

Chloroplast DNA from lettuce and *Barnadesia* (Asteraceae): structure, gene localization, and characterization of a large inversion

Robert K Jansen* and Jeffrey D. Palmer

Department of Biology, University of Michigan, Ann Arbor, MI 48109, USA

Summary. We have cloned into plasmids 17 of 18 lettuce chloroplast DNA *SacI* fragments covering 96% of the genome. The cloned fragments were used to construct cleavage maps for 10 restriction enzymes for the chloroplast genomes of lettuce (*Lactuca sativa*) and *Barnadesia caryophylla*, two distantly related species in the sunflower family (Asteraceae). Both genomes are approximately 151 kb in size and contain a 25 kb inverted repeat. We also mapped the position and orientation of 37 chloroplast DNA genes. The mapping studies reveal that chloroplast DNAs of lettuce and *Barnadesia* differ by a 22 kb inversion in the large single copy region. *Barnadesia* has retained the primitive land plant genome arrangement, while the inversion has occurred in a lettuce lineage. The endpoints of the derived lettuce inversion were located by comparison to the well-characterized spinach and tobacco genomes. Both endpoints are located in intergenic spacers within tRNA gene clusters; one cluster being located downstream from the *atpA* gene and the other upstream from the *psbD* gene. The endpoint near the *atpA* gene is very close to one endpoint of a 20 kb inversion in wheat (Howe et al. 1983; Quigley and Weil 1985). Comparison of the restriction site maps gives an estimated sequence divergence of 3.7% for the lettuce and *Barnadesia* genomes. This value is relatively low compared to previous estimates for other angiosperm groups, suggesting a high degree of sequence conservation in the Asteraceae.

Key words: Restriction maps – Gene organization – Chloroplast DNA – Inversion – Asteraceae

Introduction

Recent studies of over 200 species of flowering plants representing 30 families have revealed a high level of conservation of the chloroplast genome in terms of its size, organization, and primary sequence (reviewed in Whitfield and Bottomley 1983; Gillham et al. 1985; Palmer 1985a, b). All angiosperm chloroplast DNAs (cpDNAs) are circular molecules and most range in size from 135–160 kb. The chloroplast genomes of all but one group of legumes (Kolodner and Tewari 1979; Koller and Delius 1980; Palmer and Thompson 1981a, 1982; Chu and Tewari 1982) are organized into one large and one small single copy region separated by an inverted repeat ranging in size from 10–76 kb. Comparative gene mapping and cross-hybridization studies have shown that cpDNAs from 25 of the 30 examined families of flowering plants have the same genome arrangement (reviewed in Palmer 1985a, b). The gene order common to most angiosperms is also present in other land plants, including a fern and a gymnosperm (Palmer and Stein 1982, 1986). Rearrangements of varying complexity have been documented in five angiosperm families; the simpler alterations all appear to be inversions. The best-characterized of these with respect to its endpoint sequences is a 20 kb inversion in wheat (Howe et al. 1983; Quigley and Weil 1985; Howe 1985), which is also present in barley (Oliver and Poulson 1984) and maize (Palmer and Thompson 1982).

We are interested in comparing chloroplast DNAs of representative species of Asteraceae in order to assess phylogenetic relationships at higher taxonomic levels in this large and biologically diverse family of flowering plants. As a preliminary basis for these studies, we have constructed detailed restriction site and gene maps of the chloroplast genomes of two distantly related species, lettuce (*Lactuca sativa*; tribe Cichorieae) and *Barnadesia*

* Present address: Department of Ecology and Evolutionary Biology, University of Connecticut, CT 06268, USA

Offprint requests to: R. K. Jansen

Table 1. Sources of gene probes

Gene name	Species	Fragment ^a	Gene location ^b	Reference
3' rps12-5' rps7	tobacco	969 bp PstI-BamHI	-119/+19	d
3' rps7 ^c	tobacco	1,833 bp PstI-SalI	+19/++1379	d
rpl23-5' rpl2	tobacco	1,250 bp BamHI-PstI	-462/+486	e
3' rpl2-rps19-5' rpl22	tobacco	1,573 bp PstI	+486/+178	e
3' rpl22-5' rps3	tobacco	536 bp PstI-SalI	+178/+262	e
3' rps3-5' rpl16	tobacco	1,825 bp SalI-BamHI	+262/+29	e
3' rpl16-rpl14-5' rps8	tobacco	1,055 bp BamHI-SalI	+297/+250	e
3' rps8-infA	spinach	670 bp SalI	-279/++160	e, f
rps11	spinach	635 bp SalI-XbaI	-222/+413	f
rpoA ^c	spinach	1,040 bp XbaI	-78/+962	f
3' petD ^c	spinach	416 bp BamHI-XbaI	+134/++132	g
5' petD	spinach	296 bp BamHI	-162/+134	g
petB	spinach	2.4 kb SalI-BamHI	-900/++846	g
3' psbB	spinach	1,597 bp BamHI-SalI	+258/++846	g
5' psbB	spinach	338 bp BamHI	-80/+258	g
5' psbE	spinach	0.65 kb EcoRI	-465/+185	h
3' psbE-psbF	spinach	0.50 kb EcoRI	+185/++433	h
3' petA	pea	1.1 kb BamHI	+569/++550	i
5' petA	pea	0.9 kb HindIII-BamHI	-300/+569	i
3' rbcL	pea	0.8 kb HindIII-BamHI	+1340/++700	j
rbcL	pea	1,167 bp PstI-HindIII	+173/++1340	j
5' rbcL	pea	685 bp XbaI-PstI	-512/+173	j
atpB	pea	1,171 bp PstI-XbaI	-182/+989	j
atpE	spinach	420 bp EcoRI-XbaI	+41/++79	k
psaA	spinach	2.4 kb BamHI	+100/++300	l
psaB	spinach	1.6 kb BamHI	+600/++100	l
3' psbC	spinach	367 bp BamHI-PstI	+983/+1350	l
3' psbD-5' psbC	pea	1,150 bp PstI	+227 psbD/+369 psbC	m
5' psbD	pea	707 bp BamHI-PstI	-480/+227	m
rpoB	tobacco	1,063 bp BamHI	+1306/+2369	n
rpoC	pea	3.1 kb KpnI	+50/+3342	o
rps2-atpI-5' atpH	pea	2.9 kb KpnI-PstI	-405/+92	p
3' atpH	pea	0.8 kb PstI-BamHI	+92/++600	q
atpF-5' atpA	spinach	1.5 kb SalI-HindIII	+50 atpF/+800 atpA	r
3' atpA	spinach	0.9 kb HindIII-SalI	+800/++150	r
rps16	tobacco	1.9 kb HpaII-NdeI	-20/++800	s
5' psbA	pea	532 bp EcoRI-PstI	-58/+464	t
3' psbA	pea	1.2 kb PstI-EcoRI	+464/++700	t

^a Fragment sizes given in bp are based on complete sequence data; sizes given in kb are approximations based on electrophoretic mobility of the fragment

^b "-x" indicates gene probe starts x bp before the initiation codon; "+x" indicates gene probe either starts or ends x bp following the initiation codon; "++x" indicates gene probe ends x bp following the termination codon

^c Gel isolated Fragment

^d Fromm et al. (1986)

^e M. Shinozaki et al. (1986c)

^f Sijben-Muller et al. (1986)

^g Heinemeyer et al. (1984)

^h Herrmann et al. (1984)

ⁱ Willey et al. (1984)

^j Zurawski et al. (1986a, 1986b)

^k Zurawski et al. (1982)

^l Alt et al. (1984)

^m Rasmussen et al. (1984)

ⁿ Ohme et al. (1986)

^o Cozens and Walker (1986)

^p Cozens et al. (1986)

^q Huttly and Gray (1984, unpublished data)

^r Westhoff et al. (1985)

^s Shinozaki et al. (1986a)

^t Oishi et al. (1984)

caryophylla (tribe Mutisieae). We show that *Barnadesia* cpDNA has the same organization as the majority of land plants, whereas lettuce cpDNA contains a 22 kb inversion in its large single copy region. We locate the inversion endpoints within tRNA gene clusters situated downstream from the *atpA* gene and upstream from the

psbD gene. The lettuce and *Barnadesia* genomes are estimated to differ by 3.7% in nucleotide sequence, with most of the changes occurring in the single copy regions. The evolutionary implications of the inversion and the low level of sequence divergence within the Asteraceae are discussed.

Materials and methods

Chloroplast DNAs of lettuce (*Lactuca sativa*) and *Barnadesia caryophylla* were isolated by the sucrose gradient technique (Palmer 1982, 1986). Lettuce was purchased from a grocery store and *Barnadesia* was obtained from the Matthaei Botanical Gardens of the University of Michigan. The DNase I procedure of Kolodner and Tewari (1975) was used to isolate highly purified lettuce cpDNA for cloning experiments.

Restriction endonuclease digestions, agarose gel electrophoresis, bidirectional transfer of DNA fragments from agarose gels to Zetabind (AMF CUNO) nylon filters, labelling of recombinant plasmids by nick-translation, filter hybridizations, and autoradiography were performed as described (Palmer 1982, 1986). Restriction enzyme digests of lettuce and *Barnadesia* cpDNAs were electrophoresed in a 0.9% agarose gel. Electrophoresis of cloned spinach and tobacco cpDNA fragments was performed in a 4% gel using a mixture of 3% Nu-Sieve (FMC) and 1% Sea-Plaque (FMC) agarose. This allowed for detection of restriction enzyme fragments as small as 70 base pairs (bp) by filter hybridization. All hybridizations were performed at 65 °C for 14–16 h. Before reusing the nylon filters, the hybridized probes were washed off with 0.4 N NaOH at 42° for 1 h, followed by a 30 min wash in 0.1 × SSC (15 mM NaCl, 1.5 mM trisodium citrate), 0.5% SDS, and 0.2 M Tris (pH 7.5).

The lettuce chloroplast genome was cloned using the restriction enzyme SacI, which generates 18 restriction fragments ranging in size from 0.4–18.8 kb. Five µg of DNase I-purified lettuce cpDNA was digested with SacI and ligated with 50 ng of SacI digested pUC12. The ligation mixture was used to transform *E. coli* strain JM83 and recombinant white colonies were selected on ampicillin/X-gal plates. Insert sizes of recombinant plasmids from two hundred white colonies were determined by restriction enzyme analysis of purified (Birnboim and Doly 1979) plasmid DNA. Subclones of several cloned lettuce SacI fragments were prepared by digesting the clones with an enzyme that cuts once each within the insert and the polylinker of the vector, and then religating the remaining portion of the lettuce insert to the vector. Subclones of a 16.2 kb BamHI partial clone from tobacco (containing BamHI fragments 4, 11b, and 15; Hildebrand et al. 1985) and of a 13.5 kb PstI clone from spinach (Palmer and Thompson 1981b) were prepared by cutting the desired restriction fragment out of an agarose gel, followed by electroelution, phenol and chloroform extraction, and ethanol precipitation, prior to ligation to the appropriately digested pUC12 vector.

The 38 cpDNA fragments used as gene probes are listed in Table 1. All of these probes are cloned except for those containing the *rps7*, *rpoA*, and 3' *petD* genes. These three restriction fragments (Table 1) were sliced out of a 1% Sea-Plaque low melting temperature agarose (FMC) gel and nick-translated directly in the agarose. Recombinant plasmids containing probes 11 to 16 were constructed by K. Ko and J. Palmer by subcloning the indicated restriction fragments (Table 1) from spinach PstI clones (Palmer and Thompson 1981b). The first four of these probes were cloned into pBR322 and the other two into pDPL13. The remaining 29 gene probes were constructed by isolating the indicated restriction fragments (Table 1) from cpDNA clones of spinach and pea (Palmer and Thompson 1981b) and tobacco (gift of M. Sugiura) and ligating them to the appropriately digested pUC8, pUC12, pUC18, or pIC20H vector. Transformation, selection, and identification of recombinant plasmids were performed as described above for lettuce cpDNA cloning.

Results

Physical mapping of lettuce and Barnadesia chloroplast DNA

The shotgun strategy for cloning lettuce cpDNA yielded clones containing 17 of the 18 SacI restriction fragments (Fig. 1). Subsequent attempts to clone the single uncloned fragment after its gel-isolation were unsuccessful. However, approximately 9.0 kb of the 15.4 kb SacI fragment is located within the inverted repeat and is contained within the cloned 12.3 kb SacI fragment. Therefore, the lettuce clone bank covers 96% of the chloroplast genome.

CpDNAs from lettuce and *Barnadesia* were digested with the restriction endonucleases BstXI, EcoRV, HaeII, HindIII, KpnI, NcoI, NsiI, SacI, XbaI, and XhoI and the resulting restriction fragments were separated by agarose gel electrophoresis. Summation of the restriction fragments for all 10 enzymes (Tables 2 and 3) indicates that lettuce and *Barnadesia* have a genome size of approximately 151 kb.

The cloned SacI fragments from lettuce were used to construct restriction maps for the ten enzymes by the overlap hybridization procedure described by Palmer (1986). In addition, a 9.0 kb PstI cloned restriction fragment from petunia (Palmer et al. 1983a) was used to map restriction sites within the region of the chloroplast genome not covered by the lettuce clones. Reusable nylon filters containing single digests of lettuce and *Barnadesia* cpDNAs with the ten enzymes and double digests with SacI and each of the other nine enzymes were probed with each of the lettuce and petunia clones. The results of these hybridizations allowed us to construct complete physical maps for the ten enzymes (Fig. 2). The dominant structural feature of both genomes is a large inverted repeat, which we estimate has a maximum size of 25.7 kb for lettuce and 26.1 kb for *Barnadesia*, and a minimum size of 24.7 kb for both species.

Gene mapping of lettuce and Barnadesia chloroplast DNA

Thirty-seven genes were mapped on the chloroplast genomes of lettuce and *Barnadesia*. Thirty-five of the genes were located by hybridizing 38 gene probes (Table 1) constructed from the well-characterized cpDNAs of pea, spinach, and tobacco to the nylon filters used in the restriction site mapping. Differential hybridization of 5' and 3' gene probes enabled us to assign the direction of transcription for most of the genes (Table 1; Fig. 2). In several instances where 5' and 3' probes were not available, the orientation of genes was based on their cotranscription with genes for which we did have both 5' and

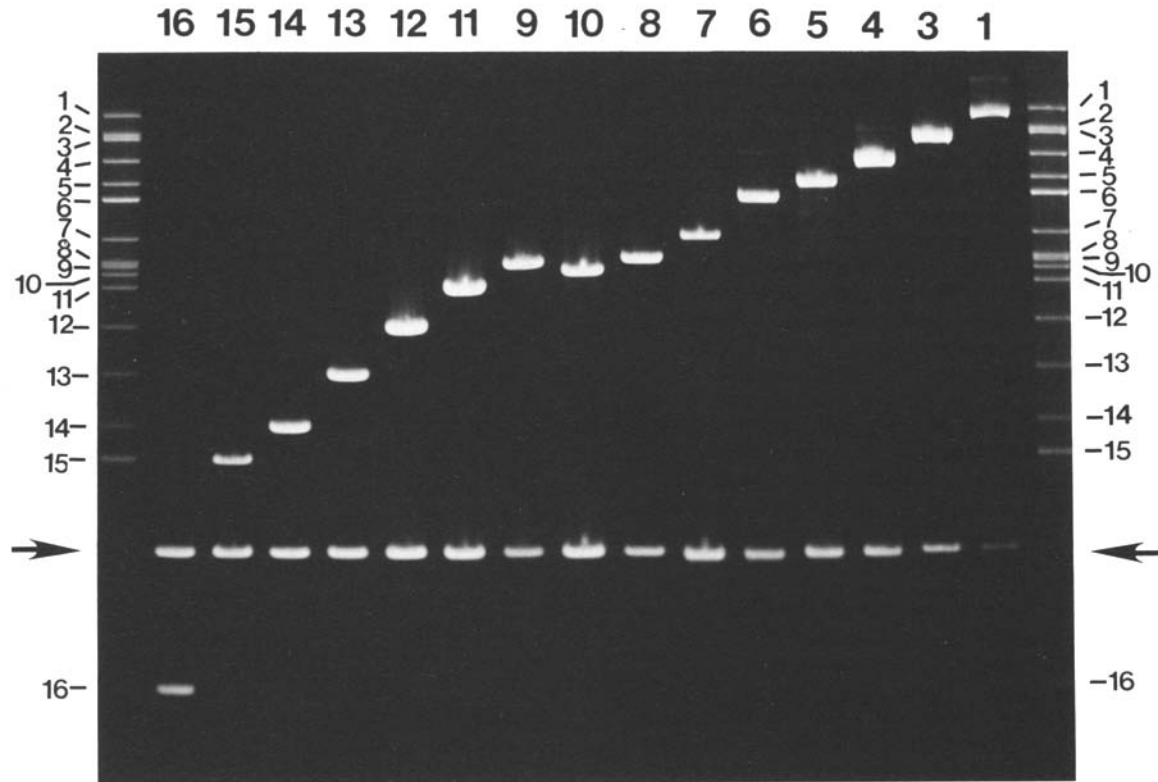


Fig. 1. Clone bank of lettuce cpDNA. *SacI* clones (inner 15 lanes) and DNase I-purified cpDNA of lettuce (outer 2 lanes) were digested with *SacI* and separated on a 1.0% agarose gel. The location of the pUC12 vector band is indicated by an arrow. Restriction fragment numbers correspond to those given in Table 2.

3' probes. For three genes, *rpoB*, *rpoC*, and *rps16*, the direction of transcription could not be determined by the above criteria, but we have indicated their orientation in Fig. 2 in the same direction relative to nearby genes as in the pea and tobacco genomes.

The 23S and 16S ribosomal RNA genes were located in the lettuce and *Barnadesia* genomes by taking advantage of highly conserved *PvuII* restriction sites present in most land plants. *PvuII* fragments of approximately 4.2 kb and 2.7 kb have been found to define the location of the 23S and 16S rRNA genes in a gymnosperm (*Ginkgo*, Palmer and Stein 1986), a fern (*Osmunda*, Palmer and Stein 1982), and many angiosperms (e.g., pea and mung bean, Palmer and Thompson 1981a; tobacco, Kusuda et al. 1980; soybean, Palmer et al. 1983b; and tomato, Palmer and Zamir 1982). The manner in which the *PvuII* fragments frame the 23S and 16S rRNA genes is described in detail by Palmer and Stein (1982). Restriction digests of lettuce and *Barnadesia* cpDNA revealed the presence of these two conserved fragments in these taxa. A complete *PvuII* restriction map (data not shown) confirmed the location of the sites in lettuce. The direction of transcription of the rRNA genes was assigned by analogy with the known orientation of the rRNA operon in all chloroplast and eubacterial genomes (reviewed in

Bohnert et al. 1982; Whitfeld and Bottomley 1983; Gillham et al. 1985; Palmer 1985a, b).

Comparative organization of lettuce and *Barnadesia* chloroplast genomes

A comparison of restriction site and gene maps of lettuce and *Barnadesia* (Figs. 2 and 3) shows that their chloroplast genomes differ by an inversion within the large single copy region. This is especially evident by the difference in order and direction of transcription of the *rpoB* through *atpA* genes. *Barnadesia* has the primitive genome organization common to most other land plants, including a fern and gymnosperm (Palmer and Stein 1986) and most angiosperms (Palmer 1985a, b), while the inversion is derived in the lettuce lineage.

The inversion endpoints are located downstream from the *atpA* gene and upstream from the *psbD* gene of *Barnadesia* (Fig. 3). A more precise localization of both endpoints was obtained by hybridizing four small lettuce subclones to *SacI* digests of *Barnadesia* cpDNA. Lettuce 0.7 kb *EcoRI* and 0.9 kb *EcoRI-SalI* fragments (see enlargement at bottom of Fig. 3) hybridize only to the *Barnadesia* 5.8 kb *SacI* fragment, while the 2.0 kb *EcoRI*-

Table 2. *Lactuca sativa* chloroplast DNA fragment identification

Fragment number	Fragment size (kb)									
	SacI	BstXI	EcoRV	HaeII	HindIII	KpnI	NcoI	NsiI	XbaI	XhoI
1	18.8	20	14.5	23	15.0	33	17.1	16.7	27	15.5
2	15.4	14.2	14.5	9.4	10.8 (2x)	27	13.6	13.3 (2x)	27	14.9 (2x)
3	14.7	9.7 (2x)	11.7	9.1	10.6	17.5	12.0	9.3	8.9	14.3
4	12.3	9.7	11.3	8.0	10.2	14.7	10.8	8.7	8.7	12.4
5	10.6	7.1 (2x)	7.6	7.7	9.0 (2x)	11.0	9.0	8.5	7.4	10.5
6	9.9 (2x)	6.5	7.3	7.2 (2x)	9.0	10.8	8.0	5.6 (2x)	5.3	10.5
7	7.7	5.9	7.0	7.0	7.5	10.3	7.7	4.5	5.3	8.7
8	7.0	5.9	5.8	6.5 (2x)	7.3	9.9	6.7	4.4	4.9	7.7
9	6.9	5.4	5.8	6.5	6.7	6.2	6.2	4.3	4.9	7.0
10	6.7	5.4 (2x)	5.6	6.3	5.4	5.0	5.7 (2x)	4.1	4.9	5.4
11	6.3	5.4	4.8	5.5	5.0	3.6	5.5	3.7	4.6	5.0
12	5.4	5.0	4.6	5.5	4.4	0.7 (2x)	4.7	3.7	4.3	3.8
13	4.6	5.0	4.5	4.3	4.2		4.3	3.3	3.6	3.7
14	3.8	4.4	3.9 (2x)	4.3	3.9		3.7 (2x)	3.1	3.5	3.3 (2x)
15	3.5 (2x)	3.8	3.7	3.2 (2x)	3.3		3.7 (2x)	3.1	3.3	3.0
16	1.8 (2x)	3.2	3.1	2.7	2.8		2.6	2.9 (2x)	3.1	2.4
17	0.4 (2x)	2.9	3.0 (2x)	2.0 (2x)	2.5		2.2	2.2	2.9	1.3
18	0.4	1.7	2.4 (2x)	1.9 (2x)	2.2 (2x)		1.9	1.9	2.6	0.8 (2x)
19		1.5	2.2 (2x)	1.7	1.6		1.7 (2x)	1.7	1.7	0.7
20		1.2	2.1	1.5	1.1		1.5	1.7	1.6	0.7
21		1.0	1.8	1.4	1.0 (2x)		0.9 (2x)	1.6	1.5	
22		0.6 (2x)	1.3	1.1 (2x)	0.9 (2x)		0.8 (2x)	1.5	1.4	
23		0.6	1.0 (2x)	0.7 (2x)	0.9		0.7 (2x)	1.5	1.4	
24		0.5 (2x)	1.0	0.6	0.8		0.7	1.4	1.3	
25			1.0	0.5 (2x)	0.6		0.6 (2x)	1.3 (2x)	1.2 (2x)	
26			0.8 (2x)	0.5	0.6			1.2	1.2	
27			0.8	0.3 (2x)	0.4			1.2	1.2	
28			0.6	0.3	0.3 (2x)			1.1	1.2	
29			0.5					1.1	0.7	
30			0.5					0.8	0.7	
31			0.3 (2x)					0.8	0.6	
32			0.3 (3x)					0.7	0.5	
33			0.3					0.6	0.5	
34								0.6	0.3	
35								0.6	0.3	
36								0.6	0.3	
37								0.6		
38								0.6		
39								0.5		
40								0.4 (2x)		
41								0.4		
42								0.4		
Sum	151.8	149.9	149.8	152.1	152.2	150.4	150.1	149.7	151.0	150.6

BglII and 1.1 kb EcoRI probes hybridize only to the *Barnadesia* 14.9 kb SacI fragment (Fig. 3). This suggests that both endpoints are very near an EcoRI site (Fig. 3) and that the inversion is approximately 22 kb in size.

Location of lettuce inversion endpoints relative to tobacco and spinach

We determined more precisely the location of the inversion endpoints by hybridizing four cloned lettuce frag-

ments (Fig. 3, enlargement at bottom) to the appropriate regions of spinach and tobacco cpDNA. These species were selected because their chloroplast genomes are not rearranged and because sequencing studies have revealed the precise location of genes and spacer sequences in the region where the lettuce inversion endpoints are located.

The tobacco chloroplast genome was used for locating the inversion endpoint downstream from the *atpA* gene. Figure 4 summarizes the location of genes and

Table 3. *Barnadesia caryophylla* chloroplast DNA fragment identification

Fragment number	Fragment size (kb)									
	SacI	BstXI	EcoRV	HaeII	HindIII	KpnI	NcoI	NsiI	XbaI	XhoI
1	18.8	15.0	22	23	15.0	32	17.1	13.3 (2x)	30	26
2	15.4	14.2	16.6	18.8	10.8 (2x)	27	13.6	11.8	22	18.5
3	14.9	13.0	11.7	16.4	10.2	22	11.8	8.7	14.9	18.5
4	13.8	13.0	11.3	8.4	10.0	14.7	11.6	7.8	6.2	14.9 (2x)
5	10.6	10.2	6.5	8.0 (2x)	9.0 (2x)	14.5	9.0	6.9	6.2	10.3
6	10.3	10.0 (2x)	6.5	8.0	9.0	10.8	8.4	6.7	5.4	8.5
7	9.9 (2x)	9.4	5.8	7.2 (2x)	8.3	10.3	8.0	5.4 (2x)	5.2	7.9
8	7.3	9.2	5.6	7.2	7.9	5.8	7.5	5.4	4.8	5.4
9	7.0	7.1 (2x)	4.8	5.5	7.7	4.9	6.2	4.9	4.4 (2x)	3.7
10	6.9	6.5	3.9 (2x)	5.5	6.8	3.6	5.6 (2x)	4.6	4.4	3.5
11	5.8	5.4	3.7	4.4	6.6	3.1	5.6 (2x)	4.6	3.8	3.3 (2x)
12	4.6	5.0	3.7	4.2	4.4	0.7 (2x)	5.3	4.4	3.3	3.1
13	3.5 (2x)	3.7	3.4	3.2 (2x)	4.1	0.4	4.3	4.2	3.1	2.8
14	2.4	2.9	3.3	1.9 (2x)	4.1		4.3	4.2	2.8	2.4
15	1.8 (2x)	1.8	3.0 (2x)	1.4	2.6		4.0	3.5	2.5	0.8 (2x)
16	1.8	1.3	2.4 (2x)	1.1 (2x)	2.5		3.7 (2x)	3.4	2.3	0.7
17	0.8	1.2	2.4	0.7 (2x)	2.2 (2x)		2.6	3.1	2.2	
18	0.4 (2x)	1.0	2.4	0.5 (2x)	1.9		1.6	3.0 (2x)	1.7	
19	0.4	0.8	2.2 (2x)	0.5	1.6		0.9 (2x)	2.3	1.6	
20		0.6 (2x)	2.2	0.4 (2x)	1.0 (2x)		0.8 (2x)	1.8	1.6	
21		0.5 (2x)	2.0	0.3 (2x)	0.9		0.6 (2x)	1.8	1.6	
22			2.0	0.3	0.6		0.6 (2x)	1.7	1.5	
23			1.7		0.4			1.6	1.4 (2x)	
24			1.5		0.3 (2x)			1.6	1.4	
25			1.1 (2x)		0.3			1.5	1.4	
26			1.0 (2x)					1.3 (2x)	1.2	
27			0.9					1.2	1.2	
28			0.9					1.1	1.1	
29			0.8					0.9	1.0	
30			0.3 (2x)					0.9	0.9	
31								0.7	0.9	
32								0.7	0.8	
33								0.6	0.7	
34								0.6	0.6	
35								0.6	0.6	
36								0.4 (2x)	0.5	
37									0.3	
Sum	152.0	150.0	149.5	150.2	151.5	150.5	150.9	150.6	151.3	149.3

selected restriction enzyme sites in this region based on several recent sequencing studies (Deno et al. 1983, 1984; Deno and Sugiura 1983, 1984; Sugita et al. 1985; Shinozaki et al. 1986a, b, c). We initially hybridized the 0.7 kb EcoRI and 0.9 kb EcoRI-SalI lettuce fragments to Southern blots containing digests of the 3.5 kb and 9.4 kb cloned BamHI fragments from tobacco (Fig. 4). The 0.7 kb probe hybridized to a 1.1 kb HindIII fragment and the 0.9 kb probe to HindIII-BamHI fragments of 0.56 kb and 1.2 kb (Fig. 4), suggesting that one inversion endpoint is located within either the 1.1 kb or 0.56 kb fragments. The location of this endpoint was mapped more precisely by digesting the 1.1 kb HindIII and 0.56

kb HindIII-BamHI restriction fragments with four four-base pair cutting restriction enzymes (Fig. 4), separating the resulting fragments on a 4% Nu-Sieve agarose gel, and hybridizing the same two lettuce fragments to filter blots of this gel. The 0.7 kb EcoRI fragment hybridized to a 630 bp region defined by a Sau3A and HpaII site and containing ORF2 and tRNA^{Ser}, while the 0.9 kb EcoRI-SalI probe hybridized to a 195 bp Sau3A restriction fragment that contains one exon and part of the intron of the tRNA^{Gly} gene (Fig. 4). Thus, the inversion endpoint is located between the tRNA^{Ser} and tRNA^{Gly} genes, which are separated by an 800 bp spacer in tobacco.

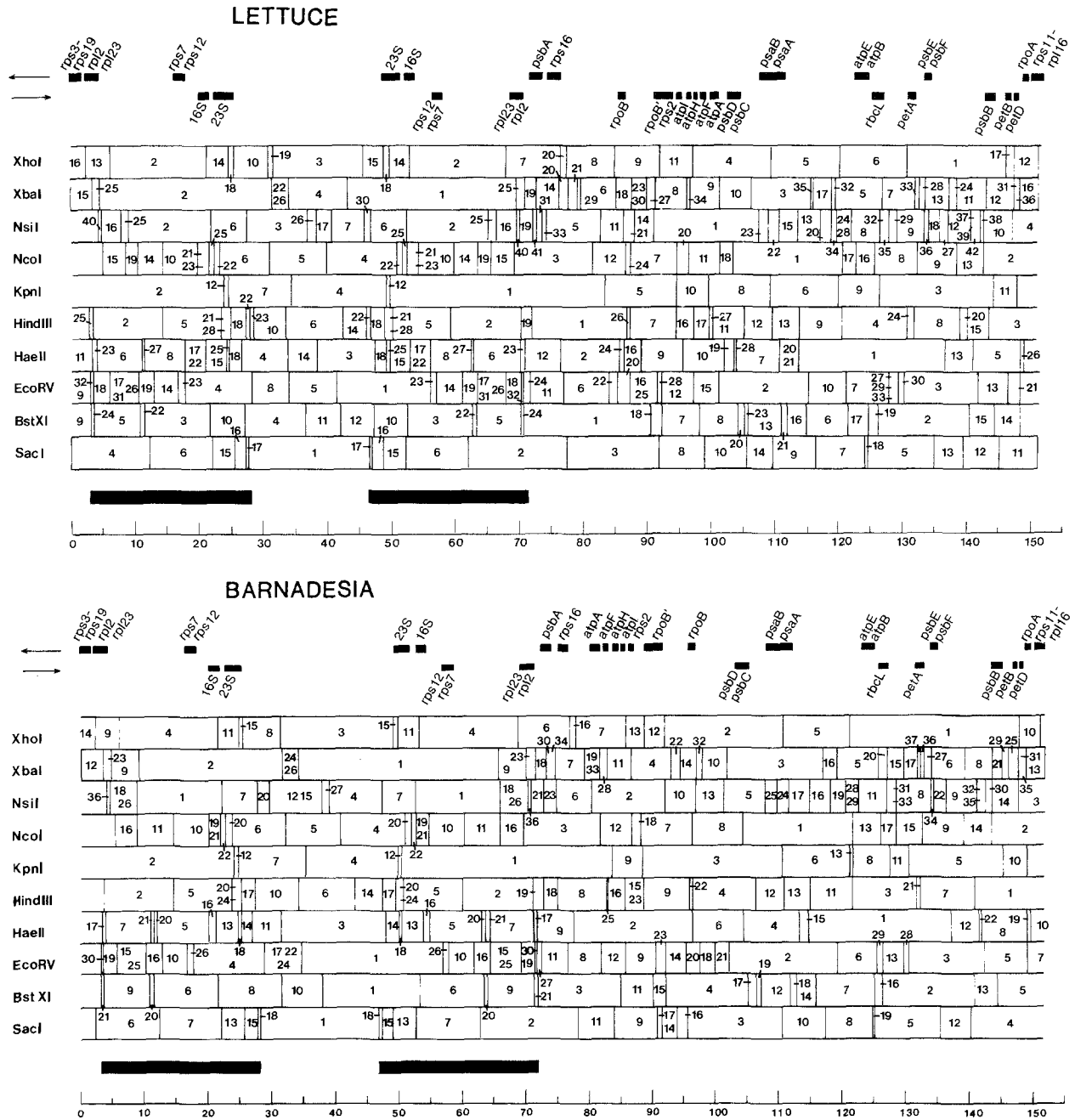


Fig. 2. Restriction maps of *Barnadesia* and lettuce cpDNA. The circular maps have been linearized at the XbaI site separating the 3.3 kb and 3.1 kb fragments (numbers 15 and 16 in lettuce and 12 and 13 in *Barnadesia*). Restriction fragment numbers correspond to those given in Tables 2 and 3. The two long, heavy black lines below each map indicate the 24.7 kb minimum extent of the inverted repeat for both species. The minimum boundaries are defined in lettuce by HindIII (fragment no. 22) and BstXI (24) and in *Barnadesia* by SacI (18) and XbaI (23). The maximum extent of the repeat is 25.7 kb for lettuce (defined by HindIII 23 and 25) and 26.1 kb for *Barnadesia* (defined by EcoRV 4 and 27). The short, heavy black lines show the position and orientation (indicated by arrows) of 35 genes listed in Table 1 and two ribosomal RNA genes. The block of genes labelled rps11-rpl16 includes five genes in the following order: rps11, infA, rps8, rpl14, and rps16. The rpl22 gene is located between rps8 and rps19. The scale at the bottom of each map is in kb. The rpoB' gene in Figs. 2 and 3 is now by convention referred to as rpoC (Shinozaki et al. 1986c). The maps in Figs. 2 and 3 portray the chloroplast genome in only one of the two orientations of the small single copy region (Palmer 1985b)

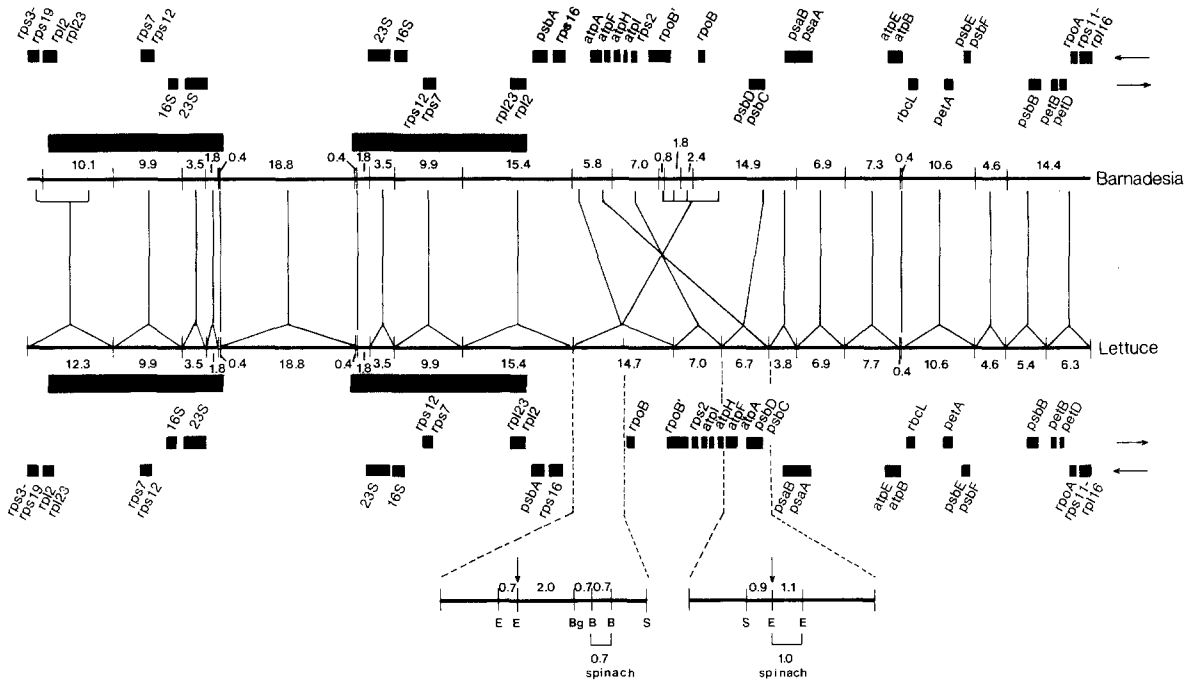


Fig. 3. Arrangement of homologous sequences in the lettuce and *Barnadesia* chloroplast genomes. Symbols and conventions for the inverted repeat and mapped genes are described in the legend of Fig. 2. The extent of lettuce *SacI* fragments used as probes is indicated by the *two lines* that converge above each fragment, while the *Barnadesia* *SacI* fragment(s) to which the probes hybridize are indicated by *lines leading* from the lettuce fragments. *Arrows* in the enlargement at the *bottom* indicate the approximate location of the inversion endpoints. Lettuce sequences homologous to the spinach 1.0 kb *HindIII-BglII* and 0.7 kb *BamHI-BglII* fragments are indicated by the *brackets* in the enlargement. Restriction sites in the enlargement are: *B* *BamHI*; *Bg* *BglII*; *E* *EcoRI*; *S* *Sall*

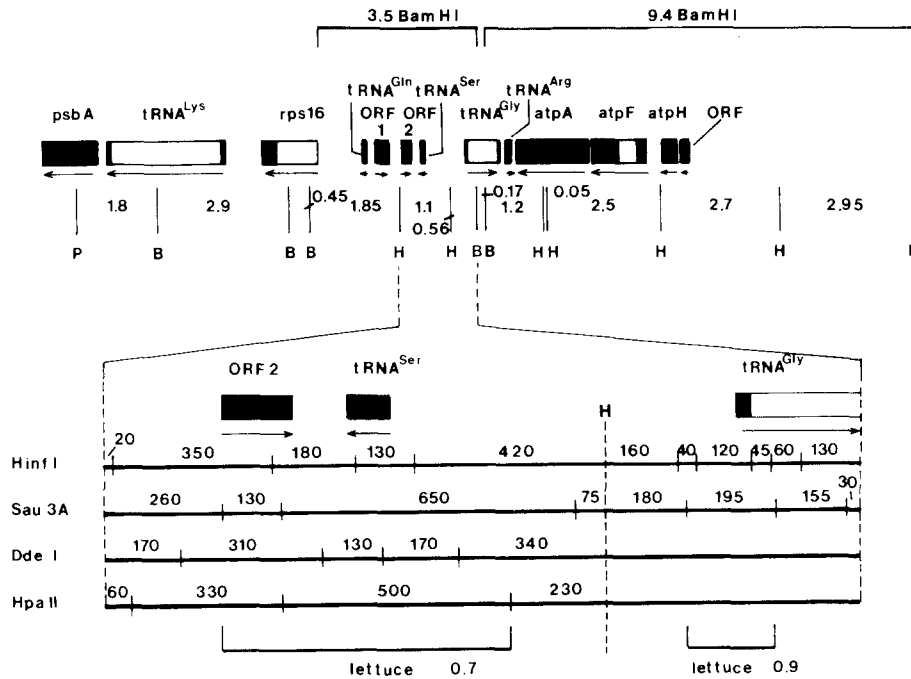


Fig. 4. Restriction site and gene map of tobacco chloroplast DNA in the region near *atpA* (Deno et al. 1983, 1984; Deno and Sugiura 1983, 1984; Sugita et al. 1985; Shinozaki et al. 1986a, 1986b). *Filled boxes* indicate exons and *open boxes* introns. The *brackets* below the enlargement show the maximum extent of homology between tobacco cpDNA and lettuce 0.7 kb *EcoRI* and 0.9 kb *EcoRI-Sall* fragments. Fragment sizes in the upper portion of the map are given in kb while those in the enlargement are in bp. Abbreviations for restriction enzymes are: *B* *BamHI*; *H* *HindIII*; *P* *PstI*

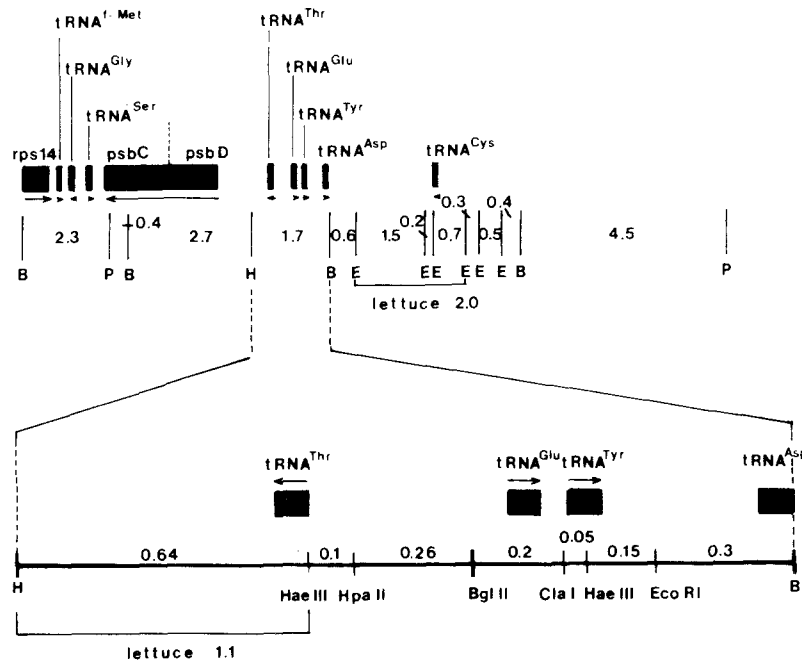


Fig. 5. Restriction site and gene map of spinach chloroplast DNA in the region near the psbC and psbD genes (Holschuh et al. 1983, 1984a, 1984b). The brackets show the maximum extent of homology between spinach cpDNA and lettuce 1.1 kb EcoRI and 2.0 kb EcoRI-BglII fragments. Sizes of fragments are given in kb. Abbreviations for restriction enzymes are: *B* BamHI; *E* EcoRI; *H* HindIII; *P* PstI

The inversion endpoint upstream from the psbD gene was located by comparison to the spinach chloroplast genome. Figure 5 shows the position of genes and selected restriction sites in this portion of the genome based on site mapping and sequencing studies by Holschuh et al. (1983, 1984a, b). A Southern blot containing BamHI, EcoRI, and HindIII digests of a cloned 13.5 kb PstI spinach fragment (Palmer and Thompson 1981b) was probed with the 1.1 kb EcoRI and 2.0 kb EcoRI-BglII lettuce fragments (Fig. 3). The 1.1 kb EcoRI fragment hybridizes to a 1.7 kb BamHI-HindIII spinach fragment and the 2.0 kb EcoRI-BglII probe to EcoRI fragments of 1.5 kb and 0.7 kb (Fig. 5) suggesting that the inversion endpoints is within one of these three spinach fragments. The 1.7 kb BamHI-HindIII fragment was subcloned, digested with BglII, ClaI, EcoRI, HaeIII, and HpaII, and the resulting restriction fragments were electrophoresed in a 4% NuSieve agarose gel, blotted, and probed with the 1.1 kb EcoRI lettuce fragment. The lettuce probe hybridizes to a 640 bp HindIII-HaeIII restriction fragment, suggesting that the inversion endpoint is between the tRNA^{Thr} and tRNA^{Glu} genes (Fig. 5), which are separated by 440 bp in spinach. The failure of both the 2.0 kb EcoRI-BglII and 1.1 kb EcoRI fragments to hybridize to the adjacent spinach fragments, which contains three tRNA genes (Fig. 5), suggests that there may have been a second rearrangement in this region of the lettuce chloroplast genome. The position of these three tRNA genes was located by hybridizing a gel-isolated 0.7 kb BglII-BamHI spinach fragment (Fig. 5) to filter blots containing digests of lettuce cpDNA. Sequences homologous to this spinach fragment are within a 0.7 kb BamHI

lettuce fragment 2.7 kb from the approximate location of the inversion endpoint (Fig. 3, enlargement at bottom).

Discussion

Structure of the lettuce and *Barnadesia* chloroplast genomes

The chloroplast genomes of lettuce and *Barnadesia* are similar in size and organization to those of other angiosperms. The only major difference is a 22 kb inversion in the large single copy portion of the lettuce genome. Our size estimates of 151 kb for the entire genome and 25 kb for the inverted repeat are very close to those obtained in two previous studies of Asteraceae species. Kolodner and Tewari's (1979) electron microscopy study of lettuce cpDNA gave estimates of 155 kb for the entire genome and 24.4 kb for the inverted repeat. Restriction site mapping of cpDNA from safflower (*Carthamus tinctorius*) gave a genome size of 151 kb and an inverted repeat size of 25 kb (Ma and Smith 1985).

The retention of the primitive land plant genome organization in *Barnadesia* cpDNA suggests that this species represents an ancestral lineage of the Asteraceae. A survey of the systematic distribution of the lettuce cpDNA inversion has revealed that this arrangement defines an ancient evolutionary split within this large and diverse flowering plant family (Jansen and Palmer unpublished). A similar situation occurs in the Onagraceae in which *Clarkia* (Sytsma and Gottlieb 1986) and *Epilobium* (Schmitz and Kowallik 1986) have retained the primitive genome organization, whereas *Oenothera* (Herrmann et al. 1983) has a 45 kb inversion.

Table 4. Sequence divergence between lettuce and *Barnadesia* chloroplast DNAs

Portion of genome compared ^a	Total number of restriction sites		Number of sites compared	Number of restriction sites shared	Percent sequence divergence ^b
	<i>Barnadesia</i>	lettuce			
Total genome (104 kb)	186	207	236	157	3.7
Single copy (79 kb)	131	146	174	103	4.9
Large single copy (61 kb)	111	123	145	89	4.6
Small single copy (18 kb)	20	23	29	14	7.1
Inverted repeat (25 kb)	56	61	63	54	1.4
Non-rDNA inverted repeat (18 kb)	34	39	41	32	2.2
Inverted repeat rDNA ^c (7 kb)	22	22	22	22	0

^a Calculations do not include restriction sites within the 22 kb inversion. Furthermore, sites in the inverted repeats are counted only once. *Numbers in parentheses* give size in kb of the region of the genome being compared

^b Values are calculated as 100p using equations 9 and 10 of Nei and Li (1979)

^c Ribosomal DNA restriction sites were delimited by conserved PvuII restriction fragments (see results of gene mapping for explanation)

Comparison of lettuce inversion with other chloroplast DNA inversions

In addition to the Asteraceae, five other angiosperm families, the Campanulaceae, Fabaceae, Geraniaceae, Onagraceae, and Poaceae, have inversions in their chloroplast genomes (reviewed in Palmer 1985a, b). It is interesting that the inversion we describe here in lettuce has one of its endpoints in the same region of the genome, just downstream from *atpA*, as do four of the best characterized of these inversions, from pea (Palmer et al. 1985), mung bean (Palmer and Thompson 1982, Palmer et al. 1987), *Oenothera* (Herrmann et al. 1983), and wheat (Howe et al. 1983; Quigley et al. 1985, Quigley and Weil 1985). The chloroplast genome of tobacco, which has the same gene order as the majority of land plants, has been completely sequenced in this region (Deno et al. 1983, 1984; Deno and Sugiura 1983, 1984; Sugita et al. 1985; Shinozaki et al. 1986a, b, c). These studies have revealed several intergenic spacers as large as 1,000 bp separating tRNA genes downstream from the *atpA* gene (Fig. 4). Any of these spacers could easily accommodate the disruptive effect of an inversion.

The most thoroughly characterized cpDNA rearrangement is a 20 kb inversion in wheat, which has also been documented in barley (Oliver and Poulson 1984) and maize (Palmer and Thompson 1982). Gene mapping and sequencing studies (Howe 1985; Quigley and Weil 1985) have shown that the inversion endpoint downstream from the *atpA* gene is located in a 170 bp spacer sequence separating the genes for tRNA^{Arg} and tRNA^{Gly} (Fig. 4 shows the location of these genes in tobacco). The lettuce inversion endpoint occurs a short distance from the wheat endpoint in a spacer between the tRNA^{Gly} and

tRNA^{Ser} genes (Fig. 4). A much greater distance separates the location of the second inversion endpoint in wheat and lettuce. In wheat this endpoint is between the tRNA^{Met} and *psaA* genes (Quigley and Weil 1985) whereas in lettuce it is between the tRNA^{Thr} and tRNA^{Glu} genes upstream from the *psbD* gene (Fig. 5). The wheat chloroplast genome has two smaller rearrangements within the 20 kb inversion (Quigley and Weil 1985). There also appears to be a second, smaller inversion in the lettuce genome located at one end of the 22 kb inversion, but more detailed investigations are needed to characterize this rearrangement.

Chloroplast DNA sequence divergence

We estimated levels of nucleotide sequence divergence between various regions of the cpDNAs of lettuce and *Barnadesia* using equations 9 and 10 of Nei and Li (1979; Table 4). Restriction sites within the 22 kb inversion were not included in our calculations because of uncertainty in aligning the rearranged sites. The remainder of the genome has diverged by 3.7% in nucleotide sequence, with the highest level of change occurring in the large and small single copy regions (Table 4). The inverted repeats show only 1.4% divergence and all of this change is due to site differences in the regions outside of the ribosomal RNA genes. A high degree of restriction site conservation within the inverted repeat has also been found in cpDNAs of *Pennisetum* (Clegg et al. 1984), several legumes (Palmer et al. 1983b), and *Brassica* (Palmer et al. 1983a).

Comparative restriction site mapping has revealed low levels of nucleotide change among closely related species

from several angiosperm families. The range of intrageneric divergence values includes 0–0.7% in *Lycopersicon* (Palmer and Zamir 1982), 0.3–2.6% in *Brassica* (Palmer et al. 1983a), 0–0.3% in *Lisianthus* (Sytsma and Schaal 1985), 0.2–1.5% in *Clarkia* (Sytsma and Gottlieb 1986), 0–6% in *Linum* (Coates and Cullis 1987), and 0.2–4.4% in *Atriplex* (J. Palmer unpublished). Intergeneric divergence values were estimated at 5.5% and 10.8% in comparisons of three legumes (Palmer et al. 1983b). Our total genome estimate of 3.7% between lettuce and *Barnadesia* is surprisingly low given these results and also since these two genera are thought to be very distantly related, as reflected by their taxonomic placement in different tribes and subfamilies of the Asteraceae (Cronquist 1955, 1977; Wagenitz 1976). This low level of sequence divergence suggests that comparative restriction site mapping should be a valuable tool for resolving phylogenetic relationships at higher taxonomic levels in the Asteraceae.

Acknowledgements. We thank L. Herbon and H. Tucker for technical assistance, M. Sugiura for providing tobacco clones used to construct several gene probes and for unpublished mapping and sequence data, K. Ko for collaboration in constructing some of the gene probes, M. Hildebrand for providing a BamHI tobacco clone, P. Srinivasan and L. Herbon for assistance in preparing the figures, and the Matthaei Botanical Gardens for providing plant material of *Barnadesia caryophylla*. This research was supported by a grant from the NSF (BSR-8415934).

References

- Alt J, Morris J, Westhoff P, Herrmann RG (1984) *Curr Genet* 8:597–606
- Birnboim HC, Doly JC (1979) *Nucleic Acids Res* 7:1513–1523
- Bohnert HJ, Crouse EJ, Schmitt JM (1982) *Encycl Plant Physiol* 148:475–530
- Chu NM, Tewari KK (1982) *Mol Gen Genet* 186:23–32
- Clegg MT, Rawson JR, Thomas K (1984) *Genetics* 106:449–461
- Coates D, Cullis C (1987) *Am J Bot*, in press
- Cozens AL, Walker JE (1986) *Biochem J* 236:453–460
- Cozens AL, Walker JE, Phillips AL, Huttly AK, Gray JC (1986) *EMBO J* 5:217–222
- Cronquist A (1955) *Am Midl Nat* 53:478–511
- Cronquist A (1977) *Brittonia* 29:137–153
- Deno H, Sugiura M (1983) *Nucleic Acids Res* 11:8407–8414
- Deno H, Sugiura M (1984) *Proc Natl Acad Sci USA* 81:405–408
- Deno H, Shinozaki K, Sugiura M (1983) *Nucleic Acids Res* 11:2185–2191
- Deno H, Shinozaki K, Sugiura M (1984) *Gene* 32:195–201
- Fromm H, Edelman M, Koller B, Goloubinoff P, Galun E (1986) *Nucleic Acids Res* 14:883–898
- Gillham NW, Boynton JE, Harris EH (1985) In: Cavalier-Smith T (ed) *DNA and evolution: natural selection and genome size*, Wiley, New York, pp 299–351
- Heinemeyer W, Alt J, Herrmann RB (1984) *Curr Genet* 8:543–549
- Herrmann RG, Westhoff P, Alt J, Winter P, Tittgen J, Bisanz C, Sears BB, Nelson N, Hurt E, Hauska G, Viebrock A, Sebald W (1983) In: Ciferri O, Dure L III (eds) *Structure and function of plant genomes*, Plenum, New York, pp 143–153
- Herrmann RG, Alt J, Schiller B, Wigger WR, Cramer WA (1984) *FEBS Lett* 176:239–244
- Hildebrand M, Jurgenson JE, Ramage RT, Bourque DP (1985) *Plasmid* 14:64–79
- Holschuh K, Bottomley W, Whitfield PR (1983) *Nucleic Acid Res* 11:8547–8554
- Holschuh K, Bottomley W, Whitfield PR (1984a) *Nucleic Acid Res* 12:8819–8834
- Holschuh K, Bottomley W, Whitfield PR (1984b) *Plant Mol Biol* 3:313–317
- Howe CJ (1985) *Curr Genet* 10:139–145
- Howe CJ, Bowman CM, Dyer TA, Gray JC (1983) *Mol Gen Genet* 190:51–55
- Huttly AK, Gray JC (1984) *Mol Gen Genet* 194:402–409
- Koller B, Delius H (1980) *Mol Gen Genet* 178:261–269
- Kolodner R, Tewari KK (1975) *Biochim Biophys Acta* 402:372–390
- Kolodner R, Tewari KK (1979) *Proc Natl Acad Sci USA* 76:41–45
- Kusuda J, Shinozaki K, Takaiwa F, Sugiura M (1980) *Mol Gen Genet* 178:1–7
- Ma C, Smith MA (1985) *Theor Appl Genet* 70:620–627
- Nei M, Li W-H (1979) *Proc Natl Acad Sci USA* 76:5269–5273
- Ohme M, Tanaka M, Chunwongse J, Shinozaki K, Sugiura M (1986) *FEBS Lett* 200:87–90
- Oishi KK, Shapiro DR, Tewari KK (1984) *Mol Cell Biol* 4:2556–2563
- Oliver RP, Poulson CR (1984) *Carlsberg Res Commun* 48:57–80
- Palmer JD (1982) *Nucleic Acids Res* 10:1593–1605
- Palmer JD (1985a) In: MacIntyre RJ (ed) *Monographs in evolutionary biology: molecular evolutionary genetics*. Plenum, New York, pp 131–240
- Palmer JD (1985b) *Ann Rev Genet* 19:325–354
- Palmer JD (1986) *Methods Enzymol* 118:167–186
- Palmer JD, Stein DB (1982) *Curr Genet* 5:165–170
- Palmer JD, Stein DB (1986) *Curr Genet* 10:823–833
- Palmer JD, Thompson WF (1981a) *Proc Natl Acad Sci USA* 78:5533–5537
- Palmer JD, Thompson WF (1981b) *Gene* 15:21–26
- Palmer JD, Thompson WF (1982) *Cell* 45:537–550
- Palmer JD, Zamir D (1982) *Proc Natl Acad Sci USA* 79:5006–5010
- Palmer JD, Shields CR, Cohen DB, Orton TJ (1983a) *Theor Appl Genet* 65:181–189
- Palmer JD, Singh GP, Pillay DTN (1983b) *Mol Gen Genet* 190:13–19
- Palmer JD, Jorgensen RA, Thompson WF (1985) *Genetics* 109:195–213
- Palmer JD, Osorio B, Aldrich J, Thompson WF (1986) *Curr Genet* 11:275–286
- Quigley F, Weil JH (1985) *Curr Genet* 9:495–503
- Quigley F, Grienenberger JM, Weil JH (1985) *Plant Mol Biol* 4:305–310
- Rasmussen OF, Bookjans G, Stummann BM, Henningsen KW (1984) *Plant Mol Biol* 3:191–199
- Schmitz UK, Kowallik KV (1986) *Plant Mol Biol* 7:115–127
- Shinozaki K, Deno H, Sugita M, Kuramitsu S, Sugiura M (1986a) *Mol Gen Genet* 202:1–5
- Shinozaki K, Deno H, Wukasugi T, Sugiura M (1986b) *Curr Genet* 10:421–423
- Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi T, Zaita N, Chunwongse J, Obokata J, Yamaguchi-Shinozaki K, Ohto C, Torazawa K, Meng BY, Sugita M, Deno H, Kamogashira T, Yamada K, Kusuda J, Takaiwa F,

- Kato M, Tohdoh N, Shimada M, Sugiura M (1986c) *EMBO J* 5:2043–2049
- Sijben-Muller G, Hallick RB, Alt J, Westhoff P, Herrmann RG (1986) *Nucleic Acids Res* 14:1029–1044
- Sugita M, Shinozaki K, Sugiura M (1985) *Proc Natl Acad Sci USA* 82:3557–3561
- Sytsma KJ, Gottlieb LD (1986) *Proc Natl Acad Sci USA* 83:5554–5557
- Sytsma KJ, Schaal BA (1985) *Evolution* 39:594–608
- Wagenitz G (1976) *Plant Syst Evol* 125:29–46
- Westhoff P, Alt J, Nelson N, Herrmann RG (1985) *Mol Gen Genet* 199:290–299
- Whitfield PR, Bottomley W (1983) *Ann Rev Plant Physiol* 34:279–310
- Willey DL, Auffret AD, Gray JC (1984) *Cell* 36:555–562
- Zurawski G, Bottomley W, Whitfield PR (1982) *Proc Natl Acad Sci USA* 79:6260–6264
- Zurawski G, Bottomley W, Whitfield PR (1986a) *Nucleic Acid Res* 14:3974
- Zurawski G, Bottomley W, Whitfield PR (1986b) *Nucleic Acid Res* 14:3975

Communicated by B. B. Sears

Received August 8/September 8, 1986