kolakovsky (1987)

Campanulaceae

Prismatocarpoideae

Prismatocarpus
Roella

Canarinoideae

Wahlenbergoideae

Wahlenbergieae

Jasione
Wahlenbergia
Codonopsis
Platyodon
Cyananthus

Azorineae

Musschieae

Musschia

Echinocodoneae

Annaea

Muehlbergelleae

Theodorovieae

Gadellieae

Ostrowskieae

Campanuloideae

Campanuleae

Adenophora
<table>
<thead>
<tr>
<th>Phytelephangion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Campanula</td>
</tr>
<tr>
<td>Symphyandra</td>
</tr>
<tr>
<td>Trachelium</td>
</tr>
<tr>
<td>Phytelephangionae</td>
</tr>
<tr>
<td>Peracarpeae</td>
</tr>
<tr>
<td>Sergieae</td>
</tr>
<tr>
<td>Michauxieae</td>
</tr>
<tr>
<td>Neocodonineae</td>
</tr>
<tr>
<td>Asyneuma</td>
</tr>
<tr>
<td>Legousia</td>
</tr>
<tr>
<td>Edraianthesis</td>
</tr>
<tr>
<td>Edraiantheae</td>
</tr>
<tr>
<td>Sachokielae</td>
</tr>
<tr>
<td>Mzymteleae</td>
</tr>
</tbody>
</table>
Table 7. Species in Campanulaceae mapped for chloroplast DNA structural rearrangements.

<table>
<thead>
<tr>
<th>Species</th>
<th>Source/Voucher</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adenophora confusa Nannf.</td>
<td>M. Cosner 177</td>
</tr>
<tr>
<td>Asyneuma virgatum (Labill.) Bornm.</td>
<td>Berlin-Dahlem 0104*</td>
</tr>
<tr>
<td>Campanula elatines L.</td>
<td>T. Ayers/BH88-287</td>
</tr>
<tr>
<td>Codonopsis viridis Wall. in Roxb.</td>
<td>T. Ayers/BH88-229</td>
</tr>
<tr>
<td>Cyananthus lobatus Wall.</td>
<td>M. Cosner 179</td>
</tr>
<tr>
<td>Edraianthus graminifolius (L.) A.DC.</td>
<td>T. Ayers/BH88-195</td>
</tr>
<tr>
<td>Jasione heldreichii Boiss. &amp; Orph.</td>
<td>T. Ayers/BH88-208</td>
</tr>
<tr>
<td>Legousia falcata (Ten.) Fritsch</td>
<td>Berlin-Dahlem 0143*</td>
</tr>
<tr>
<td>Merciera tenuifolia DC.</td>
<td>K. Steiner 2445</td>
</tr>
<tr>
<td>Musschia aurea Dum.</td>
<td>T. Ayers/BH88-274</td>
</tr>
<tr>
<td>Petromarula pinnata (L.) DC.</td>
<td>T. Ayers</td>
</tr>
<tr>
<td>Platycodon grandiflorum DC.</td>
<td>T. Ayers/BH88-216</td>
</tr>
<tr>
<td>Prisumocarpus diffusus DC.</td>
<td>K. Steiner 2448</td>
</tr>
<tr>
<td>Roella ciliata L.</td>
<td>T. Ayers</td>
</tr>
<tr>
<td>Symphyandra hofmannii Pant.</td>
<td>T. Ayers/BH88-225</td>
</tr>
<tr>
<td>Trachelium caeruleum L.</td>
<td>M. Cosner 173</td>
</tr>
<tr>
<td>Triodanuus perfoliata (L.) Nieuwl.</td>
<td>M. Cosner 178</td>
</tr>
<tr>
<td>Wahlenbergia gloriosa Loth.</td>
<td>T. Ayers/BH88-217</td>
</tr>
</tbody>
</table>

*Botanischer Garten and Botanisches Museum, Berlin-Dahlem
Table 8. Chloroplast DNA rearrangement characters in the Campanulaceae. Numbers refer to tobacco cpDNA hybridization probes. T=transposition (T'=second transposition (Chapter 3, Fig. 12) of most of 53-56), I=inversion, i=insertion (followed by approximate size in kb), D=deletion/divergence, IRc and IRe=IR contraction or expansion, respectively (followed by single copy region at which event occurred). All other characters are inversion endpoints.

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>11/60</td>
<td>22. T (53,54)</td>
<td>43. 40/56</td>
<td>64. i (15)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>56/53</td>
<td>23. T (53-56)</td>
<td>44. 39/37</td>
<td>65. IRc (LSC)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>49/37</td>
<td>24. 98/28</td>
<td>45. 37/44</td>
<td>66. T (5-8)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>40/35</td>
<td>25. 53/98</td>
<td>46. 56/61</td>
<td>67. T (6-9)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>28/11</td>
<td>26. T' (53-56)</td>
<td>47. 76/96</td>
<td>68. i (9)</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>15/76</td>
<td>27. 49/39</td>
<td>48. 77/106</td>
<td>69. i (18)</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>60/27</td>
<td>28. 37/40</td>
<td>49. i (5)</td>
<td>70. I (60-61)</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>26/44</td>
<td>29. 56/39</td>
<td>50. 39/25</td>
<td>71. T (28)</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>41/47</td>
<td>30. 53/40</td>
<td>51. 9/37</td>
<td>72. T (16-17)</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>48/36</td>
<td>31. 37/28</td>
<td>52. 48/56</td>
<td>73. i 7</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>35/25</td>
<td>32. 40/26</td>
<td>53. 49/61</td>
<td>74. IRc (SSC)</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>16/90</td>
<td>33. 27/44</td>
<td>54. 5/29</td>
<td>75. i (8)</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>T (93)</td>
<td>34. 8/40</td>
<td>55. 50/28</td>
<td>76. i (9)</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>84/76</td>
<td>35. 39/26</td>
<td>56. 26/50</td>
<td>77. i (9)</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>84/90</td>
<td>36. 27/11</td>
<td>57. 40/16</td>
<td>78. i (15)</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>10/49</td>
<td>37. 56/36</td>
<td>58. 25/10</td>
<td>79. i (16)</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>53/28</td>
<td>38. 44/10</td>
<td>59. 15/56</td>
<td>80. D (93)</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>37/60</td>
<td>39. 9/53</td>
<td>60. IRc (LSC)</td>
<td>81. i (5)</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>56/11</td>
<td>40. 49/36</td>
<td>61. IRc (SSC)</td>
<td>82. i (6)</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>I (16-17)</td>
<td>41. 8/36</td>
<td>62. IRc (LSC)</td>
<td>83. i (19)</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>56/27</td>
<td>42. 9/40</td>
<td>63. D (45-46)</td>
<td>84. i (5)</td>
<td></td>
</tr>
</tbody>
</table>
Table 9. Data matrix of 84 cpDNA rearrangement characters (Table 8). 0/1 indicates endpoints scored 0 (absent) in analyses one and two, and scored 1 (present) in analyses three and four (see text for details). ?=missing data.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Wahlenbergia</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2. Merciera</td>
<td>0/1 0/1 0/1 1 1 0/1 1 1 1 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Prismatocarpus</td>
<td>0/1 0/1 0/1 1 1 0/1 1 1 1 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Roella</td>
<td>0/1 0/1 0/1 1 1 0/1 1 1 1 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Legousia</td>
<td>0 0/1 1 1 0 1 0 1 1 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Asyneuma</td>
<td>0 0/1 1 1 0 1 0 1 1 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Petromarula</td>
<td>0 0/1 1 1 0 1 0 1 1 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Triodanthes</td>
<td>0 0/1 1 1 0 1 0 1 1 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Trachelium</td>
<td>0 1 1 1 0 1 0 1 1 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Campanula</td>
<td>0 1 0/1 1 0 1 0 1 1 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Musschia</td>
<td>0 1 1 1 0 1 0 1 1 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Symphyandra</td>
<td>0 0/1 1 1 0 1 0 1 1 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. Edraianthus</td>
<td>0 0/1 1 1 0 1 0 1 1 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. Adenophora</td>
<td>0 1 0/1 1 0 1 0 1 1 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15. Jasione</td>
<td>0 0/1 0/1 1 0 1 0 1 1 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16. Codonopsis</td>
<td>0 0 0 0 0 0 0 0 1 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17. Cyananthus</td>
<td>0 0 0 0 0 0 0 0 1 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18. Platycodon</td>
<td>0 0 0 0 0 0 0 0 1 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19. tobacco</td>
<td>0 0 0 0 0 0 0 0 0 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
<th>21</th>
<th>22</th>
<th>23</th>
<th>24</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2.</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3.</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>4.</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>5.</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>6.</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>7.</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>8.</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>9.</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10.</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>11.</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>12.</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>13.</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>14.</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>15.</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>16.</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>17.</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>18.</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>19.</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Table 9 (continued)

<table>
<thead>
<tr>
<th></th>
<th>25</th>
<th>26</th>
<th>27</th>
<th>28</th>
<th>29</th>
<th>30</th>
<th>31</th>
<th>32</th>
<th>33</th>
<th>34</th>
<th>35</th>
<th>36</th>
<th>37</th>
<th>38</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0/1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>16</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>17</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>18</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>19</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>39</th>
<th>40</th>
<th>41</th>
<th>42</th>
<th>43</th>
<th>44</th>
<th>45</th>
<th>46</th>
<th>47</th>
<th>48</th>
<th>49</th>
<th>50</th>
<th>51</th>
<th>52</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>17</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>18</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>19</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>17</td>
<td>18</td>
<td>19</td>
<td>20</td>
<td>21</td>
<td>22</td>
<td>23</td>
<td>24</td>
<td>25</td>
<td>26</td>
<td>27</td>
<td>28</td>
<td>29</td>
<td>30</td>
</tr>
<tr>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 9 (continued)
Fig. 21. Linear cpDNA maps for 18 species (Table 7) in Campanulaceae showing order of hybridized tobacco probes. Lines under maps indicate location and extent of IR. Asterisks indicate 23S rDNA duplicative transposition; (*) is partial deletion/divergence of 23S rDNA transposition. Size and location of large insertions designated by "i" followed by size in kb (insertions less than 5 kb not shown). Underlined numbers indicate sequence absent from an otherwise intact block of probes.
<table>
<thead>
<tr>
<th>Species</th>
<th>As synonym of</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>Campanula</td>
<td></td>
<td>17</td>
</tr>
<tr>
<td>Musschia</td>
<td></td>
<td>19, 115</td>
</tr>
<tr>
<td>Symphyandra</td>
<td></td>
<td>17</td>
</tr>
<tr>
<td>Edraianthus</td>
<td></td>
<td>(b)</td>
</tr>
<tr>
<td>Adenophora</td>
<td></td>
<td>(c)</td>
</tr>
<tr>
<td>Legousia</td>
<td></td>
<td>15</td>
</tr>
<tr>
<td>Asyneuma</td>
<td></td>
<td>15</td>
</tr>
<tr>
<td>Petromarula</td>
<td></td>
<td>111, 118</td>
</tr>
</tbody>
</table>

**FIGURE 21**
Figure 21 (continued)

<table>
<thead>
<tr>
<th>Triodanus</th>
<th>Wahlenbergia</th>
<th>Merciera</th>
<th>Prismatocarpus</th>
<th>Roella</th>
<th>Jasione</th>
<th>Codonopsis</th>
<th>Cyananthus</th>
<th>Platycodon</th>
</tr>
</thead>
</table>
Fig. 22. Strict consensus tree of 241 equally parsimonious trees of 94 steps generated using 84 cpDNA rearrangement characters from 18 taxa in the Campanulaceae. CI=0.76 (excluding autapomorphies). Numbers indicate changes at each node and along each branch; numbers below are bootstrap values.
Fig. 23. Strict consensus tree of 12 equally parsimonious trees of 125 steps (CI=0.78) based on 84 cpDNA rearrangement characters from 18 taxa in the Campanulaceae. The analysis gave a weight of two to non-inversion characters. Numbers indicate changes at each node and along each branch; numbers below are bootstrap values.
Fig. 24. Two equally parsimonious trees based on 84 cpDNA rearrangement characters in the Campanulaceae. The identical two trees were found in two separate cladistic parsimony analyses. The first involved recoding certain inversion endpoint characters, resulting in trees of 88 steps (CI=0.90). The second used both recoding of endpoints and weighting non-inversion characters, resulting in trees of 119 steps (CI=0.91). See text for details. Numbers indicate changes at each node and along each branch (for analysis that recoded but did not weight certain characters); numbers below are bootstrap values (identical in both analyses).
Fig. 25. Two models of seven steps each for the evolution of cpDNA structure in Platycodon from a tobacco-like ancestor, showing only inversions. Numbers in parentheses show order of hybridized tobacco cpDNA probes. Inversion endpoints are given as numbered arrows. Locations of probes 6-7, 8, and 9 believed to be the result of transposition (Chapter 3).
FIGURE 25
Fig. 26. Strict consensus tree of two equally parsimonious trees (Fig. 24) showing evolution of structural rearrangements in the Campanulaceae. I=inversion, i=insertion >5 kb, T=transposition, D=deletion/divergence. Insertions in the region of tobacco's large ORF1901 not shown. Letters following taxon names indicate Campanulinae (C), Platycodinae (P), and Wahlenberginae (W) of Schönland (1889). The genera Triodanus and Legousia were treated as Specularia, and Asyneuma and Petromarula as Phyteuma by Schönland.
Wahlenbergia (W)
Merciera (W)
Prismatocarpus (W)
Roella (W)
Legousia (C)
Asyneuma (C)
Petromarula (C)
Triodanus (C)
Trachelium (C)
Campanula (C)
Adenophora (C)
Symphyandra (C)
Edraianthus (W)
Jasione (W)
Musschia (P)
Codonopsis (W)
Cyananthus (W)
Platycodon (P)
tobacco

FIGURE 26
Fig. 27. Model for evolution of cpDNA structure in Wahlenbergia, Merciera, Prisamocarpus, and Roella from a Trachelium-like ancestor, showing only inversions. Numbers in parentheses show order of hybridized tobacco cpDNA probes. Inversion endpoints are given as numbered arrows. The final map shows only features shared by the three species; autapomorphies of individual taxa not pictured.
**Trachelium**


\[ \uparrow 1 \quad \uparrow 1 \quad \downarrow \text{Inversion 1} \]


\[ \uparrow 2 \quad \uparrow 2 \quad \downarrow \text{Inversion 2} \]


\[ \uparrow 3 \quad \uparrow 3 \quad \downarrow \text{Inversion 3} \]

**Wahlenbergia**


\[ \uparrow 4 \quad \uparrow 4 \quad \downarrow \text{Inversion 4} \]


\[ \uparrow 5 \quad \uparrow 5 \quad \uparrow 6 \quad \uparrow 6 \quad \downarrow \text{Inversions 5,6} \]

**Merciera, Prisamocarpus, Roella**


**FIGURE 27**
LITERATURE CITED


Neuhaus, H. 1989. Nucleotide sequence of the chloroplast genes for tRNA\textsubscript{Gln} and the 4 kD polypeptide of photosystem II from mustard (Sinapis alba). Nucleic Acids Res. 17:444.


