Two Chloroplast DNA Inversions Originated Simultaneously During the Early Evolution of the Sunflower Family (Asteraceae)

Ki-Joong Kim, Keung-Sun Choi,* and Robert K. Jansen†*

*School of Life Sciences and Biotechnology, Korea University, Seoul, Korea; and †Section of Integrative Biology and Institute of Cellular and Molecular Biology, University of Texas

The chloroplast DNA (cpDNA) inversion in the Asteraceae has been cited as a classic example of using genomic rearrangements for defining major lineages of plants. We further characterize cpDNA inversions in the Asteraceae using extensive sequence comparisons among 56 species, including representatives of all major clades of the family and related families. We determine the boundaries of the 22-kb (now known as 22.8 kb) inversion that defines a major split within the Asteraceae, and in the process, we characterize the second and a new, smaller 3.3-kb inversion that occurs at one end of the larger inversion. One end point of the smaller inversion is upstream of the *trnE-UUC* gene, and the other end point is located between the *trnC-GCA* and *rpoB* genes. Although a diverse sampling of Asteraceae experienced substantial length variation and base substitution during the long evolutionary history subsequent to the inversion events, the precise locations of the inversion end points are identified using comparative sequence alignments in the inversion regions. The phylogenetic distribution of two inversions is identical among the members of Asteraceae, suggesting that the inversion events likely occurred simultaneously or within a short time period shortly after the origin of the family. Estimates of divergence times based on *ndhF* and *rbcL* sequences suggest that two inversions originated during the late Eocene (38–42 MYA). The divergence time estimates also suggest that the Asteraceae originated in the mid Eocene (42–47 MYA).

Introduction

Chloroplast genome organization is highly conserved among land plants (Palmer 1991; Raubeson and Jansen 2005). Gene orders may sometimes be reversed by large inversions that are mediated by intramolecular recombination events (Ogihara, Terachi, and Sasakuma 1988; Hiratsuka et al. 1989). The low levels of homoplasy and the overall rarity of large inversions among land plant chloroplast genomes suggest that these types of rare genomic changes are very reliable phylogenetic markers (Raubeson and Jansen 2005). Several large inversions have proven to be useful phylogenetic markers in a number of land plant groups, including the three large flowering plant families: Asteraceae, Fabaceae, and Poaceae (Jansen and Palmer 1987*a*; Doyle et al. 1992, 1996). In the sunflower family (Asteraceae), Jansen and Palmer (1987*a*, 1987*b*) identified two major lineages based on the distribution of a 22-kb inversion. This ancient dichotomy in the family was later supported with morphological (Bremer 1987) and chloroplast DNA (cpDNA) sequence data (Kim et al. 1992; Kim and Jansen 1995).

The Asteraceae is one of the largest flowering plant families with approximately 1,535 genera and 23,000 species (Bremer 1994). The family includes many economically important species such as sunflower, lettuce, and artichoke, as well as many ornamentals. The Asteraceae has been the subject of intensive phylogenetic analyses using both morphological (Karis, Källersjö, and Bremer 1992) and molecular data (Kim et al. 1992; Kim and Jansen 1995). As a result, intrafamilial relationships among the major clades are relatively well established (Bremer et al. 1992; Bremer 1994; Kim and Jansen 1995). However, the times of origin and diversification of major clades of Asteraceae still remain controversial due in part to the uncertainty of the early fossil record.

The previous report of a cpDNA inversion from Asteraceae is derived from gene mapping using Southern hybridization (Jansen and Palmer 1987*a*, 1987*b*). Here we further characterize the inversion based on DNA sequence data. In addition, we identify a new 3.3-kb inversion that is coincident with one end point of the large inversion. Comprehensive sequence comparisons among 56 species of Asteraceae and related families enable the identification of the end points of the two inversions. We also estimate the times of origin for the inversion events using molecular clocks based on sequences of subunit six of chloroplast nicotinamide adenine dinucleotide (phosphate)H, NAD(P)H dehydrogenase (*ndhF*) and a large subunit of ribulose-1, 5-bisphosphate carboxylase/oxygenase (*rbcL*).

Materials and Methods

Sequence Determination and Gene Identification of *Lactuca sativa* Chloroplast Genome in the Inversion Regions

Four cloned cpDNA fragments (7.7, 7.2, 7.1, and 6.7 kb, fig. 1) containing the inversion end points of the *Lactuca sativa* chloroplast genome (Jansen and Palmer 1987*a*) were subcloned into pBluescript II vector using a combination of the four restriction enzymes *Bam*HI, *Cla*I, *Eco*RI, and *Hin*dIII. Vector-inserted cpDNA fragments were sequenced using the BigDye 3.0™ terminal cycle sequencing kit (Applied Biosystems, Foster City, Calif.) and an ABI 3700 sequencer. Sequences were assembled using Sequencher (version 4.1; Gene Codes Corporation, Ann Arbor, Mich.). Gene annotations and comparative sequence analyses were performed using Blast and open reading frame finder programs from National Center for Biotechnology Information and ClustalX (Thompson et al. 1997). Published chloroplast genome sequences of *Nicotiana* and *Panax* were used for comparative analyses (Shinozaki et al. 1986; Kim and Lee 2004). The locations and secondary structures of *trn* genes were estimated using tRNAscan-SE

Key words: chloroplast DNA inversion, nonparametric rate smoothing, molecular clock, Asteraceae.

E-mail: kimkj@korea.ac.kr.

Mol. Biol. Evol. 22(9):1783–1792. 2005
doi:10.1093/molbev/msi174
Advance Access publication May 25, 2005

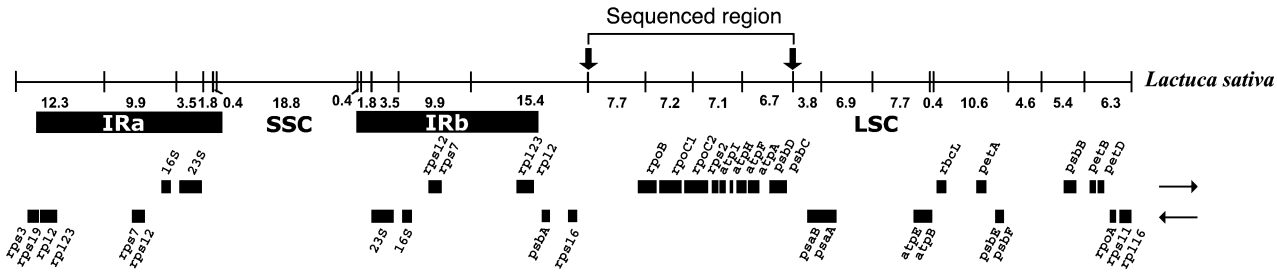

FIG. 1.—Physical map showing the cloned fragments and the gene orders of the *Lactuca sativa* chloroplast genome (adapted from Jansen and Palmer 1987*a*). Regions containing the four fragments (7.7, 7.2, 7.1, and 6.7 kb) were sequenced to characterize the two inversions. Inverted repeat (IR), small single copy (SSC), and large single copy (LSC) regions are shown below the cloned fragments. Arrows indicate the direction of transcription of chloroplast genes.

(version 1.21, Lowe and Eddy 1997) and MFOLD (version 3.0, Zuker 2003). Repeated sequences were identified using REPuter (Kurtz et al. 2001).

Sequence Determination of the Inversion End Points for 56 Species of Asteraceae and Related Families

Fifty-six species, representing all major clades of Asteraceae and seven related families, were selected for comparative sequencing of inversion end points. DNA was isolated from 1–3 g of leaf tissue using the cetyltrimethyl ammonium bromide (CTAB) method (J. J. Doyle and J. L. Doyle 1987), followed by purification using cesium chloride-ethidium bromide gradient ultracentrifugation (Palmer 1986). Six polymerase chain reaction (PCR) amplification primers were designed based on the sequence comparisons among three chloroplast genome sequences of *Lactuca* (this study), *Nicotiana* (Shinozaki et al. 1986), and *Panax* (Kim and Lee 2004). Positive and negative PCR amplifications were carried out using various combinations of the six primers. Each 50 μl reaction contained 2.5 mM $MgCl_2$, 0.2 mM deoxynucleoside triphosphate, 0.25 mM primers, 2.5 units of *Taq* polymerase, and 2–5 ng of DNA. The standard PCR amplification reactions were 30 cycles of 1 min denaturation at 94°C, 1 min annealing at 53°C, and 2 min extension at 72°C. PCR-amplified DNA was purified using the QIAquick PCR purification kit (Qiagen, Hilden, Germany) and sequenced directly using the two amplification primers. Sequence assemblies, alignments, annotations, and searches for repeats followed the above-mentioned methods.

Estimation of the Times of Origin for Two Inversions

Forty-two *ndhF* sequences representing all major clades of Asteraceae were obtained from our previously published data (Kim and Jansen 1995). Forty-two *rbcL* sequences were either obtained from our published data (Kim et al. 1992) or generated in this study. In order to make an identical taxon data set for two genes, a total of 11 new *rbcL* sequences were generated in this study by PCR amplification and sequencing (Olmstead et al. 1993) because only limited *rbcL* sequences were available. In addition, three previously published incomplete sequences were sequenced again. As a result, 14 new *rbcL* sequences were deposited in GenBank: *Barnadesia caryophylla*, AY874427; *Dasyphyllum argenteum*, AY874428; *Chuquiraga jussieui*, AY874429; *Doniophyton anomalum*, AY874430; *Schlechtendalia luzulifolia*, AY874431; *Ainsliaea acerifolia*, AY874432; *Gochnatia paucifolia*, AY874433; *Nassauvia gaudichaudii*, AY874434; *Onoseris hyssopifolia*, AY874435; *Cirsium texanum*, AY874436; *L. sativa*, AY874437; *Inula sericea*, AY874438; *Pluchea sericea*, AY874439; and *Psilostrophe gnaphalodes*, AY874440.

We utilized two different approaches for molecular clock estimates, tree distance and nonparametric rate smoothing (NPRS). For the tree distance approach, phylogenetic trees were constructed from *ndhF* and *rbcL* by neighbor-joining (NJ) using Li distance (Li 1993). The two gene trees were almost identical in topology (not shown). Nine major nodes from the *ndhF* and *rbcL* trees were selected for molecular clock assessments. One of these corresponds to the origin of the two inversions. Average of synonymous (Ks) and nonsynonymous (Ka) substitution rates were calculated for the nine major nodes in all possible pairwise combinations using MEGA2 (Kumar et al. 2001). Two independent clock calibrations from Poaceae (Wolfe et al. 1989) and Oleaceae (unpublished data) were employed to calibrate the Asteraceae molecular clock because reliable internal calibration points were not known from the fossil record.

For the NPRS method, the maximum likelihood (ML) tree using the TVM-I-T model, which was selected using Modeltest (Posada and Crandall 1998), was reconstructed from the combined *rbcL* and *ndhF* sequence data. Branch lengths of ML tree were adjusted using r8s program (Sanderson 2002) and assigned divergence times using an internal calibration fossil point of *Cornus* (Cornaceae) from the out-groups.

Results

Characterization of Two Inversions in the Chloroplast Genome of *L. sativa*

Complete sequences extending from the *rps16* intron to the *psbC* gene were generated from the chloroplast genome of *L. sativa*. The sequence (GenBank accession number AY865171) is 28,702 bp long and includes 14 protein-coding genes and 9 *trn* genes (fig. 2). Three genes, *atpF*, *rpoC1*, and *trnG-UCC*, contain introns. Based on comparisons of *Lactuca*, *Panax*, and *Nicotiana*, two inversions of 22.8 kb and 3.3 kb are present in the chloroplast genome of *L. sativa* (fig. 2). One end point of the large 22.8-kb (originally estimated as 22 kb, Jansen and Palmer 1987*a*) inversion is located between the *trnS-GCU* and

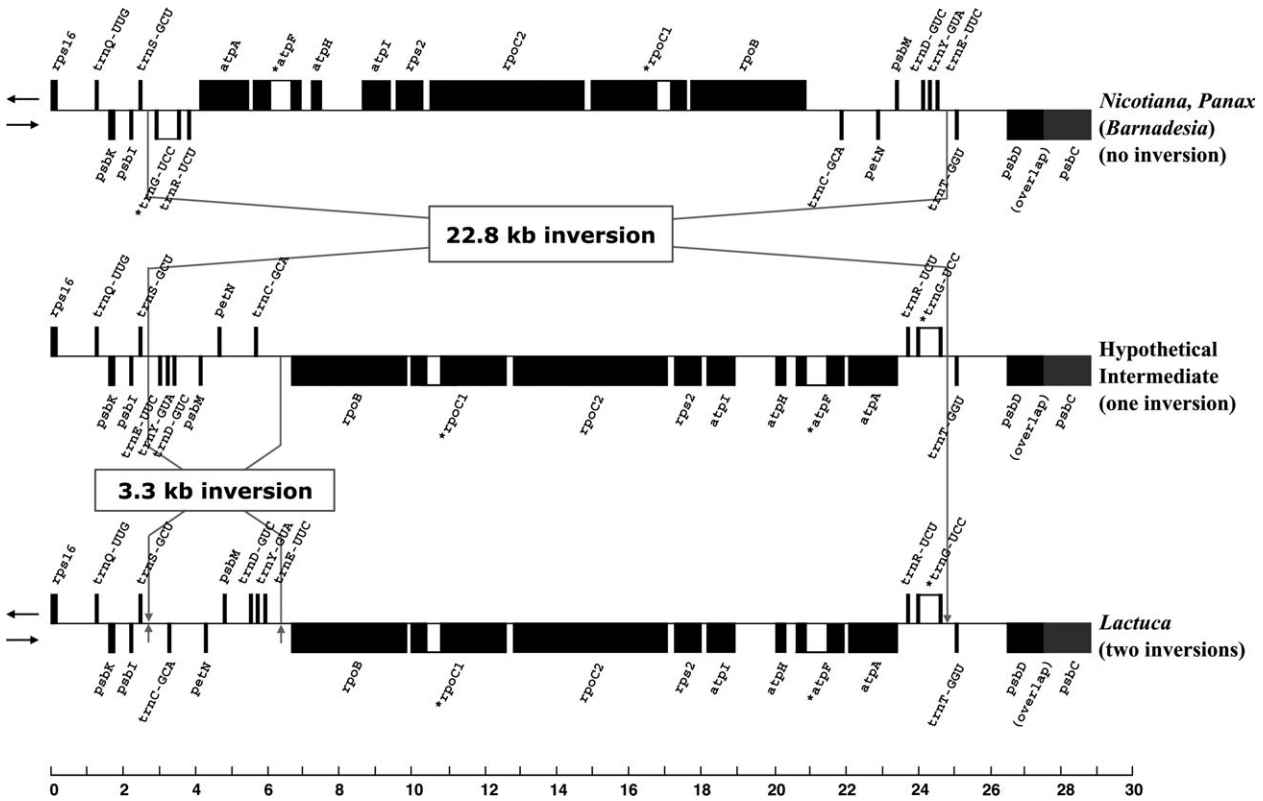

FIG. 2.—Comparative gene maps showing the two inversions in the *Lactuca sativa* chloroplast genome compared to those of *Nicotiana*, *Panax*, and *Barnadesia* genomes. The 22.8-kb inversion of *Lactuca* is located between the *trnG-UCC* and *trnE-UUC* genes. A new 3.3-kb inversion is nested within the 22.8-kb inversion and shares one inversion end point upstream of *trnS-GCU*. Horizontal arrows indicate the direction of transcription and the vertical arrows indicate the end points of inversions.

trnG-UCC genes. The other end point is located between the *trnE-UUC* and *trnT-GGU* genes. A new 3.3-kb inversion is nested within the 22.8-kb inversion, and it shares one end point just upstream of the *trnE-UUC* gene with the large inversion. The other end point of the 3.3-kb inversion is located between the *trnC-GCA* and *rpoB* genes (fig. 2).

Phylogenetic Distribution of the Two Inversions by PCR Diagnosis

We designed six primers to amplify the inversion end point regions (fig. 3). Different combinations of these primers were used in PCR reactions to determine the phylogenetic distribution of the two Asteraceae inversions. A positive PCR amplification would be expected from the primer combinations of P1/P4, P5/P3, and P2/P6 for species with both inversions, such as *Lactuca* (fig. 3, bottom). In contrast, for the species without the two inversions, such as *Nicotiana* and *Barnadesia*, a positive PCR reaction would result from the primer combinations P1/P2, P3/P4, and P5/P6 (fig. 3, top). Finally, the primer combinations of P1/P5, P3/P4, and P2/P6 would produce a positive PCR reaction if the species has only the 22.8-kb inversion. Thus, the different primer pairs produce both positive and negative results depending on the number of inversions.

Positive and negative PCR results for 11 representative species are shown in figure 4. Figure 4(*A*, *B*, and *C*) illustrates positive amplification results for species without any inversions (lanes 2–7) and negative results for species with inversions (lanes 8–12). In contrast, figure 4(*D*, *E*, and *F*) shows negative amplification results for species without inversions (lanes 2–7) and positive results for species with inversions (lanes 8–12). We attempted amplifications for the three inversion end points using all six different combinations of the primers for the 56 species of Asteraceae and related families (table 1). The results indicate that the distribution pattern of the two inversions is identical, with all related families and the subfamily Barnadesioideae lacking both inversions, whereas all other members of Asteraceae have both inversions.

Determination of the Exact Location of the Three Inversion End Points

The lengths of PCR products from the six primers flanking the inversion end points range from 650 to 1,600 bp, depending on the primer pairs used in the PCR reaction and species examined. To identify the precise location of inversion end points, we sequenced all 168 amplified DNA fragments (56 species × 3 regions, table 1).

Sequence alignments were performed in two steps. First, we divided the species into two groups based on the presence or absence of the two inversions. Alignments were subsequently performed within each group. The sequences from each of the three end point regions were aligned into six different profiles. Second, two alignment profiles for the same primer regions were combined and

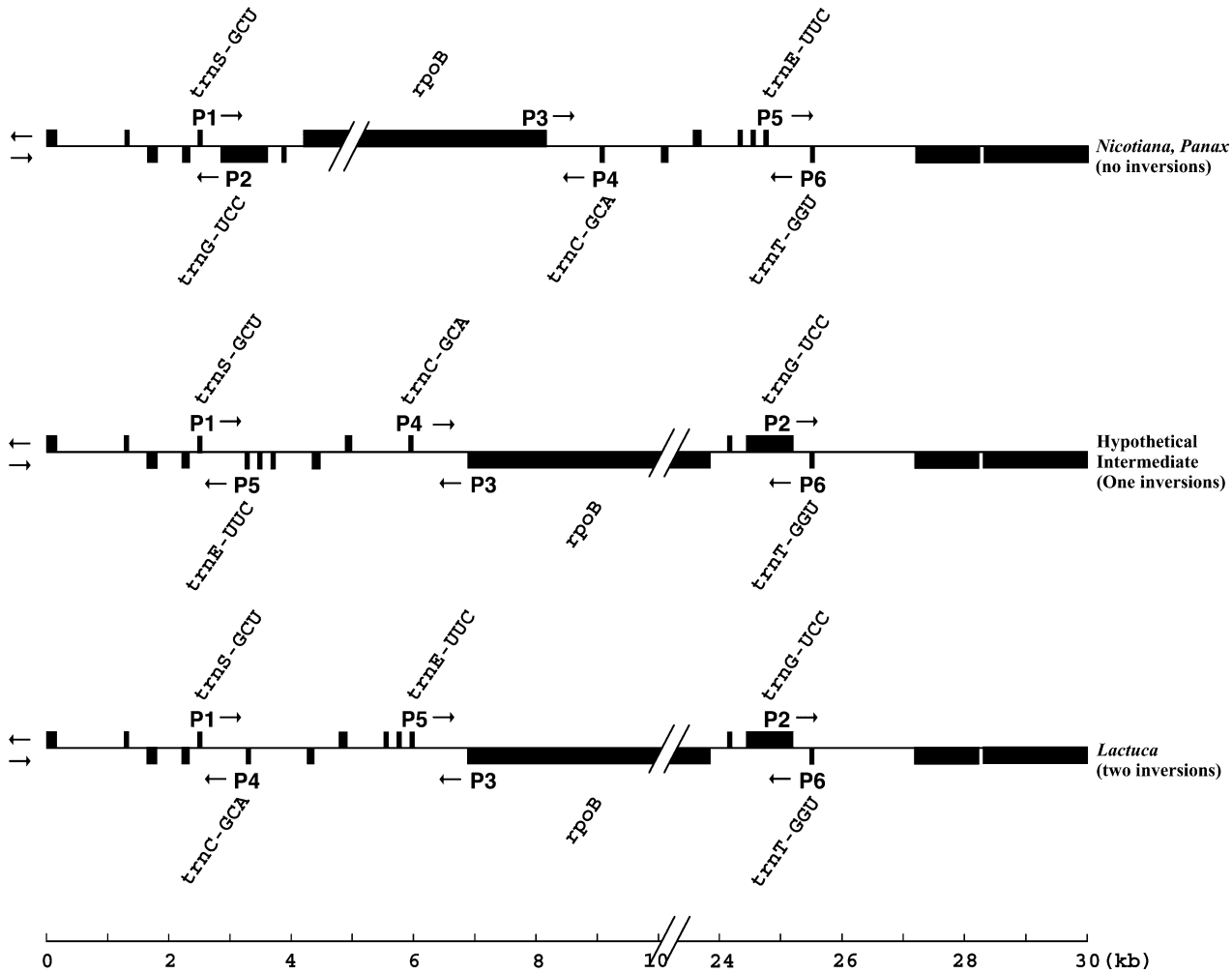

FIG. 3.—Locations of six diagnostic PCR primers for the three inversion end points. Taxa with two inversions, such as *Lactuca*, were amplified for the three end points using the primer combinations P1/P4, P5/P3, and P2/P6, respectively, while the taxa without inversions, such as *Barnadesia*, were amplified using the primer combinations P1/P2, P3/P4, and P5/P6, respectively. Primer sequences are (parentheses indicate degenerate sites): P1: 5′-AACCCTCGGTACGAATAACT-3′, P2: 5′-TT(G)C(A)TACCCATAACATCTATGTCAGCT-3′, P3: 5′-CCCTGATCAATGAACCTACA-3′, P4: 5′-GATTTGAACTGGGGAAAAAG-3′, P5: 5′-GCGTAGACATATTGC(T)CAACGAATTTACAGT-3′, and P6: 5′-AGCCCCTTATCG-GATTTGAACCGAT(G)G-3′.

realigned in both forward and reverse orientations, depending on the primers involved.

To identify the first inversion end points, sequences from primers P1/P4 for the 44 species with inversions and sequences from P1/P2 for the 12 species without inversions were aligned (see fig. 5 for aligned sequences of eight representative taxa). Sequences were aligned easily up to 283 bp upstream from primer P1 (ranging from 152 to 316 bp, depending on the species) in the *Lactuca* sequence, although several short gaps were required. However, the alignment of sequences between the two groups beyond this region was not possible because of length variation and high levels of sequence divergence. For the reverse orientation, sequences were aligned for the P1/P4 fragment of species with inversions and for the P3/P4 region of species without inversions. The sequences were aligned up to within an average of 24 bp from the P4 primer site (ranging from 432 to 604 bp, depending on the species) in the *Barnadesia* sequence. Alignment of sequences between the two groups beyond this region was not possible (fig. 5). The sequence AATTC overlaps on the two different orientations of these alignments. This overlapping sequence, which corresponds to base positions 229–233 upstream of *trnS-GCU* on the *Lactuca* chloroplast genome, is the precise location of the first inversion end point. To identify the second inversion end point, sequences from P5/P3 fragment for 44 species with the inversions and the sequences of the P5/P6 region for the 12 species without inversions were aligned. These sequences could be aligned up to an average of 235 (±49) bp from the P5 primer site (ranging from 144 to 329 bp, depending on the species). For the reverse orientation, sequences were aligned from the P5/P3 fragment for species with inversions and from the P3/P4 region for species without the inversions. These sequences were alignable up to an average of 404 (±26) bp beyond the P3 primer site (ranging from 325 to 446 bp, depending on the species). The second inversion end point cannot be located precisely due to the uncertainty of sequence alignment among the 56 species examined. This end point is located between base positions 19 and 529 upstream of *trnE-UUC* (or between

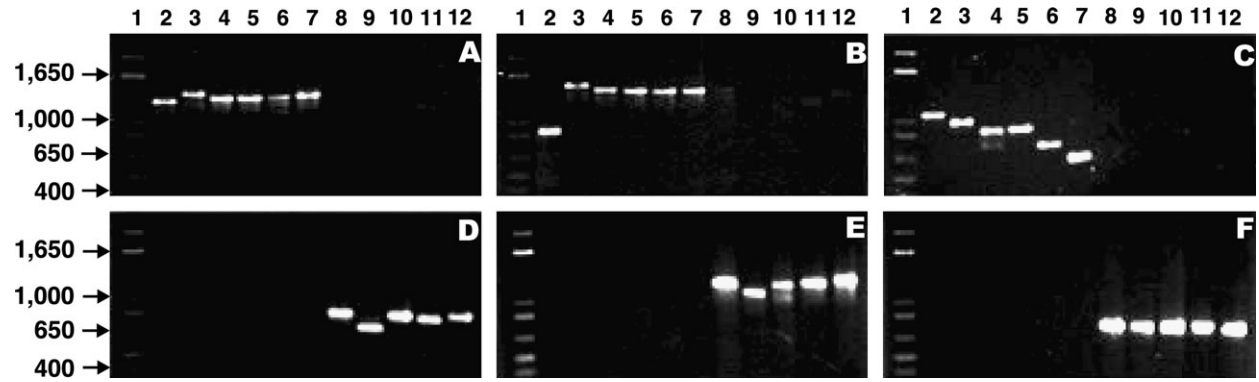

FIG. 4.—Some examples of positive/negative PCR amplifications for the presence/absence of inversions under the various combinations of primers. The primer combinations are P1/P2 in A, P3/P4 in B, P5/P6 in C, P1/P4 in D, P5/P3 in E, and P2/P6 in F, respectively. Only 11 representative taxa are shown: lane 1, 1 kb ladder; 2, *Lobelia cardinalis*; 3, *Pittosporum tobira*; 4, *Dasyphyllum argenteum*; 5, *Barnadesia caryophylla*; 6, *Chuquiraga jussieui*; 7, *Doniophyton anomalum*; 8, *Tarchonanthus camphoratus*; 9, *Cirsium texanum*; 10, *Lactuca sativa*; 11, *Dendranthemum grandiflorum*; 12, *Helianthus annuus*.

base positions 402 and 912 upstream of *rpoB*) on the *L. sativa* chloroplast genome. The broad range of uncertainty up to 510 bp is due to the high incidence of indels and base substitutions in this region.

To identify the third inversion end point, sequences from P2/P6 region for the 44 species with inversions and the sequences from P1/P2 for 12 species without inversions were aligned. These sequences were alignable up to an average 606 (±14) bp region from the P2 primer site (ranging from 570 to 634 bp, depending on the species). For the reverse orientations, the sequences from P2/P6 for species with inversions and the sequences from the P5/P6 fragment for species without inversions were aligned. The sequences were aligned up to an average 110 (±14) bp region from the primer P6 site (ranging from 66 bp to 14 bp, depending on the species). The third inversion end point is 90 bp upstream of the *trnG-UCC* gene (or 80 bp upstream of the *trnT-GGU*) on the *Lactuca* chloroplast genome.

Estimation of the Time of Origin of Inversions

Tree Distance Method

Relative rate tests using the NJ tree from both *ndhF* and *rbcL* sequences indicate significant rate heterogeneity. Therefore, the sequence data were partitioned into synonymous (Ks) and nonsynonymous (Ka) sites. The Ka sites show significant rate heterogeneity, whereas Ks sites have acceptable ranges of rate homogeneity at the 95% significance level (data not shown). Thus, we only used the Ks sites for molecular clock estimations. The Ks values of *ndhF* for the major branching events of Asteraceae are given in table 2. Because there are no unequivocal fossils for Asteraceae, two independent clocks from Poaceae and Oleaceae were used. This approach is appropriate for two reasons: (1) the fossil record of these two families is relatively well known (Muller 1981; Crepet and Feldman 1991) and (2) data from several different plant groups suggest that substitution rates may correlate with generation time (Gaut et al. 1992, 1996). The Poaceae clock (Ks = 0.1757 ± 0.0204 substitutions per 60 MYA) is derived from annual species (Wolfe et al. 1989; Crepet and Feldman 1991), while the Oleaceae clock (Ks = 0.1596 ± 0.0176 substitutions per 60 MYA) is derived primarily from woody perennials. The Asteraceae includes both annual and perennial herbs and woody species. If we accept the correlation between generation time and rates of base substitution, a clock from annual species, such as Poaceae, may result in an underestimate of the actual times of divergence for the Asteraceae. In contrast, a clock from woody species, such as Oleaceae, would overestimate the actual divergence times. The use of both of these clocks provides upper and lower bounds for estimating divergence times. Estimates of divergence times for the nine major diversification events of Asteraceae (tree not shown) are given in table 2. As expected, the Oleaceae clock always estimates older divergence times than the Poaceae clock. These estimates indicate that the Asteraceae originated in the mid Eocene (45–49 MYA, event 2 in table 2) and that the two chloroplast genome inversions occurred in the late Eocene/early Oligocene when the Barnadesioideae diverged from the rest of the Asteraceae (36–39 MYA, event 3 in table 2). In addition, most tribal splits of Asteraceae occurred during the Oligocene (28–36 MYA, events 5–8 in table 2).

NPRS Method

The branch lengths of ML trees from the combined sequences of *rbcL* and *ndhF* genes for 42 Asteraceae and related out-groups were adjusted using the NPRS method (Sanderson 2002), and evolutionary times were estimated using a calibration from one of the out-groups, Cornaceae (Takahashi, Crane, and Manchester 2002) (fig. 6). Nine major evolutionary events are also indicated in figure 6. These estimates indicate that the Asteraceae originated in the mid Eocene (42–48 MYA, event 2 in table 2 and fig. 6) and that the two chloroplast genome inversions occurred in the late Eocene/early Oligocene when the Barnadesioideae diverged from the rest of the Asteraceae (38–42 MYA, event 3 in table 2 and fig. 6). In addition, the divergence of most tribes of Asteraceae occurred during the Oligocene (24–38 MYA, events 5–8 in table 2 and fig. 6). The time estimates using the NPRS method were very similar to those of Ks distance-based method even though different calibrations were adopted (table 2).

Table 1
Taxa Used for PCR Amplication and Sequencing of Three Inversion End Points

Groups or Tribes	Scientific Names and Vouchers[a]	Amplification Primers[b]					
		P1 + P2	P3 + P4	P5 + P6	P1 + P4	P5 + P3	P2 + P6
Out-group	*Campanula ramulosa* (Jansen 984)	+w	+w	AY865321	–	–	–
	Scaevola frutescens (MBG 19426)	–	–	+	+	–	–
	Dampiera diversifolia (KEW 420-84-04494)	–	+	AY865323	+	–	–
	Platycodon grandiflorum (K.-J. Kim 13670)	AY865175	+	–	–	–	–
	Rhododendron schlippenbachii (K.-J. Kim 13671)	+	+	+	–	–	–
	Lobelia cardinalis (UCONN GH s.n.)	AY865176	+	+	–	–	–
	Pittosporum tobira (K.-J. Kim 13674)	AY865173	AY865267	AY865320	–	–	–
Barnadesieae	*Dasyphyllum argenteum* (Stuessy & Viteri 12464)	AY865177	AY865271	AY865319	–	–	–
	Barnadesia caryophylla (KEW 001-76-0038)	AY865178	AY865268	AY865322	–	–	–
	Schlechtendalia luzulifolia (Stuessy & Katinas 12810)	+w	AY865270	AY865324	–	–	–
	Chuquiraga jussieui (Stuessy *et al.* 12410)	AY865174	AY865269	AY865318	–	–	–
	Doniophyton anomalum (Stuessy & Ruiz 12780)	AY865172	AY865272	AY865317	–	–	–
Mutisieae	*Stifftia chrysantha* (KEW 386-39-38601)	–	–	–	AY865179	AY865273	AY865223
	Trixis californica (Keil 18528)	–	–	–	AY865180	AY865274	AY865224
	Perezia microcephala (UC 65-1497)	–	–	–	AY865181	AY865275	AY865225
	Acourtia microcephala (Keil 18945)	–	–	–	AY865182	AY865276	AY865226
	Mutisia acuminata (UC 64-1510)	–	–	–	AY865183	AY865277	AY865227
	Leibnitzia anandria (K.-J. Kim 13564)	–	–	–	AY865184	AY865278	AY865228
	Gerbera jamesonii (Jansen 915)	–	–	–	AY865185	AY865279	AY865229
Tarchonantheae	*Tarchonanthus camphoratus* (UC 48-0777)	–	–	–	AY865186	AY865280	AY865230
Cardueae	*Echinops exaltatus* (Jansen 1001)	–	–	–	AY865187	AY865281	AY865231
	Carlina vulgaris (F. Hellwing s.n.)	–	–	–	AY865188	AY865282	AY865232
	Atractylodes japonica (K.-J. Kim 13551)	–	–	–	AY865189	AY865283	AY865233
	Cirsium texanum (K.-J. Kim 10693)	–	–	–	AY865190	AY865284	AY865234
	Synurus deltoids (K.-J. Kim 13550)	–	–	–	AY865191	AY865285	AY865235
	Centaurea americana (K.-J. Kim 13030)	–	–	–	AY865192	AY865286	AY865236
Lactuceae	*Lactuca sativa* (no voucher, cultivated)	–	–	–	AY865193	AY865287	AY865237
	Picris altissima (KEW s.n.)	–	–	–	AY865194	AY865288	AY865238
Arctoteae	*Eremothamnus marlothianus* (Giess & van Vuux 410)	–	–	–	AY86519	AY865289	AY865239
	Arctotis stoechidifolia (Jansen 920)	–	–	–	AY865196	AY865290	AY865240
Liabeae	*Liabum glabrum* (Panero 2554)	–	–	–	AY865197	AY865291	AY865241
	Sinclairia pringlei (Panero 2437)	–	–	–	AY865198	AY865292	AY865242
Vernonieae	*Stokesia laevis* (K.-J. Kim 13591)	–	–	–	AY865199	AY865293	AY865243
Senecioneae	*Syneilesis palmata* (K.-J. Kim 13558)	–	–	–	AY865200	AY865294	AY865244
	Senecio mikanioides (MBG 72326)	–	–	–	AY865201	AY865295	AY865245
	Corymbium glabrum (Anderberg s.n., S)	–	–	–	AY865202	AY865296	AY865246
Gnaphalieae	*Antennaria neodioica* (Bayer M1386)	–	–	–	AY865203	AY865297	AY865247
	Anisothrix integra (Anderberg s.n., S)	–	–	–	AY865204	AY865298	AY865248
	Gnaphalium japonicum (K.-J. Kim 13707)	–	–	–	AY865205	AY865299	AY865249
Calenduleae	*Dimorphotheca pluvialis* (Jansen 902)	–	–	–	AY865206	AY865300	AY865250
	Calendula officinalis (Jansen 903)	–	–	–	AY865207	AY865301	AY865251
Anthemideae	*Ursinia nana* (Jansen 913)	–	–	–	AY865208	AY865302	AY865252
	Achillea millefolium (Jansen 911)	–	–	–	AY865209	AY865303	AY865253
	Dendranthemum grandiflorum (K.-J. Kim 13200)	–	–	–	AY865210	AY865304	AY865254
Astereae	*Bellis perennis* (Jansen 916)	–	–	–	AY865211	AY865305	AY865255
	Aster cordifolius (Jansen 906)	–	–	–	AY865212	AY865306	AY865256
	Erigeron x hybridus (Jansen 999)	–	–	–	AY865213	AY865307	AY865257
Inuleae	*Plunchea sericea* (UC 86-0189)	–	–	–	AY865214	AY865308	AY865258
	Nauplius sericeus (Anderberg s.n., S)	–	–	–	AY865215	AY865309	AY865259
Athroisma group	*Athroisma gracile* (Eriksson 631)	–	–	–	AY865216	AY865310	AY865260
	Blepharispermum zanguebaricum (Eriksson 602)	–	–	–	AY865217	AY865311	AY865261
Coreopsideae	*Coreopsis tinctoria* (K.-J. Kim 12007)	–	–	–	AY865218	AY865312	AY865262
Heliantheae	*Ambrosia trifida* (K.-J. Kim 10710)	–	–	–	AY865219	AY865313	AY865263
	Helianthus annuus (Price Ha89)	–	–	–	AY865220	AY865314	AY865264
Eupatorieae	*Eupatorium atrorubens* (=*Hebeclinium atrorubens*) (Jansen 908)	–	–	–	AY865221	AY865315	AY865265
	Tagetes erecta (Jansen 905)	–	–	–	AY865222	AY865316	AY865266

NOTE.—Primer sequences and positions are indicated in fig. 3.

[a] KEW, Royal Botanic Gardens at Kew, United Kingdom; MBG, Matthaei Botanical Garden, University of Michigan at Ann Arbor, USA; UC, The University of California Botanical Garden at Berkeley, USA; UCONN GH, University of Connecticut Greenhouse at Storrs, USA.

[b] Amplified and sequenced taxa are indicated by their GenBank accession numbers (AY865172–AY865324). +, amplified but not sequenced; +W, weakly amplified and not sequenced; and –, nonamplified.

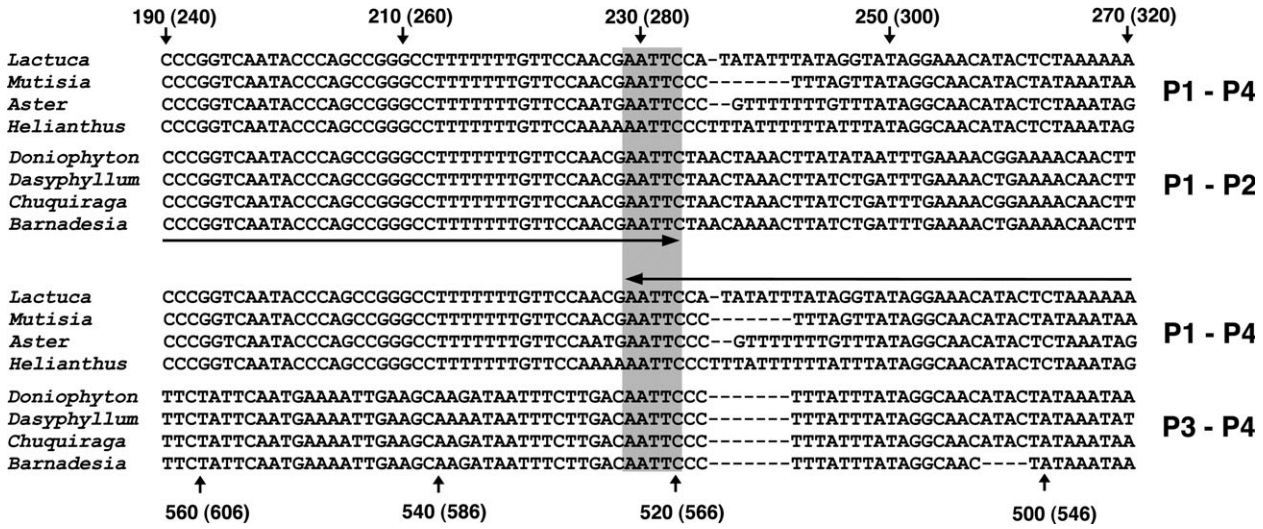

FIG. 5.—Sequence alignments showing one of three inversion end points (shaded area). Only 8 of 56 sequences are shown. Sequences of the PCR product from primers P1/P4 for the taxa with inversions and using the primers P1/P2 for the taxa without inversions were aligned in the region to the right arrow in the upper panel. Amplified sequences using the primers P1/P4 for the taxa with inversions and using the primers P3/P4 for the taxa without inversions were aligned in the region to the left arrow in the lower panel. The sequence AATTC overlaps in both directions and corresponds to the inversion end point. Numbers above the figure indicate the base positions upstream from *trnS-GCU* (coordinates from the primer P1 site are given in parentheses) on the *Lactuca* sequence, while the numbers below the figure indicate the base positions upstream from *trnC-GCA* (base positions from the primer P4 site were given in parentheses) on the *Barnadesia* sequence.

Discussion

Two Inversions Occurred Simultaneously

Two cpDNA inversions of 22.8 and 3.3 kb are shared by all major clades of Asteraceae, except members of Barnadesioideae (table 1). The larger inversion (previously estimated to be 22 kb, Jansen and Palmer 1987*a*) is 22,830 bp in length, and the second, newly identified inversion is 3.3 kb long and is nested within the large inversion (fig. 2). The phylogenetic distribution of these two inversions (table 1) is identical among members of Asteraceae, and both events occurred during a very short time period (events 3 and 5 in fig. 6) after the evolutionary split of the Barnadesioideae and the rest of the Asteraceae and prior to the subsequent rapid radiation into the other subfamilies and tribes. The identical phylogenetic distribution and brief evolutionary timescale suggest that these inversions happened simultaneously or over a very short time span.

Inversions Originated Only Once During the Early Evolution of Asteraceae

Chloroplast gene order is highly conserved among land plants (Palmer 1991; Raubeson and Jansen 2005),

Table 2
Age Estimates for the Nine Major Evolutionary Events of Asteraceae and Related Families

Average Ks Value (100×)[a]	Number of Comparisons	Estimated Times from the Ks Values (MYA)[b]	Estimated Times from the ML-NPRS Tree (MYA)[c]	Major Branching Events[d]
21.18 ± 2.02	38	80–72	53–49	Goodeniaceae/Asteraceae (1)
13.08 ± 1.56	37	49–45	48–42	Calyceraceae/Asteraceae (origin of Asteraceae) (2)
10.43 ± 0.97	155	39–36	42–38	Barnadesioideae/rest of Asteraceae (origin of two inversions) (3)
9.60 ± 0.80	184	36–33	38–35	Cardueae + Mutisieae/LALV[e] + Asteroids (5)
10.21 ± 0.79	120	38–32	35–32	LALV tribes/Asteroid tribes (6)
8.12 ± 0.89	15	31–28	29–24	LALV tribes (7)
10.30 ± 0.79	162	39–35	29–26	Asteroid tribes (8)
6.37 ± 0.94	6	24–22	35–30	Barnadesioideae genera (4)
5.65 ± 1.05	12	19–21	20–17	Helianthoid tribes (9)

NOTE.—Age estimates were derived from average synonymous substitutions (Ks) of *ndhF* gene in conventional distance-based molecular clock method and from combined whole sequence data set of *ndhF* and *rbcL* in NPRS-based molecular clock method.

[a] Ks values were calculated following Li (1993) from the NJ tree of 42 taxa.

[b] The lower and the upper limits of times were estimated by the Oleaceae and the Poaceae calibration clocks, respectively (see text), from the NJ tree.

[c] The times were estimated using the *Cornus* calibration clock from the ML-NPRS tree (see fig. 6).

[d] Major branching events were also given in figure 6 under the same numbers1.

[e] LALV = Lactuceae, Arctoteae, Liabeae, and Vernonieae.

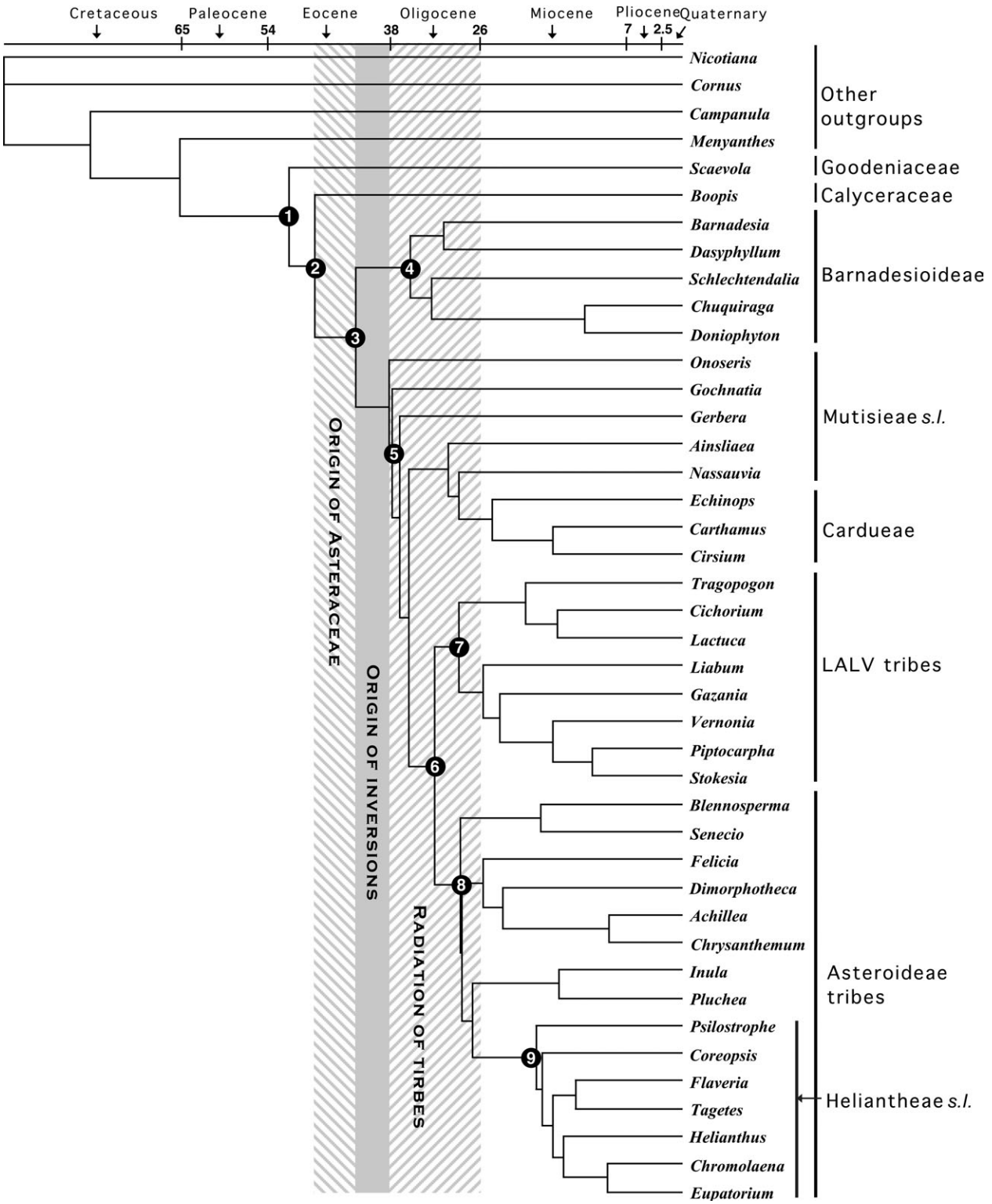

FIG. 6.—Phylogenetic tree of Asteraceae and related families with ages estimated according to the NPRS methods. The times were calibrated using the *Cornus* fossil as a reference. Events 2 and 3 correspond to the time of origin of Asteraceae and the time of origin of two chloroplast inversions, respectively.

but in most instances when changes do occur, they involve one or few inversions (Jansen and Palmer 1987*b*; Doyle et al. 1992; Raubeson and Jansen 1992). However, there are several groups of land plants that have experienced substantial numbers of cpDNA rearrangements, including conifers (Tsumura, Suyama, and Yoshimura 2000) and the angiosperm families Campanulaceae (Cosner et al. 1997; Cosner, Raubeson, and Jansen 2004), Fabaceae

(Milligan, Hampton, and Palmer 1989), Geraniaceae (Palmer, Nugent, and Herbon 1987), and Lobeliaceae (Knox, Downie, and Palmer 1993; Knox and Palmer 1999). Gene order changes in highly rearranged genomes are often associated with repeated sequences, a feature that is considered uncommon in chloroplast genomes (Palmer 1991).

The rarity of inversions in chloroplast genomes has made these characters powerful phylogenetic markers. Evidence for homoplasy in cpDNA inversions has been suggested in three groups (Downie and Palmer 1994; Hoot and Palmer 1994; Cosner, Raubeson, and Jansen 2004), and intrapopulational polymorphism has been documented in conifers (Tsumura, Suyama, and Yoshimura 2000). However, even in the highly rearranged genomes of Campanulaceae, the levels of homoplasy are extremely low and are far less than DNA sequences for the same taxa (Cosner, Raubeson, and Jansen 2004). Furthermore, the precise location of inversion end points has not been identified in any of these groups by sequence data. Thus, definitive cases of homoplasy based on DNA sequences of genomes with inversions have not been demonstrated.

Extensive comparative sequence analyses among species with and without inversions are needed for the precise identification of inversion end points. Our sequence comparisons of 56 species, including the 12 species without the two inversions and the 44 species with inversions, identified the exact location of two of the three inversion end points. The third end point could only be located within a 510-bp region because of the large number of indels and highly divergent levels of sequence variation between the *trnE-UUC* and *rpoB* genes. Thus, our sequence data indicate that the two inversions in the Asteraceae represent homologous changes that have a single origin.

The phylogenetic distribution of the two inversions in the Asteraceae is concordant with the recent molecular (Kim et al. 1992; Kim and Jansen 1995) and morphological (Bremer 1987) phylogenies, which indicate that the subfamily Barnadesioideae is sister to the rest of the family.

Two Asteraceae Inversions Originated During the Late Eocene

Divergence time estimates suggest that the basal evolutionary split in the Asteraceae occurred in the late Eocene (approximately 36–42 MYA). Thus, the two inversion events also must have originated at or near this same time period (table 2 and fig. 6). The molecular clock comparisons also suggest that the Asteraceae originated during the mid Eocene (approximately 42–49 MYA, fig. 6) and that the divergence of the major tribal lineages, with the exception of the Heliantheae group, diverged immediately after the basal split between the Barnadesioideae and the rest of the Asteraceae. Thus, the Asteraceae experienced a rapid radiation during the Oligocene.

Despite the large number of extant species, the megafossil record of the Asteraceae is extremely sparse. The identity of many fossils once considered to be members of Asteraceae remains controversial (Crepet and Stuessy 1978; DeVore and Stuessy 1995). For example, a head-like inflorescence reported from the upper Oligocene was identified initially as an Asteraceae fossil, but later investigations indicated that the fossil could not be unequivocally assigned to this family (Crepet and Stuessy 1978). There is a substantial microfossil record for the Asteraceae, which consists primarily of pollen (Graham 1996). The oldest record for Asteraceae pollen is from the upper Eocene (ca., 42 MYA), and pollen becomes increasingly common and more widely distributed in the mid to late Oligocene (Muller 1981; Graham 1996).

The fact that pollen of the Barnadesioideae is not easily differentiated from the related families Calyceraceae and Goodeniaceae (Zhao et al. 2000) makes it difficult to accurately identify the earliest pollen of the Asteraceae. The huge increase of fossil Asteraceae pollen in the Miocene on many continents suggests a rapid diversification of the family during this time period. Alternatively, the high level of pollen diversity in the Miocene could suggest that the Asteraceae is much older (Turner 1977). In contrast to the enigmatic conclusions based on fossil data, our molecular clock estimates provide evidence for the times of the origin and diversification of Asteraceae. Our results also indicate that the two cpDNA inversions in Asteraceae originated simultaneously during the late Eocene (36–42 MYA).

Acknowledgments

We thank A. Anderberg, T. Eriksson, J. Panero, F. Hellwig, and T. Stuessy for providing the plant material and H.-L. Lee for the drawings of figures. This research was supported by a grant (R01-1999-000-00063-0) from the Korea Science and Engineering Foundation and the Plant Signaling Network Research Center, Korea Science and Engineering Foundation, to K.-J.K. and an NSF grant (DEB-9020171) to R.K.J.

Literature Cited

Bremer, K. 1987. Tribal interrelationships of the Asteraceae. Cladistics **3**:210–253.

———. 1994. Asteraceae: cladistics and classification. Timber Press, Portland, Ore.

Bremer, K., R. K. Jansen, P. O. Karis, M. Källersjö, S. C. Keeley, K.-J. Kim, H. J. Michaels, J. D. Palmer, and R. S. Wallace. 1992. A review of the phylogeny and classification of the Asteraceae. Nord. J. Bot. **12**:141–148.

Cosner, M. E., R. K. Jansen, J. D. Palmer, and S. R. Downie. 1997. The highly rearranged chloroplast genome of *Trachelium caeruleum* (Campanulaceae): multiple inversions, inverted repeat expansion and contraction, transposition, insertions/deletions, and several repeat families. Curr. Genet. **31**:419–429.

Cosner, M. E., L. A. Raubeson, and R. K. Jansen. 2004. Chloroplast DNA rearrangements in Campanulaceae: phylogenetic utility of highly rearranged genomes. BMC Evol. Biol. **4**:1–17.

Crepet, W. L., and G. D. Feldman. 1991. The earliest remains of grasses in the fossil record. Am. J. Bot. **78**:1010–1014.

Crepet, W. L., and T. F. Stuessy. 1978. A reinvestigation of the fossil *Viguiera cronquistii* (Compositae). Brittonia **29**:137–153.

DeVore, M. L., and T. F. Stuessy. 1995. The place and time of origin of the Asteraceae, with additional comments on the Calyceraceae and Goodeniaceae. Pp. 23–40 *in* D. J. N. Hind, C. Jeffrey, and G. V. Pope, eds. Advances in Compositae systematics. Royal Botanic Gardens, Kew, United kingdom.

Downie, S. R., and J. D. Palmer. 1994. A chloroplast DNA phylogeny of the Caryophyllales based on structural and inverted repeat restriction site variation. Syst. Bot. **19**:236–252.

Doyle, J. J., J. I. Davis, R. I. Soreng, D. Garvin, and M. J. Anderson. 1992. Chloroplast DNA inversions and the origin of the grass family (Poaceae). Proc. Natl. Acad. Sci. USA **89**:7722–7726.

Doyle, J. J., and J. L. Doyle. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem. Bull. **19**:11–15.

Doyle, J. J., J. L. Doyle, J. A. Ballenger, and J. D. Palmer. 1996. The distribution and phylogenetic significance of a 50-kb chloroplast DNA inversion in the flowering plant family Leguminosae. Mol. Phylogenet. Evol. **5**:429–438.

Gaut, B. S., B. R. Morton, B. C. McCaig, and M. T. Clegg. 1996. Substitution rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene *Adh* parallel rate differences at the plastid gene *rbcL*. Proc. Natl. Acad. Sci. USA **93**:10274–10279.

Gaut, B. S., S. V. Muse, D. Clark, and M. T. Clegg. 1992. Relative rates of nucleotide substitution at the *rbcL* locus of monocotyledonous plants. J. Mol. Evol. **35**:292–303.

Graham, A. 1996. A contribution to the geologic history of the Compositae. Pp. 123–140 *in* D. H. N. Hind and H. J. Beentje, eds. Compositae: systematics. Proceedings of the international Compositae conference, Kew, 1994, Vol. 1. Royal Botanic Gardens, Kew, United Kingdom.

Hiratsuka, J., H. Shimada, R. Whittier et al. (15 co-authors). 1989. The complete sequence of the rice (*Oryza sativa*) chloroplast genome: intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. Mol. Gen. Genet. **217**:185–194.

Hoot, S. B., and J. D. Palmer. 1994. Structural rearrangements, including parallel inversions, within the chloroplast genome of *Anemone* and related genera. J. Mol. Evol. **38**:274–281.

Jansen, R. K., and J. D. Palmer. 1987*a*. A chloroplast DNA inversion marks an ancient evolutionary split in the sunflower family (Asteraceae). Proc. Natl. Acad. Sci. USA **84**: 5818–5822.

———. 1987*b*. Chloroplast DNA from lettuce and *Barnadesia* (Asteraceae): structure, gene localization, and characterization of a large inversion. Curr. Genet. **11**:553–564.

Karis, P. O., M. Källersjö, and K. Bremer. 1992. Phylogenetic analysis of the Cichorioideae (Asteraceae) with emphasis on the Mutisieae. Ann. Mo. Bot. Gard. **79**:416–427.

Kim, K.-J., and R. K. Jansen. 1995. *ndhF* sequence evolution and the major clades in the sunflower family. Proc. Natl. Acad. Sci. USA **92**:10379–10383.

Kim, K.-J., R. K. Jansen, R. S. Wallace, H. J. Michaels, and J. D. Palmer. 1992. Phylogenetic implications of *rbc*L sequence variation in the Asteraceae. Ann. Mo. Bot. Gard. **79**:428–445.

Kim, K.-J., and H.-L. Lee. 2004. Complete chloroplast genome sequences from Korean ginseng (*Panax schinseng* Nees) and comparative analysis of sequence evolution among 17 vascular plants. DNA Res. **11**:247–261.

Knox, E. B., S. R. Downie, and J. D. Palmer. 1993. Chloroplast genome rearrangements and the evolution of giant lobelias from herbaceous ancestors. Mol. Biol. Evol. **10**:414–430.

Knox, E. B., and J. D. Palmer. 1999. The chloroplast genome arrangement of *Lobelia thuliniana* (Lobeliaceae): expansion of the inverted repeat in an ancestor of the Campanulales. Plant Syst. Evol. **214**:49–64.

Kumar, S., K. Tamura, I. B. Jakobsen, and M. Nei. 2001. MEGA2: molecular evolutionary genetics analysis software. Bioinformatics **17**:1244–1245.

Kurtz, S., J. V. Choudhuri, E. Ohlebusch, C. Schleiermacher, J. Stoye, and R. Giegerich. 2001. REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res. **29**:4633–4642.

Li, W.-H. 1993. Unbiased estimation of the rates of synonymous and nonsynonymous substitution. J. Mol. Evol. **36**:96–99.

Lowe, T. M., and S. R. Eddy. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. **25**:955–964.

Milligan, B. G., J. N. Hampton, and J. D. Palmer. 1989. Dispersed repeats and structural reorganization in subclover chloroplast DNA. Mol. Biol. Evol. **6**:355–368.

Muller, J. 1981. Fossil pollen records of extant angiosperms. Bot. Rev. **47**:1–142.

Ogihara, Y., T. Terachi, and T. Sasakuma. 1988. Intramolecular recombination of chloroplast genome mediated by short direct-repeat sequences in wheat species. Proc. Natl. Acad. Sci. USA **85**:8573–8577.

Olmstead, R. G., B. Bremer, K. Scott, and J. D. Palmer. 1993. A parsimony analysis of the Asteridae *sensu lato* based on *rbcL* sequences. Ann. Mo. Bot. Gard. **80**:700–722.

Palmer, J. D. 1986. Isolation and structural analysis of chloroplast DNA. Pp.167–186 *in* A. Weissbach and H. Weissbach, eds. Methods in enzymology, Vol. 118. Academic Press, New York.

———. 1991. Plastid chromosomes: structure and evolution. Pp. 5–53 *in* I. K. Vasil and L. Bogorad, eds. Cell culture and somatic cell genetics in plants, Vol. 7A. The molecular biology of plastids. Academic Press, San Diego, Calif.

Palmer, J. D., J. M. Nugent, and L. A. Herbon. 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. USA **84**:769–773.

Posada, D., and K. A. Crandall. 1998. Modeltest: testing the model of DNA substitution. Bioinformatics **14**:817–818.

Raubeson, L. A., and R. K. Jansen. 1992. Chloroplast DNA evidence on the ancient evolutionary split in vascular land plants. Science **255**:1697–1699.

———. 2005. Chloroplast genomes of plants. Pp. 45–68 *in* R. Henry, ed. Diversity and evolution of plants—genotypic and phenotypic variation in higher plants. CABI Publishing, Oxfordshire, United Kingdom.

Sanderson, M. J. 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. Mol. Biol. Evol. **19**:101–109.

Shinozaki, K., M. Ohme, M. Tanaka et al. (23 co-authors). 1986. The complete nucleotide sequence of tobacco chloroplast genome: its gene organization and expression. EMBO J. **5**:2043–2049.

Takahashi, M., P. R. Crane, and S. R. Manchester. 2002. *Hironoia fusiformis* gen. et sp. nov.; a cornalean fruit from the Kamikitaba locality (upper Cretaceous, lower Coniacian) in northeastern Japan. J. Plant Res. **115**:463–473.

Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. **24**:4876–4882.

Tsumura, Y., Y. Suyama, and K. Yoshimura. 2000. Chloroplast DNA inversion polymorphism in populations of *Abies* and *Tsuga*. Mol. Biol. Evol. **17**:1302–1312.

Turner, B. L. 1977. Fossil history and geography. Pp. 19–39 *in* V. H. Heywood, J. B. Harborne, and B. L. Turner, eds. The biology and chemistry of the Compositae, Vol. 1. Academic Press, London.

Wolfe, K. H., M. Gouy, Y.-W. Yang, P. M. Sharp, and W.-H. Li. 1989. Date of the monocot-dicot divergence estimated from chloroplast DNA sequence data. Proc. Natl. Acad. Sci. USA **86**:6201–6205.

Zhao, Z., R. K. Jansen, J. Skvarla, and M. DeVore. 2000. Phylogenetic implications of pollen morphology and ultrastructure in the Barnadesioideae (Asteraceae). Lundellia **3**:26–40.

Zuker, M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. **31**:3406–3415.

Charles Delwiche, Associate Editor

Accepted May 12, 2005