To the Graduate Council:

I am submitting herewith a dissertation written by Susan M. Swensen entitled "A Phylogenetic Analysis of the Datiscaceae Based on DNA Sequences of the Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase (Large Subunit) Gene." I have examined the final copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Life Sciences.

Beth C. Mullin, Major Professor

We have read this dissertation
and recommend its ~~acceptance~~:

Accepted for the Council:

Associate Vice Chancellor
and Dean of The Graduate School

A PHYLOGENETIC ANALYSIS OF THE DATISCACEAE BASED ON DNA SEQUENCES FROM THE RIBULOSE-1,5-BISPHOSPHATE CARBOXYLASE/OXYGENASE (LARGE SUBUNIT) GENE

A Dissertation

Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Susan M. Swensen

December 1992

ACKNOWLEDGMENTS

Several people deserve thanks and appreciation for efforts extended on my behalf during the course of my dissertation work. I would like to thank my major professor, Dr. Beth C. Mullin for her careful guidance of my work and my development as a scientist and for her efforts in seeing that all the tools required for this study were available. I would like to extend my appreciation to my committee members, Dr. Edward Schilling, Dr. Karen Hughes and Dr. Gary Stacey for their support, helpful suggestions, and proofreading. The careful editing of Dr. Edward Schilling and Dr. Dwight Baker was especially valuable and greatly improved the final form of this dissertation.

I would like to express my sincere gratitude to Dr. Mark Chase for his willingness to share his data with me and for the many helpful discussions which facilitated my data analysis and interpretation.

I am grateful to Dr. Loren Rieseberg for providing plant material and DNA samples for this study and for discussing the results of his work with me. I would like to thank Mr. Don Broach and Ms. Sue Smith from the University of Tennessee Computing Center for their advice in computing matters.

I would like to acknowledge financial support for this study from the Botany Department, Life Sciences Program, and Science Alliance.

ABSTRACT

The Datiscaceae are a small angiosperm family composed of three genera and four species. Datisca consists of two morphologically similar species with a disjunct distribution in western North America and Asia. The monotypic genera Octomeles and Tetrameles are native to the Malaysian region; their tree habit is in sharp contrast to the herbaceous perennial Datisca species. Despite previous morphological, anatomical and chemical study of the Datiscaceae, the phylogenetic interrelationships of the members of the family remain uncertain and there is no consensus regarding the phylogenetic affinities of the family as a whole. This study obtained new DNA sequence data from the chloroplast gene encoding the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase (rbcL) for each member of the Datiscaceae and for three members of the Cucurbitaceae. Other rbcL sequences included in this study were provided from the collection of M. W. Chase (University of North Carolina, Chapel Hill; Chase et al., 1993). DNA sequence data were used to infer the phylogenetic affinities of the Datiscaceae relative to other dicotyledonous angiosperms and to assess whether the family is monophyletic. DNA sequence data were analyzed using two different phylogenetic inference computer programs, PAUP and PHYLIP. Phylogenetic analyses based on the principle of parsimony and on overall similarity suggested that the Datiscaceae are not a monophyletic assemblage. Phylogenetic reconstructions placed the herbaceous Datisca species closer to representatives of the Cucurbitaceae and Begoniaceae than to the other genera within the Datiscaceae. Weighted parsimony was

utilized to select among equally parsimonious trees and the resulting topology placed Begonia as the sister group to Datisca. Among the 35 other taxa included in the analysis, the sister group to the Datiscaceae/Cucurbitaceae/Begonia clade was either a group containing representatives of the genera Myrica, Morus and Rhamnus or a group containing members of the genera Celtis, Casuarina, Fagus and Chrysolepis. Both of these groups contain actinorhizal genera which suggests that the actinorhizal plants as a group may be more closely related than traditional classifications propose. Generally, the phylogenetic reconstructions showed members of the higher Hamamelidae and Dilleniidae (Violales, Theales) grouping with members of the Rosidae (Rosales, Polygalales, Rhamnales, Euphorbiales, Sapindales). The rbcL sequence analysis conducted here does not support either the monophyly of the Datiscaceae or the monophyly of some of the traditionally-recognized subclasses of dicots.

TABLE OF CONTENTS

SECTION		PAGE
I.	INTRODUCTION AND LITERATURE REVIEW	1
	A. Introduction	1
	B. Review of Related Research	5
	1. Familial relationships of the Datiscaceae	5
	2. Relationships among genera of the Datiscaceae	9
	3. The rbcL gene	11
	4. A "global" rbcL phylogeny	13
II.	METHODS	20
	A. Generation of Sequence Data	20
	1. Taxa included in this analysis	20
	2. rbcL amplification	20
	3. rbcL cloning	23
	4. rbcL sequencing	25
	B. Analysis of Sequence Data	27
	1. Sequence input and proofreading	27
	2. Phylogenetic analysis	28
	3. Choosing the best reconstruction	30
III.	RESULTS AND DISCUSSION	32
	A. Results	32
	1. rbcL sequence data	32
	2. Phylogenetic analyses using PAUP	32
	3. Data analysis using PHYLIP	58
	4. Further analysis of the data	63

B. Discussion 71

1. Methods of analysis. 71

2. Paraphyly of the Datiscaceae 74

3. Outgroups to the Datiscaceae 78

4. Patterns of change 83

5. Congruence with the "global" phylogeny 85

6. Suggestions for further work 86

REFERENCES CITED 88

VITA 94

LIST OF TABLES

TABLE		PAGE
1	Actinorhizal Host Plant Genera.	4
2	Postulated Relative Families of *Datisca*	6
3	Sources of Chloroplast DNA.	21
4	Primers Used in Double-Stranded Sequencing of *rbcL*.	26
5	UWGCG Pair-wise Distances Among Datiscaceae and Cucurbitaceae	40
6	A Comparison of Morphological and Anatomical Features	79

LIST OF FIGURES

FIGURE		PAGE
1	Map of the Distribution of Datiscaceae	2
2	"Global" Phylogeny Based on rbcL Sequence Data	15
3	Detailed Section of the "Global" Phylogeny	18
4	rbcL Sequences from Datiscaceae and Cucurbitaceae.	33
5	Preliminary Phylogenetic Analysis using Eight Taxa	41
6	Strict Consensus of Trees Generated by Phylogenetic Analysis of 13 Taxa.	45
7	Strict Consensus of Eight Trees Generated by Phylogenetic Analysis of 33 Taxa.	47
8	One of Nine Equally Parsimonious Trees Generated by Phylogenetic Analysis of 43 Taxa	50
9	50% Majority-Rule Consensus of Nine Equally Parsimonious Trees.	52
10	50% Majority-Rule Consensus of 198 Near-Minimal-Length Trees.	54
11	50% Majority-Rule Consensus of 1540 Near-Minimal-Length Trees.	56
12	Consensus of Ten Equally Parsimonious Trees Generated by PHYLIP	59
13	Phenetic Analysis of 43 Taxa	61
14	Bootstrap 50% Majority-Rule Consensus Tree	65
15	Weighted-Parsimony Analysis of 43 Taxa	67

LIST OF ABBREVIATIONS

Adenosine Triphosphate	ATP
Halogenated indolyl-β-D-galactoside	Bluogal
Bethesda Research Laboratories	BRL
Base pairs	bp
Deoxyadenosine Triphosphate	dATP
Deoxycytidine Triphosphate	dCTP
Deoxyguanosine Triphosphate	dGTP
Deoxythymidine Triphosphate	dTTP
Deoxynucleoside Triphosphate	dNTP
Deoxyribonucleic Acid	DNA
Dithiothreitol	DTT
[Ethylenedinitrilo]-Tetraacetic Acid Disodium Salt	EDTA
Ethidium bromide	EtBr
Hydrogen chloride	HCl
International Biotechnologies, Inc.	IBI
Isopropylthio-β-galactoside	IPTG
Potassium chloride	KCl
Luria-Bertani medium	LB
Magnesium chloride	$MgCl_2$
New England Nuclear	NEN
Phylogenetic Analysis Using Parsimony (Swofford, 1991)	PAUP
Phylogenetic Inference Package (Felsenstein, 1989)	PHYLIP
Polymerase chain reaction	PCR

Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase

(Large Subunit) Gene . . . . . rbcL

Sodium chloride/sodium citrate . . . . . SSC

Tris-borate-EDTA . . . . . TBE

United States Biochemical. . . . . USB

University of Wisconsin Genetics Computer Group. . . . . UWGCG

Ultraviolet. . . . . UV

I. INTRODUCTION AND LITERATURE REVIEW

A. Introduction

The Datiscaceae are a small angiosperm family composed of three genera and four species. Datisca consists of two morphologically similar herbaceous perennials which exhibit a disjunct distribution. Datisca glomerata (Presl) Baill. ranges from northern Baja California to northern California whereas Datisca cannabina L. is found in southwestern Asia and Crete (see Figure 1; Davidson, 1973). Both species of Datisca are actinorhizal and engage in a nitrogen-fixing symbiosis with the actinomycete Frankia. The other two genera of the Datiscaceae, which are not actinorhizal, are monotypic; Octomeles sumatrana Miq. and Tetrameles nudiflora R. Br. are trees native to the Malaysian region (Davidson, 1973). The Datiscaceae are of interest systematically because their relationship to other plant families is not well-defined; therefore, their phylogenetic placement is uncertain. Also, the contrast in habit between the herbaceous species of Datisca and the tree genera have prompted questions regarding the validity of the Datiscaceae as a monophyletic group. Previous studies of the Datiscaceae which have utilized morphological, anatomical, or chemical data have revealed no consensus regarding these questions. This study utilizes DNA sequence data as a new source of information to address the controversy surrounding the phylogenetic placement of the Datiscaceae and the monophyly of the group.

The primary focus of this study is the assembly and analysis of DNA sequence data from the chloroplast gene encoding the large

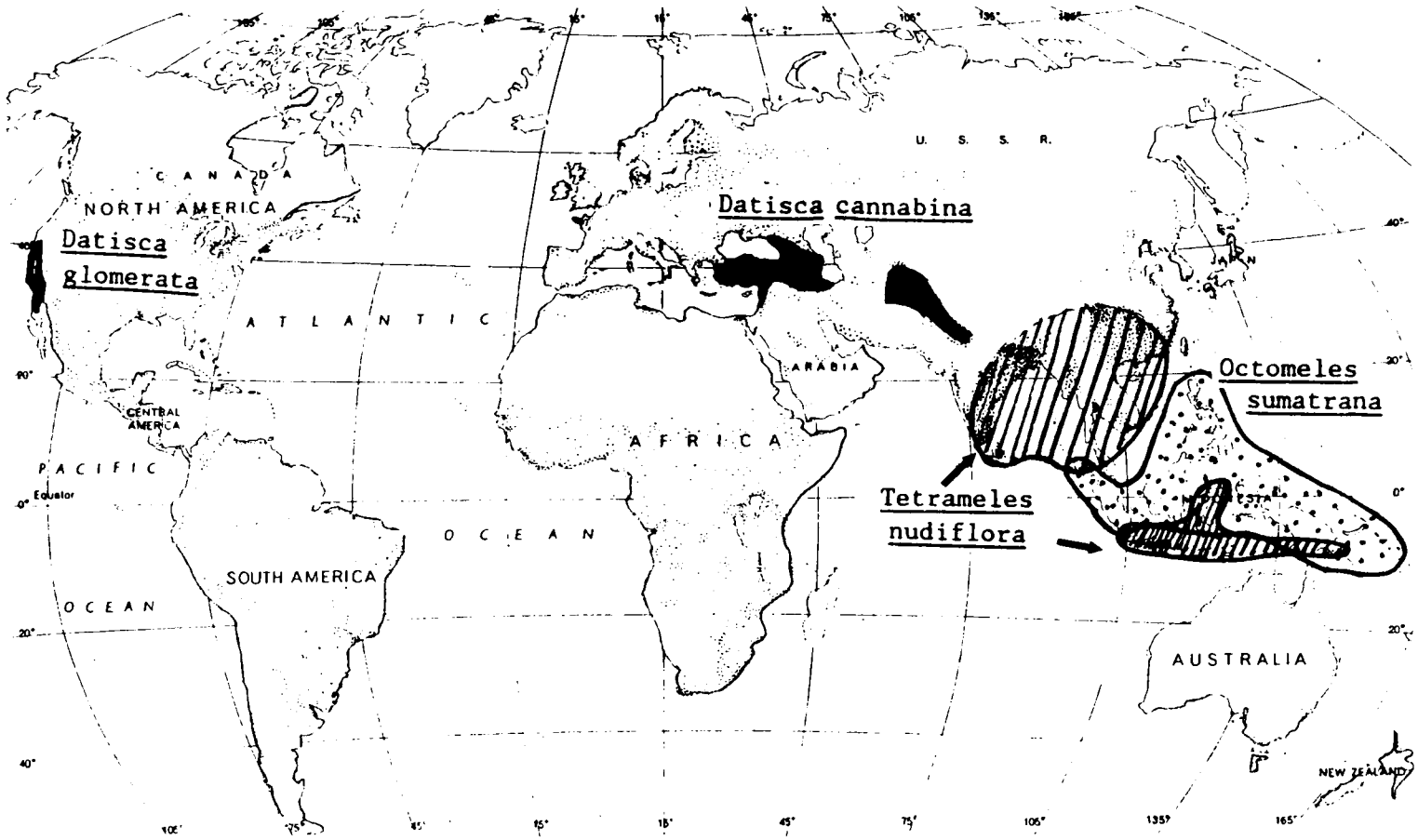


Figure 1. Map of the Distribution of Datiscaceae (modified from Davidson, [1973]).

subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase (rbcL). Sequences from the Datiscaceae and potentially related taxa have been obtained and analyzed to produce phylogenetic estimations which provide information regarding the affinities of the Datiscaceae to other taxa and to each other.

Another focus of this study is an investigation of the relationship of the actinorhizal Datisca to other actinorhizal taxa. In addition to Datisca, actinorhizal plants are found in 23 genera belonging to seven angiosperm families (see Table 1). The actinorhizal genera are a relatively diverse group, especially when compared to the symbiotic legumes. Symbioses with Rhizobium are found in many members of the Fabales which have been traditionally accepted as a monophyletic group. The only exception is the Parasponia-Rhizobium symbiosis in which the host plant is a member of the family Ulmaceae (order Urticales). The diversity of actinorhizal genera has led researchers in this area to postulate that the symbiotic relationship with Frankia was established early in evolutionary time and the hosts have since diverged. An alternative explanation suggests that different plants may have developed the ability to engage in the symbiosis at a much later time in a series of independent events (Baker and Mullin, 1992; Mullin et al., 1990). This study has considered the affinities of Datisca to other actinorhizal taxa, based on rbcL sequence data. A less distant relationship among actinorhizal taxa, than is indicated by current classification schemes, might provide support for the symbiosis having arisen once in evolutionary history. A distant relationship, consistent

Table 1. Actinorhizal Host Plant Genera[a]

Family	Genus	Numbers of Nodulated Species[b]
Betulaceae	Alnus	47
Casuarinaceae	Allocasuarina	54
	Casuarina	16
	Ceuthostoma	2
	Gymnostoma	18
Myricaceae	Comptonia	1
	Myrica	28
Elaeagnaceae	Elaeagnus	38
	Hippophaë	2
	Shepherdia	2
Rhamnaceae	Ceanothus	31
	Colletia	4
	Discaria	5
	Kentrothamnus	1
	Retanilla	2
	Talguenea	1
	Trevoa	2
Rosaceae	Cercocarpus	4
	Chamaebatia	1
	Cowania	1
	Dryas	1[c]
	Purshia	2
Coriariaceae	Coriaria	5
Datiscaceae	Datisca	2

[a]Table modified from Baker and Schwintzer, 1990.

[b]All recognized species of these genera have been observed to nodulate; however, some exceptions involve unconfirmed reports of nodulation which currently remain unresolved (D.W. Baker, pers. comm.)

[c]Although three species were originally reported to be nodulated, recent reports have confirmed nodulation in only Dryas drummondii.

with current classification, would tend to support multiple origins of the symbiotic association.

B. Review of Related Research

1. Familial relationships of the Datiscaceae

Despite previous morphological, anatomical and chemical studies no consensus has been reached on the phylogenetic position of the Datiscaceae. Datiscaceae has been allied with six different families (see Table 2) mostly within the order Violales (sensu Cronquist, 1981) near the Begoniaceae.

The relationship between the Datiscaceae and the Begoniaceae was first suggested by Lindley in 1846. Most systematic treatments have also proposed the Begoniaceae as a sister group (Cronquist, 1981; Takhtajan, 1980; Thorne, 1992; Lawrence, 1951). Similarities in ovule and seed anatomy between the Datiscaceae and the Begoniaceae (Boesewinkel and DeLange, 1983; Boesewinkel, 1984) support a close relationship for the two families. These similarities include the simple construction of the ovules and seeds, dizonate ovule primordia and two-layered integuments, operculate seeds with crushed exotestal seed coat, a relatively large, straight embryo and an endosperm that is only one cell layer thick. The main differences in the ovules and seeds are the larger nucellus in Datisca and the lack of collar cells around the operculum and the enlarged nucellar epidermal cells which are both found in the Begoniaceae (Boesewinkel, 1984). Despite the correspondences observed in ovule and seed structure between

Table 2. Postulated Relative Families of Datisca

Family	Order/Subclass	References
Begoniaceae	Violales/Dilleniidae	Corner (1976) Boesewinkel (1983, 84) Cronquist (1981) Tahktajan (1980) Thorne (1976) Lawrence (1951) Engler (1964)
Capparaceae	Capparales/Dilleniidae	Cronquist (1981) Tahktajan (1980)
Cucurbitaceae	Violales/Dilleniidae	Engler (1964) Cronquist (1981) Tahktajan (1980)
Flacourtiaceae	Violales/Dilleniidae	Davidson (1976)
Haloragaceae	Haloragales/Rosiidae	Airy-Shaw (1964)
Loasaceae	Violales/Dilleniidae	Cronquist (1981) Lawrence (1951)

Datiscaceae and Begoniaceae, morphological features of vegetative parts of the plants show few similarities.

Extensive morphological and anatomical studies of the Datiscaceae by Davidson (1973, 1976) have revealed trends and features that are also present in the Begoniaceae and in the Flacourtiaceae. Among these are tendency toward an inferior ovary, monoecy and dioecy, separate stigmas, capsular fruits, numerous anatropous ovules and seeds, inflorescences with many flowers, tendency toward anemophily, valvate sepals and nectar glands between gynoecium and androecium. Davidson (1976) mentions that the plasticity of such "traditionally reliable" characteristics in conjunction with the small size of the Datiscaceae and the Begoniaceae have made their phylogenetic relationship and placement unclear. He supports a relationship to Begoniaceae, but only through common descent from an ancestor shared with Flacourtiaceae.

A contrasting view treats the Datiscaceae as a monogeneric family (Datisca only) and allies it to the Haloragaceae in the Rosidae (Airy-Shaw, 1964). This separation of Datisca and its alliance to Haloragaceae appears to be based upon habit-related features and the presence of pinnate or pinnatifid leaves and stamens of indefinite number in Datisca. Davidson (1976) presents other similarities between Datisca and Haloragaceae but concludes that the relationship to Haloragaceae is not well-supported morphologically or anatomically.

Flavonoid analysis of the Datiscaceae (Bohm, 1988) does not provide strong evidence for relating the family to either Begoniaceae or Haloragaceae; however, the presence of kaempferol and quercetin 3-O-

mono- and diglycosides in the Datiscaceae is consistent with an alliance to the Begoniaceae.

Cucurbitacins have been found in Datisca (Restivo et al., 1973; Sasamori et al., 1983; Pohlmann, 1975, as cited in Metcalf and Rhodes, 1990) which might suggest an alliance with the Cucurbitaceae; however, cucurbitacins have also been detected in species of several other families including the Begoniaceae (Doskotch et al., 1969; Doskotch and Hufford, 1970), Brassicaceae, Euphorbiaceae and Scrophulariaceae (Metcalf and Rhodes, 1990).

The Datiscaceae have also been placed near either or both the Cucurbitaceae and the Loasaceae in different systematic treatments (Cronquist, 1981; Takhtajan, 1980; Lawrence, 1951; Thorne, 1976). The evidence supporting a relationship to these two families is even less clear than that supporting the alliance to Begoniaceae. The phylogenetic affinities of each of the three putative relative families are controversial. Cronquist (1981) places the Cucurbitaceae, Loasaceae and Begoniaceae within the Violales, but mentions the possible relationship of the Cucurbitaceae to the Grossulariaceae or Sapindaceae. Cronquist (1981) also suggests that if the Begoniaceae and Loasaceae are removed from the Violales, then the Cucurbitaceae should be as well. Hufford (1990) supports the possibility that the Loasaceae may share ancestry with the Cornales and woody saxifrages.

A relationship between the Datiscaceae and the Capparales is implied by classifications of Cronquist (1981), Takhtajan (1980) and Thorne (1992). This order was once considered, along with the Papaverales, to constitute the old, collective order Rhoeadales.

Although the Capparales have now been shown to be distinct from the Papaverales (Cronquist, 1981), which is now placed in the Magnoliidae, the phylogenetic relationship of the Capparales to other taxa remains unclear.

Because there is a lack of clear evidence for the placement of the Datiscaceae among dicotyledonous angiosperms, this study was designed to include many taxa to maximize the probability that the most closely related groups could be detected.

2. Relationships among genera of the Datiscaceae

Two views exist regarding the relationship of the three genera that comprise the Datiscaceae. The predominant view in taxonomic literature is that the genera should be considered as one family. The other view favors the separation of Octomeles and Tetrameles into a separate family, the Tetramelaceae, making the Datiscaceae a monogeneric family (Airy-Shaw, 1964).

The reasons for the separation by Airy-Shaw are based primarily on the habit-related differences between the herbaceous Datisca and the two tree genera, although other non-habit related differences are also cited. Datisca has pinnate or pinnatifid leaves with long narrow caudate-acuminate pinnae whereas in the Tetramelaceae, the leaves are simple and cordate-ovate. Differences in stamen number, arrangement and form and differences in style characteristics between Datisca and Tetramelaceae are mentioned. Although these differences and others are acknowledged by Davidson (1973, 1976), he concludes that the majority of evidence does not favor the separation of Datisca into its own family. Correspondences in ovule and seed structure among all three

genera (Boesewinkel, 1984) seem to indicate that they are best treated as a single family.

Flavonoid profiles of Datisca, Octomeles and Tetrameles (Bohm, 1988) reveal differences that support the recognition of two subgroups within the Datiscaceae. The profiles of Octomeles and Tetrameles consist only of kaempferol and quercetin 3-O-mono- and diglycosides and lack the unusual B-ring unsubstituted and 2'-hydroxylated flavonols that characterize the two species of Datisca. This difference in flavonoid profiles was not considered as support for the formation of two families; the magnitude of differences is not unusual within a family. Bohm (1988) concludes that the differences might best be acknowledged by recognizing two tribes, Datisceae and Tetrameleae, as has been done in the past (Engler, 1964).

The two species of Datisca differ in their breeding systems. Datisca glomerata has been documented as being functionally androdioecious (having both male and hermaphrodite individuals in a breeding population) and is the only species in which androdioecy has been confirmed through breeding studies to be present (Liston et al., 1990). In addition to the difference in breeding systems, the species of Datisca have been found to be genetically very different, based on the low genetic identity value (I=0.142) from isozyme analysis (Liston et al., 1989). This value is lower than any of the comparable values for congeneric species from isozyme data in flowering plants and is suggested to be evidence for the ancient divergence of the two species (Liston et al., 1989).

To test the hypothesis that the androdioecious breeding system in D. glomerata is derived from the dioecious condition, Rieseberg et al. (1992) utilized restriction site mapping of chloroplast DNA fragments to generate a molecular phylogeny of Datiscaceae. Maximum parsimony analysis resulted in trees which suggested that Datisca is monophyletic, whereas Octomeles and Tetrameles grouped together as a sister group to Datisca. In all trees, D.glomerata occupies a derived position within the Datiscaceae. The most parsimonious explanation supports the idea that androdioecy is derived from dioecy in Datiscaceae (Rieseberg et al., 1992). Only six changes (half of which are homoplasious) separate the Octomeles/Tetrameles group from the Datisca group, although since only the Datiscaceae were included in the ingroup, no comparisons could be made to address the question of monophyly of the Datiscaceae.

3. The rbcL gene

The rbcL gene, located in the chloroplast genome, encodes the large (catalytic) subunit of the enzyme ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBP carboxylase or Rubisco). The small subunit is encoded by the nuclear gene, rbcS. Together, eight large subunits and eight small subunits form an enzyme that makes up more than 50% of the total chloroplast protein and is said to be the most abundant protein in the world. RuBP carboxylase catalyzes the initial reaction of the Calvin cycle, the conversion of ribulose-1,5-bisphosphate to two molecules of 3-phosphoglycerate. In addition to its carboxylase activity, the enzyme catalyzes a competing reaction using O_2 as substrate instead of CO_2, and forming phosphoglycolic acid and 3-phosphoglycerate. Phosphoglycolic acid is oxidized during the process

of photorespiration to yield CO_2, which compromises the efficiency of photosynthesis in C_3 plants.

Several features of the rbcL gene have made it the most commonly-used gene for phylogenetic sequencing. The gene is located in the single-copy region of the chloroplast's circular genome and is present in all photosynthesizing plants. The size of rbcL, approximately 1430 bp, is long enough to provide a sufficient number of characters for phylogenetic analysis. The rate of rbcL evolution is considered to be moderate when compared to the slowly-changing ribosomal genes, or the more quickly-changing chloroplast genes rpoC (encodes RNA polymerase subunit) or ndhF (encodes NADH dehydrogenase subunit). This rate of change is appropriate for phylogenetic studies at higher taxonomic levels. rbcL is composed of one exon with no introns and has very few length variations. These features facilitate sequencing and alignment with other rbcL sequences. Additionally, the widespread interest in rbcL for phylogenetic studies has generated a large data base that now includes sequences from approximately 1500 different taxa (Chase et al., 1993).

To utilize rbcL as an indicator of phylogenetic relationships, it is important to understand as thoroughly as possible the characteristics of nucleotide change within the gene so that the most appropriate model of phylogenetic analysis can be applied. Numerous studies have shown that different classes of molecular change occur with differing probabilities. For example, nucleotide substitutions at the third codon position are more likely to be retained than substitutions at the first or second codon positions because of the degeneracy of the genetic code.

Substitutions involving transversions (purine to pyrimidine or vice versa) are less probable than transitions (purine to purine or pyrimidine to pyrimidine) because they are more complicated chemical reactions. Empirical studies such as those of Clegg et al. (1986) and Ritland and Clegg (1987) show that for rbcL, there is approximately a twofold excess of transitions over transversions and a sixfold excess of synonymous substitutions (third codon position) over second position substitutions. Based on such observations of nucleotide change, it is possible to develop a character-weighting model that estimates the cost of various nucleotide changes at each character site. Albert et al. (1993) have developed a model of character-state weighting that uses a maximum-likelihood framework for the calculation of character transformation probabilities and their weights. When this weighted-parsimony method is applied to sequences with sufficiently small rates of change, the parsimony estimate can also be a maximum-likelihood estimate (Felsenstein, 1981). If rates of change in rbcL can be considered sufficiently small, the weighted-parsimony approach may circumvent the computationally complex maximum-likelihood methods for estimating phylogeny.

4. A "global" rbcL phylogeny

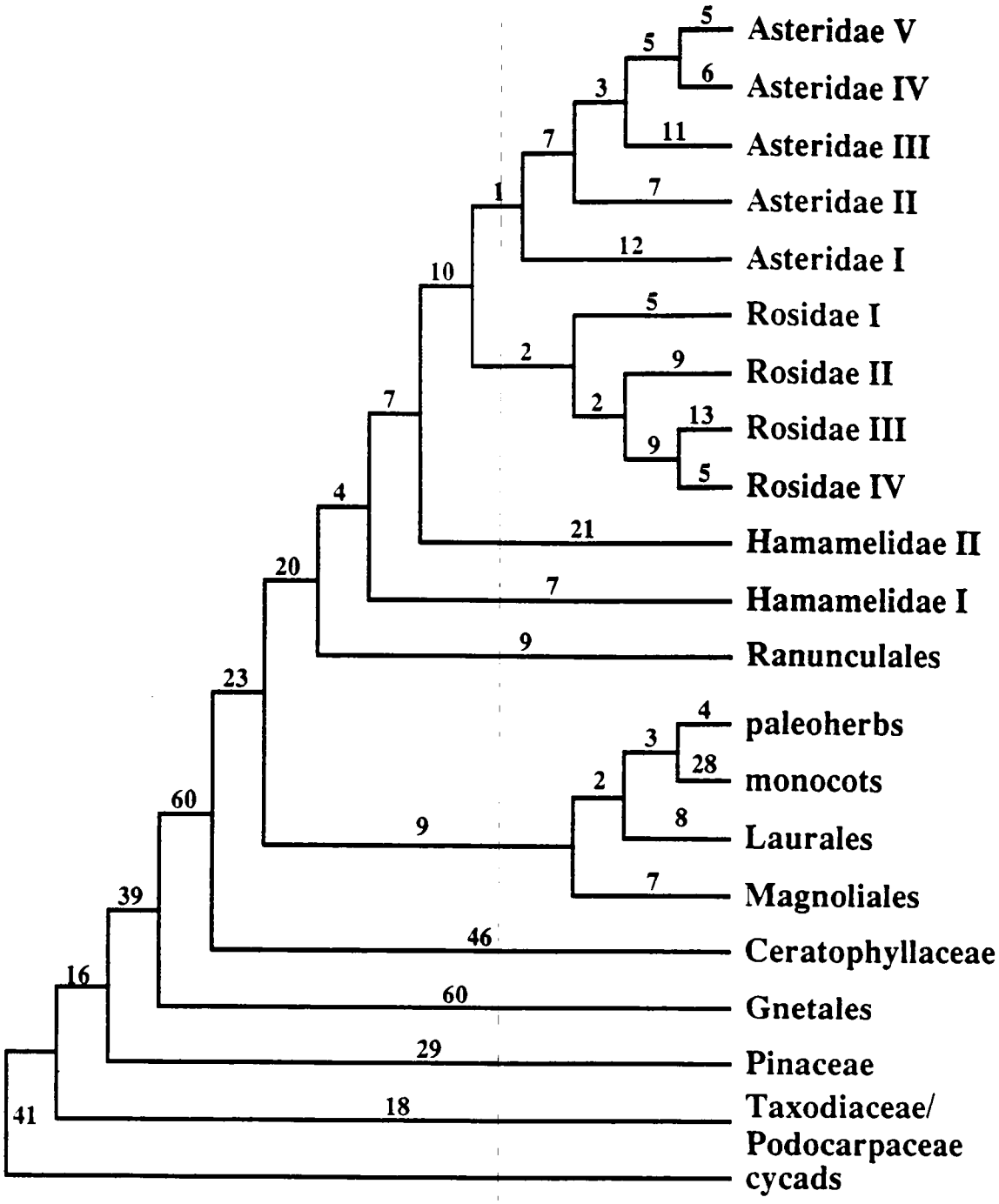
With the development of the polymerase chain reaction (PCR) the accumulation of rbcL sequence data has increased at a dramatic rate. Sequence data have been compiled for approximately 1500 species of angiosperms with all major taxonomic groups represented (Chase et al., 1993). The availability of this information has allowed expansive

comparisons to be made among seed plants, and the construction of rbcL-based phylogenies relating these taxa.

An rbcL-based maximum parsimony analysis including sequence data from 475 seed plants has been conducted by Chase et al. (1993; Figure 2). The intent of this study was to assess the utility of rbcL sequence data for higher-level phylogenetic analysis and to provide a framework to stimulate further studies that focus on restricted parts of the "global" phylogeny. Although the optimality (maximum parsimony) of their consensus reconstruction could not be guaranteed because of the computational difficulty in analyzing such a large data set, several observations could be made. Many characteristics of the tree are congruent with traditional classification and recent studies based on morphology, anatomy, or secondary chemistry. Studies focusing on restricted portions of this phylogeny have revealed relatively robust support for the interior branches. These results suggest that rbcL sequences contain relevant information that is useful for defining relationships at higher taxonomic levels.

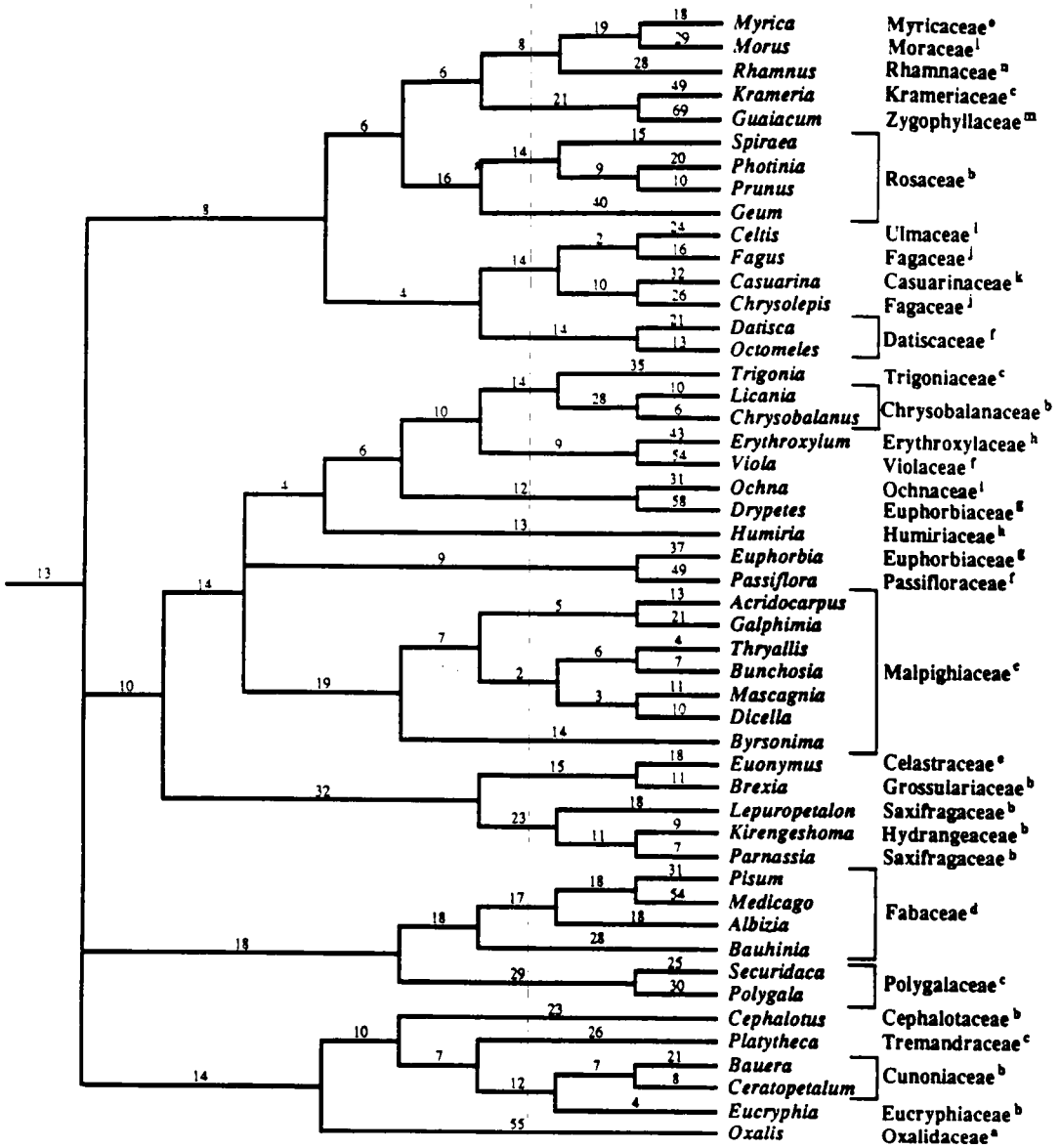
Several unexpected relationships are found in this extensive rbcL-based analysis; some have been detected by other studies and some are unique to the broad study. Of particular relevance to this project is the grouping of several dilleniid orders (including the Datiscaceae) among rosid clades. The Rosidae are split into four recognizable groups (see Figure 2): the basal-most lineage contains the Saxifragaceae sensu stricto, Crassulaceae and some of the lower hamamelids; the next-most-basal group contains the Caryophyllidae sensu lato, Droseraceae, Nepenthaceae, Dilleniaceae and Vitaceae; the other two groups form large

Figure 2. "Global" Phylogeny Based on rbcL Sequence Data. Chase et al. (1993) summary of the major clades identified in the semi-strict consensus of 500 equally parsimonious trees found using the character-state weighting method of Albert et al. (1993). The analysis included 475 species of seed plants and the ingroup is arranged with the cycads basal most. Numbers above the branches are branch lengths optimized under equal weighting. Figure taken directly from Chase et al. (1993).


Asteridae V
Asteridae IV
Asteridae III
Asteridae II
Asteridae I
Rosidae I
Rosidae II
Rosidae III
Rosidae IV
Hamamelidae II
Hamamelidae I
Ranunculales
paleoherbs
monocots
Laurales
Magnoliales
Ceratophyllaceae
Gnetales
Pinaceae
Taxodiaceae/
Podocarpaceae
cycads

sister groups, one (Figure 3) containing several families of higher Hamamelidae, the Euphorbiales, Fabales, Linales, Polygalales, Rosales plus several dilleniid families including the Datiscaceae, and the other sister group containing the orders Myrtales and Sapindales. The rosid group containing the Datiscaceae (Figure 3) is especially heterogeneous when compared to traditional classifications. The sampling of the families that may potentially belong to this group is sparse and therefore, more limited analyses could be focused on this region of the tree to better characterize the relationships depicted there.

Figure 3. Detailed Section of the "Global" Phylogeny. This section of the broad analysis represents one of the two higher rosid lineages identified by Chase et al. (1993) that contains Datiscaceae members. The numbers above the branches are the number of substitutions optimized on the general semi-strict consensus tree. Figure taken directly from Chase et al. (1993).



Rosidae III

[a] Rosidae, Geraniales
[b] Rosidae, Rosales
[c] Rosidae, Polygalales
[d] Rosidae, Fabales
[e] Rosidae, Celastrales
[f] Dilleniidae, Violales
[g] Rosidae, Euphorbiales
[h] Rosidae Linales
[i] Dilleniidae, Theales
[j] Hamamelidae, Fagales
[k] Hamamelidae, Urticales
[l] Hamamelidae, Casuarinales
[m] Rosidae, Sapindales
[n] Rosidae, Rhamnales
[o] Hamamelidae, Myricales

II. METHODS

A. Generation of Sequence Data

1. Taxa included in this analysis

Taxa chosen for sequencing included all members of the Datiscaceae and three representatives from Cucurbitaceae. No rbcL sequences were available for either family. The taxa from the Cucurbitaceae were chosen from three of the eight tribes in the Cucurbitoideae subfamily (Cucumis: Melothrieae; Cucurbita: Cucurbiteae; Luffa: Benincaseae; Jeffrey, 1990). Leaf tissue or DNA from each species in the Datiscaceae was provided by Dr. Loren Rieseberg, Rancho Santa Ana Botanic Garden (see Table 3). Leaf tissue of Cucurbita pepo (Butterstick Hybrid, Burpee Seed Company) and Luffa quinquefida (original source: Mazatlán, Mexico; C. B. Heiser) was obtained from local gardens. Tissue from Cucumis sativus was obtained from a store-bought fruit. rbcL sequences for all other taxa were obtained from Dr. Mark W. Chase, University of North Carolina, Chapel Hill (Chase et al., 1993) who has along with his lab, sequenced rbcL from many of the taxa and collected the other sequences from several labs within the U.S., Canada and Sweden.

2. rbcL amplification

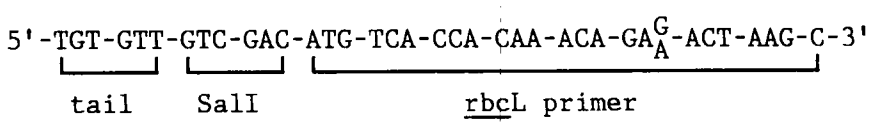
DNA extractions from leaf or fruit tissue sources were carried out using the method of Dellaporta et al. (1983). After the final precipitation step, DNA was taken up in sterile water and a 10 ng/ul solution was prepared from the original stock. The polymerase chain reaction (PCR; Mullis and Faloona, 1987; Saiki et al., 1985) was used to

Table 3. Sources of Chloroplast DNA. Vouchers at TENN unless noted otherwise. [1]Vouchers at RSA. [2]Purchased at local grocery. [3]Purchased from Burpee Seed Company.

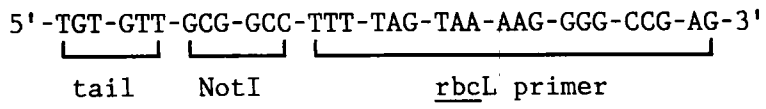
Taxon	Designation	Origin/Locality
Datisca glomerata (Presl) Baill.	Dglomerata	Inyo Co., California
Datisca cannabina L.	Dcannabina	[1]Tadzhik SSR, Soviet Union
Octomeles sumatrana Miq.	Octomeles	[1]Lundu District, Sarawak, Malaysia, Philbrick and Isnai 2273
Tetrameles nudiflora R. Br.	Tetrameles	[1]Central Botanical Garden, Sara Bori, Thailand, Philbrick and Wongpraseri 2272
Cucumis sativus L.	Cucumis	[2]Tropical Africa
Cucurbita pepo L.	Cucurbita	[3]South-central Mexico
Luffa quinquefida (Hook. & Arn.) Seeman	Luffa	Mazatlán, Mexico C. Heiser (LoMex)

amplify the section of chloroplast DNA that encodes the large subunit of the RuBP carboxylase protein. Amplification primers were synthesized by the Molecular Biology Research Facility (University of Tennessee) using an Applied Biosystems Model 381A Synthesizer. The primer sequences included restriction sites built into the 5' end with a "tail" of 6 bases and approximately 25 bases of coding sequence.

5' (forward) primer:



3' (reverse) primer:



The forward primer anneals to the first 25 bp of the rbcL gene, while the reverse primer anneals to the opposite strand approximately 1700 bp downstream from the start codon. To amplify the rbcL region, 50-100 ng of total DNA was added to a PCR cocktail which included 0.5 µM (each) amplification primers, 200 µM each of dATP, dCTP, dGTP, dTTP, 10 mM Tris-HCl pH 8.3, 50 mM KCl, 2.5 mM $MgCl_2$ in a volume of 99.5 µl. After mixing, 2.5 units of Taq DNA polymerase (AmpliTaq, Perkin Elmer Cetus) were added and mixed by vortexing. A layer of mineral oil was added on top of the PCR reaction and the tubes were placed in a Perkin Elmer DNA Thermal Cycler. The amplification program consisted of 2 min at 95°C, 1 min at 58°C, 1 min at 72°C followed by 30 cycles of 1 min at 95°C, 1 min

at 58°C, 1 min + 6 sec per cycle at 72°C. After the final cycle, the reactions were held at 72°C for 6 min to allow the final extension of any uncompleted strands followed by a soak at 4°C until reactions were removed from the cycler. The success of each amplification was determined by running a 5 µl sample of the reaction through a 1.4% agarose mini-gel in TBE buffer at 100V. The gels were stained with EtBr and visualized on a UV transilluminator. The size and quantity of the amplified product was estimated by comparison to HaeIII-cut ØX174 or HindIII-cut lambda DNA standards (BRL).

3. rbcL cloning

In preparation for cloning, the amplified product was purified by phenol/chloroform extraction and ethanol precipitation. NotI and SalI restriction endonucleases (USB) were used to generate NotI- and SalI-compatible ends on the amplified DNA and on the cloning vector pBluescript KS (Stratagene). Ligation reactions contained 200 ng each of phenol/chloroform-extracted vector and amplified DNA, 10 mM DTT, 0.5 mM ATP, 10 mM $MgCl_2$, 50 mM Tris-HCl pH 7.5 and 1 Unit of T4 DNA Ligase (BRL). Ligations were carried out at 16°C for 8-12 hours after which time more components were added so that the resulting concentrations remained the same, except that the total DNA concentration was diluted by half. The reaction was held at 16°C for 4-6 more hours.

Competent cells were prepared from E. coli strain DH5α (BRL) by treatment with $CaCl_2$ according to the procedure in Maniatis et al. (1989) except that the competent cells were kept on ice for 12 hours prior to transformation. Ligated DNA (100-200 ng) was gently mixed into 200 µl aliquots of competent cells in Eppendorf tubes which were placed

on ice for 35 min. The cells were heat-shocked in a 42°C water bath for 90 seconds then placed back on ice. 600 µl of S.O.C. media (2% Bacto-tryptone, 0.5% Bacto-yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose) were added to the tubes followed by a 1 hr incubation at 37°C with gentle shaking. Various volumes (100-400 µl) of the transformations were then spread onto LB (1% Bacto-tryptone, 0.5% Bacto-yeast extract, 0.17 M NaCl, 1.5% agar) agar plates containing 100 µg/ml ampicillin (Sigma) with 75 µl Bluogal (2%; BRL) and 25 µl IPTG (100 mM; Sigma) spread over the top 1-2 hr prior to plating. After drying, the plates were sealed with parafilm and incubated overnight (12-17 hr) at 37°C.

To identify recombinant clones containing amplified rbcL, colony lifts were prepared and hybridized with purified rbcL DNA which had been amplified previously. Colonies were lifted onto circular nylon (Amersham) filters and the bacteria were immediately lysed by placing filters (bacteria side up) in 0.75 ml puddles of denaturing solution (1.5 M NaCl, 0.5 N NaOH) on Saran Wrap for 7 min and in puddles of neutralizing solution (1.5 M NaCl, 0.5 M Tris-HCl pH 7.2, 1 mM EDTA) for 3 min, repeating the neutralization treatment (modified after Maniatis et al., 1989). The filters were rinsed in 2 changes of 2 x SSC (0.3 M NaCl, 0.03 M sodium citrate, pH 7) and were dried on Whatman 3MM paper. The DNA was bound to the filters by placing the filters (colony side down) on a UV transilluminator for 5 min. The filters were hybridized using the procedure suggested by the manufacturer. The probe used in hybridization was labeled with ^{32}P-dCTP (NEN) by the random primer method (BRL). After washing the filters according to the manufacturer's

instructions, the blots were placed in sealed bags and were exposed to x-ray film (Kodak X-Omat AR) in Kodak cassettes. After sufficient exposure time, the film was developed and the hybridizing colonies were chosen from the plates.

To verify that the selected colonies contained the proper recombinant plasmids, mini-preps (alkaline lysis method, Maniatis et al., 1989) were used to isolate the plasmid DNA which was restricted with NotI and SalI to liberate the rbcL insert. The sizes of the resulting fragments were confirmed by agarose gel electrophoresis.

4. rbcL sequencing

In preparation for sequencing, the plasmid DNA from four mini-preps was combined and treated with RNAse (Sigma) to remove RNA. The sample was extracted twice with phenol/chloroform, once with chloroform and precipitated with ethanol. The DNA was prepared for double-stranded sequencing according to the protocol supplied with the Sequenase Version 2.0 sequencing kit (USB). Once prepared, the template was sequenced using ^{35}S-dATP (NEN) according to the Sequenase instructions for single-stranded DNA . Sequencing primers used to sequence through both strands of rbcL are listed in Table 4. The M13/pUC primers anneal to vector sequences flanking the insert DNA, allowing sequences from the very ends of the rbcL DNA to be read. The internal primer sequences are based on maize rbcL and anneal to the approximate location in rbcL (relative to the ATG start codon) designated by the number of the primer. Seven primers were used for sequencing in the forward direction and eight primers were used in the reverse direction (opposite strand).

Table 4. Primers Used in Double-Stranded Sequencing of rbcL

Primer	Sequence	Source
M13/pUC	5'-GTAAAACGACGGCCAGT-3'	BRL
M13/pUCR	5'-AACAGCTATGACCATG-3'	Stratagene
Z234	5'-CGTTACAAAGGACGATGCTACCACATCGA-3'	Gerard Zurawski DNAX Research Institute Palo Alto, CA
Z427	5'-GCTTATTCAAAAACTTTCCAAGGCCCGCC-3'	
Z674	5'-TTTATAAATCACAAGCCGAAACTGGTGAAATC-3'	"
Z895	5'-GCAGTTATTGATAGACAGAAAAATCATGGT-3'	"
Z1020	5'-ACTTTAGGTTTTGTTGATTTATTGCGCGATGATT-3'	"
Z1204	5'-TTTGGTGGAGGAACTTTAGGACACCCTTGGGG-3'	"
Z1375R	5'-AATTTGATCTCCTTCCATATTTCGCA-3'	"
Z1204R	5'-CCCTAAGGGTGTCCTAAAGTTTCTCCACC-3'	"
Z1020R	5'-ATCATCGCGCAATAAATCAACAAAACCTAAAGT-3'	"
Z895R	5'-ACCATGATTCTTCTGCCTATCAATAACTGC-3'	"
Z674R	5'-GATTTCGCCTGTTTCGGCTTGTGCTTTATAAA-3'	"
Z346R	5'-AAATACGTTACCCACAATGGAAGTAAATAT-3'	"
Z153R	5'-AGAAGATTCCGCAGCCACTGCAGCCCCTGCTTC-3'	"

Sequencing reactions were electrophoresed on 6% polyacrylamide (bis:acrylamide = 1:19), 48% urea gels using 0.4 mm spacers and sharkstooth combs. The gels were held in a BRL model SO apparatus modified to accommodate an aluminum plate clamped to the back of the gel to dissipate heat and prevent smiling. After electrophoresis, the gels were soaked in a solution of 10% acetic acid, 12% methanol, for 15 min to remove urea. The gels were transferred to Whatman 3MM paper and dried in a gel dryer (Hoefer SE 1160) set at 80°C for 1 hr under vacuum. For autoradiography, gels were placed in direct contact with x-ray film (Kodak, X-Omat AR) in a cassette and compressed with heavy weights to ensure good contact with the film. After 40-50 hr, the film was developed and sequences were read manually. Sequence data were verified by manually comparing sequences from both strands and by comparing overlapping regions generated by sequencing with neighboring primers.

B. Analysis of Sequence Data

1. Sequence input and proofreading

The sequence data were transferred from an IBM PC via modem to a VAXcluster mainframe computer (University of Tennessee Computing Center) using the SEQED program of the University of Wisconsin Genetics Computer Group (UWGCG) program package (Devereux et al., 1984). To verify the accuracy of the sequences, the TRANSLATE program was used to generate peptide sequences from the DNA sequences. The peptide sequences were checked for the presence of reading frame interruptions (early "stop" codons) indicative of a mistake in the DNA sequence. The amino acid sequences were also compared to a published sequence. Each time a

discrepancy was found, the sequence data were checked by rereading the autoradiogram. Most mistakes were identified by the stop codon search whereas only one or two mistakes (usually none) per sequence were found by the comparison of amino acid sequences. All mistakes were clearly due to misreading of the autoradiograms or to mistyping of the sequence while entering the data into the computer. To determine the number of base differences between pairs of sequences, the GAP program was used to first align the sequences. A file containing the names of the aligned sequences was specified in the DISTANCES program which provided a matrix of pair-wise genetic distance values. The PRETTY program was used to display the multiple aligned sequences.

2. Phylogenetic analysis

A file containing the aligned sequences was transferred from the VAXcluster to a Macintosh SE30 computer via modem. The file was edited into the NEXUS format necessary for input files to be read by PAUP 3.0s, a phylogenetic analysis program that utilizes the principle of parsimony (Swofford, 1991). The file containing seven sequences was augmented by the addition of rbcL sequences from 36 other taxa. The additional data were selected from the data matrix of Chase et al., 1993 based on preliminary data generated by their global phylogeny containing over 400 taxa. Because of the large size of the augmented data set, the heuristic search procedure was selected as the algorithm used for searching for the most parsimonious trees. The other two algorithms available in PAUP are the exhaustive search and the branch-and-bound search. The exhaustive search generates all possible tree topologies and evaluates each one. The number of possible trees for n number of

taxa can be calculated by the formula number trees = [2n - 5]!! (Felsenstein, 1978a) where !! denotes multiplication by every other number starting with (2n - 5) and continuing to 1. For example, to find the most parsimonious solution for a data set containing 25 taxa, the computer would need to build and evaluate 2.5×10^{28} different trees. The practicality of the exhaustive search is limited to analyses of 11 or less taxa. The branch-and-bound algorithm, which is a modification of the exhaustive search, aborts searches involving rearrangements that will only generate trees whose length exceeds a certain upper bound. Although this method reduces the number of trees that must be evaluated, its usefulness is limited to about 17 taxa (Penny et al., 1990; as cited in Hamby, 1990).

Heuristic searches were performed utilizing several different options available in PAUP. From the Data menu, different outgroups were selected using the "Define Outgroup" option. The "Delete-Restore Taxa" option was used to exclude taxa from various analyses or to include them again later. By defining character sets that contained first, second and third codon position bases in the input file, the "Include-Exclude Characters" option could be used to exclude certain codon positions from the analysis. Trees that were generated from various searches were saved as output files and as tree files that could be used as input for subsequent analyses.

Another phylogenetic analysis package, PHYLIP (Felsenstein, 1989) was used to generate trees based on pair-wise distances and on maximum parsimony. The PHYLIP (Version 3.4) package of programs was run on the VAXcluster via modem from an IBM PC XT. The DNADIST, DNAPARS and FITCH

programs were used to generate distance data and estimate phylogenies. The DNAPARS program was used to generate trees using the parsimony criterion which could be compared to the most parsimonious trees generated by PAUP.

3. Choosing the best reconstruction

The bootstrapping procedure (Felsenstein, 1985) available in PAUP was used to determine the statistical support for various groupings of taxa in the trees. Using this procedure, a new bootstrap data set is produced containing the same set of species, but with some of the original characters duplicated and others dropped through the process of sampling with replacement. Each bootstrapped data set is then analyzed just as the original data set was analyzed. A majority-rule consensus tree is constructed showing all of the monophyletic groupings that occurred in a majority of the bootstrap samples. A grouping which is present at least 95% of the time is considered to be statistically significant (Felsenstein, 1985).

To select among groups of most parsimonious trees having the same number of steps, but slightly different arrangements of taxa, two strategies were employed. In PAUP, the option exists to include user-defined stepmatrices in the input data set which, when activated, impose a set of weights upon various base changes at each of the codon positions. Because this weighting option involves more complicated calculations of tree lengths, it is not feasible to use it on a large data set for the construction of trees. It is useful as a method of analyzing an existing set of trees generated without a weighting scheme. Files containing a group of equally most-parsimonious trees generated

using the heuristic search were inputted as starting trees from which the PAUP program was used to analyze and perform branch-swapping to reconstruct the most parsimonious tree under the imposed weighting scheme. Another strategy for selecting among a group of trees was the maximum likelihood procedure available in PHYLIP. The data file (matrix and trees) in PAUP was exported as a data set readable by PHYLIP using the "Export File" option available in PAUP. The file was then translated as a text file from Macintosh format to MS-DOS format using the Macintosh Apple File Transfer program. The file was transferred to an IBM PC where it was used as the input file for the DNAML program of PHYLIP. This program estimates the likelihood of each of the user-defined trees given the data and a model. The program also performs a statistical test of each of the trees against the one with the highest likelihood, to determine whether the trees are significantly different. The large size of the data set and the computational complexity of the maximum likelihood method makes the construction of trees via this method unfeasible.

III. RESULTS AND DISCUSSION

A. Results

1. rbcL sequence data

DNA sequences from the rbcL gene of seven species (Table 3) were obtained through PCR amplification, cloning, and double-stranded dideoxy (Sanger) sequencing. The sequences, displayed using the PRETTY program (UWGCG) are presented in Figure 4. For three taxa, DNA sequence data was lacking at either the 3' or 5' end (represented by the dots in Figure 4). The amplified rbcL products from these taxa were cloned via blunt-end cloning and the ends of the amplified fragments were most likely lost by the 3'→5'exonuclease activity of the T4 Polymerase enzyme which was used to generate blunt-ended termini. The sequence data from these regions was not included in the analysis and was treated by the PAUP and PHYLIP programs as missing data. Each of the sequences were of the same length, but only the first 1428 bp were included in the analysis. Using the DISTANCE program (UWGCG) the pair-wise genetic distances (Table 5) were determined as a first assessment of the amount of similarity among the seven rbcL sequences.

2. Phylogenetic analyses using PAUP

A preliminary analysis using PAUP 3.0s (Swofford, 1991) included only the seven taxa sequenced plus a Begonia sequence specified as an outgroup so that the exhaustive search procedure could be used to evaluate all possible trees. One minimal-length tree of 167 character-state changes (steps) was found (Figure 5). An overall consistency

Figure 4. rbcL Sequences From Datiscaceae and Cucurbitaceae. Sequences were aligned using PHYLIP's GAP program and are displayed here in the 5'→3' direction using the PRETTY program. Dots indicate missing data. Nucleotides shown in lower case are those that differ from the consensus sequence.

```
           1                                                        50
Dglomerata ..........  ..........  ..........  ..........  ..........
Dcannabina ATGTCACCAC  AAACAGAGAC  TgAGCAAGTG  TTGGATTCAA  AGCTGGTGTT
Octomeles  ATGTCACCAC  AAACAGAGAC  TAAGCAAGTG  TTGGATTCAA  AGCTGGTGTT
Tetrameles ..........  ........AC  TAAGCAAGTG  TTGGATTCAA  AGCTGGTGTT
Cucurbita  ATGTCACCAC  AAACAGAGAC  TAAGCAAGTG  TTGGATTCAA  AGCTGGTGTT
Cucumis    ATGTCACCAC  AAACAGAGAC  TAAGCAAGTG  TTGGATTCAA  AGCTGGTGTT
Luffa      ATGTCACCAC  AAACAGAGAC  TAAGCAAGTG  TTGGATTCAA  AGCTGGTGTT
Consensus  ATGTCACCAC  AAACAGAGAC  TAAGCAAGTG  TTGGATTCAA  AGCTGGTGTT

           51                                                      100
Dglomerata ..........  ..........  ........CT  GAcTATGAAA  CCAAAGATAC
Dcannabina AAAGATTATA  AATTGACTTA  TTATACTCCT  GAcTATGAAA  CCAAAGATAC
Octomeles  AAAGATTATA  AATTGACTTA  TTATACTCCT  GAATATGAAA  CCAAAGATAC
Tetrameles AAAGATTATA  AATTGACTTA  TTATACTCCT  GAATATcAAA  CCAAAGATAC
Cucurbita  AAAGATTATA  AATTGACTTA  TTATACTCCT  GAATATGAAA  CCAAAGATAC
Cucumis    AAAGATTATA  AATTGACTTA  TTATACTCCT  GAATATGAAA  CCAAAGATAC
Luffa      AAAGATTATA  AATTGACTTA  TTATACTCCT  GAATATcAAA  CCAAAGATAC
Consensus  AAAGATTATA  AATTGACTTA  TTATACTCCT  GAATATGAAA  CCAAAGATAC

           101                                                     150
Dglomerata TGATATCTTG  GCAGCATTCC  GAGTAACTCC  TCAACCtGGA  GTTCCACCTG
Dcannabina gGATATCTTG  GCAGCATTCC  GAGTAACTCC  TCAACCtGGA  GTTCCACCTG
Octomeles  TGATATCTTG  GCAGCATTCC  GAGTAACTCC  TCAACCcGGA  GTTCCtCCTG
Tetrameles TGATATCTTG  GCAGCATTCC  GAGTAACTCC  TCAACCcGGA  GTTCCgCCTG
Cucurbita  TGATATCTTG  GCAGCATTCC  GAGTAACTCC  TCAACCGGGA  GTTCCACCcG
Cucumis    TGATATCTTG  GCAGCATTCC  GAGTAACTCC  TCAACCGGGA  GTTCCACCTG
Luffa      TGATATCTTG  GCAGCATTCC  GAGTAACTCC  TCAACCGGGA  GTTCCACCcG
Consensus  TGATATCTTG  GCAGCATTCC  GAGTAACTCC  TCAACCGGGA  GTTCCACCTG

           151                                                     200
Dglomerata AGGAAGCAGG  tGCCGCTGTA  GCTGCTGAAT  CTTCTACTGG  TACATGGACA
Dcannabina AGGAAGCAGG  tGCCGCTGTA  GCTGCTGAAT  CTTCTACTGG  TACATGGACA
Octomeles  AGGAAGCAGG  GGCCGCTGTA  GCTGCTGAAT  CTTCTACTGG  TACATGGACA
Tetrameles AGGAAcgAGG  GGCCGCTGTA  GCTGCTGAAT  CTTCTACTGG  TACATGGACA
Cucurbita  AGGAAGCgGG  GGCCGCcGTA  GCTGCTGAAT  CTTCTACTGG  TACATGGACA
Cucumis    AGGAAGCAGG  GGCCGCTGTA  GCTGCTGAAT  CTTCTACTGG  TACATGGACA
Luffa      AGGAAGCAGG  GGCCGCTGTA  GCTGCTGAAT  CTTCTACTGG  TACATGGACA
Consensus  AGGAAGCAGG  GGCCGCTGTA  GCTGCTGAAT  CTTCTACTGG  TACATGGACA

           201                                                     250
Dglomerata ACTGTGTGGA  CCGATGGGCT  TACCAGTCTT  GATCGTTACA  AAGGACGATG
Dcannabina ACTGTGTGGA  CCGATGGtCT  TACCAGTCTT  GATCGTTACA  AAGGACGATG
Octomeles  ACTGTGTGGA  CCGATGGGCT  TACCAGTCTT  GATCGTTACA  AAGGACGATG
Tetrameles ACTGTGTGGA  CCGATGGGCT  TACCAGTCTT  GATCGTTACA  AAGGACGATG
Cucurbita  ACTGTGTGGA  CCGATGGGCT  TACCAGTCTT  GATCGTTACA  AAGGACGATG
Cucumis    ACTGTGTGGA  CCGATGGGCT  TACCAGTCTT  GATCGTTACA  AAGGACGATG
Luffa      ACTGTGTGGA  CCGATGGGCT  TACCAGTCTT  GATCGTTACA  AAGGACGATG
Consensus  ACTGTGTGGA  CCGATGGGCT  TACCAGTCTT  GATCGTTACA  AAGGACGATG
```

```
          251                                                        300
Dglomerata CTActcCATC GAGCCTGTTG CTGGAGAAGA AAATCAATAT ATTGCTTATG
Dcannabina CTAcGGCATC GAGCCTGTTG CTGGAGgAGA AAATCAATAT ATTGCTTATG
Octomeles  CTAccaCATC GAGCCTGTTG CTGGAGAAGA AAgTCAATtT ATTGCTTATG
Tetrameles CTATcaCATC GAGCCTGTTG CTGGAGAAGA AAgTCAATtT ATTGCTTATG
Cucurbita  CTATGGaATC GAGCCTGTTc CTGGAGAAGA AAATCAATAT ATTGCTTATG
Cucumis    CTATGGaATC GAGCCTGTTG CTGGAGAAGA AAATCAATAT ATTGCTTATG
Luffa      CTATGGCATC GAGCCTGTTG CTGGAGAAGA AAATCAATAT ATTGCTTATG
Consensus  CTATGGCATC GAGCCTGTTG CTGGAGAAGA AAATCAATAT ATTGCTTATG

          301                                                        350
Dglomerata TAGCTTATCC CtTAGACCTT TTTGAAGAAG GTTCTGTTAC TAACATGTTT
Dcannabina TAGCTTATCC CtTAGACCTT TTTGAAGAAG GTTCTGTTAC TAACATGTTT
Octomeles  TAGCTTATCC CCTAGACCTT TTTGAAGAAG GTTCTGTTAC TAACATGTTT
Tetrameles TAGCTTATCC CtTAGACCTT TTTGAAGAAG GTTCTGTTAC TAACATGTTT
Cucurbita  TAGCTTATCC CCTAGACCTT TTTGAAGAAG GTTCTGTTAC TAACATGTTT
Cucumis    TAGCTTATCC tCTAGACCTT TTTGAAGAAG GTTCTGTaAC TAACATGgTT
Luffa      TAGCTTATCC CCTAGACCTT TTTGAAGAAG GTTCTGTTAC TAAtATGTTT
Consensus  TAGCTTATCC CCTAGACCTT TTTGAAGAAG GTTCTGTTAC TAACATGTTT

          351                                                        400
Dglomerata ACTTCCATTG TGGGTAATGT ATTTGGGTTC AAGGCCCTGC GCGCTCTACG
Dcannabina ACTTCCATTG TGGGTAATGT ATTTGGGTTC AAGGCCCTGC GCGCTCTACG
Octomeles  ACTTCCATTG TGGGTAATGT ATTTGGGTTC AAGGCCCTGC GCGCTCTACG
Tetrameles ACTTCCATTG TGGGTAATGT ATTTGGGTTC AAGGCCCTGC GCGCTCTACG
Cucurbita  ACTTCCATTG TGGGTAATGT ATTTGGaTTC AAGGCtCTGC GtGCTCTACG
Cucumis    ACTTCCATTG TGGGTAATGT ATTTGGaTTC AAGGCtCTaC GtGCTCTACG
Luffa      ACTTCCATTG TGGGTAATGT ATTTGGaTTC AAGGCtCTGC GCGCTCTACG
Consensus  ACTTCCATTG TGGGTAATGT ATTTGGGTTC AAGGCCCTGC GCGCTCTACG

          401                                                        450
Dglomerata TCTGGAaGAT TTGCGAATCC CTcCTGCTTA TgTTAAAACT TTCCAAGGgC
Dcannabina TCTGGAaGAT TTGCGAATCC CTcCTGCTTA TgTTAAAACT TTCCAAGGgC
Octomeles  TCTGGAGGAT TTGCGAATCC CTACTGCTTA TATTAAAACT TTCCAAGGCC
Tetrameles TCTGGAGGAT TTGaGAATCC CTACTGCTTA TATTAAAACT TTCCAAGGCC
Cucurbita  TCTGGAGGAT TTGCGAATCC CTACTGCTTA TATTAAAACT TTCCAAGGCC
Cucumis    TCTGGAGGAT TTGCGAATCC CTACTGCTTA TATTAAAACT TTCCAAGGCC
Luffa      TCTGGAGGAT TTGCGAATCC CTACTGCTTA TATTAAAACT TTCCAAGGCC
Consensus  TCTGGAGGAT TTGCGAATCC CTACTGCTTA TATTAAAACT TTCCAAGGCC

          451                                                        500
Dglomerata CGCCTCATGG TATCCAAGTT GAGAGAGATA AATTGAACAA GTATGGCCGC
Dcannabina CGCCTCATGG TATCCAAGTT GAGAGAGATA AATTGAACAA GTATGGCCGC
Octomeles  CGCCTCATGG TATCCAAGTT GAGAGAGATA AATTGAACAA GTATGGCCGC
Tetrameles CGCCTCATGG TATCCAAGTT GAGAGAGATA AATTGAACAA GTATGGCCGC
Cucurbita  CGCCTCATGG TATCCAgGTT GAaAGAGATA AATTGAACAA GTATGGtCGC
Cucumis    CGCCTCATGG TATCCAgGTT GAaAGAGATA AATTGAACAA GTATGGtCGC
Luffa      CaCCTCAcGG TATCCAgGTT GAaAGAGATA AATTGAACAA GTATGGtCGC
Consensus  CGCCTCATGG TATCCAAGTT GAGAGAGATA AATTGAACAA GTATGGCCGC
```

```
            501                                                    550
Dglomerata  CCTCTATTGG GATGTACTAT TAAACCTAAA TTGGGATTAT CCGCTAAGAA
Dcannabina  CCTCTATTGG GATGTACTAT TAAACCTAAA TTGGGATTAT CCGCTAAGAA
Octomeles   CCTCTATTGG GATGTACTAT TAAACCTAAA TTGGGATTAT CCGCTAAGAA
Tetrameles  CCTCTATTGG GATGTACTAT TAAACCTAAA TTGGGATTAT CCGCTAAGAA
Cucurbita   CCTCTATTGG GATGTACTAT TAAACCaAAA TTGGGATTAT CCGCTAAGAA
Cucumis     CCTCTATTGG GATGTACTAT TAAACCaAAA TTGGGATTAT CCGCTAAGAA
Luffa       CCTCTATTGG GATGTACTAT TAAACCaAAA TTGGGATTAT CCGCTAAGAA
Consensus   CCTCTATTGG GATGTACTAT TAAACCTAAA TTGGGATTAT CCGCTAAGAA

            551                                                    600
Dglomerata  TTACGGTAGA GCTGTTTATG AATGTCTACG tGGTGGACTT GATTTTACCA
Dcannabina  TTACGGTAGA GCTGTTTATG AATGTCTACG tGGTGGACTT GATTTTACCA
Octomeles   TTACGGTAGA GCTGTTTATG AATGTCTtCG CGGTGGACTT GATTTTACCA
Tetrameles  TTACGGTAGA GCTGTTTATG AATGTCTtCG CGGTGGACTT GATTTTACCA
Cucurbita   TTAtGGTAGA GCaGTTTATG AATGTCTACG CGGTGGACTT GATTTTACCA
Cucumis     TTAtGGTAGA GCaGTTTATG AATGTCTACG CGGTGGACTT GATTTTACCA
Luffa       TTAtGGTAGA GCaGTTTATG AATGTCTACG CGGTGGACTT GATTTTACCA
Consensus   TTACGGTAGA GCTGTTTATG AATGTCTACG CGGTGGACTT GATTTTACCA

            601                                                    650
Dglomerata  AAGATGATGA AAACGTGAAT TCCCAACCAT TTATGCGTTG GAGAGACCGT
Dcannabina  AAGATGATGA AAACGTGAAT TCCCAACCAT TTATGCGTTG GAGAGACCGT
Octomeles   AAGATGATGA AAACGTtAAT TCCCAACCAT TTATGCGTTG GAGAGACCGT
Tetrameles  AAGATGATGA AAACGTtAAT TCCCAACCAT TTATGCGTTG GAGAGACCGT
Cucurbita   AAGATGATGA AAACGTGAAT TCCCAACCAT TTATGCGTTG GAGAGACCGT
Cucumis     AAGATGATGA AAACGTGAAT TCCCAACCAT TTATGCGTTG GAGAGACCGT
Luffa       AAGATGATGA AAACGTGAAT TCCCAACCAT TTATGCGTTG GAGAGACCGT
Consensus   AAGATGATGA AAACGTGAAT TCCCAACCAT TTATGCGTTG GAGAGACCGT

            651                                                    700
Dglomerata  TTCCTATTTT GTGCGGAAGC AATTTATAAA gCACAGGCTG AAACAGGTGA
Dcannabina  TTCCTATTTT GTGCGGAAcg AATTTATAAA TCACAGGCTG AAACAGGTGA
Octomeles   TTCCTATTTT GTGCGGAAGC AATTTATAAA gCACAGGCTG AAACAGGTGA
Tetrameles  TTCCTATTTT GTGCGGAAGC AATTTATAAA gCACAGGCTG AAACAGGTGA
Cucurbita   TTCCTATTTT GTGCGGAAGC cATTTATAAA TCACAGGCTG AAACAGGTGA
Cucumis     TTCCTATTTT GTGCGGAAGC tATTTtTAAA TCACAGGCTG AAACAGGTGA
Luffa       TTCCTATTTT GTGCGGAAGC cATTTATAAA TCACAGGCTG AAACAGGTGA
Consensus   TTCCTATTTT GTGCGGAAGC AATTTATAAA TCACAGGCTG AAACAGGTGA

            701                                                    750
Dglomerata  AATCAAGGGG CATTACTTGA ATGCTACTGC AGGTACATGC GAAGAAATGA
Dcannabina  AATCAAGGGG CATTACTTGA ATGCTACTGC AGGTACATGC GAAGAAATGA
Octomeles   AATCAAGGGG CATTACTTGA ATGCTACTGC AGGTACATGC GAAGAAATGA
Tetrameles  AATCAAGGGG CATTACTTGA ATGCTACTGC gGGTACATGC GAAGAAATGA
Cucurbita   AATCAAGGGa CATTACTTGA ATGCTACTGC gGGTACATGC GAAGAAATGA
Cucumis     AATCAAaGGa CATTACTTGA ATGCTACTGC gGGTACATGC GAAGAAATGA
Luffa       AATCAAGGGa CATTACTTGA ATGCTACTGC AGGTACATGC GAAGAAATGA
Consensus   AATCAAGGGG CATTACTTGA ATGCTACTGC AGGTACATGC GAAGAAATGA
```

```
           751                                                       800
Dglomerata TcAAAAGGGC TGTATTTGCC AGAGAATTGG GAGTTCCTAT CGTAATGCAT
Dcannabina TcAAAAGGGC TGTATTTGCC AGAGAATTGG GAGTTCCTAT CGTAATGCAT
Octomeles  TcAAAAGGGC TGTATTTGCC AGAGAATTGG GAGTTCCTAT CaTAATGCAT
Tetrameles TGAAAAGGGC TGTATTTGCC AGAGAATTGG GAGTTCCTAT CaTAATGCAT
Cucurbita  TGAAAAGGGC TaTATTTGCt cGAGAATTGG GAGcTCCTAT CGTAATGCAT
Cucumis    TGAAAAGGGC TaTATTTGCC cGAGAATTGG GAGcTCCTAT CGTAATGCAT
Luffa      TGAAAAGGGC TaTATTTGCC AGAGAATTGG GAGTTCCTAT CGTAATGCAT
Consensus  TGAAAAGGGC TGTATTTGCC AGAGAATTGG GAGTTCCTAT CGTAATGCAT

           801                                                       850
Dglomerata GACTACTTAA CAGGGGGcTT CACTGCAAAT ACTAGCTTGG CTCATTATTG
Dcannabina GACTACTTAA CAGGGGGcTT CACTGCAAAT ACTAGCTTGG CTCATTATTG
Octomeles  GACTACTTAA CAGGGGGATT CACTGCAAAT ACTAGCTTGG CTCATTATTG
Tetrameles GACTACTTAA CAGGGGGATT CACTGCAAAT ACTAGCTTGG CTCATTATTG
Cucurbita  GACTACTTAA CAGGtGGATT CACTGCAAAT ACTAGCTTGG CTCATTATTG
Cucumis    GACTACTTAA CAGGtGGATT CACTGCAAAT ACTAGCTTGG CTCATTATTG
Luffa      GACTACTTAA CAGGtGGATT CACTGCAAAT ACTAGCTTGG CTCATTATTG
Consensus  GACTACTTAA CAGGGGGATT CACTGCAAAT ACTAGCTTGG CTCATTATTG

           851                                                       900
Dglomerata CCGAGATAAT GGTCTACTTC TTCACATTCA CCGTGCAATG CATGCAGTTA
Dcannabina CCGAGATAAT GGTCTACTTC TTCACATTCA CCGTGCAATG CATGCAGTTA
Octomeles  CCGAGATAAT GGTCTACTTC TTCACATTCA CCGTGCAATG CATGCAGTTA
Tetrameles CCGAGATAAT GGTCTACTTC TTCACATTCA CCGTGCAATG CATGCAGTTA
Cucurbita  CCGAGATAAT GGTCTACTTC TTCACATTCA CCGTGCAATG CATGCcGTTA
Cucumis    CCGAGATAAT GGTCTACTTC TTCACATTCA CCGTGCAATG CATGCtGTTA
Luffa      CCGAGATAAT GGTCTACTTC TTCACATTCA CCGTGCAATG CATGCcGTTA
Consensus  CCGAGATAAT GGTCTACTTC TTCACATTCA CCGTGCAATG CATGCAGTTA

           901                                                       950
Dglomerata TTGATAGACA GAAGAATCAc GGTATaCACT TTCGTGTACT AGCTAAAGCG
Dcannabina TTGATAGACA GAAGAATCAc GGTATaCACT TTCGTGTACT AGCTAAAGCG
Octomeles  TTGATAGACA GAAGAATCAT GGTATGCACT TTCGTGTACT AGCTAAAGCG
Tetrameles TTGATAGACA GAAGAATCAT GGTATGCACT TTCGTGTACT AGCTAAAGCG
Cucurbita  TTGATAGACA GAAGAATCAT GGTATGCACT TcCGTGTACT AGCTAAAGCG
Cucumis    TTGATAGACA GAAGAATCAT GGTATGCACT TcCGTGTACT AGCTAAAGCG
Luffa      TTGATAGgCA GAAGAATCAT GGTATGCACT TcCGTGTACT AGCTAAAGCG
Consensus  TTGATAGACA GAAGAATCAT GGTATGCACT TTCGTGTACT AGCTAAAGCG

           951                                                      1000
Dglomerata TTACGTATGT CTGGTGGAGA TCATATTCAC tCTGGTACCG TAGTAGGTAA
Dcannabina TTACGTATGT CTGGTGGAGA TCATATTCAC tCTGGTACCG TAGTAGGTAA
Octomeles  TTACGTATGT CTGGTGGAGA TCATATTCAC GCTGGTACCG TAGTAGGTAA
Tetrameles TTACGTcTGT CTGGTGGAGA TCATATTCAC GCTGGTACCG TAGTAGGTAA
Cucurbita  TTACGTcTGT CTGGTGGAGA cCATATTCAC GCTGGTACCG TAGTAGGTAA
Cucumis    TTACGTATGT CTGGTGGAGA cCATATTCAC GCTGGTACCG TAGTAGGTAA
Luffa      TTACGTATGT CTGGTGGAGA cCATATTCAC GCTGGTACCG TAGTAGGTAA
Consensus  TTACGTATGT CTGGTGGAGA TCATATTCAC GCTGGTACCG TAGTAGGTAA
```

```
          1001                                                1050
Dglomerata ACTTGAAGGG GAAAGAGAaA TCACTTTAGG CTTTGTTGAT TTACTACGTG
Dcannabina ACTTGAAGGG GAAAGAGAaA TCACTTTAGG CTTTGTTGAT TTACTACGTG
Octomeles  ACTTGAAGGG GAAAGAGAcA TCACTTTAGG CTTTGTTGAT TTACTACGcG
Tetrameles ACTTGAAGGG GAAAGAGAgA TCACTTTAGG tTTTGTTGAT TTACTACGcG
Cucurbita  ACTTGAAGGG GAAAGAGAcA TCACTTTAGG CTTTGTTGAT TTACTACGTG
Cucumis    ACTTGAAGGG GAAAGAGAaA TCACTTTAGG CTTTGTTGAT TTACTACGTG
Luffa      ACTTGAAGGG GAAAGAGAcA TCACTTTAGG CTTTGTTGAT TTACTACGTG
Consensus  ACTTGAAGGG GAAAGAGA-A TCACTTTAGG CTTTGTTGAT TTACTACGTG

          1051                                                1100
Dglomerata ATGATTTTAT TGAAAAAGAT CGAAGCCGCG GTATTTATTT CACTCAAGAT
Dcannabina ATGATTTTAT TGAAAAAGAT CGAAGCCGCG GTATTTATTT CACTCAAGAT
Octomeles  ATGATTaTAT TGAAAAAGAc CGAAGCCGCG GTATTTtTTT CACTCAAGAT
Tetrameles ATGATTaTgT TGAAAAAGAT CGAAGCCGCG GTATTTATTT CACTCAAGAT
Cucurbita  ATGATTTTgT TGAAAAAGAT CGAAGCCGCG GTATTTATTT CACTCAAGAT
Cucumis    ATGATTTTgT TGAAAAAGAc CGAAGCCGCG GTATTTATTT CACTCAAGAT
Luffa      ATGATTTTAT TGAAAAAGAT CGAAGCCGCG GTATTTATTT CACTCAAGAT
Consensus  ATGATTTTAT TGAAAAAGAT CGAAGCCGCG GTATTTATTT CACTCAAGAT

          1101                                                1150
Dglomerata TGGGTCTCTC TACCAGGTGT TCTGCCcGTG GCTTCAGGGG GTATTCACGT
Dcannabina TGGGTCTCTC TACCAGGTGT TCTGCCtGTG GCTTCAGGGG GTATTCACGT
Octomeles  TGGGTCTCTC TACCAGGTGT TtTGCCgGTG GCTTCAGGGG GTATTCACGT
Tetrameles TGGGTCTCTC TACCAGGTGT TtTGCCcGTG GCTTCAGGGG GTATTCACGT
Cucurbita  TGGGTCTCTt TACCAGGTGT TCTGCCAGTG GCTTCcGGtG GTATTCACGT
Cucumis    TGGGTCTCTt TACCAGGTGT TCTGCCAGTG GCTTCcGGtG GTATTCACGT
Luffa      TGGGTCTCTt TACCAGGTGT TCTGCCAGTG GCTTCcGGtG GTATTCACGT
Consensus  TGGGTCTCTC TACCAGGTGT TCTGCCAGTG GCTTCAGGGG GTATTCACGT

          1151                                                1200
Dglomerata TTGGCATATG CCTGCTCTGA CCGAGATCTT TGGAGATGAT TCCGTACTAC
Dcannabina TTGGCATATG CCTGCTCTGA CCGAGATCTT TGGAGATGAT TCCGTACTAC
Octomeles  TTGGCATATG CCTGCTCTGA CCGAGATCTT TGGAGATGAT TCCGTACTAC
Tetrameles TTGGCATATG CCcGCTCTGA CCGAGATCTT TGGAGATGAT TCCGTACTAC
Cucurbita  TTGGCATATG CCTGCTCTaA CCGAGATtTT TGGAGATGAT TCtGTACTAC
Cucumis    TTGGCATATG CCTGCTCTaA CCGAGATtTT TGGAGATGAT TCtGTACTAC
Luffa      TTGGCATATG CCTGCTCTaA CCGAGATtTT TGGAGATGAT TCtGTACTAC
Consensus  TTGGCATATG CCTGCTCTGA CCGAGATCTT TGGAGATGAT TCCGTACTAC

          1201                                                1250
Dglomerata AATTCGGTGG AGGAACTTTA GGGCACCCTT GGGGTAATGC ACCcGGTGCC
Dcannabina AATTCGGTGG AGGAACTTTA GGGCACCCTT GGGGTAATGC ACCcGGTGCC
Octomeles  AATTCGGTGG AGGAACTTTA GGaCACCCTT GGGGaAATGC ACCTGGTGCC
Tetrameles AATTCGGTGG AGGAACTTTA GGaCACCCTT GGGGaAATGC ACCTGGTGCC
Cucurbita  AATTCGGcGG AGGAACTTTg GGGCACCCTT GGGGTAATGC ACCaGGTGCC
Cucumis    AATTCGGcGG AGGAACTTTg GGGCACCCTT GGGGTAATGC ACCTGGTGCC
Luffa      AATTCGGcGG AGGAACTTTg GGGCACCCTT GGGGTAATGC ACCTGGTGCC
Consensus  AATTCGGTGG AGGAACTTTA GGGCACCCTT GGGGTAATGC ACCTGGTGCC
```

```
            1251                                                    1300
Dglomerata  GTAGCTAACC GAGTAGCTCT AGAAGCATGT GTACAAGCTC GTAATGAGGG
Dcannabina  GTAGCTAACC GAGTAGCTCT AGAAGCATGT GTACAAGCTC GTAATGAGGG
Octomeles   GTAGCTAACC GAGTAGCTCT AGAAGCATGT GTACAAGCTC GTAATGAGGG
Tetrameles  GTAGCTAACC GAGTAGCTCT AGAAGCATGT GTACAAGCTC GTAATGAGGG
Cucurbita   GTAGCTAACC GAGTAGCTCT AGAAGCATGT GTACAAGCTC GTAATGAGGG
Cucumis     GTAGCTAACC GAGTAGCTCT AGAAGCATGT GTACAAGCTC GTAATGAGGG
Luffa       GTAGCTAACC GAGTAGCTCT AGAAGCATGT GTACAAGCTC GTAATGAGGG
Consensus   GTAGCTAACC GAGTAGCTCT AGAAGCATGT GTACAAGCTC GTAATGAGGG

            1301                                                    1350
Dglomerata  ACGTGATCTT GCTCGTGAGG GTAATGAAAT TATCCGTGAG GCTAcTAAAT
Dcannabina  ACGTGATCTT GCTCGTGAGG GTAATGAAAT TAaCCGTGAG GCTAcTAAAT
Octomeles   ACGTGATCTT GCTCGTGAGG GTAATGAAAT TATCCGTGAG GCTAGTAAAT
Tetrameles  ACGTGATCTT GCTCGTGAGG GTAATGAAAT TATCCGTGAG GCTAGTAAAT
Cucurbita   ACGTGATCTT GCTCGTGAGG GTAATGAAAT TATCCGTGAG GCTAGTAAAT
Cucumis     ACGTGATCTT GCTCGTGAGG GTAATGAAAT TATCCGTGAG GCTAGTAAAT
Luffa       ACGTGATCTT GCTCGTGAGG GTAATGAAAT TATCCGTGAG GCTAGTAAAT
Consensus   ACGTGATCTT GCTCGTGAGG GTAATGAAAT TATCCGTGAG GCTAGTAAAT

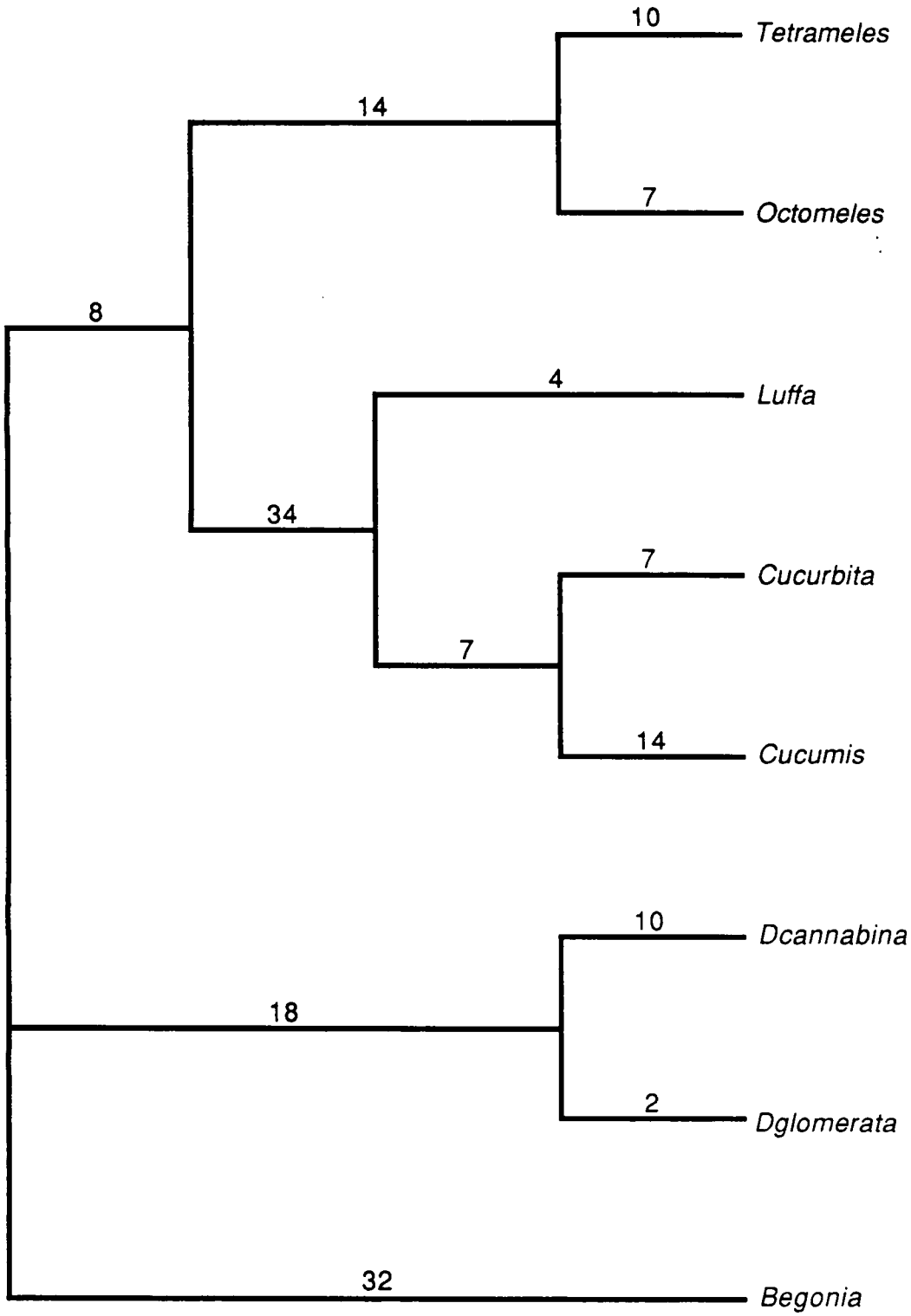
            1351                                                    1400
Dglomerata  GGAGTCCaGA ACTAGCTGCT GCTTGTGAAG TATGGAAGGA GATCAAATTT
Dcannabina  GGAGTCCaGA ACTAGCTGCT GCTTGTGAAG TATGGAAGGA GATCAAATTT
Octomeles   GG..... ..... ..... ..... .....
Tetrameles  GGAGTCCgGA ACTAGCTGCT GCTTGTGAAG TATGGAAGGA aATCAAATTT
Cucurbita   GGAGTCCTGA ACTAGCTGCT GCTTGTGAAG TATGGAAGGc GATCAAATTc
Cucumis     GGAGTCCTGA ACTAGCTGCT GCTTGTGAAG TATGGAAGGA GATCAAATTT
Luffa       GGAGTCCTGA ACTAGCTGCT GCTTGTGAAG TATGGAAGGc GATCAAATTT
Consensus   GGAGTCCTGA ACTAGCTGCT GCTTGTGAAG TATGGAAGGA GATCAAATTT

            1401                                  1435
Dglomerata  GAATTTccaG CAATGGATAC TTTGTAATCC AGTAA
Dcannabina  GAATTTGAaG CAATGGATAC TTTGTAATCC AGTAA
Octomeles   ..... ..... ..... .....
Tetrameles  GAATTTccaG CAATGGATAC TTTGTAATCC AGTAA
Cucurbita   GAATTTGAgG CAgTaGATAC TTTGTAATCC AGTAA
Cucumis     GAATTTGAgG CAATGGATAC TTTGTAAgCC AGTAA
Luffa       GAATTTGAgG CAgTaGATAC TTTGTAATCC AGTAA
Consensus   GAATTTGA-G CAATGGATAC TTTGTAATCC AGTAA
```

Table 5. UWGCG Pair-wise Distances Among Datiscaceae and Cucurbitaceae. Matrix calculated using the DISTANCES Program (Devereux et al., 1984).

	1	2	3	4	5	6	7
1. Dglomerata	1.0000						
2. Dcannabina	0.9912	1.0000					
3. Octomeles	0.9186	0.9704	1.0000				
4. Tetrameles	0.9698	0.9633	0.9741	1.0000			
5. Cucurbita	0.9521	0.9533	0.9586	0.9534	1.0000		
6. Cucumis	0.9543	0.9554	0.9593	0.9541	0.9847	1.0000	
7. Luffa	0.9573	0.9582	0.9638	0.9562	0.9875	0.9819	1.0000

Figure 5. Preliminary Phylogenetic Analysis Using Eight Taxa. Most parsimonious unrooted tree generated with PAUP using the exhaustive search and specifying Begonia as the outgroup. Tree length=167; CI=0.832; RC=0.638.


10
Tetrameles
14
7
Octomeles
8
4
Luffa
34
7
Cucurbita
7
14
Cucumis
10
Dcannabina
18
2
Dglomerata
32
Begonia

index (CI; Kluge and Farris, 1969) value of 0.832 was calculated by PAUP as a measure of how well the data "fits" the particular topology. This value is defined as the minimal number of character-state changes required by the data set divided by the total number of character-state changes required to most parsimoniously fit all of the characters onto a tree under construction. Reconstructions requiring fewer homoplasious changes in the data matrix would have CI values that are closer to 1 whereas topologies requiring more homoplasy would have CI values closer to 0. Although the ensemble CI is a common measure of the fit between the data and a tree, some problems with its performance have been detected. The number of autapomorphies present in the data and the number of taxa in the data set have both been shown to influence the CI value (Brooks et al., 1986; Archie, 1989; Sanderson and Donoghue, 1989). To circumvent the effect of autapomorphies in the data, Farris (1989) proposed the rescaled consistency index (RC) which excludes autapomorphous characters that do not contribute to the fit of the tree. The RC value for the preliminary search was 0.638.

The topology generated by the preliminary search revealed that the four species belonging to the Datiscaceae did not form a monophyletic group. Cucurbitaceae appeared as a sister group to Tetrameles/Octomeles with Datisca grouping to the outside. Additional taxa were added to the data matrix in a stepwise manner to determine whether this topology would persist. Additional taxa were chosen from the Chase et al. (1993) data set based on their position in the global phylogeny relative to the Datiscaceae. Outgroup taxa were selected from the closest sister group to the ingroup (Datiscaceae) in the global

phylogeny. Several searches were conducted that included sets of data with between eight to 43 taxa.

In the subsequent analysis, 13 taxa were analyzed using the branch and bound search. This time, Viola was specified as the outgroup and three minimal-length trees of 415 steps each were found. A strict consensus tree (Figure 6) shows a different topology than that obtained in the preliminary search. In this tree, Datiscaceae still do not form a monophyletic group, but Datisca now groups with Cucurbitaceae and Begonia with Tetrameles/Octomeles as the sister group to the outside. To determine whether Datiscaceae might form a monophyletic group in less optimal topologies, trees up to 15 steps longer than the most parsimonoius trees of 415 steps were saved during the branch and bound search. Of the 396 trees saved, none showed Datiscaceae as a monophyletic group; however, Begonia was observed to group with Datisca in trees only one step longer than the minimal-length trees. The difference in topologies identified in the exhaustive search and in the branch and bound search is indicative of the fact that Begonia could be grouped with either the Cucurbitaceae or Datisca with a difference of only one character-state change. In the exhaustive search using eight taxa, the character-state changes were optimized in such a way as to place Begonia closer to Datisca, thus pulling Datisca away from the Cucurbitaceae when Begonia was specified as the outgroup.

The next analysis was performed using the heuristic search with 33 taxa and Pisum as the outgroup. This analysis generated eight minimal-length trees of 1338 steps each. A strict consensus (Figure 7) shows a branching pattern (regarding Datiscaceae) similar to the topology

Figure 6. Strict Consensus of Trees Generated by Phylogenetic Analysis of 13 Taxa. Length of each of the three equally parsimonious trees=415; CI=0.712; RC=0.429. Trees were generated with PAUP using the branch and bound search. The unrooted trees were oriented with Viola as the outgroup. Length of the consensus tree = 425; CI=0.689; RC=0.394.

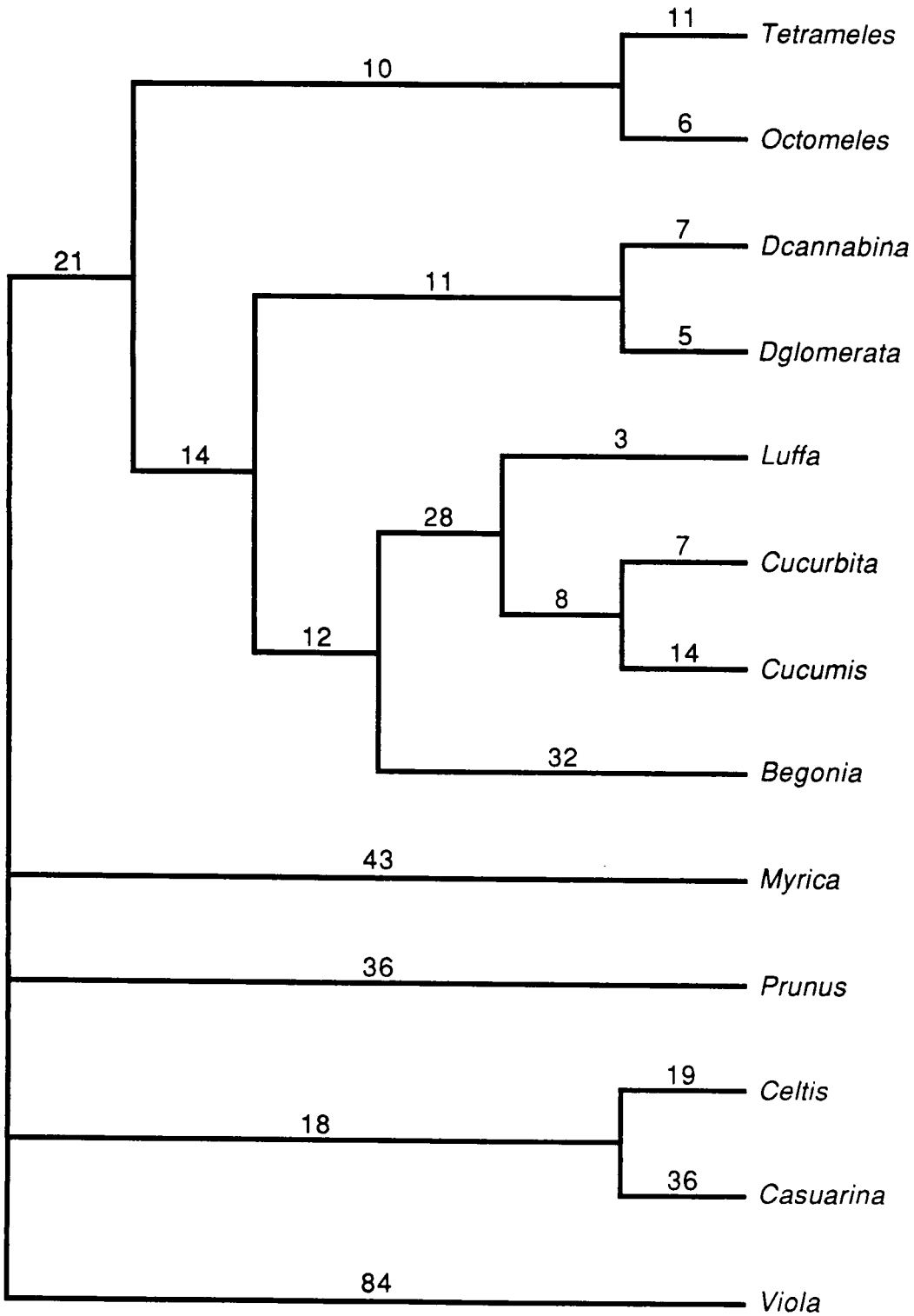
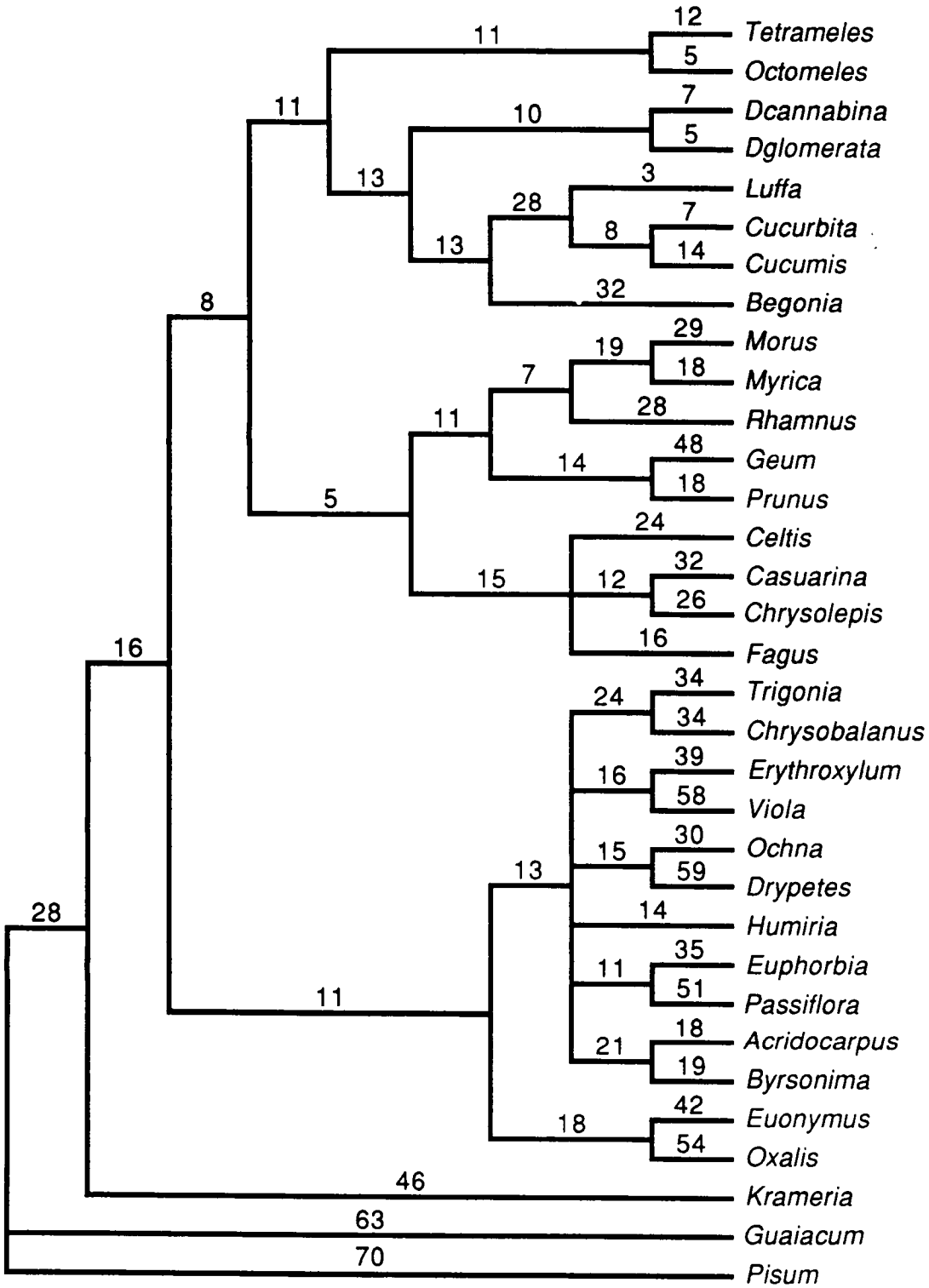

11
Tetrameles
10
6
Octomeles
7
Dcannabina
21
11
5
Dglomerata
3
14
Luffa
28
7
Cucurbita
8
12
14
Cucumis
32
Begonia
43
Myrica
36
Prunus
19
Celtis
18
36
Casuarina
84
Viola

Figure 7. Strict Consensus of Eight Trees Generated by Phylogenetic Analysis of 33 Taxa. Trees reconstructed by PAUP using simple step-wise addition of sequences and TBR branch swapping. The unrooted trees were oriented with Pisum specified as the outgroup. The eight equally parsimonious trees had lengths of 1338 steps; CI=0.469; RC=0.213. The consensus tree = 1348; CI=0.465; RC=0.208.


Tetrameles
Octomeles
Dcannabina
Dglomerata
Luffa
Cucurbita
Cucumis
Begonia
Morus
Myrica
Rhamnus
Geum
Prunus
Celtis
Casuarina
Chrysolepis
Fagus
Trigonia
Chrysobalanus
Erythroxylum
Viola
Ochna
Drypetes
Humiria
Euphorbia
Passiflora
Acridocarpus
Byrsonima
Euonymus
Oxalis
Krameria
Guaiacum
Pisum

found in the branch and bound search. An analysis of the same taxa, but designating Oxalis as the outgroup resulted in no change in the topology regarding the Datiscaceae and nearby groups (data not shown).

The final data matrix analyzed included 43 taxa with Carica, Capparis, and Gossypium specified as the outgroup. Approximately 6.5 hours were required to complete ten random-addition-sequence replications with TBR branch swapping. Nine minimal-length trees of 1801 steps each were retained. One of the most parsimonious trees is shown in Figure 8 while a 50% majority-rule consensus tree of all nine trees is shown in Figure 9. In all trees, the Datisca species remain the sister group to the clade containing the Cucurbitaceae and Begonia. Octomeles and Tetrameles together form the sister group to Datisca, Cucurbitaceae and Begonia.

Another heuristic search was conducted to determine if near-minimal length trees would show any differences in the grouping of the Datiscaceae. In addition to the nine trees of 1801 steps, 198 trees of 1802 steps were found. The majority-rule consensus tree (Figure 10) shows no change in the topology near the Datiscaceae compared to the 1801-length consensus. The search also retained trees 1803 steps long, but had to be aborted after approximately 24 hours because the program had found 1540 trees, but still had to swap on over 900 trees. Because of the risk of data loss caused by the exceeding the memory capacity of the computer, the search was terminated. A consensus for the 1803-length trees was calculated (Figure 11) but its optimality cannot be guaranteed since not all trees were found. As with the 1802-length trees, the same topology regarding the Datiscaceae was found. The

Figure 8. One of Nine Equally Parsimonious Trees Generated by Phylogenetic Analysis of 43 Taxa. Trees reconstructed with PAUP by heuristic search using ten replications of random-sequence-addition and TBR branch swapping. Tree length=1801; CI=0.407; RC=0.176. Numbers above the branches are estimated branch lengths optimized by accelerated transformation (ACCTRAN). The unrooted tree is oriented with Gossypium, Carica and Capparis as the outgroup. This tree topology was also selected by maximum likelihood analysis (PHLYIP DNAML program) as that with the lowest log likelihood value (-12283.39703) among the nine trees.

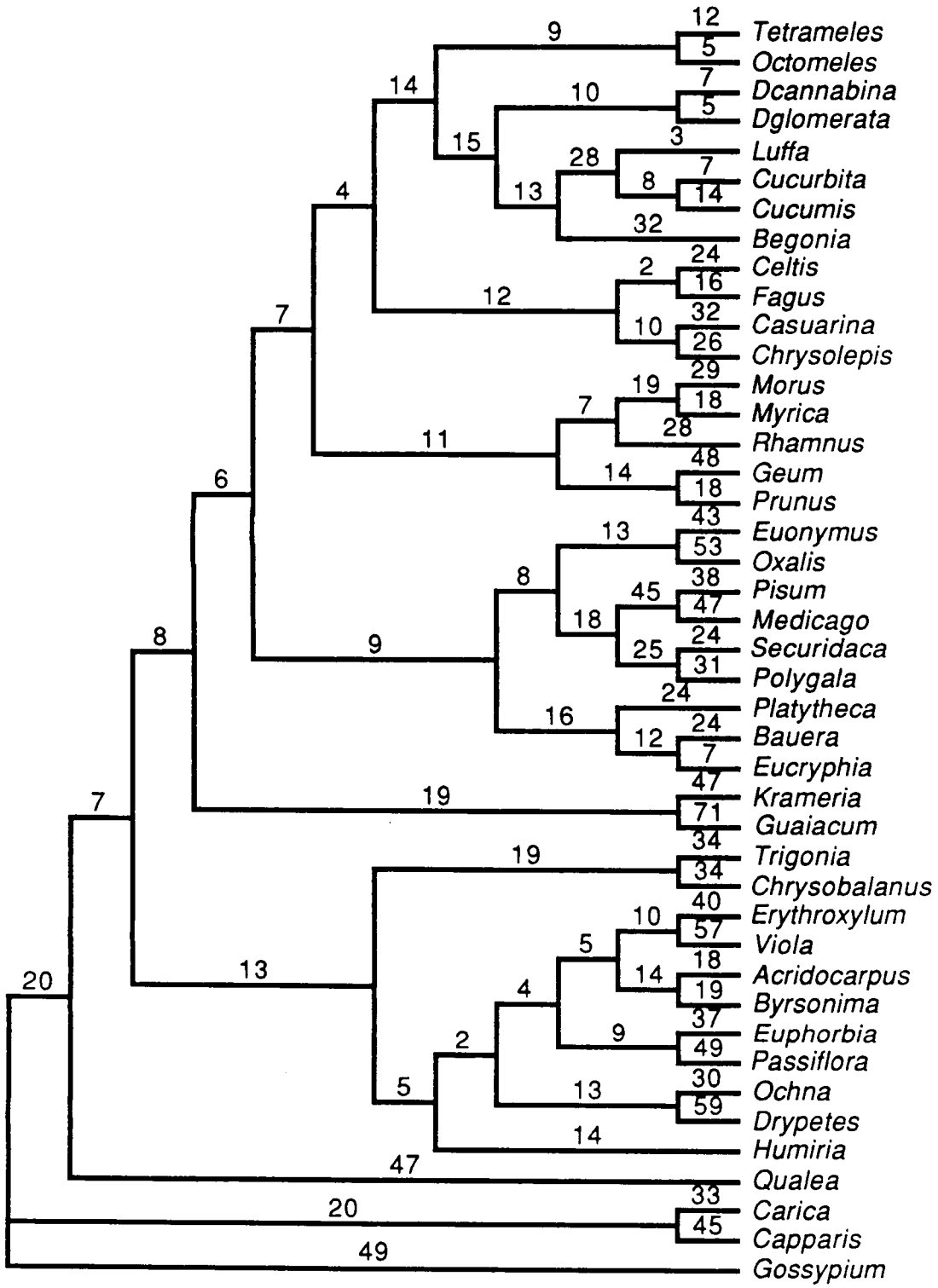

Tetrameles
Octomeles
Dcannabina
Dglomerata
Luffa
Cucurbita
Cucumis
Begonia
Celtis
Fagus
Casuarina
Chrysolepis
Morus
Myrica
Rhamnus
Geum
Prunus
Euonymus
Oxalis
Pisum
Medicago
Securidaca
Polygala
Platytheca
Bauera
Eucryphia
Krameria
Guaiacum
Trigonia
Chrysobalanus
Erythroxylum
Viola
Acridocarpus
Byrsonima
Euphorbia
Passiflora
Ochna
Drypetes
Humiria
Qualea
Carica
Capparis
Gossypium

Figure 9. 50% Majority-Rule Consensus of Nine Equally Parsimonious Trees. Original trees generated by phylogenetic analysis of 43 taxa using the heuristic search in PAUP. Consensus tree calculated using the "Consensus" option under the TREES menu. Italicized numbers below branches reflect the percentage of original trees that support the group descending from that branch. Consensus tree length = 1802; CI=0.407; RC= 0.716.

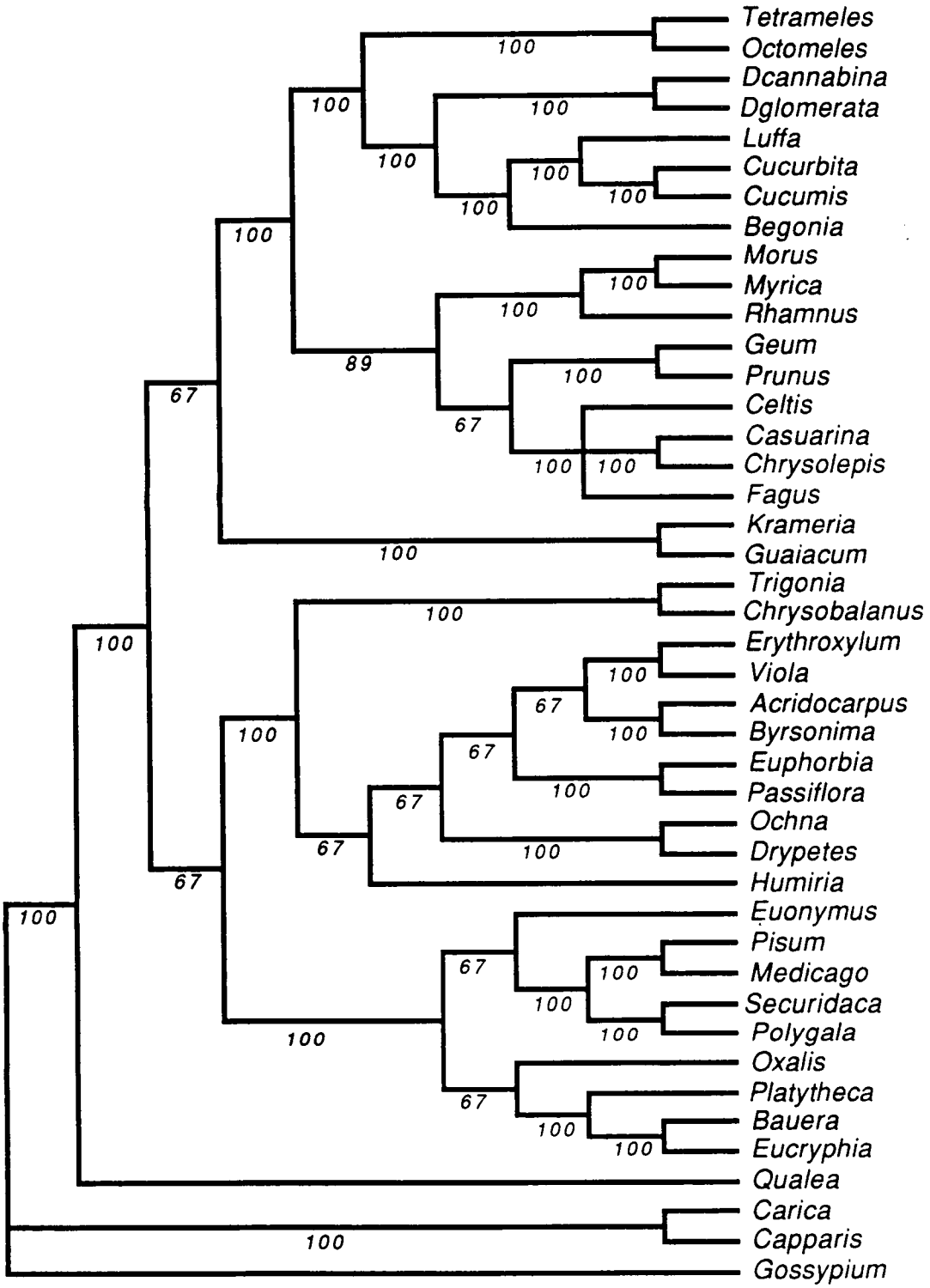

Tetrameles
Octomeles
Dcannabina
Dglomerata
Luffa
Cucurbita
Cucumis
Begonia
Morus
Myrica
Rhamnus
Geum
Prunus
Celtis
Casuarina
Chrysolepis
Fagus
Krameria
Guaiacum
Trigonia
Chrysobalanus
Erythroxylum
Viola
Acridocarpus
Byrsonima
Euphorbia
Passiflora
Ochna
Drypetes
Humiria
Euonymus
Pisum
Medicago
Securidaca
Polygala
Oxalis
Platytheca
Bauera
Eucryphia
Qualea
Carica
Capparis
Gossypium

Figure 10. 50% Majority-Rule Consensus of 198 Near-Minimal-Length Trees. Trees generated using the heuristic search in PAUP and specifying "Keep all trees<=length 1803". Trees of length 1802 (one step longer than most parsimonious tree) were selected using the "Filter Trees" option in the TREES menu and are summarized here in a 50% majority-rule consensus tree. Numbers above branches reflect branch lengths optimized to the consensus topology. Italicized numbers below the branches reflect the percentage of the original trees that support the group descending from that branch. Branches without numbers below are supported in 100% of the original trees.

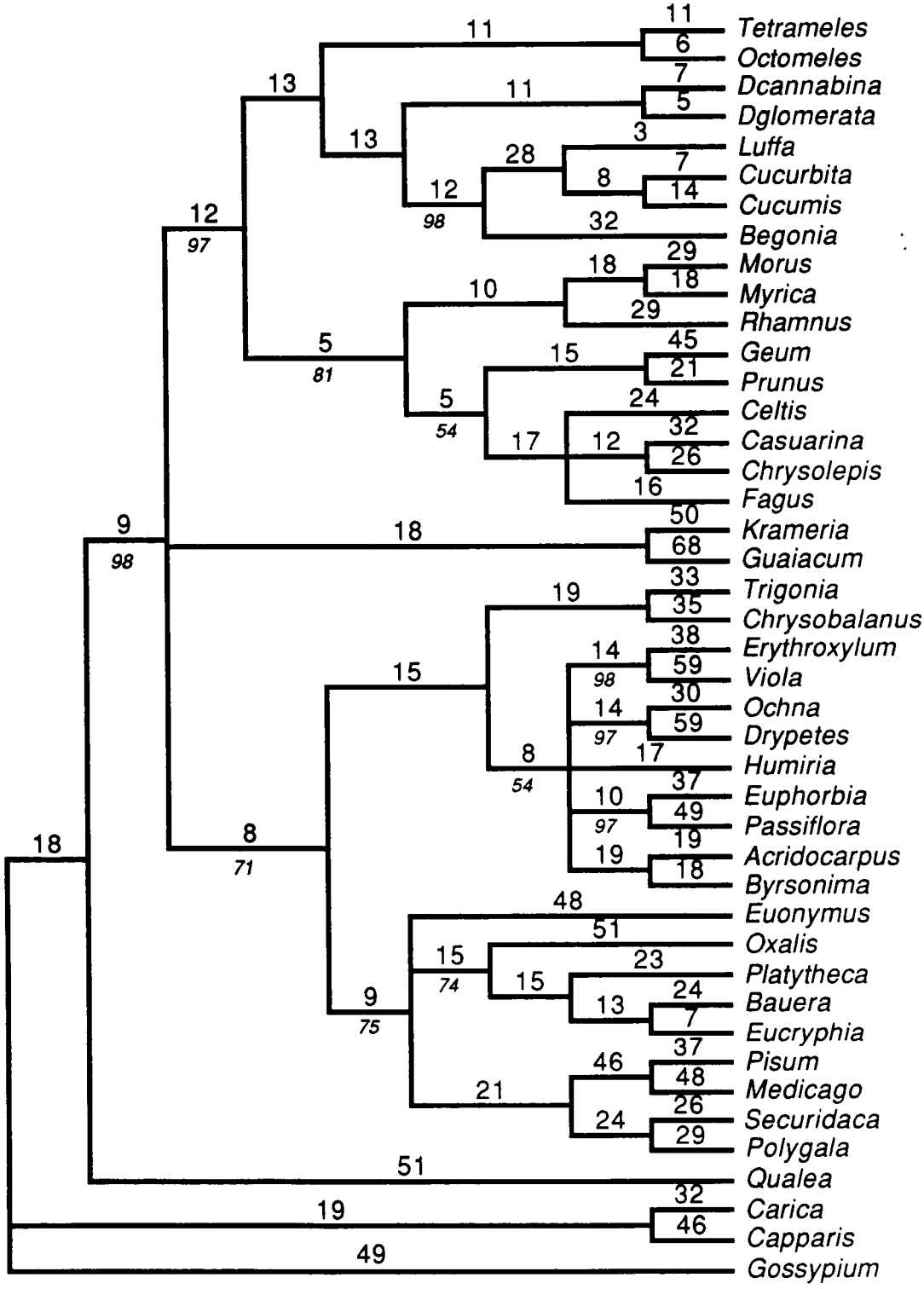
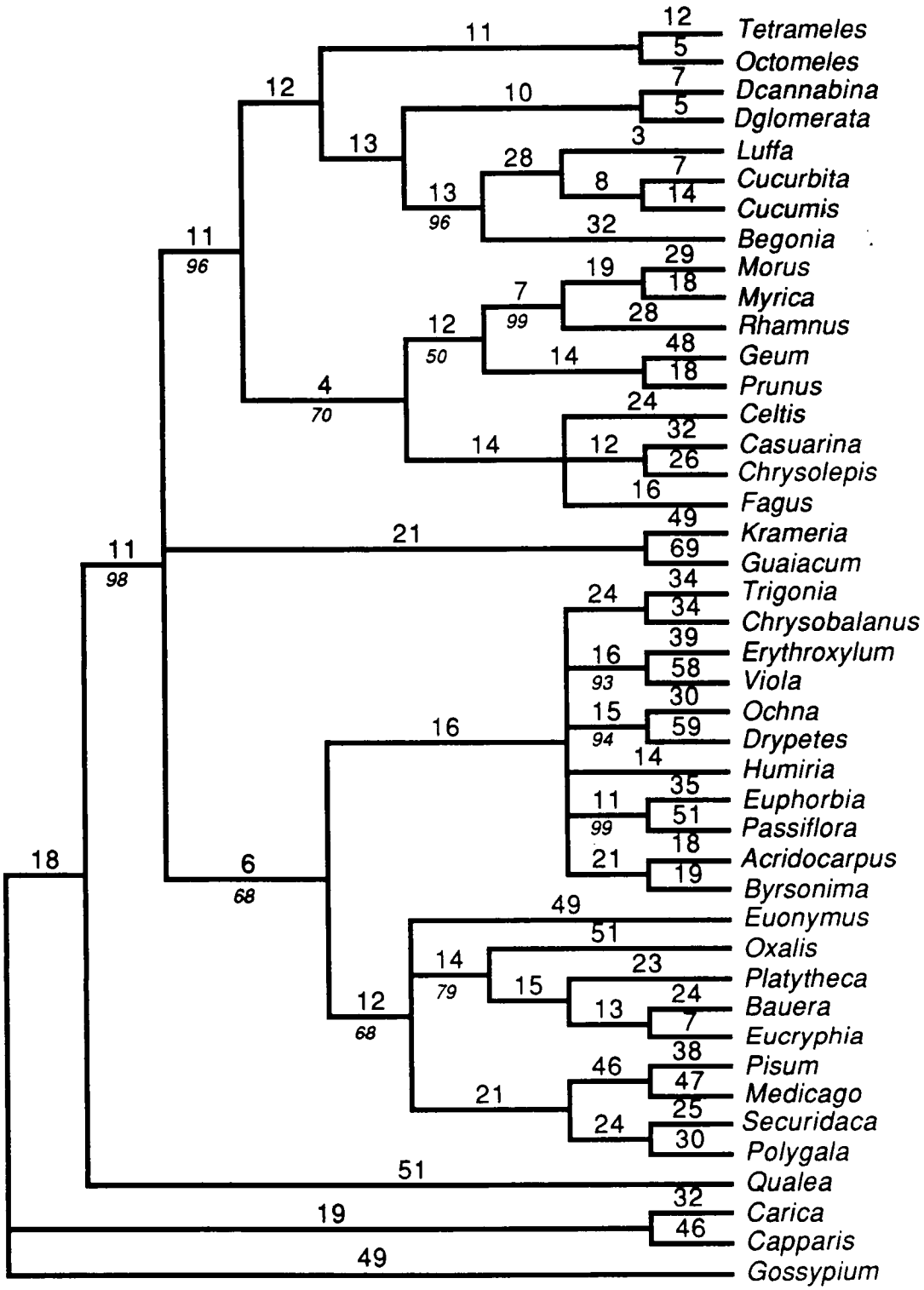

Tetrameles
Octomeles
Dcannabina
Dglomerata
Luffa
Cucurbita
Cucumis
Begonia
Morus
Myrica
Rhamnus
Geum
Prunus
Celtis
Casuarina
Chrysolepis
Fagus
Krameria
Guaiacum
Trigonia
Chrysobalanus
Erythroxylum
Viola
Ochna
Drypetes
Humiria
Euphorbia
Passiflora
Acridocarpus
Byrsonima
Euonymus
Oxalis
Platytheca
Bauera
Eucryphia
Pisum
Medicago
Securidaca
Polygala
Qualea
Carica
Capparis
Gossypium

Figure 11. 50% Majority-Rule Consensus of 1540 Near-Minimal-Length Trees. Original trees generated by heuristic search using 43 taxa. Trees of 1815 steps were retained using the "Filter Trees" option in PAUP. Numbers above the branches reflect the branch length optimized to the consensus topology. Italicized numbers below branches reflect the percentage of original trees that support the group descending from that branch. Branches without numbers below are supported in 100% of the original trees.


Tetrameles
Octomeles
Dcannabina
Dglomerata
Luffa
Cucurbita
Cucumis
Begonia
Morus
Myrica
Rhamnus
Geum
Prunus
Celtis
Casuarina
Chrysolepis
Fagus
Krameria
Guaiacum
Trigonia
Chrysobalanus
Erythroxylum
Viola
Ochna
Drypetes
Humiria
Euphorbia
Passiflora
Acridocarpus
Byrsonima
Euonymus
Oxalis
Platytheca
Bauera
Eucryphia
Pisum
Medicago
Securidaca
Polygala
Qualea
Carica
Capparis
Gossypium

stability of the topology near Datiscaceae is supported by the fact that the branching patterns did not decay in the trees that were one or two steps longer than the most parsimonious trees.

A comparison of the CI and RC values of the trees generated by the different searches reveals that as the number of taxa in the analysis increased, both the CI and RC value decreased. This phenomenon has been observed for the CI index by others (Sanderson and Donoghue, 1989; Archie, 1989) and the results presented here underscore the need for caution when comparing homoplasy indices from data sets with different numbers of taxa.

3. Data analysis using PHYLIP

To determine whether different parsimony programs would produce different topologies with the same data set, the 43-taxa data matrix was entered into the DNAPARS parsimony program available in PHYLIP. Because it is not easy to specify multiple outgroup taxa in PHYLIP, the outgroup for this analysis consisted of only Gossypium. Using the global branch-swapping option, ten minimal-length trees were generated. A majority-rule consensus tree was generated using the CONSENSE program (Figure 12). The consensus topology is the same for Datiscaceae, Cucurbitaceae and Begonia as that generated by the unweighted parsimony program in PAUP.

A phenetic analysis (based on overall similarity) was also performed using the DNADIST and FITCH programs of PHYLIP. Distances were computed from the sequence data using DNADIST and these were entered into the FITCH program for phylogeny estimation. The resulting phenogram (Figure 13) is different from all the parsimony-based trees

Figure 12. Consensus of Ten Equally Parsimonious Trees Generated by PHYLIP. Sequence data from 43 taxa were analyzed using the DNAPARS program which carries out unrooted analysis on DNA sequences. The consensus tree was generated using the CONSENSE program which displays all groups of taxa occurring in more than 50% of the input trees plus any groups occurring in less than 50% but which continue to resolve the tree without contradicting more frequent groups. The tree is oriented with Gossypium as the outgroup. Tree length = 1803 steps. The DRAWTREE program was used to generate the figure.

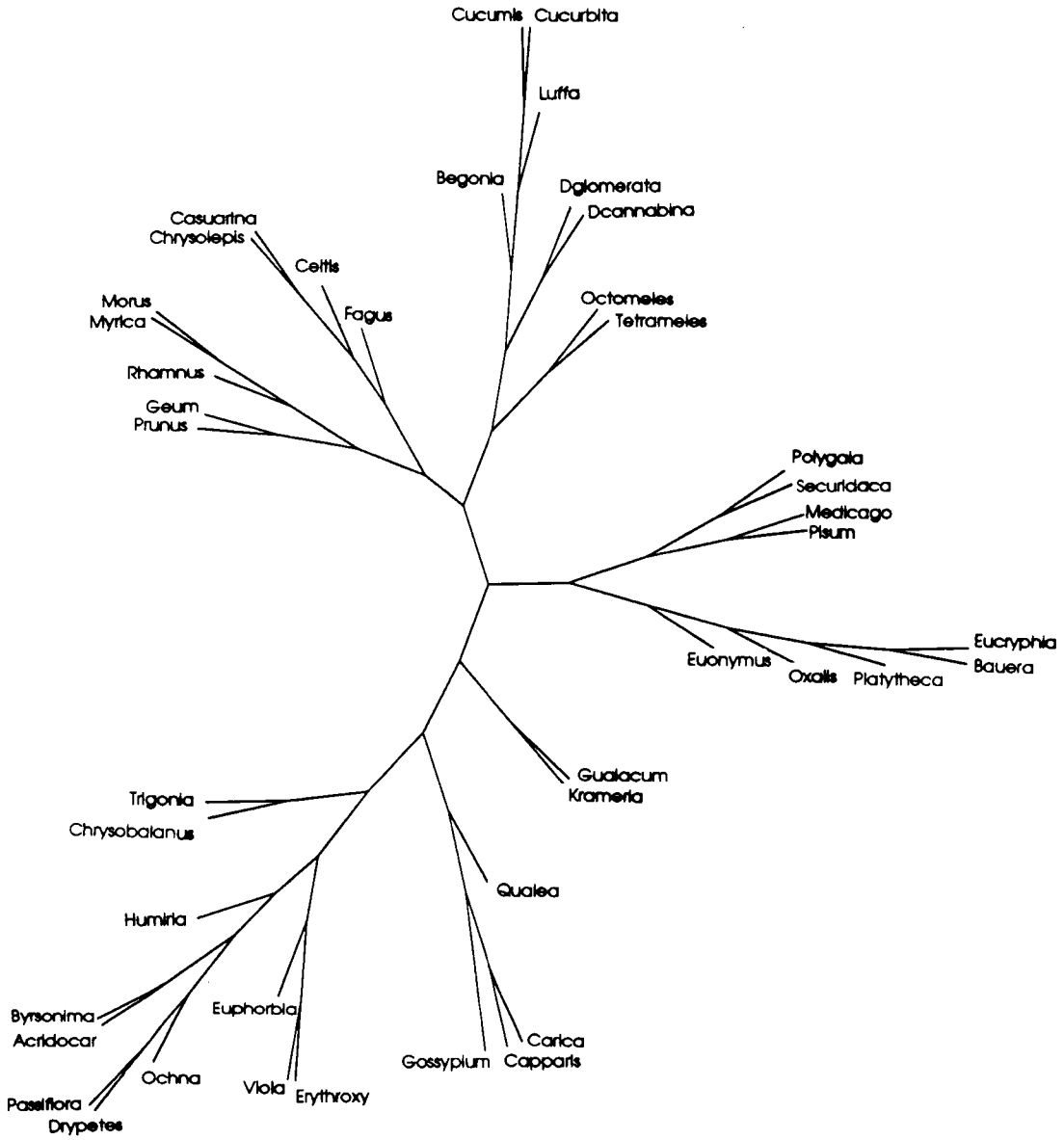
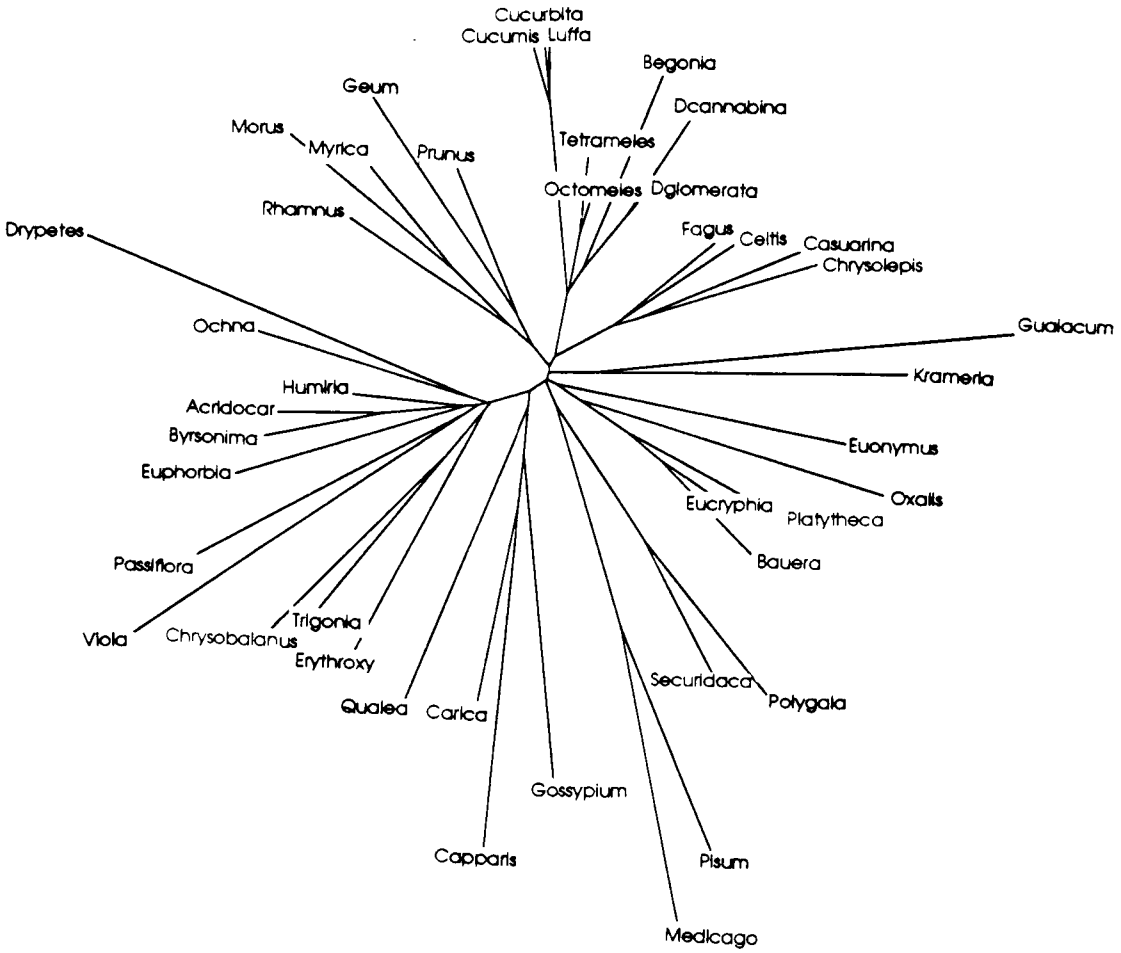

Cucumis
Cucurbita
Luffa
Begonia
Dglomerata
Dcannabina
Casuarina
Chrysolepis
Celtis
Fagus
Octomeles
Tetrameles
Morus
Myrica
Rhamnus
Geum
Prunus
Polygala
Securidaca
Medicago
Pisum
Eucryphia
Bauera
Euonymus
Oxalis
Platytheca
Guaiacum
Krameria
Trigonia
Chrysobalanus
Qualea
Humiria
Byrsonima
Acridocar
Euphorbia
Carica
Gossypium
Capparis
Ochna
Viola
Erythroxy
Passiflora
Drypetes

Figure 13. Phenetic Analysis of 43 Taxa. The DNADIST program in PHYLIP was used to generate pair-wise distances for the 43 sequences. Distance data were entered into the FITCH program which estimates phylogenies under the additive tree model. The global rearrangement option was selected to optimize the tree search by allowing subtree rearrangement. The unrooted tree is oriented with Gossypium as the outgroup. The DRAWTREE program was used to generate the figure.


Cucurbita
Cucumis Luffa
Begonia
Geum
Dcannabina
Morus
Myrica
Prunus
Tetrameles
Octomeles
Dglomerata
Rhamnus
Drypetes
Fagus
Celtis
Casuarina
Chrysolepis
Ochna
Guaiacum
Krameria
Humiria
Acridocar
Byrsonima
Euphorbia
Euonymus
Oxalis
Eucryphia
Platytheca
Bauera
Passiflora
Viola
Chrysobalanus
Trigonia
Erythroxy
Securidaca
Polygala
Qualea
Carica
Gossypium
Capparis
Pisum
Medicago

regarding the position of Octomeles and Tetrameles relative to the Datisca, Begonia and the Cucurbitaceae. In the phenetic arrangement, the Cucurbitaceae appear as a sister group to the Datiscaceae which includes Begonia grouping most closely with Datisca. Comparing the phenogram with the parsimony-based tree generated by PHYLIP (Figure 12), there is less resolution among the interior nodes of the phenogram. In the phenogram, the group containing Celtis and Casuarina are closer to the Datiscaceae than the Myrica/Rhamnus/Prunus group whereas in the parsimony-based tree, the Celtis and Myrica clades are arranged more closely to each other.

4. Further analysis of the data

To estimate the variability associated with the occurrence of specific groupings of taxa in the 43-taxa analysis, the bootstrap resampling method, available in the PAUP program, was employed according to Felsenstein (1985). By resampling from the original data, a set of bootstrapped data sets were constructed and phylogenies estimated from each data set. The variability among these phylogenetic estimations was used as a means of indicating the amount of error associated with the phylogeny estimated from the original data. As suggested by Felsenstein (1988), prior to conducting the bootstrapping, a specific group of interest in the analysis was identified. The Datiscaceae (including Octomeles, Tetrameles, and the two Datisca species) was selected as the group for which monophyly was to be tested.

The bootstrap resampling was repeated 100 times and the heuristic search option was selected for tree construction. The analysis required approximately 38 hours and produced a bootstrap 50% majority-rule

consensus tree (Figure 14) that shows all the groups that occurred in more than half of the bootstrap estimates of the tree. The bootstrap consensus tree shows that 68% of the estimates placed the Cucurbitaceae and Begonia in the same clade as Datisca, indicating that 32% of the estimates do not support this topology. Of these 32%, numerous arrangements are possible, many of which would not place the Datiscaceae together as a monophyletic group. From the 50% majority-rule bootstrap consensus tree, it cannot be determined if any of the estimates grouped the Datiscaceae as a monophyletic group, but the majority of estimates do not support the monophyly of the Datiscaceae.

In addition to the bootstrapping procedure, a parsimony analysis that imposed various weights upon different character-state changes was used to select among the nine equally-parsimonious minimal-length trees generated in the PAUP analysis of 43 taxa. The nine trees were used as a starting point for a heuristic search using a weighting scheme similar to that of Albert et al. (1993). Three stepmatrices were used in the input data set to differentially weight substitutions at each of the three codon positions (Swofford, 1991). The values of these weights were as follows: for transitions, 5520 (first position), 6368 (second position), 4039 (third position); for transversions, 6620 (first position), 7470 (second positions), 5127 (third position). These weights result in transversions being weighted 1.2 times more heavily than transitions and third codon position changes being weighted less than first or second position changes. Using the TBR branch-swapping option, the search found one minimal-length tree having 1802 steps (Figure 15). This tree differs from the previous nine in the

Figure 14. Bootstrap 50% Majority-Rule Consensus Tree. Bootstrapping was performed by PAUP using the heuristic search option for tree reconstruction from each of the 100 bootstrapped data sets. Gossypium, Carica and Capparis were selected as outgroup taxa. Numbers below branches reflect the percentage of bootstrap replications (50% or higher) that support the group descending from that branch.

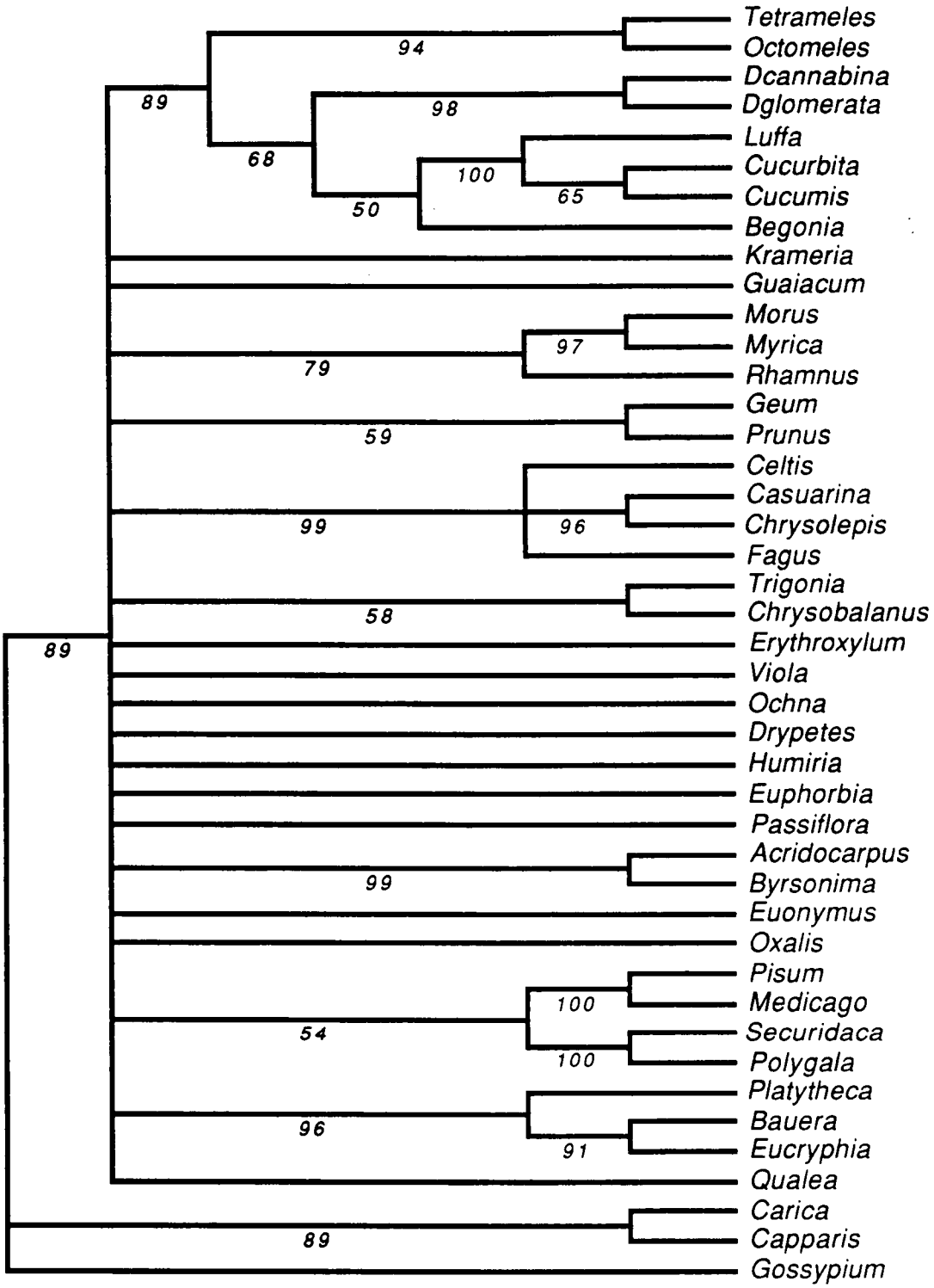
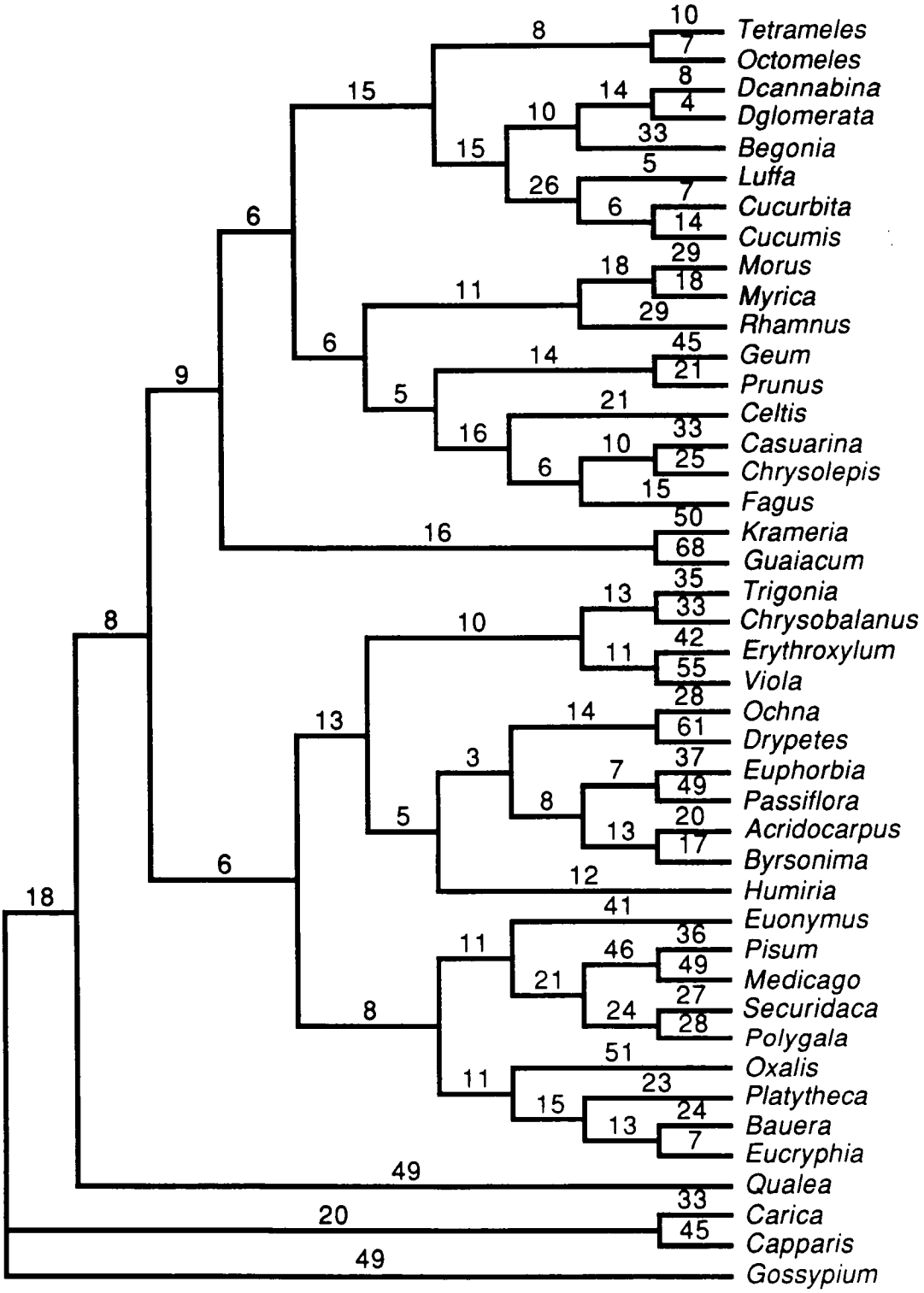

Tetrameles
Octomeles
Dcannabina
Dglomerata
Luffa
Cucurbita
Cucumis
Begonia
Krameria
Guaiacum
Morus
Myrica
Rhamnus
Geum
Prunus
Celtis
Casuarina
Chrysolepis
Fagus
Trigonia
Chrysobalanus
Erythroxylum
Viola
Ochna
Drypetes
Humiria
Euphorbia
Passiflora
Acridocarpus
Byrsonima
Euonymus
Oxalis
Pisum
Medicago
Securidaca
Polygala
Platytheca
Bauera
Eucryphia
Qualea
Carica
Capparis
Gossypium
94
89
98
68
100
65
50
97
79
59
99
96
58
89
99
100
54
100
96
91
89

Figure 15. Weighted Parsimony Analysis of 43 Taxa. Tree resulting from PAUP heuristic search employing character-state weights similar to those of Albert et al. (1993). Nine equally parsimonious trees previously obtained through unweighted parsimony analysis were used as starting trees for this reconstruction.


Tetrameles
Octomeles
Dcannabina
Dglomerata
Begonia
Luffa
Cucurbita
Cucumis
Morus
Myrica
Rhamnus
Geum
Prunus
Celtis
Casuarina
Chrysolepis
Fagus
Krameria
Guaiacum
Trigonia
Chrysobalanus
Erythroxylum
Viola
Ochna
Drypetes
Euphorbia
Passiflora
Acridocarpus
Byrsonima
Humiria
Euonymus
Pisum
Medicago
Securidaca
Polygala
Oxalis
Platytheca
Bauera
Eucryphia
Qualea
Carica
Capparis
Gossypium

positioning of Begonia relative to Datisca and is only one step longer. With character-weighting employed, Begonia groups more closely to Datisca than to the Cucurbitaceae, similar to its position in the phenetic tree generated by PHYLIP and in the near-minimal trees saved during the branch and bound analysis of 13 taxa under equal weighting.

The file containing the nine trees was also analyzed using a maximum likelihood program available in PHYLIP. The data file containing the 43 sequences with the nine trees was transferred to the IBM for use as the input file for the DNAML program. The user-defined tree option was selected to prompt the program to determine the likelihood estimates of the nine trees rather than to construct trees from the data. Of the nine trees analyzed, the one with the lowest log-likelihood value (-12283.397) is the tree shown in Figure 8. With user-defined trees supplied, DNAML performs a statistical test of each of the trees against the one with the highest likelihood (lowest log-likelihood). This test (Kishino and Hasegawa, 1989) uses the mean and variance of log-likelihood differences between trees (across all sites) and provides the same result as a bootstrap analysis, but is computationally more simple. A mean that is more than 1.96 standard deviations different indicates that the trees are significantly different. The results of this test showed that none of the eight trees were significantly worse than the tree with the highest likelihood.

It is of interest to compare the results of the sequence analysis with the restriction site analysis of the Datiscaceae performed by Rieseberg et al. (1992). Rieseberg et al. mapped restriction sites for 30 different restriction enzymes using PCR-amplified chloroplast DNA.

The mapped region included the rbcL gene and an adjacent open reading frame (ORF106) and thus, a portion of the restriction site data could be compared to a DNA sequence-generated map. The majority of sites were in agreement with the sequence-generated map. It appeared, however, that some enzyme sites were mapped incorrectly, especially those created by frequently-cutting enzymes. The positions of HhaI, HaeIII, Sau96I, BstNI, and EcoRV sites were not congruent with sites found in the actual sequence. Digestions performed with this combination of enzymes produced several small fragments (50 bp and less) whose sizes would be difficult to estimate using EtBr-stained agarose gels. The discrepancies between the two data sets might also be caused by misinterpretation of contiguous or homologous fragments in the restriction site analysis because no hybridizations were performed to verify fragment homology. Additionally, some of the restriction enzymes chosen for the analysis (BstNI, Sau96I, HaeIII, HhaI, BssHII) could recognize nested sites within the DNA, causing some sites to be counted more than once. It could not be determined from the data presented in Rieseberg et al. (1992) which of the mutations defining the branches were within rbcL and whether these mutations involved suspect sites, and thus, the accuracy of the tree could not be assessed. Nevertheless, the comparison of these two data sets demonstrates the need for caution in relying on agarose gel electrophoresis for interpreting the existence and sizes of restriction fragments for phylogenetic purposes.

Rieseberg et al. (1992) utilized several representatives from the Begoniaceae and Cucurbitaceae as outgroups based on previous suggestions (Cronquist, 1981; R.F.Thorne, pers. comm.) that these families may be

the most closely related to Datiscaceae. By forcing these taxa as the outgroup, Datiscaceae (including Octomeles and Tetrameles) necessarily formed a monophyletic group. Rieseberg has since analyzed the data using a different outgroup and maintains that the restriction site data support the monophyly of Datiscaceae sensu lato (L. Rieseberg, pers. comm.).

B. Discussion

1. Methods of analysis

The overall conclusion based on different analyses of the rbcL sequence data is that the Datiscaceae are not a monophyletic assemblage. This conclusion is supported by the results of parsimony-based analyses, phenetic analysis and weighted parsimony analysis. Some differences in the branching patterns relating Datisca, Begonia and the Cucurbitaceae were detected and these differences depended on the type of analysis employed. To determine which result is the best estimation, it is useful to consider the various attributes of each type of analysis.

In the unweighted parsimony-based analyses, a group formed by Begonia and the Cucurbitaceae is the sister group to Datisca and a group formed by Octomeles and Tetrameles is the nearest outgroup. These results differ from those obtained in the phenetic analysis where Begonia is a sister group to Datisca and the Cucurbitaceae are placed completely outside the Datiscaceae (Figure 13). The parsimony algorithm employed by PAUP and PHYLIP for DNA sequence data (unordered characters) works by assigning character-states to internal nodes of the tree so as to minimize the total number of character-state changes

according to Fitch (1971). For sequence data, this approach is considered to be preferable to either the Wagner or Dollo parsimony algorithms because it imposes the fewest constraints on permissible character-state changes. Wagner parsimony (Kluge and Farris, 1969) is appropriate when character-state changes follow an ordered pattern from one state to the next. Dollo parsimony assumes that the probabilities of character-state transformations are asymmetrical and some types of changes are not allowed (Swofford and Olsen, 1990). Fitch parsimony allows unordered character-state changes and assumes that all changes are equally likely. Although the Fitch algorithm is the best choice for use with sequence data, it does not account for any differences in types of character-state changes that occur. One criticism of parsimony methods in general has been that they fail to produce consistent phylogeny estimations when the amounts of evolutionary change in different lineages are sufficiently unequal (Felsenstein, 1978). In trees where long, unbranched lineages exist, multiple changes cannot be detected and homoplasious changes can accumulate to the point that taxa at the ends of these long branches appear to be associated. The potential for inconsistency increases when the data contain sequences that are relatively divergent with few interconnecting taxa (Swofford and Olsen, 1990).

The phenetic analysis resulted in a different branching pattern regarding Datisca, Begonia and the Cucurbitaceae than the parsimony analyses and also provides less resolution in the interior portion of the tree when compared to the PHYLIP parsimony-based tree (Figures 12 and 13). The phenetic analysis performed using PHYLIP utilizes the

Kimura 2-parameter model (Kimura, 1980) to generate a distance matrix and the Fitch-Margoliash (1967) method of phylogeny estimation. An assumption of this model is that the data are additive, meaning that the lengths of the branches between any pair of taxa can be summed to yield a measure of the amount of change. When the branch length sum separating two taxa is large, the expected distance will not match the branch length sum because overlaid substitutions cannot be detected. This method is therefore inconsistent when differences between DNA sequences are large, and branch lengths will be underestimated (Felsenstein, 1988). A more general criticism of distance methods is that substantial information is lost when comparisons among sequences are reduced to a single number.

The tree shown in Figure 8 was selected as the most likely topology of nine equally-parsimonious trees given the data using the maximum likelihood method and model. Some researchers consider this method to be generally more consistent and better than either parsimony or distance methods because it is based on an explicit rather than an implicit model. If the model is found to be inappropriate, it can be changed to better fit the evolutionary process. Unfortunately, the computational complexity of maximum likelihood estimations prevents the analysis of data sets containing more than about ten taxa unless one has access to a supercomputer. Maximum likelihood was used in this analysis only to determine the log-likelihoods of some user-defined trees.

As noted by Felsenstein (1978), there is a close relationship between maximum likelihood and maximum parsimony methods. This relationship has been useful in the construction of a weighted parsimony

method which becomes the maximum likelihood method if the characters under analysis have small probabilities of change (Felsenstein, 1981). This condition may be met by slowly-evolving genes such as rbcL. The weighted-parsimony method is difficult to use for large data sets, and was used in this study only as a means of selecting among a group of equally-parsimonious unweighted trees.

It is important to point out that none of the methods of estimating phylogenies is optimal. Each method has its own inconsistencies and therefore results obtained may not represent actual evolutionary relationships. The weighted parsimony method of selecting trees may be the most reasonable approach, however, to estimating the relationships of the Datiscaceae. All of the sequence information is utilized, different types of nucleotide change (transition, transversion, codon position) are allowed and can be weighted under a maximum likelihood framework. Problems associated with long branch lengths and spurious homology cannot be eliminated, but can be reduced by not including widely divergent taxa in the analysis and by choosing outgroup taxa that are as close as possible to the ingroup. Given the attributes of this weighted parsimony method, the results shown in Figure 15 are considered to be the best representation of the relationships concerning the Datiscaceae.

2. Paraphyly of the Datiscaceae

The results presented here, suggesting that the Datiscaceae form a paraphyletic group, are in disagreement with the conclusions of Davidson (1973, 1976) and Boesewinkel (1984) based on morphological and anatomical evidence which support the monophyly of the Datiscaceae. The molecular

data are in agreement, however, with the conclusions of Airy-Shaw (1964), based on morphological data. The different conclusions drawn from studies of morphology and anatomy appear to be caused by different interpretations of these data rather than by a lack of basic information or by the presence of conflicting data.

The bulk of morphological and anatomical data for the Datiscaceae comes from detailed studies by Davidson of floral and vegetative parts (1973) and wood anatomy (1976). Several features are noted to differ between the Datisca species and Octomeles and Tetrameles. The long, pendant, indeterminate inflorescences in Octomeles and Tetrameles are in contrast with the determinate flower clusters, or dichasia, of Datisca. The floral vasculature of Octomeles/Tetrameles differs from Datisca by having single style traces continuing from a large stylar bundle whereas Datisca styles are supplied by two traces derived from adjacent lateral bundles in the ovary wall. The female flowers of Octomeles and Tetrameles develop into tubular ovaries that are surmounted by a "crown" composed of a nectary cup (in Octomeles), the perianth tube, and the styles. The Datisca ovary is angular in transection and there is no crown structure or nectary. The pollen grains of Datisca are often shed in tetrads (never united in tetrads at maturity for Octomeles or Tetrameles) and have a more conspicuously sculptured exine. Octomeles and Tetrameles are similar in petiolar anatomy, but Datisca differs in having 3-trace unilacunar nodes with a very simple pattern of bundle arrangement at the proximal end of the petiole. Differences in trichome ontogeny are mentioned as important distinguishing features between Datisca and the other genera. These differences are minimized by

Davidson (1973) whereas features such as the similarity in dispersal mechanisms of Datisca and Tetrameles are emphasized.

In the study of wood anatomy, Davidson (1976) concludes that Datisca wood differs both quantitatively and qualitatively from that of the two trees. This observation would be expected since Datisca produces much less wood and the mature pattern is more slowly obtained than in the arboreous genera (Davidson, 1976). The outer wood produced in large Datisca stems is different than that close to the pith, however, the characteristics of the outer wood do not seem to resemble closely the wood of the other two genera. Davidson (1976) points out that a major difference in the wood of Datisca, when compared to Octomeles and Tetrameles, is the conspicuous arrangement of vessels in concentric or tangential bands which do not correspond to growth rings. Ring porous wood was observed only in the buttresses of Tetrameles (Davidson, 1976). Despite these differences in wood anatomy, Davidson (1976) concludes that the evidence supports the inclusion of the three genera in a single family.

A study of the ovule and seed structure of the Datiscaceae (Boesewinkel, 1984) concluded that Datisca, Octomeles and Tetrameles are closely related and seem to be correctly placed in the same family. The only observation supporting this conclusion is a comparison of seed shape among D. cannabina, Octomeles and either an immature or an abnormal seed sample from Tetrameles (Boesewinkel, 1984). The study lacks the corresponding analyses of ovule ontogenesis and seed development for Octomeles and Tetrameles, that was done for Datisca. Based on the comparison of a limited number of characteristics among the

genera, and the lack of suitable seed material from Tetrameles, the conclusion of Boesewinkel (1984) seems weakly supported.

The separation of Octomeles and Tetrameles into a separate family, (Tetramelaceae (Warb.) Airy-Shaw), as was done by Airy-Shaw (1964), is based on a number of differences between Datisca and the tree genera. The difference in habit is mentioned as a major reason for the separation of the genera and this emphasis on habit was criticized by Davidson (1976). Airy-Shaw (1864) admits that such differences in habit are found within other families, but in those cases the extremes are connected by a series of intermediates. Airy-Shaw (1964) cites differences between Datisca and the tree genera that go far beyond general habit. These differences include the pinnate or pinnatifid leaves with long, narrow caudate-acuminate pinnae in Datisca which differ from the simple, cordate-ovate leaves of the Tetramelaceae. In Datisca, the stamens are of indefinite number and are without definite arrangement in regard to the sepals; filaments are very short and slender; anthers are large, straight, and oblong. In Tetramelaceae, the stamens are isomerous with and opposite to the sepals; the filaments are elongate, somewhat flattened-subulate, and the anthers are relatively short and ovoid, or if larger and oblong, then conspicuously recurved. The styles of Datisca are elongate-subulate, densely papillose from base to apex, the apex being acute; those of Tetramelaceae are shorter or very short, stout, with large capitate or obliquely oblong-decurrent stigmas (Airy-Shaw, 1964). These differences, which are acknowleged by Davidson (1973) are interpreted by Airy-Shaw (1964) as support for the separation of Datisca from Octomeles and Tetrameles. The opposite

conclusions arrived at by Davidson (1973) and Airy-Shaw (1964) were based on observations of essentially the same characteristics, but with different characteristics being emphasized in each case. Many of the characters mentioned above have been incorporated into a table (Table 6) for separate interpretation. Based on the features included in Table 6, it is difficult to find support for the monophyly of the Datiscaceae.

Despite the extensive morphological and anatomical analyses performed on the Datiscaceae, none of the characters have been interpreted cladistically. An important focus of future study should be the characterization of features describing Datisca, Octomeles and Tetrameles as apomorphies or plesiomorphies and the construction of a tree based on derived characters. This tree could then be compared to the rbcL-based phylogeny and the congruence between morphological and molecular data could be assessed more effectively.

The difference in opinion regarding the monophyly of the Datiscaceae illustrates the difficulty in the interpretation of morphological and anatomical features and in the application of any conclusion to actual changes in classification. The results of the rbcL analysis, however, strongly support Airy-Shaw's (1964) interpretation and recognition of two families. The choice of rank for the separate groups is largely a matter of opinion but the molecular data presented here support the division at the familial level.

3. Outgroups to the Datiscaceae

In the phylogenetic estimations presented here, it is of interest to look beyond the group which includes the Datiscaceae, Cucurbitaceae and Begonia (the "ingroup" for this discussion), and consider the

Table 6. A Comparison of Morphological and Anatomical Features[a]

Character	Datisca[b]	Octomeles	Tetrameles	Begoniaceae	Cucurbitaceae
inflorescences	greatly shortened determinate	long indeterminate	long indeterminate	sometimes long determinate	various indeterminate
♀ calyx	3-4 lobes	6-8 teeth	4-5 sepals	2-5 petaloid tepals	5 sepals
♂ calyx	4-10 lobes	6-8 teeth	4-5 sepals	2-5 petaloid tepals	5 sepals
♂ corolla	no petals	6-8 petals	no petals	no petals	5 petals
style/stigma	long, bifid	short, capitate or peltate stigmas	short, grooved spatulate stigmas	bifid	usually bilobed stigmas
anthers	6-20 (indefinite)	6-8	4-5	indefinite	five, often modified
ovary placentation	parietal	parietal	parietal; often meeting at center	axile	parietal
fruit	capsule	capsule	capsule	capsule	berry or pepo, less often capsule
dehiscence	apical pore	longitudinal	apical pore	loculicidal	various
embryo sac	Allium[c]	N.A.	N.A.	Polygonum	Polygonum

Table 6. (continued)

Character	Datisca[b]	Octomeles	Tetrameles	Begoniaceae	Cucurbitaceae
seed operculum	present[c]	present	present	present	absent
pollen tectum	sculptured	unornamented	N.A.	various	various
leaf shape	pinnatifid to pinately compound	simple, obtuse to subcordate, ovate	simple, cordate-ovate	simple; sometimes palmately compound	palmately lobed to palmately compound
leaf venation	pinnate	pinnate and palmate	pinnate and palmate	palmate	palmate
stipules	absent	absent	absent	present	absent
chromosome number	x = 11	N.A.	N.A.	x = 10 - 21+	x = 7 - 14
habit	herb	tree	tree	herb	herb

[a]Characters taken from previous studies in the literature: Cronquist (1981), Davis (1966), Davidson (1973), Boesewinkel (1984), Boesewinkel and DeLange (1983), Bates et al. (1990), van Steenis (1953).

[b]Characters based on Datisca glomerata unless noted otherwise. Most characters are common to both species.

[c]Characters based on Datisca cannabina.

N.A. = information not available.

closest outgroup. In most reconstructions, the next closest outgroup was either a group containing members of the genera Rhamnus, Myrica, and Morus or a group consisting of representatives of Chrysolepis, Casuarina, Celtis, and Fagus. The phenetic analysis suggests that the latter group is closer to the ingroup than the former group, but the parsimony analyses reveal little resolution regarding the relative proximity of either group to the ingroup.

Although the Chrysolepis group contains only members of the higher Hamamelidae, the Myrica group contains Rhamnus, which is placed in the subclass Rosidae (Cronquist, 1981). The mixing of taxa traditionally placed within the Rosidae with members of the higher Hamamelidae (excluding the Trochodendrales and the Hamamelidales; Cronquist, 1981) is consistently apparent in these reconstructions. Specifically, Geum, Prunus, Krameria and Guaiacum are shown to group near the hamamelid taxa. Also, taxa traditionally placed within the subclass Dilleniidae such as Viola, Passiflora and Ochna, in addition to the Datiscaceae, are shown interspersed among rosid taxa. Although not a specific aim of this study, the reconstructions provide evidence to support the idea that the higher hamamelids may actually be rosids (Doyle & Donoghue, 1987) and that the Dilleniidae are not a monophyletic group.

It is important to note that both of the closest outgroups to the ingroup contain actinorhizal genera. Both Myrica and Casuarina contain actinorhizal species and Rhamnus belongs to a family that contains seven actinorhizal genera. Traditionally, the actinorhizal genera have been separated from each other and placed in three different subclasses (Cronquist, 1981). The proximity of these taxa in the rbcL-based

phylogenies lends support to the view that the actinorhizal symbiosis may have originated once in a single ancestor, and the hosts have since diverged with the loss of the symbiotic association occurring in various lineages. Additional evidence to support this evolutionary view comes from a recent rbcL-based analysis which included a representative of another taxonomically isolated actinorhizal genus, Coriaria (M.W. Chase; pers. comm.). This analysis placed Coriaria as a sister group to Datisca. The genus Coriaria has been placed in the Ranunculales by Cronquist (1981) but no agreement exists among systematists as to its proper placement. Although more actinorhizal sequences need to be analyzed, these data suggest that the actinorhizal taxa may be more closely related than previously thought.

From the families listed as potential relatives of Datiscaceae based on previous studies (Table 2), only the alliances to Cucurbitaceae and Begoniaceae are supported by the rbcL sequence data. Representative taxa from Capparales (Capparis, Brassica, Moringa), Haloragales (Myriophyllum, Gunnera) and Violales (Carica, Passiflora, Viola) were included in either this analysis or the broad analysis of Chase et al., (1993) and were not observed to group near Datiscaceae. In fact, in the broad analysis, the positions of Capparis, Brassica and Moringa in the closest outgroup to the Datiscaceae made them suitable choices as outgroup taxa for the present analyses. Myriophyllum and Gunnera were observed to group even further from Datiscaceae with other members of the Rosiidae (Chase et al., 1993).

4. Patterns of change

The PAUP reconstructions presented here show branch length estimations which were calculated by PAUP (using ACCTRAN optimization) and reflect the number of apomorphies that define particular branching points. Using the "Describe Trees" command in PAUP and specifying the apomorphy list, a list of the character state changes at each node was obtained and analyzed to investigate the nature of the changes defining important branch points in PAUP trees.

Inspection of the apomorphy list from the unweighted analysis of 43 taxa revealed that the majority of nucleotide changes occur at the third position of the codon, although the percentage of third position changes varied. Only 25% of the changes that separate D. cannabina from D. glomerata are third position changes whereas over 90% of the changes defining most interior nodes of the tree are third position changes. The fact that the majority of substitutions occur at the third position was also revealed when this position was dropped from the data matrix and only first and second positions were used for tree reconstruction. The resulting consensus tree (not shown) had numerous polychotomies and lacked the resolution seen with the other reconstructions. The information from the third position substitutions is therefore important in defining the relationships of the Datiscaceae to each other and to Cucurbitaceae, Begonia and other groups.

Among the nucleotide substitutions examined, almost all first and second position substitutions lead to amino acid substitutions whereas only a few third position changes result in amino acid substitutions. Of the nine substitutions defining the Octomeles/Tetrameles branch, three

result in amino acid substitutions (including one third-position change). Octomeles and Tetrameles are separated by 17 substitutions, eight of which result in amino acid changes (including two third-position changes). The branch leading to Datisca is defined by 10 substitutions, resulting in three amino acid changes (one third-position change caused one of the amino acid changes). The Datisca species are separated by 12 nucleotide substitutions which result in six amino acid changes. The Datisca/Cucurbitaceae/Begonia clade is defined by 13 nucleotide substitutions resulting in seven amino acid changes. In general, most branch points are defined by several substitutions, and the substitutions are distributed evenly throughout the rbcL gene although a few more amino acid substitutions are found at the extreme 3' end. The active sites in the large subunit of the RuBP carboxylase enzyme include four amino acid residues: lysine at residue number 175; lysine at 334; cysteine at 459 and lysine at the 201 (numbered from the amino-terminal end). The first three residues are binding sites for the RuBP substrate whereas the lysine 201 is the residue that binds the activator CO_2 (Miziorko and Lorimer, 1983). The amino acid substitutions changes mentioned above do not involve any of these active site residues and all are found outside of the more conserved regions which surround lysines 175, 201 and 334.

The apomorphy list also provides a measure of homoplasy for each of the character-state changes. Although a complete analysis of homoplasy at each character is impractical, levels of homoplasy for selected characters and branch points can be assessed. The average of the individual character consistency index (ci) values for the nine

substitutions defining the Octomeles/Tetrameles branch is 0.289. The average ci value for the 10 substitutions defining the Datisca branch is 0.395. Such low values indicate that few of the substitutions are unique; however, several terminal nodes are defined by unique substitutions (autapomorphies). Datisca cannabina is defined by five unique nucleotide sustitutions, four of which resulted in amino acid changes. Tetrameles has two unique substitutions resulting in a single amino acid change. Begonia has eight unique substitutions, Cucumis has two and Cucurbita has one. None of the interior nodes are defined by unique substitutions.

The nature of the substitutions observed in these data is consistent with changes that might be expected to occur in a highly conserved gene. Although the sequences being analyzed involve interfamilial and interordinal comparisons, the majority of differences are synonymous substitutions which reflect the selective constraint imposed on the gene. Of the amino acid changes analyzed, approximately half are highly conservative changes involving substitutions of amino acids with similar properties. Such conservative changes coupled with the fact that these changes are not close to any of the active sites suggest that the changes are unlikely to greatly alter the activity of the protein product expressed.

5. Congruence with the "global" phylogeny

The results of this limited analysis of the Datiscaceae and related taxa are generally congruent with the topology presented by Chase et al., 1992, and therefore provide support for their broad analysis. In such a broad analysis, the greater number of outgroup taxa relative to a

group of interest may change the way that character-states are assessed at ingroup branch points. Also, if more taxa are added in addition to those in a specific section of the "global" tree, the original branching patterns may change. With the selection of 36 taxa from the broad analysis and the addition of seven new sequences, the major branching patterns did not change. In comparison with the broad analysis (Figure 3), however, the positions of some groups of taxa were inconsistent with the present analysis. Among these, the position of the Casuarina/Celtis clade as the closest sister group of Datisca could not be supported by this analysis (Figure 9). The position of Krameria and Guaiacum near the Myrica group in the broad analysis was also not supported. Krameria and Guaiacum together were observed in the present analyses to associate with many different groups and did not appear to be firmly supported in any position. The differences in the placement of these groups in the broad analysis of Chase et al. (1993) and in the analysis presented here may reflect the need for more potentially related taxa to be added to the analysis. The addition of more taxa would serve to interrupt some of the longer branches in the topology and would allow the character-state changes to be optimized more accurately.

6. Suggestions for further work

Future work on this project should include studies focusing on three areas: 1) further analysis of the morphological and anatomical data, 2) analysis of sequence data from other genes, and 3) analysis of sequence data from additional actinorhizal taxa.

A cladistic analysis of the morphological and anatomical characters already identified by previous studies and the construction of a tree

based on derived charateristics would help in the interpretation of these characters. Through this analysis, the congruence between the morphological data and the molecular data could be assessed more effectively.

The relationships of the Datiscaceae to each other and to Begonia and Cucurbitaceae are further defined by the rbcL-based analysis presented here and it is doubtful that the addition of taxa to this analysis would change the branching patterns in this part of the tree. The addition of more sequence data, however, from another conserved gene such as ribosomal DNA would be valuable as an independent corroboration of the rbcL-based phylogeny.

Relationships among five actinorhizal taxa revealed by the rbcL-based analysis contradict relationships presented in traditional classification schemes. Clearly, more taxa need to be examined to fully address the relationships among actinorhizal plants and the questions related to the evolution of the actinorhizal symbiosis. Subsequent studies should include actinorhizal taxa from the unsampled families Elaeagnaceae, Betulaceae (Alnus), Rhamnaceae, and Rosaceae.

REFERENCES CITED

Airy-Shaw, H. K. 1964. Diagnoses of new families, new names, etc. for the seventh edition of Willis's "Dictionary." Kew Bull. 18: 249-273.

Archie, J. W. 1989. Homoplasy excess ratios: new indices for measuring levels of homoplasy in phylogenetic systematics and a critique of the consistency index. Syst. Zool. 38: 253-269.

Albert, V. A., M. W. Chase, and B. D. Mishler. 1993. Character-state weighting for cladistic analysis of protein-coding DNA sequences: prospects, limitations, and an example from the plastid rbcL gene. Ann. Missouri Bot. Gard. 80: in press.

Baker, D. D. and B. C. Mullin. 1992. Actinorhizal symbioses. In: Biological Nitrogen Fixation. G. Stacey, R. H. Burris and H. J. Evans, Eds. Chapman and Hall, New York.

Baker, D. D. and C. R. Schwintzer. 1990. Introduction. In: The Biology of Frankia and Actinorhizal Plants. C. R. Schwintzer and J. D. Tjepkema, Eds. Academic Press, Inc., New York.

Bates, D. M., R. W. Robinson and C. Jeffrey. 1990. Biology and Utilization of the Cucurbitaceae. Cornell University Press, Ithaca, New York.

Boesewinkel, F. D. 1984. Ovule and seed structure in Datiscaceae. Acta Bot. Neerl. 33: 419-429.

_____ and A. DeLange. 1983. Development of ovule and seed in Begonia squamulosa Hook. F. Acta Bot. Neerl. 32: 417-425.

Bohm, B. A. 1988. Flavonoid systematics of the Datiscaceae. Biochem. Syst. Ecol. 16: 151-155.

Brooks, D. R., R. T. O'Grady and E. O. Wiley. 1986. A measure of the information content of phylogenetic trees, and its use as an optimality criterion. Syst. Zool. 35: 571-581.

Chase, M. W., O. E. Soltis, R. G. Olmstead, D. Morgan, D. H. Les, M. R. Duvall, R. Price, H. G. Hills, Y. Qiu, K. A. Kron, J. H. Rettig, E. Conti, J. D. Palmer, M. T. Clegg, J. R. Manhart, K. J. Sytsma, H. J. Michaels, W. J. Kress, M. J. Donoghue, W. D. Clark, M. Hedrén, B. S. Gaut, R. K. Jansen, K.-J. Kim, C. F. Wimpee, J. F. Smith, G. R. Furnier, S. H. Straus, Q. Xiang, G. M. Plunkett, P. S. Soltis, L. E. Eguiarte, G. H. Learn, Jr., S. C. Barrett, S. Graham, and V. A. Albert. 1993. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene rbcL. Ann. Missouri Bot. Gard. 80: in press.

Clegg, M. T., K. Ritland, and G. Zurawski. 1986. Processes of chloroplast DNA evolution. In: Evolutionary Processes and Theory. S. Karlin and E. Nevo, Eds. Academic Press, Inc., New York.

Corner, E. J. H. 1976. The Seeds of Dicotyledons. Vol. 1. Cambridge University Press, New York.

Cronquist, A. 1981. An Integrated System of Classification of Flowering Plants. Columbia University Press, New York.

Davidson, C. 1973. An anatomical and morphological study of Datiscaceae. Aliso 8: 49-110.

_____. 1976. Anatomy of xylem and phloem of the Datiscaceae. Contributions in Science, Natural History Museum of Los Angeles County 280: 1-28.

Davis, G. L. 1966. Systematic Embryology of the Angiosperms. J. Wiley and Sons, Inc., New York.

Dellaporta, S. L., J. Wood, and J. B. Hicks. 1983. A plant DNA minipreparation: version II. Plant Mol. Biol. Rep. 1: 19-21.

Devereux, J., P. Haeverli, and O. Smithies. 1984. A comprehensive set of sequence analysis programs for the VAX. Nucl. Acids Res. 12: 387-395.

Doskotch, R. W., M. Y. Malik, and J. L. Beal. 1969. Cucurbitacin B, the cytotoxic principle of Begonia tuberhybrida var. alba. Lloydia 2: 115-122.

_____ and C. D. Hufford. 1970. Hexanor-cucurbitacin D, a degraded cucurbitacin from Begonia tuberhybrida var. alba. Can. J. Chem. 48: 1787-1788.

Doyle, J. A. and M. J. Donoghue. 1987. The importance of fossils in elucidating seed plant phylogeny and macroevolution. Review of Palaeobot. Palynol. 50: 63-95.

Engler, A. 1964. Syllabus der Pflanzen-Familien II Band. Gebruder Bortraeger, Berlin.

Farris, J. S. 1989. The retention index and rescaled consistency index. Cladistics 5: 417-419.

Felsenstein, J. 1978a. The number of evolutionary trees. Syst. Zool. 27: 27-33.

_____. 1978b. Cases in which parsimony and compatibility methods will be positively misleading. Syst. Zool. 27: 401-410.

_____. 1981. A likelihood approach to character weighting and what it tells us about parsimony and compatibility. Biol. J. Linn. Soc. 16: 183-196.

_____. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39: 783-791.

_____. 1988. Phylogenies from molecular sequences: inference and reliability. Ann. Rev. Genet. 22: 521-565.

_____. 1989. PHYLIP--Phylogeny Inference Package (version 3.2). Cladistics 5: 164-166.

Fitch, W. M. 1971. Toward defining the course of evolution: minimum change for a specific tree topology. Syst. Zool. 20: 406-416.

Fitch, W. M. and E. Margoliash. 1967. Construction of phylogenetic trees. Science 155: 279-284.

Hamby, R. K. 1990. Ribosomal RNA and the early evolution of flowering plants. Ph.D. dissertation. Louisiana State University, Baton Rouge.

Hufford, L. 1990. Androecial development and the problem of monophyly of Loasaceae. Can. J. Bot. 68: 402-419.

Jeffrey, C. 1990. Appendix: an outline classification of the Cucurbitaceae. In: Biology and Utilization of the Cucurbitaceae. Cornell University Press, Ithaca, New York.

Kimura, M. 1980. A simple model for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. 16: 111-120.

Kishino, H. and M. Hasegawa. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. J. Mol. Evol. 29: 170-179.

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

Lawrence, G. H. M. 1951. Taxonomy of Vascular Plants. The McMillan Company, New York.

Lindley, J. 1846. The Vegetable Kingdom. Vol. I. Bradbury and Evans, London.

Liston, A., L .H. Rieseberg, and T. S. Elias. 1989. Morphological stasis and molecular divergence in the intercontinental disjunct genus Datisca (Datiscaceae). Aliso 12: 525-542.

Liston, A., L. H. Rieseberg, and T. S. Elias. 1990. Functional androdioecy in the flowering plant Datisca glomerata. Nature 343: 641-642.

Maniatis, T., J. Sambrook, and E. F. Fritsch. 1989. Molecular Cloning. A Laboratory Manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Metcalf, R. L. and A. M. Rhodes. 1990. Coevolution of the Cucurbitaceae and Luperini (Coleoptera: Chrysomelidae): Basic and applied aspects. In: Biology and Utilization of the Cucurbitaceae. D. M. Bates, R. W. Robinson and C. Jeffrey, Eds. Cornell Univ. Press, Ithaca, New York.

Miziorko, H. M. and G. H. Lorimer. 1983. Ribulose-1,5-bisphosphate carboxylase-oxygenase. Ann. Rev. Biochem. 52: 507-535.

Mullin, B. C., S. M. Swensen, and P. Goetting-Minesky. 1990. Hypotheses for the evolution of actinorhizal symbioses. In: Nitrogen Fixation: Achievement and Objectives. Chapman and Hall, New York.

Mullis, K. B. and F. A. Faloona. 1987. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. Meth. Enzym. 155: 335-350.

Penny, D., M. D. Hendy, E. A. Zimmer, and R. K. Hamby. 1990. Trees from sequences: panacea or Pandora's box? Austral. Syst. Bot. 3: 21-38.

Pohlmann, J. 1975. Die Cucurbitacine in Bryonia alba und Bryonia dioica. Phytochemistry 14: 1587-1589.

Restivo, R. J., R. F. Bryan, and S. M. Kupchan. 1973. Stereochemistry of the cucurbitacins. Crystal and molecular structure and absolute configuration of datiscoside bis-(p-iodobenzoate) dihydrate. J. Chem. Soc. Perkin Trans. 2: 892-897.

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

Ritland, K. and M. T. Clegg. 1987. Evolutionary analysis of plant DNA sequences. Am. Nat. 130S: 74-100.

Saiki, R. K., S. J. Scharf, F. Faloona, K. B. Mullis, G. T. Horn, H. A. Erlich, and N. Arnheim. 1985. Enzymatic amplification of β-globin genomic sequences and restriction site analysis of sickle cell anemia. Science 230: 1350-1354.

Sanderson, M. J. and M. J. Donoghue. 1989. Patterns of variation in levels of homoplasy. Evolution 43: 1781-1795.

Sasamori, H., K. S. Reddy, M. P. Kirkup, J. Shabanowitz, D. G. Lynn, S. M. Hecht, K. A. Woode, R. F. Bryan, J. Campbell, W. S. Lynn, E. Egert, and G. M. Sheldrick. 1983. New cytotoxic principles from Datisca glomerata. J. Chem. Soc. Perkin Trans. 1: 1333-1347.

Swofford, D. L. 1991. PAUP: Phylogenetic Analysis Using Parsimony, Version 3.0s. Computer program distributed by the Illinois Natural History Survey, Champaign, IL.

_____ and G. J. Olsen. 1990. Phylogeny reconstruction. In: Molecular Systematics. D. M. Hillis and C. Moritz, Eds. Sinauer Associates, Inc., Sunderland, MA.

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

Thorne, R. F. 1976. A phylogenetic classification of the Angiospermae. Evol. Biol. 9: 35-106.

_____. 1992. An updated phylogenetic classification of the flowering plants. Aliso 13: 365-389.

vanSteenis, C. G. G. J. 1953. Datiscaceae. Flora Malesiana I(4): 382-387.

VITA

Susan Mildred Swensen was born in Grinnell, Iowa, on November 24, 1962. She attended elementary through high school in the Grinnell Community School System. In the fall of 1985, she entered Grinnell College and in May 1988, she received a Bachelor of Arts degree in Biology. In the fall of 1985 she entered the Botany Department at The University of Tennessee, Knoxville, as a graduate teaching assistant and in August 1988 she received a Master of Science degree in Botany. In September 1988 she entered the Plant Physiology and Genetics Program at The University of Tennessee, Knoxville, and received a Ph.D. in December 1992.

She is presently working as a postdoctoral researcher at the Rancho Santa Ana Botanic Gardens in Claremont, California.