

PARALLEL EVOLUTION OF GLUCOSINOLATE BIOSYNTHESIS INFERRED FROM CONGRUENT NUCLEAR AND PLASTID GENE PHYLOGENIES¹

JAMES E. RODMAN,^{2,6} PAMELA S. SOLTIS,³ DOUGLAS E. SOLTIS,³
KENNETH J. SYTSMAN,⁴ AND KENNETH G. KAROL⁵

²Division of Environmental Biology, National Science Foundation, Arlington, Virginia 22230;

³Department of Botany, Washington State University, Pullman, Washington 99164-4238;

⁴Department of Botany, University of Wisconsin, Madison, Wisconsin 53706-1381; and

⁵Laboratory of Molecular Systematics, Smithsonian Institution, Washington, DC 20560

The phytochemical system of mustard-oil glucosides (glucosinolates) accompanied by the hydrolytic enzyme myrosinase (β -thioglucosidase), the latter usually compartmented in special myrosin cells, characterizes plants in 16 families of angiosperms. Traditional classifications place these taxa in many separate orders and thus imply multiple convergences in the origin of this chemical defense system. DNA sequencing of the chloroplast *rbcL* gene for representatives of all 16 families and several putative relatives, with phylogenetic analyses by parsimony and maximum likelihood methods, demonstrated instead a single major clade of mustard-oil plants and one phylogenetic outlier. In a further independent test, DNA sequencing of the nuclear 18S ribosomal RNA gene for all these exemplars has yielded the same result, a major mustard-oil clade of 15 families (Akaniaceae, Bataceae, Brassicaceae, Bretschneideraceae, Capparaceae, Caricaceae, Gyrostemonaceae, Koerberliniaceae, Limnanthaceae, Moringaceae, Pentadiplandraceae, Resedaceae, Salvadoraceae, Tovariaceae, and Tropaeolaceae) and one outlier, the genus *Drypetes*, traditionally placed in Euphorbiaceae. Concatenating the two gene sequences (for a total of 3254 nucleotides) in a data set for 33 taxa, we obtain robust support for this finding of parallel origins of glucosinolate biosynthesis. From likely cyanogenic ancestors, the “mustard oil bomb” was invented twice.

Key words: Capparales s.l.; DNA sequencing; glucosinolates; phylogeny; rDNA (18S).

Mustard-oil glucosides (also named glucosinolates; Ettliger and Kjaer, 1968) are oxime-derived sulfur-containing compounds whose breakdown products include the familiar pungent principles of mustard, radish, and capers (Fenwick, Heaney, and Mullin, 1983). The compounds usually are accompanied in the plant by a hydrolytic enzyme, myrosinase (a β -thioglucoside glucohydrolase, E.C. 3.2.3.1), which may be compartmented in special myrosin cells (Fig. 1). Hypothesized to deter herbivores or pathogens, this “mustard oil bomb” (Lüthy and Matile, 1984) is characteristic of all Brassicaceae, including the genomic model *Arabidopsis*, but occurs as well in 15 other angiosperm families (Table 1). The phytochemical system of glucosinolates with accompanying myrosinase enzyme, the latter compartmented in myrosin cells, is believed to characterize all the species in these families, except for Euphorbiaceae, where *Drypetes* (interpreted to include *Guya* and *Putranjiva*) is the only established source of mustard oils (Ettliger and Kjaer,

1968; Rodman, 1981; Ettliger, 1987). Traditional classifications like Cronquist's (1981, 1988) place these 16 families in several widely separate taxonomic orders and thus imply multiple origins for the glucosinolates-with-myrosinase system. In turn, this multiple-origin viewpoint has fragmented the study of host fidelity and evolution by herbivores and pathogens adapted to glucosinolate-producing plants (Chew, 1988).

In his early attempt to evaluate taxonomic relationships of angiosperms, Dahlgren (1975, 1977) challenged orthodox classifications of mustard-oil plants by expanding the order Capparales to encompass nearly all families of glucosinolate taxa. Later he retreated from this position (Dahlgren, 1980, 1983; Dahlgren, Rosendal-Jensen, and Nielsen, 1981), and emphasized the considerable morphological and habitat diversity among these taxa. In a cladistic analysis of 90 morphological and phytochemical characters, Rodman (1991b) found weak support for Dahlgren's (1975) expanded Capparales. To test Dahlgren's radical reclassification of glucosinolate taxa with an independent molecular data set, representatives of all 16 families were sequenced for the chloroplast *rbcL* gene and analyzed with parsimony and maximum likelihood methods (Gadek et al., 1992; Rodman et al., 1993, 1994, 1996a). Comparisons were made with numerous putative relatives, sequenced as part of a large collaborative analysis (Chase et al., 1993; Price and Palmer, 1993). With one exception, all mustard-oil taxa united into a single major clade or lineage (Fig. 2), nested within a large “rosid” plexus (Chase et al., 1993). The sole exception is the genus *Drypetes*, which is usually placed within the spurge family Euphorbiaceae and is affiliated with other

¹ Manuscript received 28 July 1997; revision accepted 19 November 1997.

The authors thank Ihsan Al-Shehbaz, Ray Cranfield, Robert Hirano, Hugh Iltis, Ching-I Peng, Robert Price, and Christopher Quinn for plant samples, Kandis Elliot for graphics, and Peter Raven and George Johnson for the idea behind our Fig. 1 cartoon. Support from the Andrew Mellon Foundation (to P. S. Soltis and D. E. Soltis) and the National Science Foundation (DEB 9307000 to D. E. Soltis and DEB 9407270 to K. J. Sytsma), advice from Elizabeth Zimmer, and the facilities of the Smithsonian Institution's Laboratory of Molecular Systematics are appreciated. The authors also thank reviewers Neil Harriman and Kathleen Kron for advice.

⁶ Author for correspondence (e-mail: jrodman@nsf.gov); agency citation is solely for purposes of identification.

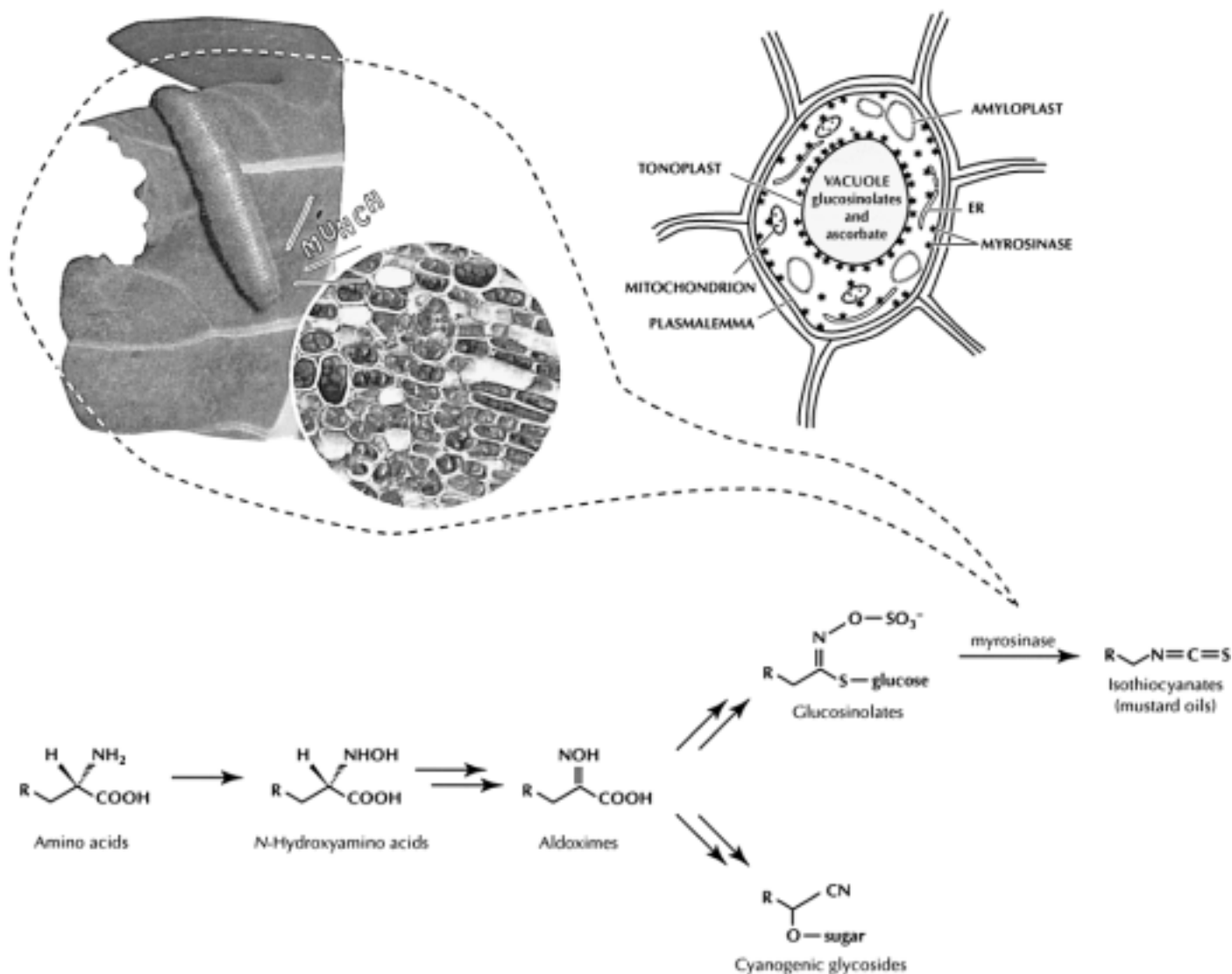


Fig. 1. Biosynthesis of glucosinolates (mustard-oil glucosides) and of cyanogenic glycosides from amino acids via aldoximes (Du et al., 1995), and hydrolysis of glucosinolates to isothiocyanates (mustard oils; not shown are glucose, sulfate, other potential breakdown products) catalyzed by myrosinases, a family of thioglucoside glucohydrolases (Xue et al., 1992, 1995) associated with cell membranes (Lüthy and Matile, 1984) and concentrated in myrosin cells (stained cells in insert: Werker and Vaughan, 1974; Höglund, Lenman, and Rask, 1992). Herbivore feeding is hypothesized to bring enzyme and substrate into contact, thereby releasing toxic mustard oils (Ehrlich and Raven, 1965), analogous to cyanide release from cyanogenic plants (Saupe, 1981).

rosids in lineages thus far only sparsely sampled for gene sequence diversity (Chase et al., 1993; Conti, Litt, and Sytsma, 1996; Soltis et al., 1997). A maximum likelihood analysis (see Felsenstein, 1985, 1995) yielded a single fully resolved tree, identical in topology with the corresponding portion in Fig. 2 of the major mustard-oil clade. For this analysis, the data set was reduced to 20 taxa by paring the outgroups to just two, *Ailanthus* and *Gossypium*, and deleting *Drypetes*. Congruence between the results from morphology and from DNA sequences of the *rbcl* gene gave strong support to Dahlgren's (1975) radically expanded Capparales (Rodman et al., 1996a). In turn, that congruence reinforced the hypothesis of dual or parallel origins of mustard-oil glucoside biosynthesis (Rodman et al., 1993).

The considerable diversity of floral form and growth

habit among the Capparales thus circumscribed by Dahlgren (1975; amended by Rodman et al., 1996a) has elicited skepticism from specialists in systematic anatomy and morphology (Tobe and Peng, 1990; Tobe and Raven, 1991, 1995; Carlquist, 1996; Doweld, 1996). Moreover, reliance upon a single gene for phylogenetic reconstruction can be theoretically problematic (Pamilo and Nei, 1988; Doyle, 1992). Phenomena of introgression, lineage sorting, and mistaken orthology can confound results (Doyle, 1992), although none of these is likely to obtain in the case of family-level comparisons of *rbcl* among mustard-oil taxa (Rodman et al., 1993). Testing phylogenetic conclusions with additional independent molecular data constitutes an appropriate corroboration or refutation. We initiated this second test with DNA sequence data from a nuclear gene, the 18S ribosomal RNA gene

TABLE 1. Cronquist's (1981) classification of the 16 known mustard-oil families (see Ettliger, 1987), with exemplar species used in sequencing of the plastid *rbcL* and nuclear 18S ribosomal RNA genes.^a

Family, order exemplar species	GenBank accession or voucher	
	<i>rbcL</i>	18S nrDNA
Akaniaceae (1/1), Sapindales <i>Akania bidwillii</i> (Hogg) Mabb.	L12568	<i>Fernando & Quinn s.n.</i> (UNSW 21606)
Bataceae (1/2), Batales <i>Batis maritima</i> L.	L22438	U42504
Brassicaceae (350/3000), Capparales <i>Arabidopsis thaliana</i> (L.) Heynh.	U91966	X16077
<i>Brassica hirta</i> Moench.	NA	X17062
<i>Brassica juncea</i> (L.) Coss.	unpublished	NA
Bretschneideraceae (1/1), Sapindales <i>Bretschneidera sinensis</i> Hemsl.	M95753	<i>Leu & Lin 726</i> (WIS)
Capparaceae (45/800), Capparales <i>Capparis hastata</i> Jacq.	M95754	NA
<i>Capparis sandwichiana</i> DC.	NA	<i>Iltis 30502</i> (WIS)
<i>Cleome hassleriana</i> Chodat	M95755	U42511
<i>Setchellanthus caeruleus</i> T. Brandeg.	U41455	<i>Iltis & Lasseigne 100</i> (WIS)
Caricaceae (4/30), Violales <i>Carica papaya</i> L.	M95671	U42514
Euphorbiaceae (317/8000), Euphorbiales <i>Drypetes roxburghii</i> (Wall.) Hurusawa	M95757	U42534
Gyrostemonaceae (5/17), Batales <i>Gyrostemon tepperi</i> (H. Walter) A. George	L22440	<i>Thomson 2243</i> (MO)
(Koeberliniaceae) (1/1), Capparales <i>Koeberlinia spinosa</i> Zucc.	L14600	U42512
Limnanthaceae (2/11), Geraniales <i>Floerkea proserpinacoides</i> Willd.	L12679	U42784
<i>Limnanthes douglasii</i> R. Br.	L14700	<i>Price s.n.</i> (IND)
Moringaceae (1/10), Capparales <i>Moringa oleifera</i> Lam.	L11359	U42786
(Pentadiplandraceae) (1/1), Capparales <i>Pentadiplandra brazzeana</i> Baill.	U38533	<i>Hart 180</i> (MO)
Resedaceae (6/70), Capparales <i>Reseda alba</i> L.	M95756	<i>Price s.n.</i> (IND)
Salvadoraceae (3/12), Celastrales <i>Salvadora angustifolia</i> Turill	U38532	<i>Zarucchi 7554</i> (F)
Tovariaceae (1/2), Capparales <i>Tovaria pendula</i> Ruiz & Pav.	M95758	<i>Smith & Smith 1831</i> (WIS)
Tropaeolaceae (3/92), Geraniales <i>Tropaeolum majus</i> L.	L14706	L31796

^a Number of known genera/species after family name (from Cronquist, 1981). *Koeberlinia* and *Pentadiplandra* are placed within Capparaceae by Cronquist but are accorded family rank by other taxonomists; *Koeberlinia*, although not known to produce mustard oils, is reported to have myrosin cells (Gibson, 1979); *Setchellanthus* is placed within Capparaceae but is under consideration for family rank (H. H. Iltis, personal communication). *Drypetes* is the only known mustard-oil producing genus in Euphorbiaceae (Ettliger, 1987). Unless qualified in the table, the same species was sequenced for both genes. Voucher information includes collector name, number, and Index Herbariorum code. NA = not applicable.

(Rodman et al., 1996b; see Nickrent and Soltis, 1995), in comparisons with a wide range of putative relatives of mustard-oil plants (Soltis et al., 1997). Here we report the completion of that test, with sequence data for 18S nrDNA from representatives of all 16 families of mustard-oil plants. Combining the two gene data sets, we find robust support for Dahlgren's (1975) radical classification of an expanded Capparales, encompassing all glucosinolate taxa except *Drypetes*. Near relatives of the two glucosinolate lineages include plants that biosynthesize cyanogenic glycosides, a finding that reinforces Kjaer's (1973) hypothesis that pathways of cyanoglycoside metabolism have been recruited and modified to yield mustard-oil glucosides.

MATERIALS AND METHODS

Materials—For all 16 families of mustard-oil plants and for numerous putative relatives suggested by traditional classifications of these

taxa (see Rodman, 1991a), exemplar species or placeholders were sequenced for the two genes (Table 1). Methods, vouchers, and GenBank accessions for all *rbcL* sequences used here have been described (Gadek et al., 1992; Chase et al., 1993; Rodman et al., 1993, 1994, 1996a). Several species of Brassicaceae have been sequenced for *rbcL*, and parsimony analyses show robust support for monophyly of the family and sister-group affiliation with cleomoid Capparaceae (Price, Palmer, and Al-Shehbaz, 1994). We chose one sequence from a species of the type genus as placeholder in the *rbcL* analysis (Fig. 2) and added sequences from *Arabidopsis thaliana* for the 18S (Fig. 3) and combined genes study (Fig. 4). Three species from Gyrostemonaceae have been analyzed for *rbcL* and the family shown to be monophyletic (Rodman et al., 1994); we sequenced *Gyrostemon tepperi* for 18S nrDNA as placeholder for our analyses. Likewise, three species from Salvadoraceae have been analyzed for *rbcL* and the family shown to be monophyletic (Rodman et al., 1996a); we sequenced *Salvadora angustifolia* as placeholder for the 18S and combined genes analyses. Methods, vouchers, and GenBank accessions for outgroup sequences of 18S nrDNA are given in Soltis and Soltis (1997) and in Soltis et al. (1997).

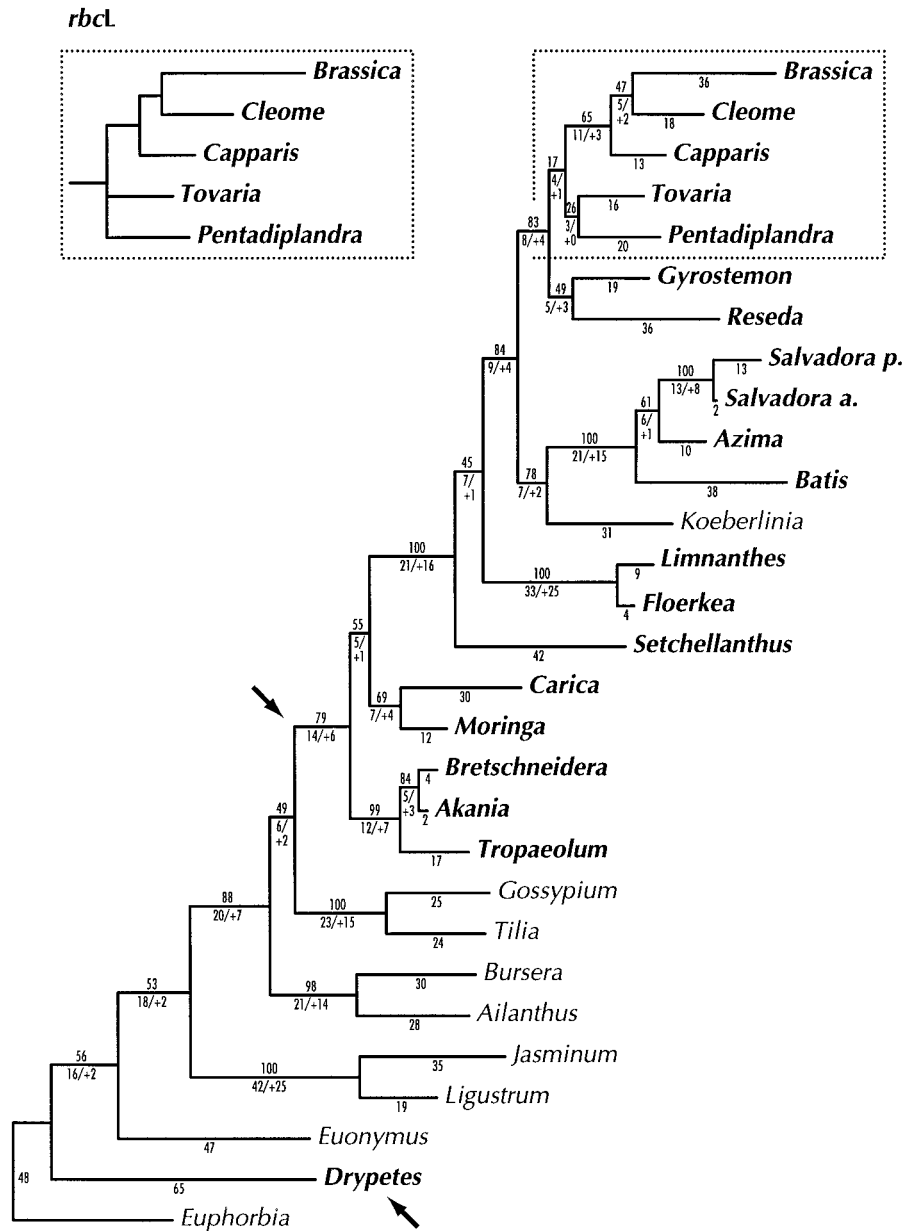


Fig. 2. Relationships of mustard-oil taxa (names in boldface) and putative relatives inferred from *rbcl* gene sequences of exemplar species; length 1035 steps with consistency index 0.55 (0.47 without autapomorphies at 878 steps) and retention index 0.61. One of three equally parsimonious trees (inset shows strict consensus of the variable portion) with bootstrap support indicated above branches. Lengths and decay steps (where applicable) are indicated below branches; *Euphorbia* arbitrarily designated as root. *Salvadora a.* and *p.* are *S. angustifolia* and *S. persica*. Arrows mark the two mustard-oil clades. A maximum likelihood analysis of 18 mustard-oil taxa (excluding *Drypetes*) plus two outgroups, *Ailanthus* and *Gossypium* (K. G. Karol, unpublished data), yielded the same topology, with \ln likelihood = -5738.675 .

DNA protocols—For the 18S ribosomal RNA gene, amplification and sequencing protocols are as described by Soltis and Soltis (1997) and Soltis et al. (1997). We used GenBank accessions X16077, X17062, X16600, L28137, L24145, L24046, L24398, and L31796 for *Arabidopsis thaliana*, *Brassica hirta*, *Euonymus alatus*, *Francoa sonchifolia*, *Gossypium hirsutum*, *Malpighia coccigera*, *Morus alba*, and *Tropaeolum majus*, respectively. For 26 other taxa (14 mustard-oil species of Table 1, except *Gyrostemon tepperi*, *Pentadiplandra brazzeana*, and *Reseda alba*, and for outgroups *Acer rubrum*, *Bombax ceiba*, *Brexia madagascariensis*, *Celtis yunnanensis*, *Clarkia xantiana*, *Citrus aurantium*, *Euphorbia pulcherrima*, *Geranium* sp., *Koelreuteria paniculata*,

Pilea cadierei, *Punica granatum*, and *Turnera ulmifolia*), we conducted cycle sequencing with primers described by Soltis et al. (1997), in many cases with the DNA extracts used in previous *rbcl* analyses. For these taxa, sequences were obtained with an ABI[®] Model 373A automated sequencer, with assembly facilitated by the Sequencher[®] program (Soltis et al., 1997). For three mustard-oil taxa (the *Gyrostemon*, *Pentadiplandra*, and *Reseda* samples noted above), manual sequencing with described primers (Soltis et al., 1997) was performed. Alignment of the 1814-base sequences was done visually, with single base insertions relative to the sequence for soybean, *Glycine max* (Eckenrode, Arnold, and Meagher, 1985), made at base positions 233 (T in *Koerberlinia*),

18S nrDNA

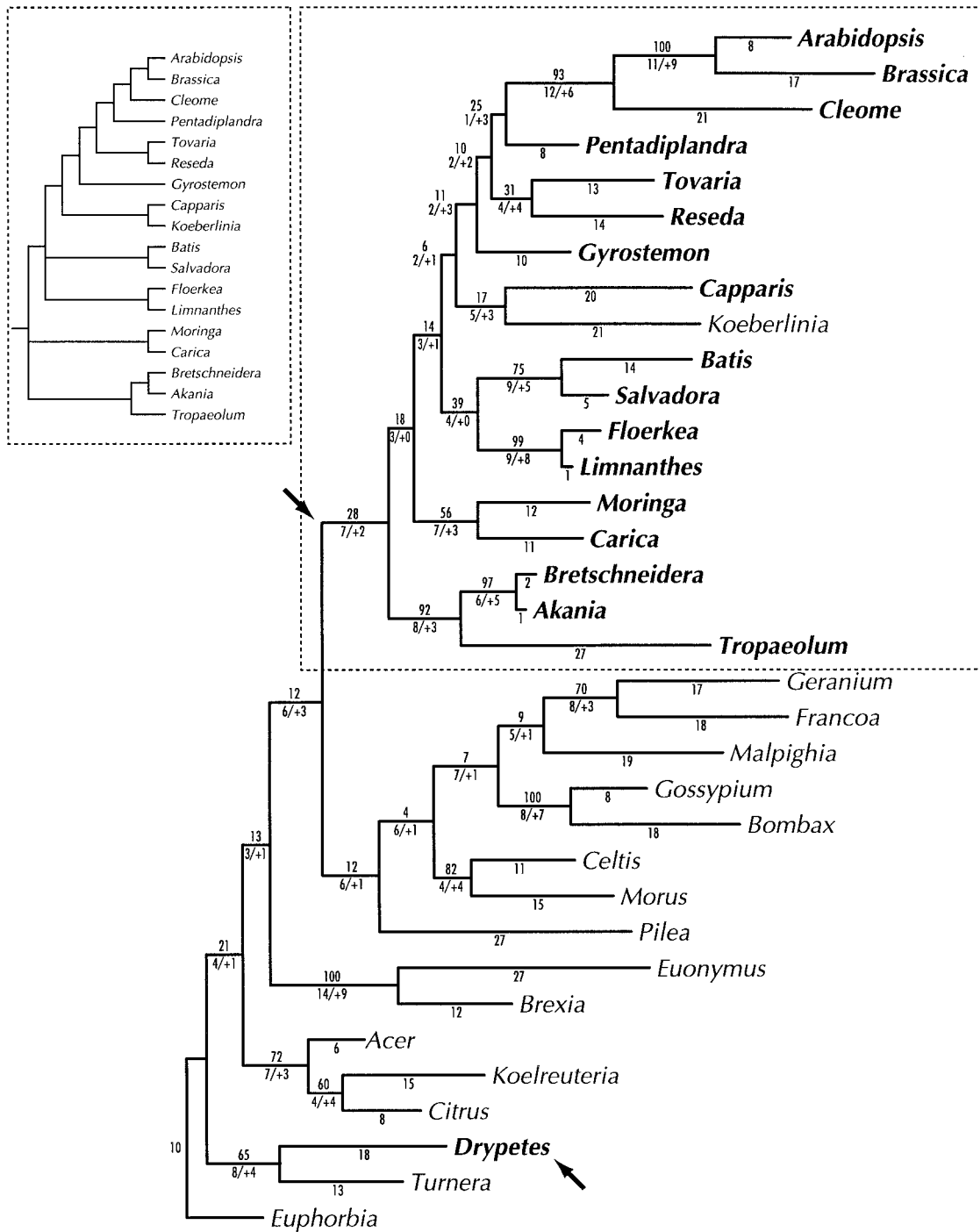


Fig. 3. Relationships of mustard-oil taxa (names in boldface) and putative relatives inferred from nuclear 18S ribosomal RNA gene sequences of exemplar species; length 636 steps with consistency index 0.53 (0.40 without autapomorphies at 498 steps) and retention index 0.51. One of five equally parsimonious trees (inset shows strict consensus of the variable portion) with bootstrap support indicated above branch, lengths and decay steps below. Branch lengths proportional to number of nucleotide substitutions; *Euphorbia* arbitrarily designated as root. Arrows mark the two mustard-oil clades.

498 (C in *Arabidopsis*, T in *Gyrostemon*), 669 (A in *Batis*, C in *Cleome* and *Drypetes*), 784 (G in *Salvadora*, T in *Sechellanthus*), 1185 (T in all Fig. 3 taxa), and 1360 (G in *Sechellanthus*). Deletions affecting multiple taxa were placed at positions 542 (T in *Glycine*, gap in all other Fig. 3 taxa), 1364 (gap in *Batis* and *Salvadora*), 1369 (gap in

Floerkea, *Limnanthes*, and *Tovaria*), and 1500 (gap in *Arabidopsis*, *Brassica*, and *Cleome*). Over a broader range of taxa (Soltis et al., 1997), alignments are problematic in the region of helices E10-1, 17, E23-1, and 43 relative to the sequence for soybean, for which a secondary structure has been proposed (Eckenrode, Arnold, and Meagher,

1985; Nickrent and Soltis, 1995). We encountered little difficulty, however, in aligning the entire 18S sequence with the taxa under study and in creating a matrix of 1852 positions (including gaps). Nucleotide substitutions among the mustard-oil taxa, relative to the *G. max* sequence, were divided equally between loop and stem regions (46 and 54%, respectively).

Data analysis—Aligned sequences were exported to PAUP 3.1.1 (Swofford, 1993) for phylogenetic reconstruction using parsimony methods. The Heuristic Search, TBR branch swapping, MULTIPARS, and ACCTRAN options were employed using the random taxon addition option (100 replicates with Steepest Descent). Characters were run unordered, equivalent to equal weighting at all nucleotide sites (Fitch, 1971). To assess robustness of linkages, bootstrapping over 100 replicates (Felsenstein, 1985) and decay analyses (Bremer, 1988) were performed with PAUP programs. Bootstrap analyses (with Heuristic Search, simple taxon addition, and TBR branch swapping) were summarized in majority-rule consensus trees. Decay analyses were performed by invoking Topological Constraints (with Heuristic Search, simple taxon addition, TBR branch swapping) and saving trees that breached the constraint, or by relaxing the parsimony criterion by five steps, summarizing results by strict consensus, and repeating sequentially with the Filter Trees option down to one step.

Initial analyses of the 18S nrDNA data set produced numerous (>60) equally parsimonious trees, and a “jackknife” approach (Lanyon, 1985) was used to find taxa causing instability. The mustard-oil plant *Setchellanthus* was found to cause most of the instability (in Lanyon’s sense), and this taxon was dropped from the 18S analysis reported in Fig. 3. *Setchellanthus* was reinstated when the two gene sequences were concatenated for the combined genes analysis. For outgroups, the number and choice reflect a compromise between representation of numerous putative relatives suggested by traditional classifications (see Cronquist, 1981, 1988) and the maintenance of a computationally tractable data set. Bootstrap analyses, in particular, become intractable beyond 40 or so taxa. Congruence between topologies derived from the two separate genes, *rbcL* and 18S RNA, was assessed with Rohlf’s (1982) consensus index as implemented in PAUP.

In constructing an exemplar for two ingroup (mustard-oil) taxa, *Brassica* and *Capparis*, for the combined genes study, it was necessary to concatenate sequence data from two different species in each genus (Table 1). Although not an ideal strategy, the compromise relies upon the small differences observed in gene sequence between closely related species, for both *rbcL* (Doebley et al., 1990; Hudson et al., 1990) and 18S nrDNA (Nickrent and Soltis, 1995). Also, in constructing an exemplar for four of the 13 outgroups, it was necessary to concatenate sequences from species of different genera for the family Rutaceae (*Coelonema pulchellum* for *rbcL* and *Citrus aurantium* for nrDNA) and the family Sapindaceae (*Cupaniopsis anacardioides* for *rbcL* and *Koelreuteria paniculata* for nrDNA), and from species of different families for the order Myrtales (*Lythrum hyssopifolia* for *rbcL* and *Punica granatum* for nrDNA) and the order Passiflorales (*Passiflora quadrangularis* for *rbcL* and *Turnera ulmifolia* for nrDNA). For all the rest, sequences for the two genes were available from the same species. Such compromises are not ideal and would be unacceptable in seeking to resolve relationships within the outgroup clades. Effects on resolution of mustard-oil taxa are likely to be negligible because nucleotide differences within the outgroup families and orders are smaller than differences between them and the major mustard-oil clade.

RESULTS

The phylogenetic signal expected from 18S nrDNA is weaker than that from *rbcL* as a consequence of fewer variable and hence informative sites for the taxa under study (see Nickrent and Soltis, 1995). For the *rbcL* data set of 29 taxa of mustard-oil plants and putative relatives

(Fig. 2), 20.4% of the nucleotide sites (286 of 1402 bases) proved informatively variable; three equally parsimonious trees were found, with a consistency index of 0.47 (not counting autapomorphies) and a retention index of 0.61 and with many clades enjoying robust bootstrap and decay-step support. For the same mustard-oil taxa (excluding only *Setchellanthus*) along with putative relatives in a data set of 34 taxa (Fig. 3), only 8.1% of the nrDNA sites (147 of 1814 bases) show informative variation. Five equally parsimonious trees were recovered, with lower consistency (0.40) and retention indices (0.51), and with most clades showing only weak support. Despite this weaker support, the main results from 18S nrDNA match those from *rbcL*: a single major mustard-oil clade and a second, unrelated lineage comprising the genus *Drypetes*. Intervening between the two are numerous lineages of rosoid taxa (see Chase et al., 1993; Soltis et al., 1997), including Geraniales, Malvales, and Sapindales.

A broad analysis of 18S nrDNA sequences for 223 species of flowering plants yielded over 5000 equally parsimonious trees, the strict consensus of which nonetheless preserved a major mustard-oil clade (represented by eight families in the data set) separate from the genus *Drypetes* (Soltis et al., 1997). The sheer size of that data set precluded tests of robustness such as bootstrap and decay analysis. With all 16 families of mustard-oil plants represented in our data set, the large number of taxa also precludes an exhaustive search that would guarantee retrieval of all most-parsimonious trees, but tests of robustness can be employed to assess the strength of linkages (Fig. 3). Strong support from nrDNA is evident for only a few mustard-oil clades: the triads *Cleome* + (*Arabidopsis* + *Brassica*) and *Tropaeolum* + (*Akania* + *Bretschneidera*), and the sister taxa *Batis* + *Salvadora* and *Floerkea* + *Limnanthes*. The major mustard-oil clade itself is weakly supported (28% bootstrap, two decay steps). Branch swapping, performed with the MacClade program (Maddison and Maddison, 1992), measured increases in tree length (decay in parsimony) when traditional taxonomic associations are enforced, such as *Akania* linked with Sapindales (11 extra steps), Limnanthaceae with Geraniales (seven extra), or *Salvadora* with Celastrales (four extra). Although evidence from 18S nrDNA for monophyly of a major mustard-oil clade is weak, the phylogenetic signal remains strikingly congruent with that from the chloroplast *rbcL* gene.

For a subset of 25 species common to the two analyses, *rbcL* (Fig. 2) and 18S nrDNA (Fig. 3), we computed Rohlf’s (1982) consensus index to be 0.677, which is moderately high. Where 18S-based clades (Fig. 3) within the major mustard-oil lineage are incongruent with *rbcL* clades, the former are weakly supported and collapse to polytomies within a few decay steps. We conclude that the two data sets do not yield strongly incongruent results (see Bull et al., 1993). Adopting a “combined evidence” approach (Hoot, Culham, and Crane, 1995; Kron, 1996; Rodman et al., 1996a), therefore, we concatenated sequences for the nuclear and plastid genes where these had been analyzed from the same species or from two species of the same genus. The combined genes data set included 20 mustard-oil taxa and 13 putative relatives, and yielded four equally parsimonious trees (Fig. 4) with a consis-

Combined genes tree

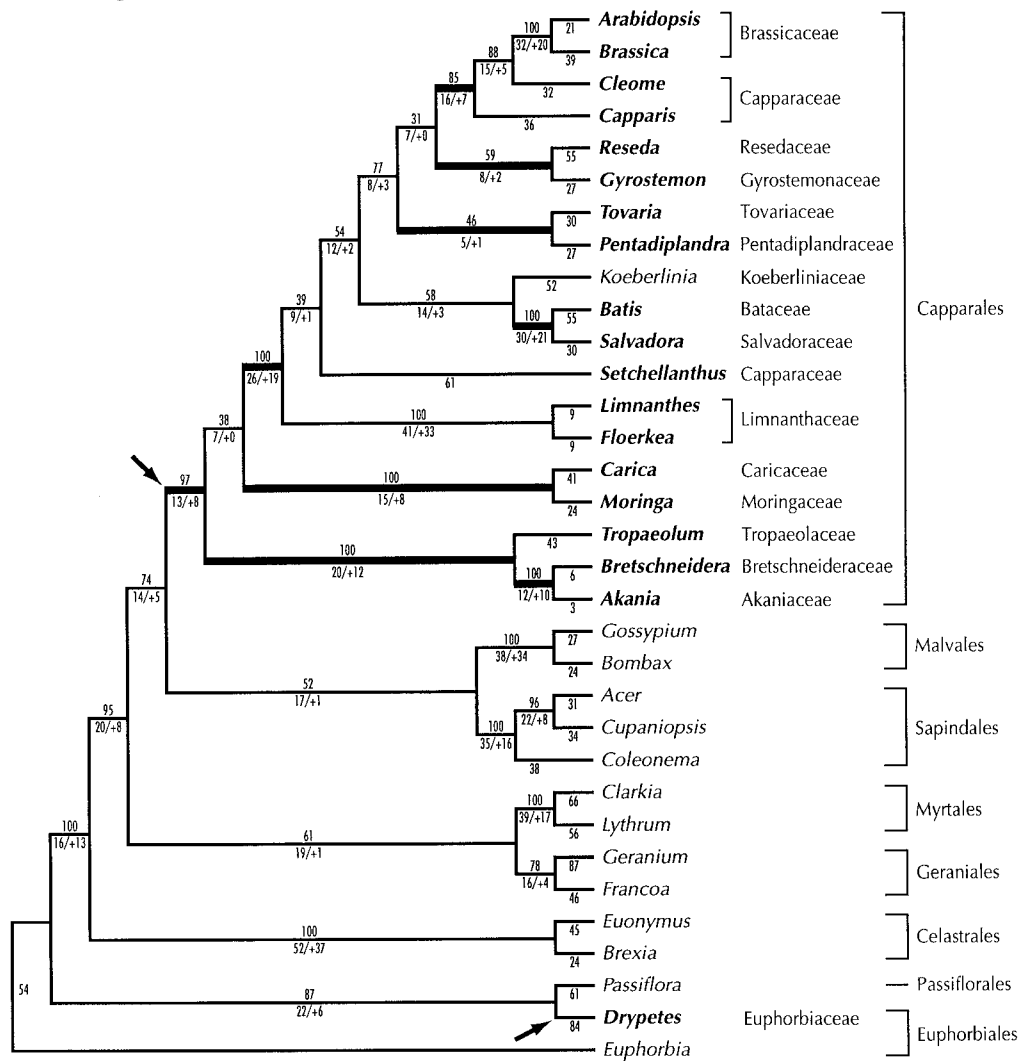


Fig. 4. Relationships of mustard-oil taxa (genus names in boldface) and putative relatives inferred from combined *rbcL* and 18S nrDNA sequences; length 1877 steps with consistency index 0.52 (0.42 without autapomorphies at 1580 steps) and retention index 0.55. Four equally parsimonious trees were found; the two branches with +0 decay values collapse to trichotomies in the strict consensus tree. The tree with the lowest f-value (Swofford and Maddison, 1987) is shown here with bootstrap values indicated above the branch, lengths and decay steps below. *Euphorbia* arbitrarily designated as root. Arrows mark the major mustard-oil clade and the second, unrelated lineage of *Drypetes*. Branches with increased bootstrap or decay-step values, over those found in the single gene analyses (Figs. 2, 3), indicated by thick lines.

cy index of 0.42 and a retention index of 0.55. In this combined genes tree, *rbcL* and 18S rDNA sequence data reinforce each other to produce a major clade of mustard-oil taxa, excluding *Drypetes*, with robust bootstrap (97%) and decay (eight steps) support. Several internal branches as well gain enhanced support from the combined genes; this is evident in bootstrap values higher than those found in the single gene analyses and, for clades already with high bootstrap values, in decay values higher than previously found. Clades with this strengthened support include Tropaeolaceae + (Akaniaceae + Bretschneideraceae) with 100% bootstrap and 12 decay steps; Akaniaceae + Bretschneideraceae, 100% bootstrap and 10 decay steps; Bataceae + Salvadoraceae, 100% bootstrap and 21 decay steps; Caricaceae + Moringaceae, 100% bootstrap and eight decay steps; and the sister group to

this latter clade, with 100% bootstrap and 19 decay steps, a lineage diagnosed by a C-terminal extension or tail to the *rbcL* molecule (Rodman et al., 1994, 1996a). One clade suffers some diminution of support: *Cleome* + (*Arabidopsis* + *Brassica*), from 93% bootstrap and six decay steps in the 18S analysis, to 88% and five steps.

DISCUSSION

Three lines of evidence now are concordant in confirming Dahlgren's (1975) radical union of a morphologically diverse order Capparales. The first comprised a set of morphological and phytochemical characters, many of these cited over the years in the variant classifications of these plants (Rodman, 1991a, b). The second comprised a data set of DNA sequences from the chloroplast gene

rbcL for exemplars of all 16 mustard-oil families and numerous putative relatives (Rodman et al., 1996a). The third comprises a comparable data set of DNA sequences from the nuclear 18S ribosomal RNA gene for many of the same exemplars and putative relatives. All three data sets were analyzed with the same cladistic methods. Fifteen families (Akaniaceae, Bataceae, Brassicaceae, Bretschneideraceae, Capparaceae, Caricaceae, Gyrostemonaceae, Koerberliniaceae, Limnanthaceae, Moringaceae, Pentadiplandraceae, Resedaceae, Salvadoraceae, Tovariaceae, and Tropaeolaceae) constitute a robustly supported monophyletic group, marked by the syndrome of mustard-oil glucosides and myrosinase enzyme. This clade is coincident with Dahlgren's (1975) early circumscription of an expanded order Capparales when amended to include Akaniaceae and Caricaceae (Gadek et al., 1992; Rodman et al., 1993, 1994, 1996a). Anatomical or morphological similarities previously considered indicators of close relationship for certain mustard-oil families, for example, Akaniaceae and Bretschneideraceae with Sapindales (Tobe and Peng, 1990; Thorne, 1992; Tobe and Raven, 1995; Doweld, 1996), Bataceae and Gyrostemonaceae with Centrospermae or with Sapindales (Cronquist, 1981; Tobe and Raven, 1991), Limnanthaceae and Tropaeolaceae with Geraniales (Cronquist, 1988), Salvadoraceae with Celastrales or Oleales (Thorne, 1992), must now be considered morphological convergences or retentions of primitive features (Rodman, 1991b). Within this major mustard-oil clade, the Capparaceae, even when pared of *Koerberlinia*, *Pentadiplandra*, and *Tovaria*, remain paraphyletic (see also Judd, Sanders, and Donoghue, 1994) or, if including *Setchellanthus*, polyphyletic.

The sole other mustard-oil taxon, the genus *Drypetes*, is phylogenetically distant from the major mustard-oil clade and is nested among rosoid lineages still sparsely sampled for DNA sequence diversity (Conti, Litt, and Sytsma, 1996; Wurdack and Chase, 1996; Soltis et al., 1997). The few species of this large (~150 species) pantropical genus that have been screened for mustard-oil glucosides give positive results (Ettlinger and Kjaer, 1968). Ultrastructural studies reveal protein-accumulating phloem cells similar to developing myrosin cells (Jørgensen, Behnke, and Mabry, 1977), although tests for ascorbate-activated myrosinase enzymes proved negative (Ettlinger, 1987). Unfortunately, no studies have been reported on the enzymology of mustard-oil glucoside biosynthesis in *Drypetes*, in contrast to the situation in several capparalean taxa (Du et al., 1995). Despite limited species sampling, and with severe constraints on homology assessment, the evidence suggests that glucosinolate biosynthesis and myrosin cell formation in *Drypetes* are remarkably similar to those of the Capparales. Hence, the congruent nuclear and plastid gene phylogenies require an inference of parallel evolution for this phytochemical defense system. The "mustard oil bomb" (Lüthy and Matile, 1984) was invented twice.

Addressing the evolutionary origins of glucosinolates, Kjaer (1973) emphasized similarities in biosynthesis between cyanogenic glycosides, cyanide-releasing compounds that are widespread in angiosperms (Saupe, 1981), and mustard-oil glucosides, which are restricted, we now find, to two lineages. He hypothesized that glucosinolate biosynthesis evolved through the recruitment

and modification of cyanogen biosynthesis, with the significant addition of sulfur to the molecule. His hypothesis is reinforced by the recent report (Du et al., 1995) of a common enzymological formation of intermediary aldoximes from precursor amino acids (Fig. 1) demonstrated both in cyanogenic plants and in mustard-oil taxa. Other workers, however, have challenged the universality of that oxime-forming pathway on other enzymatic grounds (Bennett et al., 1995). Compatible with Kjaer's (1973) hypothesis, the combined-genes phylogeny (Fig. 4) links taxa of cyanogenic plants like Euphorbiaceae, Passifloraceae, and Sapindaceae (Saupe, 1981) as close relatives of the two mustard-oil clades. As phylogenetic analyses incorporate denser sampling among these rosoid lineages (Conti et al., 1996; Gadek et al., 1996; Wurdack and Chase, 1996), we should gain a clearer indication of the sister groups to the two glucosinolate clades. That phyletic framework, in turn, can guide the choice of likely cyanogenic relatives for comparative sequencing of biosynthetic enzymes, to test Kjaer's (1973) hypothesis of recruitment and modification. Alternative targets for such comparative sequencing are the gene families for the hydrolyzing enzymes unique to the two chemical defense systems, the cyanogen β -glycosidases (Henrissat, 1993; Møller and Poulton, 1993) and the myrosinases or β -thioglucosidases (Xue et al., 1992; Thangstad et al., 1993). Already, an impressive 44% similarity in amino acid sequence has been reported (Lenman et al., 1993) between a cyanogen β -glycosidase in clover (*Trifolium repens*, Leguminosae) and a myrosinase in turnip (*Brassica napus*). The molecular tools are at hand for dissecting and characterizing the parallel origins of glucosinolate biosynthesis, in the genus *Drypetes* and in Dahlgren's expanded order Capparales.

LITERATURE CITED

- BENNETT, R. N., A. J. HICK, G. W. DAWSON, AND R. M. WALLSGROVE. 1995. Glucosinolate biosynthesis. Further characterization of the aldoxime-forming microsomal monooxygenases in oilseed rape leaves. *Plant Physiology* 109: 299–305.
- BREMER, K. 1988. The limits of amino acid sequence data in angiosperm phylogenetic reconstruction. *Evolution* 42: 795–803.
- BULL, J. J., J. P. HUELSENBECK, C. W. CUNNINGHAM, D. L. SWOFFORD, AND P. J. WADDELL. 1993. Partitioning and combining data in phylogenetic analysis. *Systematic Biology* 42: 384–397.
- CARLQUIST, S. 1996. Wood anatomy of Akaniaceae and Bretschneideraceae: a case of near-identity and its systematic implications. *Systematic Botany* 21: 607–616.
- CHASE, M. W., ET AL. 1993. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbcL*. *Annals of the Missouri Botanical Garden* 80: 528–580.
- CHEW, F. S. 1988. Are crucivores controlled by mustard oils? No. In H. G. Cutler [ed.], *Biologically active natural products*, 155–181. American Chemical Society, Washington, DC.
- CONTI, E., A. LITT, AND K. J. SYTSMA. 1996. Circumscription of Myrtales and their relationships to other rosoids: evidence from *rbcL* sequence data. *American Journal of Botany* 83: 221–233.
- CRONQUIST, A. 1981. An integrated system of classification of flowering plants. Columbia University Press, New York, NY.
- . 1988. The evolution and classification of flowering plants. Second edition. New York Botanical Garden, Bronx, NY.
- DAHLGREN, R. 1975. A system of classification of the angiosperms to be used to demonstrate the distribution of characters. *Botaniska Notiser* 128: 119–147.
- . 1977. A commentary on a diagrammatic presentation of the

- angiosperms in relation to the distribution of character states. *Plant Systematics and Evolution* (Supplement 1): 253–283.
- . 1980. A revised system of classification of the angiosperms. *Botanical Journal of the Linnean Society* 80: 91–124.
- . 1983. General aspects of angiosperm evolution and macrosystematics. *Nordic Journal of Botany* 3: 119–149.
- , S. ROSENDAL-JENSEN, AND B. J. NIELSEN. 1981. A revised classification of the angiosperms with comments on correlation between chemical and other characters. In D. A. Young and D. S. Seigler [eds.], *Phytochemistry and angiosperm phylogeny*, 149–204. Praeger, New York, NY.
- DOEBLEY, J., M. DURBIN, E. M. GOLEBERG, M. T. CLEGG, AND D. P. MA. 1990. Evolutionary analysis of the large subunit of carboxylase (*rbcl*) nucleotide sequence among the grasses (Gramineae). *Evolution* 44: 1097–1108.
- DOWELD, A. B. 1996. The systematic relevance of fruit and seed anatomy and morphology of *Akania* (Akaniaceae). *Botanical Journal of the Linnean Society* 120: 379–389.
- DOYLE, J. J. 1992. Gene trees and species trees: molecular systematics as one-character taxonomy. *Systematic Botany* 17: 144–163.
- DU, L., J. LYKKESFELDT, C. E. OLSEN, AND B. A. HALKIER. 1995. Involvement of cytochrome P450 in oxime production in glucosinolate biosynthesis as demonstrated by an *in vitro* microsomal enzyme system isolated from jasmonic acid-induced seedlings of *Sinapis alba* L. *Proceedings of the National Academy of Sciences, USA* 92: 12505–12509.
- ECKENRÖDE, V. K., J. ARNOLD, AND R. B. MEAGHER. 1985. Comparison of the nucleotide sequence of soybean 18S rDNA with the sequences of other small-subunit rRNAs. *Journal of Molecular Evolution* 21: 259–269.
- EHRlich, P. R., AND P. H. RAVEN. 1965. Butterflies and plants: a study in coevolution. *Evolution* 18: 586–608.
- ETTLINGER, M. G. 1987. Systematic distribution and biochemical properties of glucosinolates (mustard oil glucosides). *XIV International Botanical Congress Abstracts*, 280. Botanical Museum, Berlin-Dahlem.
- , AND A. KJAER. 1968. Sulfur compounds in plants. *Recent Advances in Phytochemistry* 1: 59–144.
- FELSENSTEIN, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 42: 795–803.
- . 1995. PHYLIP manual. Version 3.57c. University of California Herbarium, Berkeley, CA.
- FENWICK, G. R., R. K. HEANEY, AND W. J. MULLIN. 1983. Glucosinolates and their breakdown products in food and food plants. *CRC Critical Reviews in Food Science and Nutrition* 18: 123–201.
- FITCH, W. M. 1971. Toward defining the course of evolution: minimum change for a specific tree topology. *Systematic Zoology* 20: 406–416.
- GADEK, P. A., C. J. QUINN, J. E. RODMAN, K. G. KAROL, E. CONTI, R. A. PRICE, AND E. S. FERNANDO. 1992. Affinities of the Australian endemic Akaniaceae: new evidence from *rbcl* sequences. *Australian Systematic Botany* 5: 717–724.
- , E. S. FERNANDO, C. J. QUINN, S. B. HOOT, T. TERRAZAS, M. C. SHEAHAN, AND M. W. CHASE. 1996. Sapindales: molecular delimitation and infraordinal groups. *American Journal of Botany* 83: 802–811.
- GIBSON, A. C. 1979. Anatomy of *Koeberlinia* and *Canotia* revisited. *Madroño* 26: 1–12.
- HENRISSAT, B. 1993. New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochemical Journal* 293: 781–788.
- HÖGLUND, A.-S., M. LENMAN, AND L. RASK. 1992. Myrosinase is localized to the interior of myrosin grains and is not associated to the surrounding tonoplast membrane. *Plant Science* 85: 165–170.
- HOOT, S. B., A. CULHAM, AND P. R. CRANE. 1995. The utility of *atpB* gene sequences in resolving phylogenetic relationships: comparison with *rbcl* and 18S ribosomal DNA sequences in the Lardizabaleaceae. *Annals of the Missouri Botanical Garden* 82: 194–207.
- HUDSON, G. S., J. D. MAHON, P. A. ANDERSON, M. J. GIBBS, M. R. BADGER, T. J. ANDREWS, AND P. R. WHITFIELD. 1990. Comparisons of *rbcl* genes for the large subunit of ribulose-bisphosphate carboxylase from closely related C₃ and C₄ plant species. *Journal of Biological Chemistry* 265: 808–814.
- JØRGENSEN, L. B., H.-D. BEHNKE, AND T. J. MABRY. 1977. Protein-accumulating cells and dilated cisternae of the endoplasmic reticulum in three glucosinolate-containing genera: *Azorella*, *Capparis*, *Drypetes*. *Planta* 137: 215–224.
- JUDD, W. S., R. W. SANDERS, AND M. J. DONOGHUE. 1994. Angiosperm family pairs: preliminary phylogenetic analyses. *Harvard Papers in Botany* 5: 1–51.
- KJAER, A. 1973. The natural distribution of glucosinolates: a uniform group of sulfur containing glucosides. In G. Bendz and J. Santesson [eds.], *Chemistry in botanical classification*, 229–234. Academic Press, New York, NY.
- KRON, K. A. 1996. Phylogenetic relationships of Empetraceae, Epacridaceae, Ericaceae, Monotropaceae, and Pyrolaceae: evidence from nuclear ribosomal 18S sequence data. *Annals of Botany* 77: 293–303.
- LANYON, S. M. 1985. Detecting internal inconsistencies in distance data. *Systematic Zoology* 34: 397–403.
- LENMAN, M., A. FALK, J. XUE, AND L. RASK. 1993. Characterization of a *Brassica napus* myrosinase pseudogene: myrosinases are members of the BGA family of β -glycosidases. *Plant Molecular Biology* 21: 463–474.
- LÜTHY, B., AND P. H. MATILE. 1984. The mustard oil bomb: rectified analysis of the subcellular organisation of the myrosinase system. *Biochemie und Physiologie der Pflanzen* 179: 5–12.
- MADDISON, W. P., AND D. R. MADDISON. 1992. MacClade, analysis of phylogeny and character evolution, version 3.01. Sinauer, Sunderland, MA.
- MØLLER, B. L., AND J. E. POULTON. 1993. Cyanogenic glycosides. In P. J. Lea [ed.], *Methods in plant biochemistry*, vol. 9, 183–207. Academic Press, San Diego, CA.
- NICKRENT, D. L., AND D. E. SOLTIS. 1995. A comparison of angiosperm phylogenies from nuclear 18S rDNA and *rbcl* sequences. *Annals of the Missouri Botanical Garden* 82: 208–234.
- PAMILO, P., AND M. NEI. 1988. Relationships between gene trees and species trees. *Molecular Biology and Evolution* 5: 568–583.
- PRICE, R. A., AND J. D. PALMER. 1993. Phylogenetic relationships of the Geraniaceae and Geraniales from *rbcl* sequence comparisons. *Annals of the Missouri Botanical Garden* 80: 661–671.
- , ———, AND I. A. AL-SHEHBAZ. 1994. Systematic relationships of *Arabidopsis*: a molecular and morphological perspective. In E. M. Meyerowitz and C. R. Somerville [eds.], *Arabidopsis*, 7–19. Cold Spring Harbor Press, New York, NY.
- RODMAN, J. E. 1981. Divergence, convergence, and parallelism in phytochemical characters: the glucosinolate-myrosinase system. In D. A. Young and D. S. Seigler [eds.], *Phytochemistry and angiosperm phylogeny*, 43–79. Praeger, New York, NY.
- . 1991a. A taxonomic analysis of glucosinolate-producing plants, part 1: Phenetics. *Systematic Botany* 16: 598–618.
- . 1991b. A taxonomic analysis of glucosinolate-producing plants, part 2: Cladistics. *Systematic Botany* 16: 619–629.
- , K. G. KAROL, R. A. PRICE, E. CONTI, AND K. J. SYTSMA. 1994. Nucleotide sequences of *rbcl* confirm the Capparalean affinity of the Australian endemic Gyrostemonaceae. *Australian Systematic Botany* 7: 57–69.
- , ———, ———, AND K. J. SYTSMA. 1996a. Molecules, morphology, and Dahlgren's expanded order Capparales. *Systematic Botany* 21: 289–307.
- , R. A. PRICE, K. KAROL, E. CONTI, K. J. SYTSMA, AND J. D. PALMER. 1993. Nucleotide sequences of the *rbcl* gene indicate monophyly of mustard oil plants. *Annals of the Missouri Botanical Garden* 80: 686–699.
- , P. S. SOLTIS, D. E. SOLTIS, AND K. J. SYTSMA. 1996b. Dual origin of mustard oil biosynthesis inferred from congruent nuclear 18S ribosomal RNA and plastid *rbcl* gene phylogenies. *American Journal of Botany* 83: 188 (Abstract).
- ROHLF, F. J. 1982. Consensus indices for comparing classifications. *Mathematical Biosciences* 59: 131–144.
- SAUPE, S. G. 1981. Cyanogenic compounds and angiosperm phylogeny. In D. A. Young and D. S. Seigler [eds.], *Phytochemistry and angiosperm phylogeny*, 80–116. Praeger, New York, NY.
- SOLTIS, D. E., AND P. S. SOLTIS. 1997. Phylogenetic relationships in Saxifragaceae sensu lato: a comparison of topologies based on 18S

- rDNA and *rbcL* sequences. *American Journal of Botany* 84: 504–522.
- , ET AL. 1997. Angiosperm phylogeny inferred from 18S ribosomal DNA sequences. *Annals of the Missouri Botanical Garden* 84: 1–49.
- SWOFFORD, D. L. 1993. PAUP, phylogenetic analysis using parsimony, version 3.1.1. Illinois Natural History Survey, Champaign, IL.
- , AND W. P. MADDISON. 1987. Reconstructing ancestral character states under Wagner parsimony. *Mathematical Biosciences* 87: 199–229.
- THANGSTAD, O. P., P. WINGE, H. HUSEBYE, AND A. BONES. 1993. The myrosinase (thioglucoside glucohydrolase) gene family in Brassicaceae. *Plant Molecular Biology* 23: 511–524.
- THORNE, R. F. 1992. Classification and geography of the flowering plants. *Botanical Review* (Lancaster) 58: 225–348.
- TOBE, H., AND C. PENG. 1990. The embryology and relationships of *Bretschneidera* (Bretschneideraceae). *Journal of the Linnean Society, Botany* 103: 139–152.
- , AND P. H. RAVEN. 1991. The embryology and relationships of Gyrostemonaceae. *Australian Systematic Botany* 4: 407–420.
- , AND ———. 1995. Embryology and relationships of *Akania* (Akaniaceae). *Botanical Journal of the Linnean Society* 118: 261–274.
- WERKER, E., AND J. G. VAUGHAN. 1974. Anatomical and ultrastructural changes in aleurone and myrosin cells of *Sinapis alba* during germination. *Planta* 116: 243–255.
- WURDACK, K. J., AND M. W. CHASE. 1996. Molecular systematics of Euphorbiaceae *sensu lato* using *rbcL* sequence data. *American Journal of Botany* 83: 203 (Abstract).
- XUE, J., M. LENMAN, A. FALK, AND L. RASK. 1992. The glucosinolate-degrading enzyme myrosinase in Brassicaceae is encoded by a gene family. *Plant Molecular Biology* 18: 387–398.
- , M. JØRGENSEN, U. PIHLGREN, AND L. RASK. 1995. The myrosinase gene family in *Arabidopsis thaliana*: gene organization, expression and evolution. *Plant Molecular Biology* 27: 911–922.