Proc. Natl. Acad. Sci. USA
Vol. 92, pp. 10379–10383, October 1995
Evolution

ndhF sequence evolution and the major clades in the sunflower family

KI-JOONG KIM* AND ROBERT K. JANSEN†

Department of Botany, University of Texas, Austin, TX 78713-7640

Communicated by Peter H. Raven, Missouri Botanical Garden, St. Louis, MO, June 21, 1995

ABSTRACT **An extensive sequence comparison of the chloroplast *ndhF* gene from all major clades of the largest flowering plant family (Asteraceae) shows that this gene provides ≈3 times more phylogenetic information than *rbcL*. This is because it is substantially longer and evolves twice as fast. The 5′ region (1380 bp) of *ndhF* is very different from the 3′ region (855 bp) and is similar to *rbcL* in both the rate and the pattern of sequence change. The 3′ region is more A+T-rich, has higher levels of nonsynonymous base substitution, and shows greater transversion bias at all codon positions. These differences probably reflect different functional constraints on the 5′ and 3′ regions of *ndhF*. The two patterns of base substitutions of *ndhF* are particularly advantageous for phylogenetic reconstruction because the conserved and variable segments can be used for older and recent groups, respectively. Phylogenetic analyses of 94 *ndhF* sequences provided much better resolution of relationships than previous molecular and morphological phylogenies of the Asteraceae. The *ndhF* tree identified five major clades: (*i*) the Calyceraceae is the sister family of Asteraceae; (*ii*) the Barnadesioideae is monophyletic and is the sister group to the rest of the family; (*iii*) the Cichorioideae and its two basal tribes Mutisieae and Cardueae are paraphyletic; (*iv*) four tribes of Cichorioideae (Lactuceae, Arctoteae, Liabeae, and Vernonieae) form a monophyletic group, and these are the sister clade of the Asteroideae; and (*v*) the Asteroideae is monophyletic and includes three major clades.**

The chloroplast *rbcL* gene has been the most widely sequenced gene for phylogeny reconstruction in plants during the past decade (1, 2). The availability of over 1000 sequences from all major plant lineages has provided valuable information for both phylogenetic and molecular evolutionary comparisons. Thus, the tempo and mode of *rbcL* evolution are well characterized (2, 3) compared with other plant genes. Unfortunately, the phylogenetic utility of this gene is sometimes limited to studies above the family level because of its slow evolutionary rate (2, 4, 5). Furthermore, even above the family level, *rbcL* data sometimes fail to generate fully resolved and well-supported phylogenetic trees in rapidly radiated groups (6, 7).

For recently derived groups, rapidly evolving sequences may provide many more phylogenetically informative characters. For example, internal transcribed spacer sequences of nuclear ribosomal RNA have been used successfully at the intergeneric and interspecific levels (8). However, this marker is often not useful at higher taxonomic levels because high divergence and length variation make sequence alignment very problematical (8). Thus, in plants few if any genes have been shown to be appropriate for addressing phylogenetic issues in relatively young and rapidly radiated families.

We have explored the use of alternative chloroplast genes for phylogeny reconstruction in the large and relatively recent angiosperm family Asteraceae. Most chloroplast genes are either too short or too conserved to provide adequate numbers of characters in recently evolved families. A number of alternative genes have been suggested as potential candidates for phylogenetic comparisons at lower taxonomic levels (9). The phylogenetic utility of one of these, *matK*, has been recently demonstrated (10). Comparison of sequences of two chloroplast genomes (rice and tobacco), however, revealed only two genes, *rpoC1* and *ndhF*, that are considerably longer and evolve faster than *rbcL* (9, 11). We selected *ndhF* because it is longer and evolves slightly faster than *rpoC1* (11), because *rpoC1* has an intron that may require additional effort in DNA amplification and sequencing, and because *rpoC1* is variably present in angiosperm chloroplast DNAs. In the chloroplast genomes of most angiosperms, *ndhF* is located at one end of the small single-copy region (12). It was identified as encoding the ND5 protein of chloroplast NADH dehydrogenase based on weak amino acid sequence identity (31%) to human mitochondrial NADH dehydrogenase (13–15). Its transcription product was identified from tobacco (15), but not from rice (16), even though genes from other subunits of the enzyme are actively transcribed in the rice chloroplast. The *ndhF* gene is 2133 bp long in tobacco (13, 17), which is 1.5 times longer than *rbcL*. Amino acid sequence divergence of *ndhF* is 4 times greater than *rbcL* in comparisons between tobacco and rice (18). Thus, the longer size and higher sequence divergence of *ndhF* indicate that this gene may provide many more informative characters than *rbcL* for phylogeny reconstruction.

The Asteraceae is an ideal family for examining rates and patterns of evolution and phylogenetic utility of *ndhF* sequences for three reasons. (*i*) It is the largest flowering plant family with ≈2500 genera and 30,000 species (19). Although there is controversy concerning its age, fossil evidence indicates that the Asteraceae originated recently in the upper Oligocene (20–22). The appearance of abundant and diverse Asteraceae pollen in the lower Miocene (22) and the large number of extant species clearly indicate that the family has experienced a recent and rapid radiation. (*ii*) Although the Asteraceae has been the focus of intensive phylogenetic studies during the past 10 years using both molecular (4, 23, 24) and morphological (25–27) data, the circumscription and relationships of several major clades remain unresolved. (*iii*) The availability of *rbcL* sequences for the same species of Asteraceae (4) enables direct comparisons of the phylogenetic utility and sequence evolution of this well-characterized gene with *ndhF*.

We have sequenced *ndhF* from 89 species of Asteraceae and from 5 species from related families.‡ The results have important implications for understanding patterns and rates of evolution of chloroplast genes and the identification of the major clades of the largest family of flowering plants. Further-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: Ks, number of synonymous substitutions; Ka, number of nonsynonymous substitutions; Ko, total number of substitutions; Ts, transition(s); Tv, transversion(s); indel, insertion/deletion.

*Present address: Department of Biology, Yeung-Nam University, Kyeung-Book, South Korea.
†To whom reprint requests should be addressed.

‡The sequences reported in this paper have been deposited in the GenBank data base (accession nos. L39373–L39467).

more, we demonstrate that *ndhF* has excellent potential for resolving phylogenies within and among plant families that have experienced recent and rapid radiation.

MATERIALS AND METHODS

Eighty-nine species representing all tribes and major clades of the Asteraceae were selected for comparative sequencing of the *ndhF* gene. Five species were also selected as putative outgroups and sister groups from closely related families (25, 26). Genomic DNA was isolated using hexadecyltrimethylammonium bromide extraction methods (28).

A 3.6-kb recombinant plasmid (29) containing the entire region of the lettuce *ndhF* gene was subcloned into M13 and pBluescript vectors and sequenced using universal (United States Biochemical) and some specific *ndhF* primers (see below). The lettuce *ndhF* sequence was aligned with sequences from tobacco (13, 17) and rice (30). Conserved regions were selected from the three aligned sequences, and 22 primers were designed. To sequence the entire gene, amplification primers were designed outside of the coding region.

Two overlapping fragments were amplified using two pairs of primers for each species. Amplifications were performed in 100 μl with 1 unit of *Tfl* polymerase (Epicentre Technologies, Madison, WI), 8 mM $MgCl_2$, 0.2 mM of each dNTP, 0.25 μM of each primer, and 2–8 μg of DNA. Each of 30 thermal cycles consisted of 1 min of denaturation at 95°C, 1 min of annealing at 55°C, and 1 min and 45 sec of extension at 72°C with a 3-sec extension per cycle. The amplified DNA was purified with agarose gel electrophoresis and glass powder and directly sequenced using Sequenase (United States Biochemical) and ^{35}S-labeled ATP. Largely overlapping sequences were generated for the forward strand using at least 12 forward primers. Sequences for the reverse strand for 11 species were confirmed using up to 8 primers, but this was unnecessary in most cases because of the high degree of overlap of the forward primers.

Both DNA and amino acid sequences were aligned using the PILEUP program of the Genetics Computer Groups package in a UNIX computer system, and some gaps were manually adjusted. Numbers of synonymous (Ks), nonsynonymous (Ka), and total (Ko) substitutions (31) were calculated using MEGA (32). Transition (Ts)/transversion (Tv) ratios and G+C content were calculated using MACCLADE (33). We performed the same statistical analyses for *rbcL* data (4) to compare the evolutionary parameters of *ndhF* to *rbcL*.

Phylogenetic analyses were performed on the 94 taxon *ndhF* data matrix and on a reduced matrix that included the 28 species for which both *rbcL* and *ndhF* sequences were available. Synapomorphic gaps were removed from the *ndhF* data matrix before phylogenetic analyses. All parsimony analyses were performed with the heuristic search option with equal weighting of characters using PAUP (34). Equally parsimonious tree islands (35) were searched for by using 1000 random sequence additions with MULPARS on, STEEPEST DESCENT on (500 replications) and off (500 replications), and TBR branch swapping. Bootstrap analysis (36) was also performed to evaluate the degree of support for internal branches using 300 replicates with simple sequence addition, MULPARS off, STEEPEST DESCENT off, and TBR branch swapping.

RESULTS

All 94 *ndhF* sequences were complete except *Campanula* and *Scaevola*. Twenty-four base pairs at the 5′ end of *Campanula* and ≈30 bp at the 3′ end of *Scaevola* were missing because of the use of internal primers for DNA amplification. Twenty-eight of the species sequenced for *ndhF* have also been sequenced for *rbcL* (4, 13, 17). After alignment, the sequences were truncated at the shortest stop codon for both *rbcL* (1428 bp) and *ndhF* (2235 bp) for further analyses.

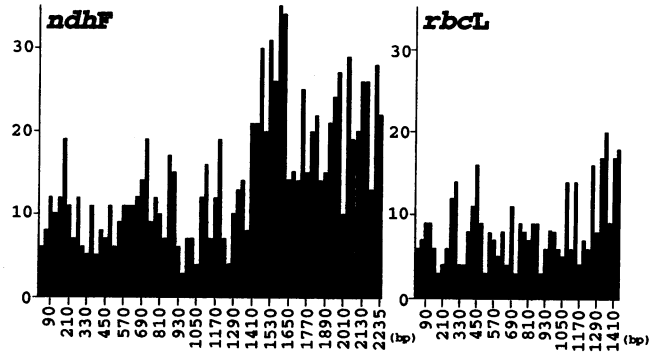

FIG. 1. Distribution of base substitutions across the *ndhF* and *rbcL* genes. Each bar represents the number of base substitutions per 30 bp. Numbers of base substitutions are inferred from all the shortest trees, and average values are shown in the histogram.

Patterns and Rates of Base Substitution in *ndhF* and *rbcL*. Fig. 1 compares the patterns of base substitutions in *ndhF* and *rbcL* for the same 28 taxa. The average number of base substitutions per 30 bp was 14.7 in *ndhF* vs. 8.6 in *rbcL*. The substitution frequency was markedly uneven within *ndhF*. The 5′ end (first 1380 bp) of the gene has a slightly higher frequency of substitutions (average 10.0) than *rbcL* (8.6). In contrast, the 3′ end (remaining 855 bp) of the gene showed an average of 22.0 substitutions, which was 2.6 times higher than *rbcL* and twice as high as the 5′ region of *ndhF*. Because of this, we

Table 1. Comparison of evolutionary features of two chloroplast genes, *rbcL* and *ndhF*

	rbcL	5′ *ndhF**	3′ *ndhF**	Total *ndhF*	*ndhF* + *rbcL*
Nucleotide sites					
Aligned	1428	1380	855	2235	3663
Variable	342	363	432	795	1137
Synapomorphic	120	144	231	375	495
G+C content†					
1st codon	0.573	0.398	0.317	0.366	0.448
2nd codon	0.434	0.381	0.312	0.354	0.385
3rd codon	0.302	0.258	0.277	0.247	0.267
All positions	0.437	0.346	0.285	0.324	0.367
Ts/Tv ratios†					
1st codon	0.790	0.751	0.738	0.737	0.771
2nd codon	0.584	0.653	0.827	0.758	0.749
3rd codon	1.164	1.641	0.871	1.252	1.202
All positions	0.972	1.272	0.822	0.991	0.998
Ks, Ka, and Ko per 100 sites‡					
Ks × 100	11.33	12.52	17.93	14.44	13.15
Ka × 100	2.21	1.87	8.35	4.24	3.44
Ko × 100	3.69	4.13	10.56	6.41	5.55
Phylogenetic analyses					
No. of trees	8	47	48	2	2
Tree length (TL)	711	674	949	1631	2340
TL by codon					
1st	173	152	261	413	580
2nd	90	54	232	258	377
3rd	448	468	456	935	1383
Consistency index	0.608	0.685	0.669	0.673	0.652
g1 statistics§	−0.536	−0.766	−0.865	−1.009	−0.771

**ndhF* sequences were divided into the 5′ and 3′ regions at coordinate 1380.
†Average values of 28 sequences and all equally parsimonious trees were used.
‡All possible pairwise comparisons were performed.
§Values were derived from 10,000 randomly sampled trees (37).

divided the *ndhF* sequence into two regions (5′ and 3′) for further comparisons (Table 1). The 28 sequences of *ndhF* had 795 variable and 375 phylogenetically informative positions. The number of informative sites was 3 times greater in *ndhF* than in *rbcL* (Table 1). This was due to the larger gene size and the more rapidly evolving 3′ region of *ndhF*. Indeed, 231 of 375 (61.6%) informative sites occurred in the 3′ region.

The Ts/Tv ratio was similar for *ndhF* and *rbcL*; however, there were substantial differences in this ratio between the 5′ and 3′ regions of *ndhF*. Furthermore, the Ts/Tv ratio differed markedly at the third codon position between the two regions of *ndhF*. The Ks value was slightly higher in *ndhF* than that in *rbcL*, and that difference can be attributed largely to the 3′ region of *ndhF*. In contrast, Ka in *ndhF* was almost twice as high as that in *rbcL*. Again, the different Ka values between two genes were primarily due to the 3′ region of *ndhF*. The Ka value in the 3′ region of *ndhF* was 5.5 times greater than that in the 5′ region and 4 times higher than that in *rbcL*. There were also substantial differences in base composition between the two genes (Table 1). The G+C content in *rbcL* was 43.7% and shows a more balanced A+T/G+C ratio. In contrast, the *ndhF* gene was A+T-rich, with a G+C content of only 32.4%. Within *ndhF*, the 3′ region showed a greater A+T bias (72%) than the 5′ portion (65%).

Length Variation in *ndhF*. The length of *ndhF* ranged from 2211 bp in *Calendula* to 2271 bp in *Blennosperma*. Gene lengths of 2235 and 2226 bp occurred in the majority (76%) of Asteraceae. A length of 2235 bp occurred commonly in the Barnadesieae, Mutisieae, and Cardueae, whereas a length of 2226 bp was most frequently encountered in other tribes. Length variation of *ndhF* was due to insertions/deletions (indels) in the first 500 bp of the 3′ region and near the stop codon of the gene. In contrast, length variation was only encountered near the stop codon of *rbcL* (4). The 94 aligned *ndhF* sequences required seven insertions and nine deletions in the middle of the gene, and most of these were concentrated between coordinates 1400 and 1600 bp. The indels were short (3, 6, and 9 bp), and all were in-frame. Only two indels were phylogenetically informative at or above the tribal level. A 9-bp deletion at coordinate 1494 was shared by members of the tribe Calenduleae, and a second 9-bp deletion at coordinate 1717 defined a major tribal clade of Asteraceae (Fig. 2). There were 27 additional indels near the stop codon. Some of these were shared among closely related genera; however, none were phylogenetically informative at the intertribal level.

Phylogenetic Analysis of *ndhF* Sequences. To investigate the differences in levels of homoplasy and tree topology between the two regions of *ndhF*, we conducted phylogenetic analyses for each region of the gene with a 28-taxon data matrix. The 5′ and 3′ regions showed similar levels of homoplasy (Table 1). Trees from the two regions of *ndhF* were congruent (data not shown). Parsimony analysis using the entire *ndhF* gene resulted in fewer equally parsimonious trees and showed greater resolution than any portion of the data. The trees from the entire

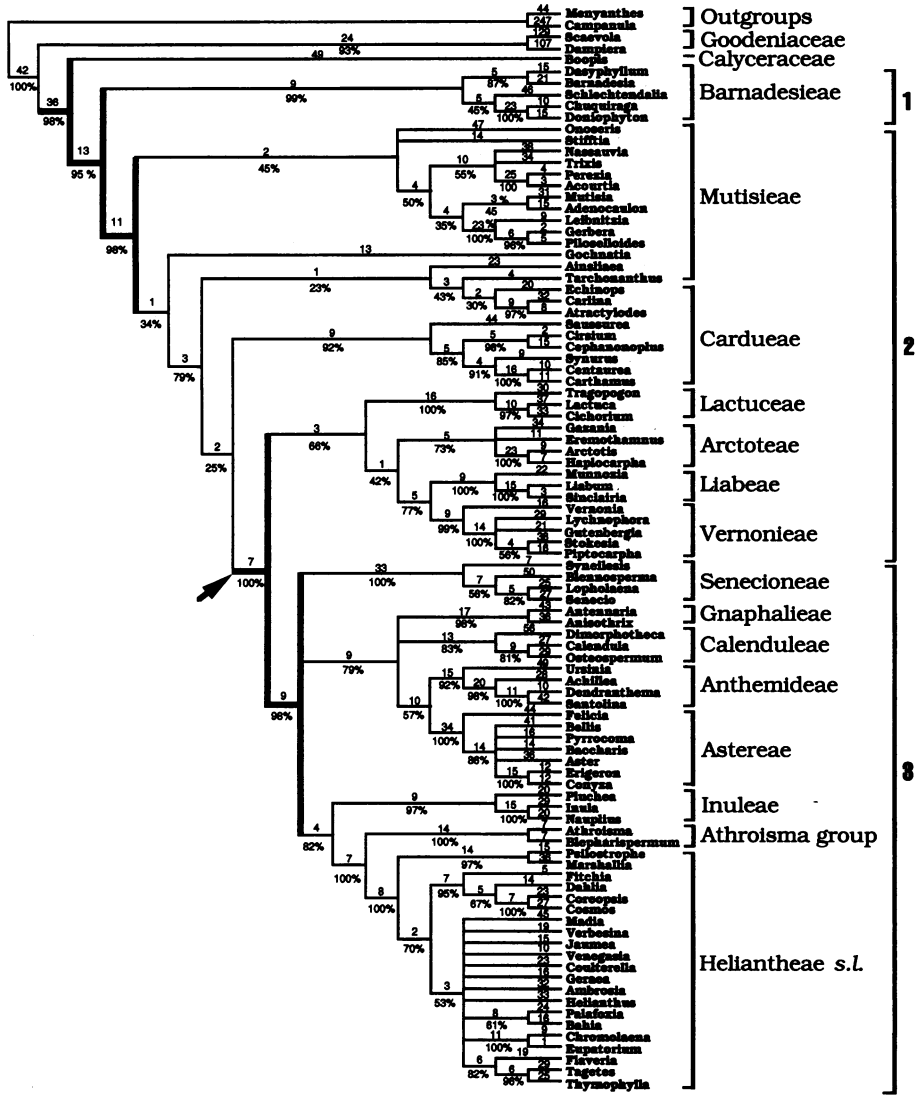

FIG. 2. Phylogenetic tree of Asteraceae showing the relationships among major clades of the family. This is the strict consensus tree of 8235 equally parsimonious trees with 3247 steps (consistency index of 0.467) based on 94 *ndhF* sequences. The numbers above and below the nodes indicate numbers of supporting characters and percentages of bootstrap support, respectively. Branches in bold indicate the major clades. The taxa below the bold arrow have a 9-bp deletion mutation at coordinate 1717. Brackets show the traditional tribal and subfamilial circumscriptions (26): 1 = Barnadesioideae, 2 = Cichorioideae, and 3 = Asteroideae.

ndhF gene also showed less homoplasy and more phylogenetic signal than *rbcL* as indicated by a higher consistency index (0.673 vs. 0.608) and more negative g1 values (−1.009 vs. −0.536) (Table 1). Combined *ndhF* and *rbcL* data generated the same two tree topologies as that produced by *ndhF* alone. Furthermore, the strict consensus tree for 28 *ndhF* sequences showed the same major clades as the 94 taxon tree (Fig. 2).

The *ndhF* data matrix for 94 taxa consisted of 2238 sites, including 1031 variable and 575 synapomorphic positions. A heuristic search with TBR branch swapping and a single entry of taxa identified more than 7000 equally parsimonious trees of 3247 steps and a consistency index of 0.467. Additional parsimony analyses with 1000 random entries of taxa failed to find shorter trees. Seventeen seeds of the 1000 entries found an additional 17,000 equally parsimonious trees with 3247 steps. Condensation of all equally parsimonious trees (7000 + 17,000) from the two analyses resulted in 8235 unique trees. In spite of the large number of equally parsimonious trees, the strict consensus tree provided excellent resolution of the major clades of the Asteraceae, and many of these are supported by high bootstrap values (Fig. 2).

DISCUSSION

Evolution of *ndhF*. The 5′ region of *ndhF* showed evolutionary patterns similar to those of *rbcL* (Table 1). In contrast, the 3′ portion of *ndhF* was substantially different from the 5′ region in a number of features, including higher A+T content, Tv-biased base substitutions, and an increased Ka. Most chloroplast genes show Tv bias at the first and second codon positions, whereas Ts predominate at the third position (2). In the 3′ portion of *ndhF*, however, all three codon positions showed similar levels of Tv bias. This pattern is not due to multiple substitutions at the third position because the amount of homoplasy was very similar at all codon positions. Similar patterns of Tv bias in *ndhF* (Table 1) may reflect unbiased codon utilization among codon positions. More than 50% of the changes on the *ndhF* tree occurred in first and second codon positions, which are more likely to result in amino acid substitutions. This indicates that the 3′ region of *ndhF* may have experienced a more relaxed selection pressure. Thus, the distinct evolutionary patterns in the 5′ and 3′ portions of *ndhF* probably reflect differential functional constraints on the gene. Unfortunately, we cannot confirm a relationship between sequence evolution and functional aspects of the gene product because the three-dimensional protein structure of *ndhF* is unknown. The 3′ region of *ndhF* also exhibits a high frequency of indels, which also indicates relaxed evolutionary constraints. Most of these indels are associated with short direct repeats of 3, 6, and 9 bases, suggesting that slipped-strand mispairing events during replication are the primary contributing factors (38).

Phylogenetic Utility of *ndhF*. The *ndhF* gene provides more phylogenetic information in the Asteraceae than *rbcL* because it is longer and one portion of it evolves more rapidly. Two recent phylogenetic studies in the angiosperm families Solanaceae (39) and Acanthaceae (40) also indicated that *ndhF* has more useful phylogenetic information than *rbcL*, even with limited taxon sampling. One concern with the use of fast-evolving sequences for phylogeny reconstruction is that phylogenetic information accumulated during radiation may be destroyed by multiple substitutions at the same sites. As a result, homoplasy is likely to increase, and it may eventually overwhelm the phylogenetic signal (41). Two factors indicate that this is not the case for our *ndhF* data. First, homoplasy in the 3′ region of *ndhF* is lower than that of *rbcL* and nearly identical to that of the 5′ region of *ndhF*. The g1 statistic, which provides a measure of the amount of phylogenetic signal (37), shows that the 3′ region of *ndhF* has more signal than either *rbcL* or the 5′ region of *ndhF*. Second, the tree topology based on only the 3′ portion of *ndhF* is congruent with previous molecular (4, 23) and morphological (25–27) trees for the Asteraceae. Thus, we believe that the 3′ region of *ndhF* provides useful phylogenetic information.

Circumscription and Relationships of Major Clades of Asteraceae. The Asteraceae has been subjected to extensive phylogenetic studies using both molecular (4, 23, 24) and morphological (25–27) characters. In many cases, both DNA and morphological data support the recognition of the same clades. Several taxonomic conflicts still remain, and a number of outstanding phylogenetic issues were not addressed by previous analyses. Our *ndhF* data represent the most extensive sampling of the Asteraceae at the higher taxonomic levels. Thus, the comprehensive taxon sampling combined with the large number of phylogenetically informative characters in *ndhF* provides compelling evidence for higher level relationships in the family. Phylogenetic relationships suggested by the *ndhF* tree can be grouped into three categories: (*i*) relationships not suggested by earlier studies, (*ii*) resolution of current controversies, and (*iii*) additional support for previous hypotheses of relationships.

Calyceraceae is the sister family of Asteraceae. Several families have been suggested as the sister family of the Asteraceae (25, 26). Recent morphological (25) and molecular studies (42) have narrowed down the possible sister groups to two families, the Calyceraceae and Goodeniaceae. The Calyceraceae forms a monophyletic group with the Asteraceae in the *ndhF* tree (Fig. 2). Expanded *ndhF* analyses, which include most major lineages of higher dicots, strongly support the sister group relationship between Calyceraceae and Asteraceae (unpublished results). This is the first phylogenetic tree to strongly suggest that the sister group of Asteraceae is the Calyceraceae. A previous *rbcL* tree (42) suggested that the Goodeniaceae and Calyceraceae are equally probable as sister groups to the Asteraceae. Our data are congruent with morphological hypotheses using both pollen and floral characters (20, 43).

Recognition of basal clades of Asteraceae. The first major clade of Asteraceae in the *ndhF* tree is composed of the Barnadesioideae. The basal position of this group was first suggested based on the lack of a 22-kb chloroplast DNA inversion (24). Morphological (25), chloroplast DNA restriction site (44), and *rbcL* (4) data supported a basal position for this subfamily. Our *ndhF* data also strongly support the basal dichotomy of the Barnadesioideae and the rest of the family. In addition, the *ndhF* tree indicates that the Mutisieae and Cardueae form paraphyletic assemblages near the base of the family (Fig. 2). The paraphyly of the Mutisieae is in agreement with recent morphological cladograms (27). The Cardueae is sometimes divided into the four subtribes, Carlininae, Echinopsidinae, Carduinae, and Centaureinieae (26). The first two subtribes are more closely related to the members of Mutisieae than to the Cardueae in the *ndhF* tree (Fig. 2), supporting previous suggestions that the Cardueae should be divided into more than one tribe (45).

Paraphyly of the Cichorioideae. The Cichorioideae traditionally includes the six tribes Mutisieae, Cardueae, Lactuceae, Arctoteae, Vernonieae, and Liabeae (26). Morphological data (25, 27) indicated that this subfamily is paraphyletic, whereas chloroplast DNA data (4, 23) weakly supported its monophyly. The *ndhF* tree (Fig. 2) strongly supports the morphological data in confirming the paraphyly of the Cichorioideae. The classification of this subfamily remains unresolved, but clearly additional subfamilies will have to be erected to accommodate the diverse tribes of the Cichorioideae. The Lactuceae, Arctoteae, Liabeae, and Vernonieae (LALV) of the Cichorioideae form a monophyletic group, and this clade is strongly supported as the sister group of the Asteroideae (nine characters with 100% bootstrap value). In addition, all members of the LALV-Asteroideae clade share a 9-bp deletion at coor-

dinate 1770 bp of the *ndhF* gene (Fig. 2), which provides additional support for its monophyly. Even though the *ndhF* tree is in agreement with morphological trees (27) with respect to the paraphyly of the Cichorioideae, phylogenetic relationships among the tribes are somewhat different from the morphological cladograms. For example, *ndhF* data provide modest support (three characters, 66% bootstrap value) for the monophyly of the LALV tribes, which form the sister group of the Asteroideae, while morphological trees (27) suggest that they are paraphyletic.

Monophyletic Asteroideae has three clades. Monophyly of the Asteroideae has been suggested by both morphological (25) and molecular (4, 23) data. More recent morphological cladograms cast some doubt on the monophyly of the Asteroideae (46). All molecular trees (4, 23) consistently indicate the monophyly of the subfamily. The *ndhF* tree (Fig. 2) strongly supports the monophyly of the Asteroideae (nine characters and 100% bootstrap value). The *ndhF* tree also provides better resolution of tribal relationships in the Asteroideae than heretofore presented. Three major clades are suggested: (*i*) the Senecioneae; (*ii*) the Gnaphalieae, Calenduleae, Anthemideae, and Astereae; and (*iii*) the Inuleae, *Athroisma* group, and Heliantheae *sensu lato* including the Helenieae, Coreopsideae, Eupatorieae, and Tageteae. Two novel relationships in the Asteroideae are apparent in the *ndhF* tree. First, a close relationship between the Inuleae (including Plucheeae) and Heliantheae *sensu lato* has not been suggested by previous workers. Our *ndhF* tree strongly places two former genera of Inuleae, *Blepharispermum* and *Athroisma*, into an intermediate position between these tribes. These two genera, along with a third genus *Leucoblepharis* (not sequenced), are traditionally grouped together in the Inuleae. However, these genera were excluded from the Inuleae (47) and placed in the subtribe Ecliptinae of Heliantheae (48) because of a number of putative synapomorphies, including carbonized achene walls and glandular apical anther appendages. Our *ndhF* data do not support the placement of these genera in the Ecliptinae but instead suggest that they should be a sister group to Heliantheae *sensu lato*. Second, the Gnaphalieae is positioned in clade 2 of the Asteroideae, although its relationships were not fully resolved. The *ndhF* tree agrees with morphological studies (47) that segregate the Gnaphalieae from the Inuleae.

In summary, *ndhF* sequence comparisons have provided important insights into the evolution and phylogenetic utility of this gene. The gene shows two distinct patterns of evolution. The first 1380 bp in the 5′ region of *ndhF* are highly conserved in length and pattern of base substitution and thus are quite similar to *rbcL*. In contrast, the 855 bp at the 3′ end are clearly under more relaxed functional constraints as evidenced by the presence of indels, higher A+T base composition, nearly identical transversion bias at all three codon positions, and more than 4 times as many nonsynonymous substitutions than the 5′ portion. These two patterns of sequence evolution make *ndhF* an excellent phylogenetic marker at a wider range of taxonomic levels than *rbcL*. For comparisons of more divergent taxa, the 5′ region of the gene can be utilized, whereas studies of more recently diverged groups could focus on the 3′ region of *ndhF*. Phylogenetic analyses of *ndhF* in the Asteraceae indicate that this gene provides ≈3 times as many informative characters as *rbcL*. Thus, *ndhF* is well suited to address phylogenetic questions in angiosperm families that are young and have experienced rapid radiation.

We thank A. Anderberg, G. Anderson, K. Bremer, J. Francisco-Ortega, C. Ferguson, V. Grant, F. Hellwig, A. Hempel, P. Karis, S. Keeley, G. Nesom, R. Olmstead, J. Palmer, J. Panero, T. Stuessy, and B. Turner for plant material, primers, or critical reading of the manuscript. This study was supported by National Science Foundation Grant DEB-9020171 to R.K.J.

1. Chase, M. W., Soltis, D. E., Olmstead, R. G., Morgan, D., Les, D. H., *et al.* (1993) *Ann. Mo. Bot. Gard.* **80,** 528–580.
2. Clegg, M. T. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 363–367.
3. Morton, B. R. (1994) *Mol. Biol. Evol.* **11,** 231–238.
4. Kim, K.-J., Jansen, R. K., Wallace, R. S., Michaels, H. J. & Palmer, J. D. (1992) *Ann. Mo. Bot. Gard.* **79,** 428–445.
5. Doebley, J., Durbin, M., Golenberg, E. M., Clegg, M. T. & Ma, D. P. (1990) *Evolution* **43,** 1137–1156.
6. Smith, J. F., Kress, W. J. & Zimmer, E. A. (1993) *Ann. Mo. Bot. Gard.* **80,** 620–630.
7. Olmstead, R. G., Bremer, B., Scott, K. M. & Palmer, J. D. (1993) *Ann. Mo. Bot. Gard.* **80,** 700–722.
8. Baldwin, B. G., Sanderson, M. J., Porter, J. M., Wojciechowski, M., Campbell, C. S. & Donoghue, M. J. (1995) *Ann. Mo. Bot. Gard.* **82,** 247–277.
9. Olmstead, R. G. & Palmer, J. D. (1994) *Am. J. Bot.* **81,** 1205–1224.
10. Johnson, L. A. & Soltis, D. E. (1995) *Ann. Mo. Bot. Gard.* **82,** 149–175.
11. Wolfe, K. H. (1991) in *The Photosynthetic Apparatus: Molecular Biology and Operation*, eds. Vogorad, L. & Vasil, I. K. (Academic, San Diego), Vol. 7B, pp. 467–482.
12. Palmer, J. D. (1991) in *The Molecular Biology of Plastids*, eds. Bogorad, L. & Vasil, I. (Academic, San Diego), Vol. 7A, pp. 5–53.
13. Shinonzaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hayashida, N., *et al.* (1986) *EMBO J.* **5,** 2043–2049.
14. Ohyama, K., Fukuzawa, H., Kohchi, T., Shirai, H., Sano, T., Sano, S., Umesono, K., Shiki, Y., Takeuchi, M., Chang, Z., Aota, S., Inokuchi, H. & Ozeki, H. (1986) *Nature (London)* **322,** 572–574.
15. Matsubayashi, T., Wakasugi, T., Shinozaki, K., Yamaguchi-Shinoxaki, K., Zaita, N., Hidaka, T., Meng, B. Y., Ohto, C., Tanaka, M., Kato, A., Maruyama, T. & Sugiura, M. (1987) *Mol. Gen. Genet.* **210,** 385–393.
16. Kanno, A. & Hirai, A. (1993) *Curr. Genet.* **23,** 166–174.
17. Olmstead, R. G., Sweere, J. A. & Wolfe, K. H. (1993) *Plant Mol. Biol.* **22,** 1191–1193.
18. Sugiura, M. (1989) *Annu. Rev. Cell Biol.* **5,** 51–70.
19. Turner, B. L. & Nesom, G. L. (1993) in *Biological Diversity of Mexico*, eds. Ramamoorthy, T. P., Bye, R., Lot, A. & Fa, J. (Oxford Univ. Press, Oxford), pp. 559–577.
20. Turner, B. L. (1977) in *Biology and Chemistry of the Compositae*, eds. Heywood, V. H., Harborne, J. B. & Turner, B. L. (Academic, London), Vol. 1, pp. 21–39.
21. Raven, P. H. & Axelrod, D. I. (1974) *Ann. Mo. Bot. Gard.* **61,** 539–673.
22. Muller, J. (1981) *Bot. Rev.* **47,** 1–142.
23. Jansen, R. K., Michaels, H. J. & Palmer, J. D. (1991) *Syst. Bot.* **16,** 98–115.
24. Jansen, R. K. & Palmer, J. D. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 5818–5822.
25. Bremer, K. (1987) *Cladistics* **3,** 210–253.
26. Bremer, K. (1994) *Asteraceae: Cladistics and Classification* (Timber Press, Portland, OR).
27. Karis, P. O., Källersjö, M. & Bremer, K. (1992) *Ann. Mo. Bot. Gard.* **79,** 416–427.
28. Doyle, J. J. & Doyle, J. A. (1987) *Phytochem. Bull.* **19,** 11–15.
29. Jansen, R. K. & Palmer, J. D. (1987) *Curr. Genet.* **11,** 553–564.
30. Hiratsuka, J., Shimada, H., Whittier, R., Ishibashi, T., Sakamota, M., Mori, M., Kondo, C., Honji, Y., Sun, C.-R., Meng, B.-Y., Li, Y.-Q., Kanno, A., Nishizawa, Y., Hirai, A., Shinonzaki, K. & Sugiura, M. (1989) *Mol. Gen. Genet.* **217,** 185–194.
31. Li, W.-H., Wu, C.-I. & Luo, C.-C. (1985) *Mol. Biol. Evol.* **2,** 150–174.
32. Kumar, S., Tamura, K. & Nei, M. (1992) *MEGA: Molecular Evolutionary Genetics Analysis* (Pennsylvania State Univ., University Park), Version 1.01.
33. Maddison, W. P. & Maddison, D. R. (1992) *MACLADE: Analysis of Phylogeny and Character Evolution* (Sinauer, Sunderland, MA), Version 3.0.
34. Swofford, D. L. (1991) *PAUP: Phylogenetic Analysis Using Parsimony* (Illinois Natural History Survey, Champaign), Version 3.1.
35. Maddison, D. R. (1991) *Syst. Zool.* **40,** 315–328.
36. Felsenstein, J. (1985) *Evolution* **39,** 783–791.
37. Hillis, D. M. (1987) in *Phylogenetic Analysis of DNA Sequences*, eds. Miyamoto, M. M. & Cracraft, J. (Oxford Univ. Press, New York), pp. 278–294.
38. Levinson, G. & Gutman, G. A. (1987) *Mol. Biol. Evol.* **4,** 203–221.
39. Olmstead, R. G. & Sweere, J. A. (1994) *Syst. Biol.* **43,** 467–481.
40. Scotland, R. W., Sweere, J. A., Reeves, P. A. & Olmstead, R. G. (1995) *Am. J. Bot.* **82,** 266–275.
41. Donoghue, M. J., Doyle, J. A., Gauthier, J., Kluge, A. G. & Rowe, T. (1989) *Annu. Rev. Ecol. Syst.* **20,** 431–460.
42. Michaels, H. J., Scott, K. M., Olmstead, R. G., Szaro, T., Jansen, R. K. & Palmer, J. D. (1993) *Ann. Mo. Bot. Gard.* **80,** 742–751.
43. Skvarla, J. J., Turner, B. L., Patel, V. C. & Tomb, A. S. (1977) in *Biology and Chemistry of the Compositae*, eds. Heywood, V. H., Harborne, J. B. & Turner, B. L. (Academic, London), Vol. 1, pp. 141–248.
44. Jansen, R. K. & Palmer, J. D. (1988) *Am. J. Bot.* **75,** 751–764.
45. Dittrich, M. (1977) in *Biology and Chemistry of the Compositae*, eds. Heywood, V. H., Harborne, J. B. & Turner, B. L. (Academic, London), Vol. 2, pp. 999–1015.
46. Karis, P. O. (1993) *Plant Syst. Evol.* **186,** 69–93.
47. Anderberg, A. A. (1989) *Can. J. Bot.* **67,** 2277–2296.
48. Eriksson, T. (1991) *Taxon* **40,** 33–39.