8 Chloroplast DNA and Phylogenetic Studies in the Asteridae

RICHARD G. OLMSTEAD[1], ROBERT K. JANSEN[2], HELEN J. MICHAELS[3], STEPHEN R. DOWNIE[1], AND JEFFREY D. PALMER[1]

[1]*Department of Biology, Indiana University, Bloomington, IN 47405, USA*
[2]*Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT 06268, USA*
[3]*Department of Biology, University of Michigan, Ann Arbor, MI 48109, USA*

I. Introduction

The quest for understanding angiosperm phylogeny has led plant biosystematists to pursue many avenues of research. An abbreviated chronology of plant systematics since Linnaeus indicates that the study of floral structure and plant anatomy predominated through the 18th, 19th, and early 20th centuries, followed by the introduction of cytology and palynology in the mid-twentieth century, secondary plant chemicals in the 1960's, and protein variation in the 70's. This diverse array of approaches has in common one important link; all involve the study of phenotypes arising from the underlying genetic material through various biochemical and developmental pathways. Only in this decade have advances in molecular biology enabled plant systematists to readily examine the genetic material itself, DNA, to investigate phylogenetic relationships.

Of the three genomes in plants (nuclear, chloroplast, and mitochondrial), the chloroplast genome has proved to be the most useful for phylogenetic analyses to date. Its presence in high copy number, often 5,000 genomes per cell, makes it relatively easy to extract, and even total DNA extracts from 2-3 grams of fresh leaf material are rich enough in chloroplast DNA (cpDNA) for most systematic purposes. The chloroplast genome is small and varies little in size in green land plants, 120-217 kb, with much of the size variation accounted for by difference in the size of a large inverted repeat (Fig. 1). The chloroplast reproduces clonally with little or no recombination in most plants and is inherited maternally in most flowering

BIOLOGICAL APPROACHES AND EVOLUTIONARY TRENDS IN PLANTS ISBN 0-12-402960-4

plants. The structure and function of the chloroplast genome have been the subject of several recent reviews (Whitfield and Bottomley, 1983; Palmer, 1985; Zurawski and Clegg, 1987). The low rate of nucleotide substitution in cpDNA relative to nuclear DNA and animal mtDNA (Wolfe *et al.*, 1987), combined with a highly conserved gene content and arrangement, makes both restriction site mapping (using chloroplast probes from even distantly related species) and nucleotide sequencing feasible approaches for comparative studies. However, the much larger size and greater gene content of the nuclear genome and the apparently slower rate of nucleotide substitution (Wolfe *et al.*, 1987) in the plant mitochondrial genome suggest that each will have valuable phylogenetic utility in the future.

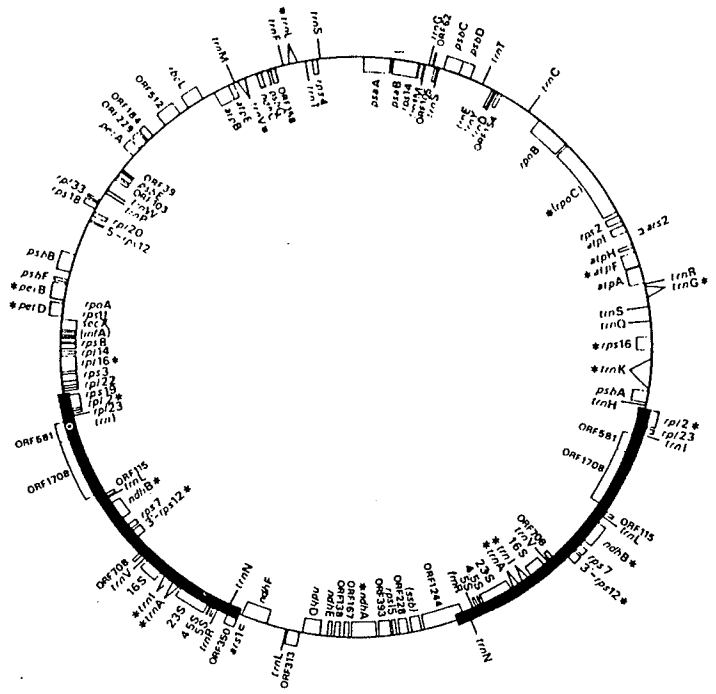

FIG. 1 Circular map of the cpDNA chromosome in tobacco (*Nicotiana tabacum*) showing the densely packed arrangement of genes and open reading frames. The total genome size is 155,844 base pairs and is organized into large (87 Kb) and small (18 Kb) single copy regions separated by two copies of the inverted repeat (25 Kb) indicated by the dark bars. Genes transcribed clockwise are shown on the inside of the circle; those transcribed in the reverse direction are on the outside.

Evolutionary change in cpDNA can be categorized into two distinct classes, nucleotide substitution (i.e. point mutation) and structural rearrangements (i.e. insertions, deletions, and inversions). Both classes can be exploited for phylogenetic inference. In our research on the systematics of the subclass Asteridae, we are assessing variation from both sources to address questions concerning phylogenetic relationships at different levels.

II. Nucleotide Substitutions

Point mutations resulting in nucleotide substitutions are the most common source of DNA variation among species. Nucleotide substitutions can be detected by restriction site analysis when mutations occur in restriction endonuclease recognition sites and by direct sequence comparison of homologous sequences such as genes, introns, or conserved spacer regions. Most mutations observed in restriction sites will be at sites in non-coding regions of the genome and at "silent"sites within protein genes, where nucleotide substitution rates are highest, making restriction site analysis most appropriate for phylogenetically closely related organisms. Rates of nucleotide substitution in genes and some introns are lower than in non-coding regions due to functional constraints on sequence evolution, making DNA sequencing of specific genes more appropriate for comparisons at greater phylogenetic distance.

A. Restriction site analysis

The analysis of restriction site variation involves the digestion of the cpDNA by restriction enzymes followed by the comparison of DNA fragments separated by gel electrophoresis. Restriction site mutations are detected by the presence or absence of fragments on the gel. Two small fragments indicate the presence of a restriction site in one sample, whereas a single larger fragment in another sample comprised of the two smaller fragments indicates its absence. Early studies involved the visual inspection of the gels themselves, however multiple mutational differences and the presence of length mutations (insertions/deletions) between samples create problems for interpreting homologous fragments. This approach is sufficient only for closely related taxa cut with restriction enzymes yielding few (5-20) fragments and requires purified cpDNA. The preferred method today is the filter hybridization approach, in which the DNA is transferred from the gel to nylon filters, which are then probed successively with cloned cpDNA fragments from a completely mapped reference genome. By this method more divergent taxa can be examined and more frequently cutting enzymes used (this reduces the number of enzymes needed to sample the same number of bases)with little problem in determining homology among fragments, because only a small portion of the genome is examined with each successive round of hybridization. For our current cpDNA analysis of the Solanaceae we have constructed a set of 40 clones, ranging in size from 2-5 kb, of tobacco to use as probes in hybridization, from larger clones kindly provided by M. Sugiura. In selecting restriction enzymes for a systematic survey, it is worth noting that enzymes with A/T-rich recognition sites (eg. DraI, TTTAAA) are more likely to

provide a greater proportion of variable sites, because the non-coding regions of land plant cpDNA are A/T rich.

Carrying out systematic research using cpDNA requires that attention be paid to many of the same considerations important in studies using "conventional" data. In studies of large groups it will not be possible to sample all taxa, so representative specimens must be chosen carefully. It is important to include more than one outgroup, so that character polarity can be assessed and the hypothesis of monophyly for the study group can be tested adequately. The conservative nature of cpDNA evolution suggests that single individuals will adequately represent most taxa at the generic-level and above (unless doubts exist concerning the monophyly of the group). Chloroplast DNA polymorphism within species is often very low (Banks and Birky, 1985; Michaels, unpublished data), but can be considerable (Soltis *et al.*, 1989) and may extend beyond species limits among very closely related species, thereby creating conflicting phylogenetic inferences (Doyle *et al.*, 1989). Reconstructing phylogeny from any data source requires that homologs be identified for the variable characters observed. The best way to assess homology of restriction sites is to construct restriction maps for each taxon for each enzyme used. Invariant sites provide reference points for aligning maps even though they do not provide phylogenetic information. Likewise, sites which differ in only a single taxon are phylogenetically uninformative. Only those sites present in two or more taxa and missing in at least two other taxa are phylogenetically informative. The phylogenetic reconstruction is then carried out using a cladistic analysis based on either Wagner parsimony (site gains and losses treated equally) or Dollo parsimony (single gain permitted for each site, but multiple losses allowed). For studies at the intrageneric level, little conflict is expected in the data and either method of parsimony is likely to produce the same tree (e.g. Palmer and Zamir, 1982; Sytsma and Gottlieb, 1986). However, as more divergent taxa are surveyed, more conflict in the data will arise and Dollo parsimony will be preferred based on the premise that parallel restriction site gains are expected to be much less likely than parallel site losses (DeBry and Slade, 1985; Jansen and Palmer, 1988).

Two examples of interspecific studies in the Solanaceae will serve to illustrate the advantages and limitations of cpDNA analysis among congeneric species. *Lycopersicon* was the subject of the first cladistic analysis of cpDNA restriction site variation (Palmer and Zamir, 1982). Included were eight species of *Lycopersicon* and two species of *Solanum* as outgroups. The cpDNAs were cut with 25 restriction enzymes and 39 variable sites were observed, of which 14 were phylogenetically informative. The resulting cladistic analysis (Fig. 2) suggests that all of the restriction site variation is consistent with the most parsimonious tree except for a single site, which is implied by the analysis to have been lost in parallel in two lineages, thus giving a consistency index (Kluge and Farris, 1969) of 0.93. This extremely high level of consistency within the cpDNA data suggests that the resulting tree provides a reliable estimate of phylogenetic relationships within *Lycopersicon*. The cpDNA phylogeny confirms that *Solanum pennellii* belongs in *Lycopersicon*

and that red fruits are a derived feature within *Lycopersicon*, since the three species bearing red fruits included in this study, *L. esculentum*, *L. pimpinellifolium*, and *L. cheesmanii*, form a monophyletic group. At the same time, however, the data are insufficient to completely resolve relationships within *Lycopersicon* as evidenced by two internal trichotomies and identical cpDNA for *L. chilense* and three accessions of *L. peruvianum* for the restriction sites surveyed. Also, cpDNA polymorphism is observed within the widely distributed and morphologically variable species, *L. peruvianum*, for which six accessions were examined.

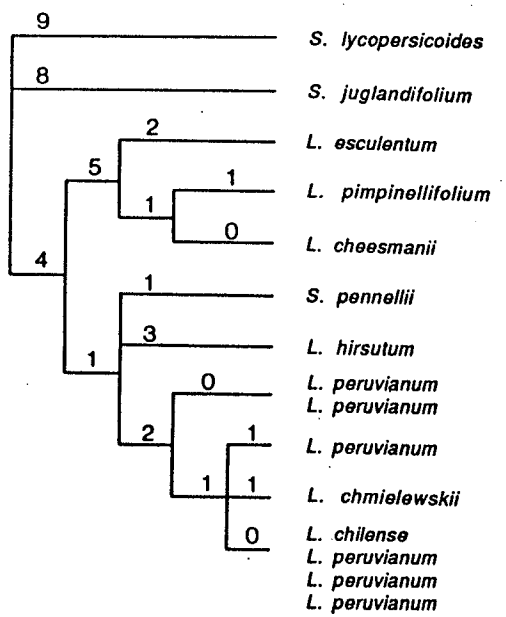

FIG. 2 Chloroplast DNA phylogeny of eight species of *Lycopersicon* (including *Solanum pennellii*). Numbers indicate the number of restriction site mutations defining each branch.

A more recent study of the genus *Nicotiana* (Olmstead and Palmer, unpublished data) investigates 21 species of *Nicotiana*, with one species of *Lycopersicon* and three species of *Petunia* included as outgroups. The analysis of restriction site variation in the large single copy region of the chloroplast genome yielded 187 variable sites, of which 108 were informative. A consensus tree of the twelve most parsimonious trees is shown in Fig.3. This analysis differs from that of *Lycopersicon* in having a greater number of taxa and informative restriction sites, but is similar in having highly consistent data for a study of its size (CI = 0.80) and in yielding an incompletely resolved tree. Note that in both of these studies Wagner and Dollo parsimony produce identical trees. In *Nicotiana* the incomplete resolution of the cpDNA tree carries an interesting implication for the evolution of that genus

in Australia. Relationships among the six species of *Nicotiana* from Australia, *N. velutina, N. rotundifolium, N. megalosiphon, N. excelsior, N. gossei,* and *N. exigua,* cannot be resolved on the basis of cpDNA, because so little variation exists, suggesting a very recent radiation there. Also, the polyploid hybrid species, *N. tabacum,* has identical cpDNA with that of one of its putative progenitors, *N. sylvestris,* confirming the identity of the latter as the maternal hybrid parent. The association of highly consistent data with incomplete resolution of phylogeny emerges as a pattern in many studies of cpDNA of congeneric species (Palmer and Zamir, 1982; Palmer *et al.*, 1983; Hosaka *et al.*, 1984; Sytsma and Schaal 1985; Perl-Treves and Galun, 1985; Coates and Cullis, 1987; Doebley *et al.*, 1987).

The conclusions drawn from interspecific studies suggest that much more divergent chloroplast genomes may be mapped and that their restriction site variation may be used effectively for phylogeny reconstruction. With increasing evolution-

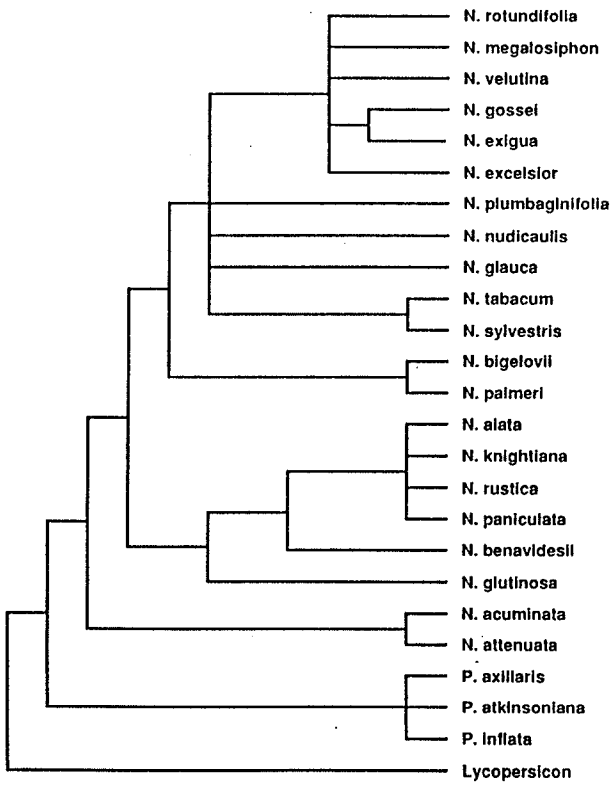

FIG. 3 Chloroplast DNA consensus tree of 21 species of *Nicotiana* with *Lycopersicon esculentum* and three species of *Petunia* as outgroups.

ary divergence among taxa in a cpDNA survey, two predictions can be made: 1) restriction site mutations will be more numerous and will enable complete resolution of phylogenetic relationships and 2) a greater number of inconsistencies will arise in the cpDNA data, particularly parallel losses of restriction sites. The first study undertaken to examine intrafamilial relationships focused on the family Asteraceae (Jansen and Palmer, 1988; Jansen *et al.*, submitted, a). A restriction site analysis using 11 enzymes was carried out on 57 genera representing 5 tribes. The bootstrap method of Felsenstein (1985) was used in conjunction with Dollo parsimony to analyze the large (927 variable sites and 328 informative sites) and less consistent (CI=0.39) data set produced by this analysis (Fig. 4). The bootstrap analysis provides a means of assessing the support for groups within the tree as indicated by the number of bootstrap replicates which support each group on the tree. As predicted on the basis of studies at the species-level, there are a sufficient number of informative restriction site mutations to provide a completely resolved estimate of phylogeny. However, also as predicted, the diminished consistency of the data, relative to data from studies at the species-level, results in conflicting estimates of phylogeny for some groups as reflected in the low bootstrap figures on some branches of the tree. The consistency index of 0.39 (note that Dollo parsimony will always produce lower CI values than Wagner parsimony) is low relative to studies among congeneric species cited above. However, other considerations, including the distribution of inconsistent characters (Jansen, et al., submitted, a) and an inverse correlation between consistency index and number of taxa in phylogenetic analyses in general (Sanderson and Donoghue, 1989; Archie, 1989), suggest that the cpDNA analysis is still a strong estimate of phylogenetic relationships in the Asteraceae.

The cpDNA phylogeny of the Asteraceae provides a test of the several hypotheses of tribal relationships in the Asteraceae that have been advanced by systematists studying the family, along with the predictions of those hypotheses regarding which tribe retains ancestral elements of the family. The results identify the tribe Mutisieae as the ancestral tribe in the family, finally settling a long-standing debate. However, the results go further to show that the Mutisieae is not a monophyletic tribe and that the subtribe Barnedesiinae is the sister group to the rest of the family (Jansen and Palmer, 1988; Jansen et al., submitted, b). In addition, a rigorous cladistic analysis of the Asteraceae using non-molecular characters (Bremer, 1987) is available for comparison with the cpDNA analysis (Jansen et al., submitted, b). The correct tribal placement of problem genera, whose tribal affinities have been obscure due to unique morphological attributes, often can be made readily based on cpDNA relationships (Keeley and Jansen, 1989; Jansen et al., submitted, b).

A second important Asteridae family, the Solanaceae, is currently the subject of an extensive restriction site survey (Olmstead and Palmer, unpublished data) including 133 species and 11 restriction enzymes. Unlike the Asteraceae, there has been no rigorous analysis of phylogenetic relationships in the Solanaceae. The cpDNA

analysis will provide the first reliable estimate of tribal relationships in the family and, with a broad representation of the large genus *Solanum* (31 species included), will furnish information concerning the relationship of that important genus to many other allied genera.

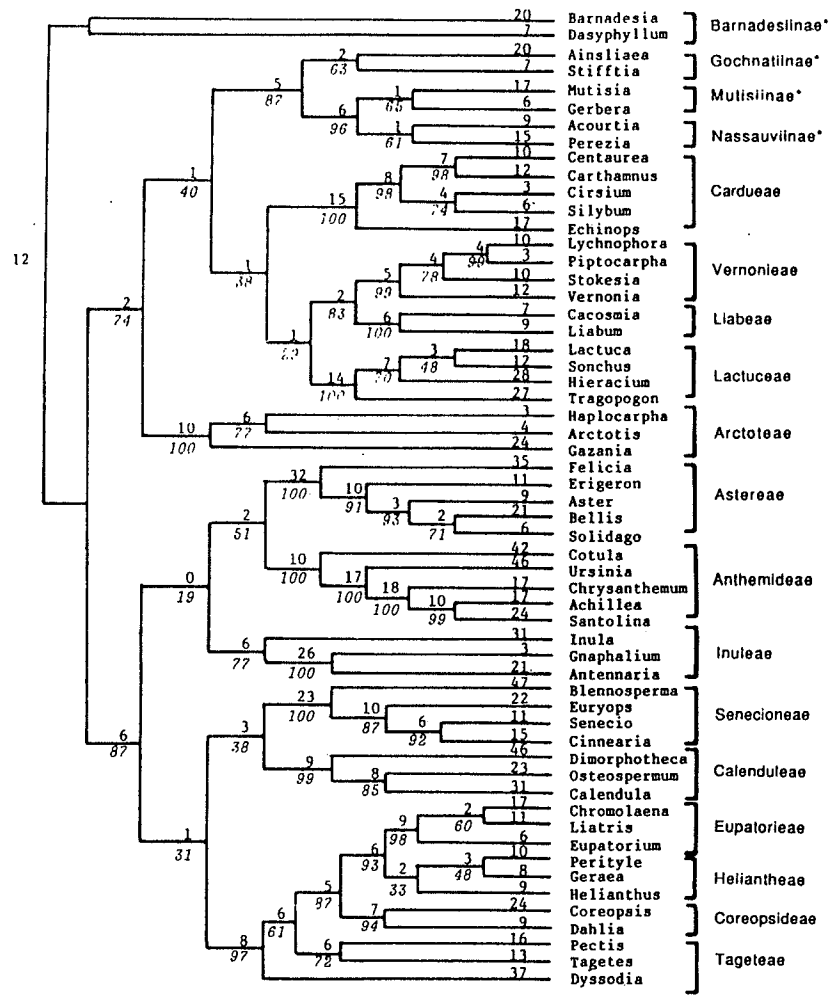

FIG. 4 Chloroplast DNA majority rule bootstrap tree of 57 genera of Asteraceae constructed using Dollo parisimony. Numbers above the line represent the number of restriction site mutations and the numbers below the line represent the percent of bootstrap replicates supporting each branch. Genera and tribal designations are at right.

B. DNA sequencing

The selection of a sequence for comparison should take into consideration the degree of phylogenetic divergence between the taxa under study. The length of the sequence and degree of divergence are important variables to consider when choosing an appropriate sequence. Published compilations of lengths and percent sequence divergence comparisons for the full complement of chloroplast genes between *Nicotiana tabacum* and *Marchantia polymorpha* (Wolfe and Sharp, 1988) and *N. tabacum* and *Oryza sativa* (Sugiura, 1989), the only three completely sequenced chloroplast genomes, provide an excellent starting point for the choice of a sequence for a phylogenetic study. As with restriction site analysis, identification of homologs when using sequence data for phylogenetic analysis is critical. All chloroplast genes, including those on the inverted repeat, evolve as single copy genes, thereby avoiding problems of gene homology inherent in the study of many nuclear genes. In sequence comparisons, individual nucleotides are the characters and correct identification of homologs depends on proper sequence alignment. If all sequences are the same length then alignment usually is trivial, however insertions, deletions, and highly divergent sequences present problems. As with mapping restriction sites, the invariant nucleotides are important for proper alignment and only those sites in which at least two nucleotides are shared by two or more taxa are phylogenetically informative in a parsimony analysis. For very distantly related organisms, in which silent nucleotide positions (e.g. most third codon positions in protein coding genes) may approach saturation with substitutions, the translated amino acid sequences of chloroplast protein genes may provide a more reliable indication of relationships. The simplest and most common approach to analysing sequence data entails weighting all sequence substitutions equally, however, some justification exists for applying weights to different substitutions. Two proposed systems of weighting include weighting transversions (purine-pyrimidine) over transitions (purine-purine or pyrimidine-pyrimidine) and applying a weighting scheme in which any substitution is weighted as the inverse of the number of other character states (nucleotides or amino acids) at that position, thereby reducing the impact of substitutions at those positions most likely to exhibit convergent evolution (Felsenstein, 1981). Alternative methods of phylogenetic reconstruction using sequence data are available that take all nucleotide positions into account, including distance matrix and maximum likelihood methods (Felsenstein, 1988).

Most of the interest in sequencing chloroplast genes for phylogenetic analysis of angiosperms has centered on the large subunit of ribulose bisphosphate carboxylase (*rbc*L), although no specific phylogenetic studies using *rbc*L sequences have been published (but see Palmer *et al.*, 1988 for a phenetic comparison of published sequences). The *rbc*L sequence similarity at the amino acid level for comparisons of tobacco with *Marchantia* and tobacco with rice are 91% and 93%, respectively (Wolfe and Sharp, 1988; Sugiura, 1989). A slowly evolving gene such as *rbc*L may be expected to exhibit insufficient nucleotide substitution to be useful for

inferring phylogeny among closely related species, but should be useful at higher taxonomic levels. In the Asteridae we are using *rbc*L sequence data for phylogenetic studies at the interfamilial level, however enough sequences are available within two families, the Asteraceae and Solanaceae, to examine the effectiveness of *rbc*L sequence data for inferring intrafamilial relationships as well. This work-in-progress has focused on the families most closely related to the Asteraceae, but will be expanded to include 30-40 species representing all of the major families in the Asteridae, along with outgroups chosen from the Rosidae. To date 19 sequences have been determined for *rbc*L within the Asteridae. A phylogenetic analysis of these sequences was conducted with spinach as the outgroup using the bootstrap and Wagner parsimony (Fig. 5). Of the 1437 nucleotide positions in the typical *rbc*L sequence in the Asteridae, 384 were variable and the resulting tree has a CI of 0.56. The results show that numerous nucleotide substitutions distinguish representatives of different families whereas relatively few substitutions distinguish genera within each of the two families represented by more than one species. The greater interfamilial resolution is reflected in the greater number of bootstrap replicates supporting family-level clades, thus lending confidence to the use of *rbc*L sequences for phylogenetic reconstruction in the Asteridae. The interfamilial relationships depicted in Fig. 5 should be viewed as

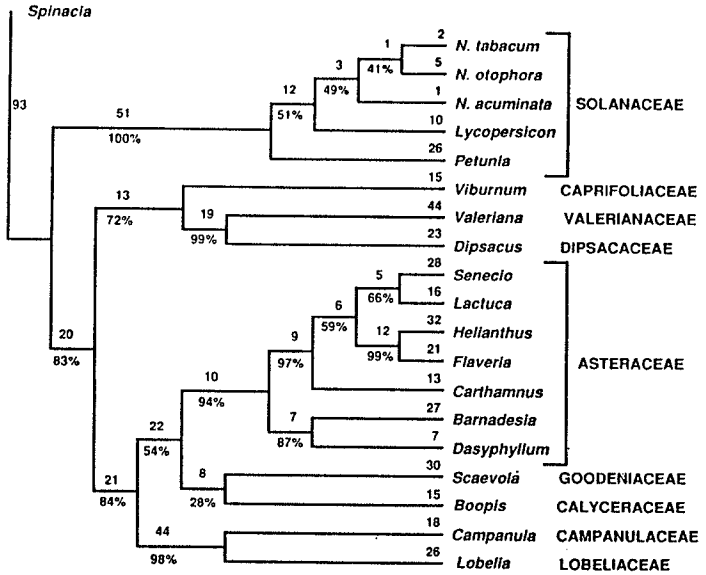

FIG. 5 Chloroplast DNA majority rule bootstrap tree using Wagner parsimony of 17 species representing 7 families of the subclass Asteridae based on *rbc*L sequence data. Numbers above the line represent the number of nucleotide substitutions along the branch and the numbers below the line are the percent of bootstrap replicates supporting each branch.

preliminary until more families have been sampled, but these preliminary results identify the Goodeniaceae, or a clade comprised of the Goodeniaceae and Calyceraceae, as the best candidate for the sister group to the Asteraceae of the families represented thus far.

III. Structural Rearrangements

Structural rearrangements are the result of several kinds of mutations including insertions and deletions in either coding or non-coding DNA and sequence inversions. Most of the previously documented rearrangements of the chloroplast genome have been discovered through research aimed at understanding its structure and function. The importance of rearrangements for understanding phylogeny has been recognized only recently and the first efforts aimed at surveying chloroplast genome structure specifically for systematic ends are now underway, with little published data yet available. Insertions and deletions of a small number of bases (1-20) may be very common, though difficult to detect, and the assessment of homology is difficult when they occur in non-coding regions. Larger insertions and deletions (100 + bp) occur less frequently and are readily detected by restriction site mapping, making the assessment of homology easier. However, drawing phylogenetic inference from insertions and deletions in non-coding sequences must proceed with caution, because these mutations tend to occur most frequently in "hotspots" (Tassopulu and Kung, 1984; Palmer, 1985; Palmer *et al.*, 1988). Exact size estimates and map locations must be confirmed using several restriction enzymes in order for confidence to be placed in the homology of such insertions and deletions. An example of a deletion in a non-coding region is found in the Solanaceae, where a 650 base pair deletion distinguishes *Nicotiana* from the rest of the family (Fig. 6). The size of the deletion is confirmed by mapping 10 restriction enzymes and the site of the deletion can be localized to within a 285 base pair region of the tobacco sequence.

The loss of genes or introns from the chloroplast genome is a special class of deletions, which occur rarely, thereby making the shared absence of a given gene or intron a powerful phylogenetic statement. The presence of a particular gene or intron can be detected by hybridization using cloned fragments of DNA specific to that particular gene or intron from another species of plant. Recently documented cases of chloroplast gene loss include the *rpl*22 gene, which is absent in all members of the legume family surveyed (Palmer *et al.*, 1988; Palmer and Doyle, unpublished data), and the *tuf*A gene, which is missing from the chloroplasts of all land plants (Baldauf and Palmer, submitted). Introns, although non-coding, also are conserved evolutionarily and the absence of an intron can furnish a useful phylogenetic marker. The intron in the *rpl*2 gene is missing from all sampled members of the Caryophyllales (Palmer and Zurawski, unpublished data), providing additional

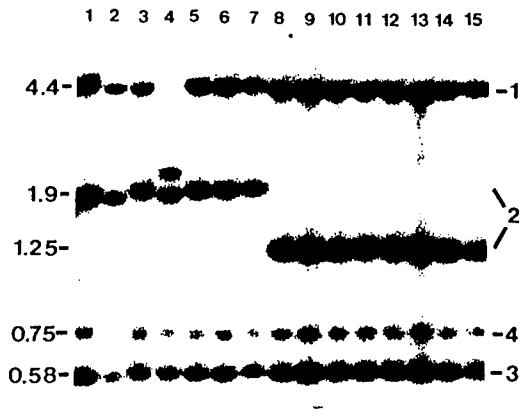

FIG. 6 Evidence of an approximately 650 bp deletion in *Nicotiana*. Numbers on the right indicate the sequence of fragments on the cpDNA molecule, numbers on the left indicate size of fragments in kb. Lanes 1-4 are various genera of Solanaceae, lanes 5-7 *Petunia*, and lanes 8-15 *Nicotiana*. All DNAs were digested with EcoRI. The approximately 1.9 kb fragment in *Petunia* is conserved with only minor size variation throughout the Solanaceae, implying that the smaller fragment in *Nicotiana* is the result of a ca. 650 bp deletion.

molecular evidence for that well-supported clade. In addition, the *rpl*2 intron apparently has been lost independently in other dicot lineages and perhaps in some monocots as well (Downie and Olmstead, unpublished data). A survey for missing chloroplast genes, introns, and conserved open reading frames (ORFs) in 88 species representing 36 families of the Asteridae (Table 1, Downie, unpublished data) indicates a non-random distribution of this class of deletions, suggesting that some yet unknown factors may trigger rearrangements and gene losses in certain chloroplast lineages.

TABLE 1
Genes, introns, and open reading frames (ORFs) absent from members of the Asteridae as represented by 94 species and 35 families (Downie, unpublished data)

Gene/intron/ORF	Missing from (species missing/species sampled):
*rpl*2 intron	Convolvulaceae/Cuscutaceae (4/4)
5' ORF 512	Oleaceae (3/4); Campanulaceae/Lobeliaceae (8/8)
3' ORF 512	Oleaceae (1/4); Campanulaceae/Lobeliaceae (7/8)
ORF 1244	Campanulaceae/Lobeliaceae (8/8)
ORF 581	Campanulaceae/Lobeliaceae (5/8)
ORF 75-ORF350	Convolvulaceae/Cuscutaceae (4/4); Campanulaceae (2/8)

The loss of one copy of the inverted repeat is another special case of a deletion which has been documented in three plant families, Pinaceae, Leguminosae, and Geraniaceae (Palmer *et al.*, 1987; Strauss *et al.*, 1988; Lavin 1990; Calie and Palmer, unpublished data). Whereas the loss of one copy of the inverted repeat is an extremely rare event, variation in size of the inverted repeat is common and usually consists of an expansion or contraction of the inverted repeat into, or out of, a single copy region, thereby making homology of inverted repeat size variants difficult to assess. Large insursions into what is normally single copy DNA by the inverted repeat have been documented in *Pelargonium* and *Geranium* (Palmer *et al.*, 1987; Calie and Palmer, unpublished data) and in *Nicotiana acuminata* (Shen *et al.*, 1982), where a sister species, *N. attenuata*, which differs in only two of over 600 restriction sites, has a typical size inverted repeat (Olmstead and Palmer, unpublished data).

Inversions are rare, but like gene losses make excellent systematic markers and can be detected by hybridization experiments. A fragment of DNA containing the end point of an inversion will hybridize to two widely separated fragments in a species without the inversion. For a more complete discussion of the methods for detecting cpDNA rearrangements see Palmer *et al.* (1988). Inversions have been documented in several plant families and are usually recognizable through restriction mapping as a unique event delineating all or part of a family as a monophyletic unit (Palmer, 1985; Jansen and Palmer, 1987; Palmer *et al.*, 1988). One such well-characterized inversion occurs in the Asteraceae, where all members of the family except those in the subtribe Barnedesiinae carry a 22 kb inversion relative to all other known land plant cpDNAs (Jansen and Palmer, 1987).

Deletions, insertions, and inversions represent a class of phylogenetic data distinct from that derived from the analysis of nucleotide substitutions. No generally accepted means exists of incorporating these two disparate forms of molecular evidence into a single phylogenetic reconstruction. Unusual rearrangements, for which a strong argument can be made for their uniqueness, may best be treated as undisputed evidence of monophyly for a group (Jansen and Palmer, 1987), whereas more commonly occurring rearrangements, such as small insertions and deletions, or "length mutations" in restriction site analyses, may be variously weighted (Morden and Golden, 1989), treated equivalent to mucleotide substitutions (Coates and Cullis, 1987; Doebley *et al.*, 1987), or considered unreliable for phylogenetic inference (Perl-Treves and Galun, 1985; Sytsma and Gottlieb, 1986; Jansen and Palmer, 1988; Soltis *et al.*, 1989). The fact that no firm theoretical justification exists for treating disparate forms of molecular data means that judgements inevitably must be made on a case-by-case basis. In this respect, systematists incorporating disparate sorts of conventional data, such as cytology and morphology, are confronted with similar problems.

IV. Conclusions

Chloroplast DNA systematics is a multi-faceted research program which is suited to all levels of taxonomic hierarchy in flowering plants, although limitations in the amount of useful variation may be experienced in within-species studies. Species-level phylogeny is best studied using restriction site analysis, whereas higher-level relationships are best studied by direct gene sequencing. Structural rearrangements, because of their infrequent occurrence, provide excellent systematic markers and can be assayed easily. As with any source of data used for phylogenetic analysis, proper attention must be paid to phylogenetic principles when it comes to selecting taxa, sampling within taxa, and data analysis. Many of the problems that plague analyses relying on other sources of data also affect cpDNA, including the correct assessment of homology, inconsistent data arising from parallel evolution, and incomplete resolution of phylogenetic trees. However, cpDNA offers many advantages over conventional sources of data, including the ready accessibility of numerous independent characters, characters that are independent of morphology (enabling tests of hypotheses concerning phenotypic evolution), and highly consistent data sets, especially at the species-level. The use of cpDNA in systematics is emerging as a mature research program and while not the panacea of early predictions, it will be an important tool for many years and will contribute to the resolution of many outstanding systematic problems.

Acknowledgements

This research was supported by NSF Grants BSR-8415934 to J.D.P. and R.K.J. and BSR-8717600 to J.D.P.

References

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.

Baldauf, S. L. and Palmer, J. D. (1990). Evolutionary transfer of the chloroplast *tuf*A gene to the nucleus. *Nature* (in press).

Banks, J. A. and Birky, C. W. (1985). Chloroplast DNA diversity is low in a wild plant, *Lupinus texensis*. *Proc. Natl. Acad. Sci. U.S.A.* **82**, 6950-6954.

Bremer, K. (1987). Tribal interrelationships of the Asteraceae. *Cladistics* **3**, 210- 253.

Coates, D. and Cullis, C. A. (1987). Chloroplast DNA variability among *Linum* species. *Am. J. Bot.* **74**, 260-268.

DeBry, R. W. and Slade, N. A. (1985). Cladistic analysis of restriction endonuclease cleavage maps within a maximun-likelihood framework. *Syst. Zool.* **34**, 21-34.

Doebley, J., Renfroe, W. and Blanton, A. (1987). Restriction site variation in the *Zea* chloroplast genome. *Genetics* **117**, 139-147.

Doyle, J. J., Doyle, J., and Brown, A. H. D. (1989). The limits of chloroplast DNA in phylogeny reconstruction: polymorphism and phylogeny in the B genome of *Glycine*. *Am. J. Bot.* **76**(suppl.), 239.

Felsenstein, J. (1981). A likelihood approach to character weighting and what it tells us about parsimony and compatibility. *Biol. J. Linn. Soc.* **16**, 183-196.

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**, 783-791.

Felsenstein, J. (1988). Phylogenies from molecular sequences: inference and reliability. *Annu. Rev. Genet.* **22**, 521-565.

Hosaka, K., Ogihara, Y., Matsubayashi, M. and Tsunewaki, K. (1984). Phylogenetic relationship between the tuberous *Solanum* species as revealed by restriction endonuclease analysis of chloroplast DNA. *Jpn. J. Genet.* **59**, 349-369.

Jansen, R. K. and Palmer, J. D. (1987). A chloroplast inversion marks an ancient evolutionary split in the sunflower family (Asteraceae). *Proc. Natl. Sci. U.S.A.* **84**, 5818-5822.

Jansen, R. K. and Palmer, J. D. (1988). Phylogenetic implications of chloroplast DNA restriction site variation in the Mutisieae (Asteraceae). *Am. J. Bot.* **75**, 751-764.

Jansen, R. K., Holsinger, K. E., Michaels, H. J. and Palmer, J. D. Phylogenetic analysis of chloroplast DNA restriction site data at higher taxonomic levels: an example from the Asteraceae. *Evolution* (submitted, a).

Jansen, R. K., Michaels, H. J. and Palmer, J. D. Phylogeny and character evolution in the Asteraceae based on chloroplast DNA restriction site mapping. *Evolution* (submitted, b).

Keeley, S. C. and Jansen, R. K. (1989). Chloroplast DNA evidence for tribal placement of *Brachylaena*, *Tarchonanthus*, and *Pluchea* (Asteraceae). *Am. J. Bot.* **76**(suppl.), 250.

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

Lavin, M. *et al.* (1990). Evolutionary significance of the loss of the chloroplast DNA inverted repeat in the Leguminosae subfamily Papillionidae. *Evolution* **44**(2), 390-402.

Morden, C. W. and Golden, S. S. (1989). *psb*A genes indicate common ancestry of prochlorophytes and chloroplasts. *Nature* **337**, 382-385.

Palmer, J. D. (1985). Comparative organization of chloroplast genomes. *Annu. Rev. Genet.* **19**, 325-354.

Palmer, J. D. and Zamir, D. (1982). Chloroplast DNA evolution and phylogenetic relationships in *Lycopersicon*. *Proc. Natl. Acad. Sci. U.S.A.* **79**, 5006-5010.

Palmer, J. D., Nugent, J. M. and Herbon, L. A. (1987). Unusual structure of geranium chloroplast DNA: a triple-sized inverted repeat, extensive gene duplications, multiple inversions, and two repeat families. *Proc. Natl. Acad. Sci. U.S.A.* **84**, 769-773.

Palmer, J. D., Shields, C. R., Cohen, D. B. and Orton, T. J. (1983). Chloroplast DNA evolution and the origin of amphidiploid *Brassica*. *Theor. Appl. Genet.* **65**, 181-169.

Palmer, J. D., Jansen, R. K., Michaels, H. J., Chase, M. W. and Manhart, J. R. (1988). Chloroplast DNA variation and phylogeny. *Ann. Mo. Bot. Gard.* **75**, 1180-1206.

Perl-Treves, R. and Galun, E. (1985). The *Cucumis* plastome: physical map, intrageneric variation and phylogenetic relationships. *Theor. Appl. Genet.* **71**, 417-429.

Sanderson, M. J. and Donoghue, M. J. (1989). Patterns of variation and levels of homoplasy. *Evolution* **43**, 1781-1795.

Shen, G. F., Chen, K., Wu, M. and Kung, S. D. (1982). *Nicotiana* chloroplast genome IV. *N. acuminata* has larger inverted repeats and genome size. *Mol. Gen. Genet.* **187**, 12-18.

Soltis, D. E., Soltis, P. S. and Ness, B. D. (1989). Chloroplast DNA variation and multiple origins of autopolyploidy in *Heuchera micrantha* (Saxifragaceae). *Evolution* **43**, 650-656.

Strauss, S. H., Palmer, J. D., Howe, G. T. and Doerksen, A. H. (1988). Chloroplast genomes of two conifers lack a large inverted repeat and are extensively rearranged. *Proc. Natl. Acad. Sci. U.S.A.* **85**, 3898-3902.

Sugiura, M. (1989). The chloroplast chromosomes in land plants. *Annu. Rev. Cell Biol.* **5**, 51-70.

Sytsma, K. J. and Gottlieb, L. D. (1986). Chloroplast DNA evolution and phylogenetic relationships in *Clarkia* sect. *Peripetasma* (Onagraceae). *Evolution* **40**, 1248-1261.

Sytsma, K. J. and Schaal, B. A. (1985). Phylogenetics of the *Lisianthius skinneri* (Gentianaceae) species complex in Panama utilizing DNA restriction fragment analysis. *Evolution* 39, 594-608.

Tassopulu, D. and Kung, S. D. (1984). *Nicotiana* chloroplast genome 6. Deletion and hot spot - a proposed origin of the inverted repeats. *Theor. Appl. Genet.* 67, 185-193.

Whitfield, P. R. and Bottomley, W. (1983). Organization and structure of chloroplast genes. *Annu. Rev. Plant Physiol.* 34, 279-310.

Wolfe, K. H. and Sharp, P. M. (1988). Identification of functional open reading frames in chloroplast genomes. *Gene* 66, 215-222.

Wolfe, K. H., Li, W. H. and Sharp, P. M. (1987). Rates of nucleotide substitutions vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc. Natl. Acad. Sci. U.S.A.* 84, 9054-9058.

Zurawski, G. and Clegg, M. T. (1987). Evolution of higher plant chloroplast DNA-encoded genes: implication for structure-function and phylogenetic studies. *Annu. Rev. Plant Physiol.* 38, 391-418.