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

JAMES E. RODMAN,2,6 PAMELA S. SOLTIS,3 DOUGLAS E. SOLTIS,3 KENNETH J. SYTSMA,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 (Akanthaceae, Bataceae, Brassicaceae, Bretschneideraceae, Capparaceae, Caricaceae, Gymnostemonaceae, Koeberliniaceae, Limnanthaceae, Moringaceae, Pentadactylidaeae, 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: Ettlinger 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 glucohydro-lase, 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 Guia and Putranjiva) is the only established source of mustard oils (Ettlinger and Kjaer, 1968; Rodman, 1981; Ettlinger, 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...
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) radially 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...
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 sequenced for rbcL, and the family shown to be monophyletic (Rodman et al., 1996a); we sequenced Gyrostemon tepperi for 18S nrDNA as placeholder for our analyses. Likewise, three species from Salvadora-caceae have been analyzed for rbcL and the family shown to be monophyletic (Rodman et al., 1994); 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).

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

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unpublished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

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. Itis, personal communication). Drypetes is the only known mustard-oil producing genus in Euphorbiaceae (Ettlinger, 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.
Fig. 2. Relationships of mustard-oil taxa (names in boldface) and putative relatives inferred from \textit{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. Branch lengths proportional to number of nucleotide substitutions; \textit{Euphorbia} arbitrarily designated as root. \textit{Salvadora} \textit{a.} and \textit{p.} are \textit{S. angustifolia} and \textit{S. persica}. Arrows mark the two mustard-oil clades. A maximum likelihood analysis of 18 mustard-oil taxa (excluding \textit{Drypetes}) plus two outgroups, \textit{Ailanthus} and \textit{Gossypium} (K. G. Karol, unpublished data), yielded the same topology, with In likelihood = $-5738.675$.

\textbf{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 \textit{Arabidopsis thaliana}, \textit{Brassica hirta}, \textit{Euonymus alatus}, \textit{Francoa sonchifolia}, \textit{Gossypium hirsutum}, \textit{Malpighia coccigera}, \textit{Morus alba}, and \textit{Tropaeolum majus}, respectively. For 26 other taxa (14 mustard-oil species of Table 1, except \textit{Gyrostromen tepperi}, \textit{Pentadiplandra brazzeana}, and \textit{Reseda alba}, and for outgroups \textit{Acer rubrum}, \textit{Bombax ceiba}, \textit{Brevia madagascariensis}, \textit{Celtis yunnanensis}, \textit{Clarkia xantiana}, \textit{Citrus aurantium}, \textit{Euphorbia palcherrima}, \textit{Geranium} sp., \textit{Koelreuteria paniculata}, \textit{Pilea cadierei}, \textit{Punica granatum}, and \textit{Turnera ulmifolia}), we conducted cycle sequencing with primers described by Soltis et al. (1997), in many cases with the DNA extracts used in previous \textit{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 \textit{Gyrostromen}, \textit{Pentadiplandra}, and \textit{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, \textit{Glycine max} (Eckenrode, Arnold, and Meagher, 1985), made at base positions 233 (T in \textit{Koelreuteria}),
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 *Gyrosteremon*), 669 (A in *Batis*, C in *Cleome* and *Drypetes*), 784 (G in *Salvadora*, T in *Setchellanthus*), 1185 (T in all Fig. 3 taxa), and 1360 (G in *Setchellanthus*). 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,
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, MULPARS, 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 (Feltenstein, 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 major-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, Brasica 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 (\( Coleonema pulchellum \) for \( rbcL \) and \( Citrus aurantium \) for nrDNA) and the family Sapindaceae (\( Cupaniopsis anacardioides \) for \( rbcL \) and Koeleria 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 \( Turnerera udmijofolia \) 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 taxa. Congruence between topologies derived from the two separate genes was found with a consistency index of 0.61 and with many 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 rosid 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 preserves 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 precludes 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 + (\( Arbidopsis + Brassica \)) 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 Salvadorea with Celastraceae (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 consisten-
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 (Akaniaeae, Bataceae, Brassicaceae, Breschneideraeae, Capparaceae, Caricaceae, Gynostemonaceae, Koeberliniaceae, 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 Akaniaeae 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, Akaniaeae and Breschneideraeae with Sapindales (Tobe and Peng, 1990; Thorne, 1992; Tobe and Raven, 1995; Doweld, 1996). Bataceae and Gynostemonaceae with Centrospermae or with Sapindales (Cronquist, 1981; Tobe and Raven, 1991). Limnanthaceae and Tropaeolaceae with Geraniaceae (Cronquist, 1988). Salvadoraceae with Celastrales or Oleales (Thorne, 1992), must now be considered morphological convergences or reten-

pions of primitive features (Rodman, 1991b). Within this major mustard-oil clade, the Capparaceae, even when pared of Koeberlinia, 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 rosid lineages still sparsely sampled for DNA sequence diversity (Conti et al., 1996; Wurdack and Chase, 1996; Solits 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 plroom cells similar to developing myrosin cells (Jorgensen, 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 cappara-
lean 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 phytotoxic chemical defense system. The “mustard oil bomb” (Luthy 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 glu-
cosinolate 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 aldo-

xes 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, Passifloraeae, and Sapindaceae (Saupe, 1981) as close relatives of the two mustard-oil clades. As phylogenetic analyses incorporate denser sampling among these rosid 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 phy-
letic framework, in turn, can guide the choice of likely cyanogenic relatives for comparative sequencing of bio-
synthetic enzymes, to test Kjaer’s (1973) hypothesis of recruitment and modification. Alternative targets for such comparative sequencing are the gene families for the hy-
drolyzing enzymes unique to the two chemical defense systems, the cyanogen β-glucosidases (Henriissat, 1993; Möller and Poulton, 1993) and the myrosinases or β-thioglycosidases (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-glucosidase in clover (Trifolium repens, Leguminosae) and a myrosinase in turnip (Brasica 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


Carquisset, S. 1996. Wood anatomy of Akaniaeae and Breschneider-


Conti, E., A. Litt, and K. J. Sytsma. 1996. Circumscription of Myr-


———. 1988. The evolution and classification of flowering plants. Sec-


———. 1988. A commentary on a diagrammatic presentation of the


