Curr Genet (1987) 11:275–286

Current Genetics

© Springer-Verlag 1987

Chloroplast DNA evolution among legumes: Loss of a large inverted repeat occurred prior to other sequence rearrangements

Jeffrey D. Palmer[1], **Bernardita Osorio**[2,4], **Jane Aldrich**[3], **and William F. Thompson**[2,5]

[1] Department of Biology, University of Michigan, Ann Arbor, MI 48109, USA
[2] Carnegie Institution of Washington, Department of Plant Biology, 290 Panama Street, Stanford, CA 94305, USA
[3] The Standard Oil Company, 4440 Warrensville Center Road, Cleveland, OH 44128, USA

Summary. We have compared the sequence organization of four previously uncharacterized legume chloroplast DNAs – from alfalfa, lupine, wisteria and subclover – to that of legume chloroplast DNAs that either retain a large, ribosomal RNA-encoding inverted repeat (mung bean) or have deleted one half of this repeat (broad bean). The circular, 126 kilobase pair (kb) alfalfa chloroplast genome, like those of broad bean and pea, lacks any detectable repeated sequences and contains only a single set of ribosomal RNA genes. However, in contrast to broad bean and pea, alfalfa chloroplast DNA is unrearranged (except for the deletion of one segment of the inverted repeat) relative to chloroplast DNA from mung bean. Together with other findings reported here, these results allow us to determine which of the four possible inverted repeat configurations was deleted in the alfalfa-pea-broad bean lineage, and to show how the present-day broad bean genome may have been derived from an alfalfa-like ancestral genome by two major sequence inversions. The 147 kb lupine chloroplast genome contains a 22 kb inverted repeat and has essentially complete colinearity with the mung bean genome. In contrast, the 130 kb wisteria genome has deleted one half of the inverted repeat and appears colinear with the alfalfa genome. The 140 kb subclover genome has been extensively rearranged and contains a family of at least five dispersed repetitive sequence elements, each several hundred bp in size; this is the first report of dispersed repeats of this size in a land plant chloroplast genome. We conclude that the inverted repeat has been lost only once among legumes and that this loss occurred prior to all the other rearrangements observed in subclover, broad bean and pea. Of those lineages that lack the inverted repeat, some are stable and unrearranged, other have undergone a moderate amount of rearrangement, while still others have sustained a complex series of rearrangements either with or without major sequence duplications and transpositions.

Key words: Chloroplast genome evolution – Inverted repeat – Inversion – Repeated sequence

Introduction

One of the more remarkable features of the chloroplast genome is the conservation of its most prominent structural feature, a large, ribosomal RNA-encoding inverted repeat of between 10 kb and 76 kb in size. This inverted repeat is present in the chloroplast genomes of representatives from all 33 angiosperm families examined to date (Whitfeld and Bottomley 1983; Gillham et al. 1985; Palmer 1985a, b) as well as those from the gymnosperm *Ginkgo biloba* (Palmer and Stein 1986), three species in the fern genus *Osmunda* (Palmer and Stein 1982; Stein et al. 1986), and the liverwort *Marchantia polymorpha* (Ohyama et al. 1983). The only documented loss of the ancient inverted repeat structure among land plants is within one section of the family Leguminosae – represented so far by the species pea (*Pisum sativum*), broad bean (*Vicia faba*) and chickpea (*Cicer arietinum*) – whose chloroplast genomes have deleted one entire segment of the inverted repeat (Koller and Delius 1980; Palmer and Thompson 1981a; Chu and Tewari 1982). With this single exception, then, the basic inverted repeat structure has been a feature of chloroplast DNAs throughout some 400,000,000 years of land plant evolution.

[4] *Present address:* 271 Hill Crest Road, Stafford, PA 19087, USA
[5] *Present address:* Department of Botany, North Carolina State University, Raleigh, NC 27695, USA

Offprint requests to: J. D. Palmer

A strong correlation exists between the presence of the inverted repeat and a stable chloroplast genome in which major sequence rearrangements are rare (Fluhr and Edelman 1981; Palmer and Thompson 1982; Palmer et al. 1983a, b; de Heij et al. 1983; Palmer and Stein 1986). In contrast, numerous rearrangements have occurred in the plastid genomes of those legumes that have lost the inverted repeat structure (Palmer and Thompson 1981a, 1982; Mubumbila et al. 1984; Palmer et al. 1985a). Two equally tenable hypotheses have been postulated to explain these observations (Palmer and Thompson 1982): (1) The inverted repeat may stabilize the chloroplast genome against rearrangements. In this case the loss of the inverted repeat in certain legumes would lead directly to an increased frequency of rearrangements. (2) The chloroplast genome of certain legumes somehow acquired the ability to rearrange, and the deletion of the inverted repeat simply represents one of many rearrangements following this acquisition.

In order to clarify the relationship, if any, between the loss of the inverted repeat and the apparent destabilization of the chloroplast genome we have investigated chloroplast DNAs from four additional legume species and compared them in organization to previously characterized chloroplast DNAs. Our results show that the inverted repeat deletion has preceded all known legume chloroplast DNA rearrangements, but that the deletion does not always produce an unstable genome. Instead, we show that legume chloroplast genomes lacking the inverted repeat may display quite different patterns of structural evolution and rearrangement. Finally, we present the first demonstration of large dispersed repetitive sequences in a land plant chloroplast DNA.

Materials and methods

Chloroplast DNAs from alfalfa (*Medicago sativa* cv. Regen S; seed obtained from T. Bingham), mung bean (*Vigna radiata* cv. berken; seed obtained from W. Atlee Burpee Co.), lupine (*Lupinus polyphyllus* var. *regalis*, Russell lupines, plants obtained from a local nursery), wisteria (*Wisteria floribunda*, leaf material obtained from a plant growing on the Carnegie Institution grounds), and subclover (*Trifolium subterraneum*, cv. Tallarook; seed obtained from Regional Plant Introduction Station, Experiment, Georgia, 30212) were prepared from sucrose gradient-purified chloroplasts as described (Palmer 1982, 1986). Broad bean (*Vicia faba* cv. long pod; seed obtained from W. Atlee Burpee Co.) chloroplast DNA was prepared from DNase I-treated chloroplasts as described (Kolodner and Tewari 1975). *E. coli* plasmid DNA was isolated as described by Birnboim and Doly (1979).

Alfalfa chloroplast DNA PstI fragments were ligated to PstI-digested pBR322 as described (Palmer and Thompson 1981b), transformed into *E. coli* strain LE392 (rk^-, mk^-, $recA^+$, SuII, SuIII) according to Dagert and Ehrlich (1979), and recombinant colonies selected on the basis of a tetracycline resistant, ampicillin sensitive phenotype. Subclover chloroplast DNA PstI fragments were ligated to PstI-digested pUC8 (Vieira and Messing 1982), and recombinant (white) colonies selected on X-gal indicator plates. Mung bean chloroplast DNA clones used in this study are those originally described in Palmer and Thompson (1981b) and subclones thereof whose construction will be described elsewhere (J. Palmer and W. Thompson in preparation). Gene probes are described in Table 1.

Restriction endonuclease digestions, agarose gel electrophoresis, bidirectional nitrocellulose filter transfers of DNA, labeling of recombinant plasmids with ^{32}P by nick translation, and filter hybridizations were performed exactly as described (Palmer 1982, 1986). All filters were washed at 65 °C in 2 x SSC, 0.1% SDS prior to autoradiography. Tobacco chloroplast 16S and 23S rRNAs were hydrolyzed by alkali and labeled at the 5′ end with ^{32}P using polynucleotide kinase (Maizels 1976).

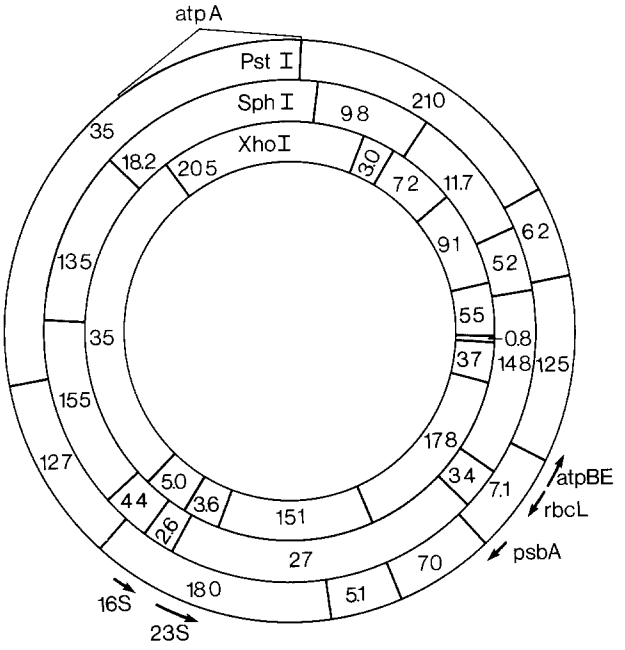

Fig. 1. Physical map of the alfalfa chloroplast chromosome. Gene positions and orientations are based on the hybridization data of Table 1, except for the orientation of the rRNA genes which is by analogy to other chloroplast rRNA operons (Whitfeld and Bottomley 1983)

Results

Organization of the alfalfa chloroplast genome

In order to facilitate restriction mapping, and in particular to allow direct examination of sequence homologies with other legume chloroplast DNAs, we cloned all but one of the nine PstI fragments of alfalfa chloroplast DNA into the PstI site of pBR322. The alfalfa PstI sites, and also sites for XhoI and SphI, were mapped by hybridizing each of the eight PstI clones to replica nitrocellulose filters containing alfalfa chloroplast DNA digested with PstI, SphI-PstI, SphI, SphI-XhoI, and XhoI [see Palmer (1982) for full description and illu-

Table 1. Summary of gene mapping hybridizations

Gene probe				Filter-bound fragment hybridized			
	Probe fragment	Source species	Reference[a]	Alfalfa		Mung bean	
				SphI-PstI	XhoI	PstI	SacI
*rbc*L 5′	574 bp PstI	Maize	(1)	2.6	17.8	b	b
*rbc*L 3′	821 bp PstI-KpnI	Spinach	(2)	3.4, 2.6	17.8	b	b
*psb*A 5′	1,800 bp BglII	Mung bean	(3)	7.0, 1.0	17.8	b	b
*psb*A 3′	850 bp HindIII	Spinach	(4)	7.0	17.8	b	b
*atp*BE 5′	1,977 bp EcoRI	Spinach	(5)	12, 2.6	17.8	11.1, 1.2	17.4, 9.8
*atp*BE 3′	1,670 bp EcoRI	Spinach	(5)	12	17.8	11.1	17.4
*atp*A	2,400 bp SalI	Spinach	(6)	16.2	20.5	34	16.2, 4.1
16S rRNA	1,489 b RNA	Tobacco	(7)	2.6	5.0	b	b
23S rRNA	2,804 b RNA	Tobacco	(8)	13.0	15.1, 3.6	b	b

[a] (1) McIntosh et al. (1980); (2) Zurawski et al. (1981); (3) Palmer et al. (1982) and J. Palmer and W. Thompson, unpublished data; (4) Zurawski et al. (1982a); (5) Zurawski et al. (1982b); (6) Westhoff et al. (1981); (7) Tohdoh and Sugiura (1982); (8) Takaiwa and Sugiura (1982)

[b] Gene mapping already published (Palmer and Thompson 1981a; Palmer et al. 1982)

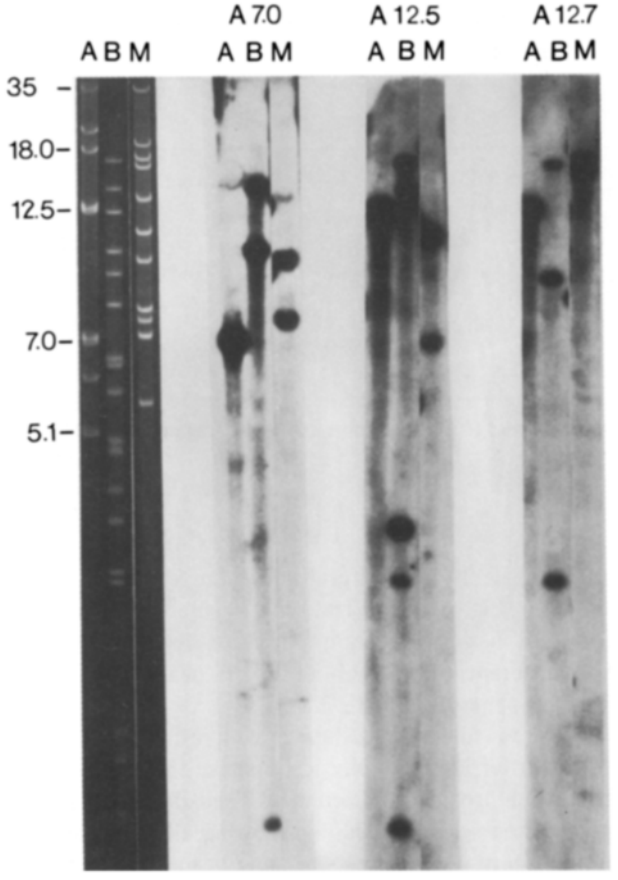

Fig. 2. Hybridization of alfalfa chloroplast DNA restriction fragments to filter-bound broad bean and mung bean chloroplast DNAs. Three cloned alfalfa chloroplast DNA PstI fragments, A7.0, A12.5 and A12.7, were each labeled with ^{32}P by nick-translation and hybridized to replica nitrocellulose filters containing alfalfa PstI fragments (*A*), broad bean SalI-KpnI fragments (*B*) and mung bean PstI fragments (*M*) separated on a 0.7% agarose gel. Numbers at left indicate size in kb of certain of the alfalfa PstI fragments

stration of this mapping strategy]. The single uncloned PstI fragment (of 35 kb) was isolated from a preparative agarose gel and hybridized to a similar filter. Construction of a complete map of the PstI, SphI and XhoI sites reveals that alfalfa chloroplast DNA exists as a single circular molecule 126 kb in length (Fig. 1). None of the PstI fragments hybridized to any restriction fragments (including all other PstI fragments) that did not directly overlap them on the map, and thus no large repeated sequences (except for repeats which may be clustered within an individual PstI fragment) are present on the alfalfa genome (see Fig. 2).

Seven different chloroplast genes were mapped onto the alfalfa chloroplast genome using hybridization probes from maize, spinach, mung bean and tobacco (Table 1). The *rbc*L and rRNA probes are gene specific, while the *psb*A, *atp*BE and *atp*A probes contain additional, small extra-genic sequences (see references to Table 1). These gene mapping hybridizations reveal that alfalfa chloroplast DNA contains only a single set of rRNA genes and that the *psb*A, *rbc*L and *atp*BE genes are tightly linked (Fig. 1). In addition, the use of two different probes for each of the *psb*A, *rbc*L and *atp*BE genes allows us to determine the direction of transcription of these genes with respect to mapped PstI and SphI restriction sites (Table 1, Fig. 1).

Table 2. Summary of mung bean, alfalfa and broad bean rearrangement hybridizations. The cloned alfalfa and mung bean restriction fragments listed in the left column were each labeled with ^{32}P and hybridized to nitrocellulose filters containing the indicated restriction fragments from broad bean, mung bean and alfalfa chloroplast DNA. Fragment sizes are given in kb

Probe fragment	Filter-bound fragment hybridized		
Mung bean	Alfalfa		
PstI[a]	SphI	PstI	XhoI
18.8	27, 4.4, 2.6	18.0	15.1, 5.0, 3.6
17.2	15.5, 4.4	35, 12.7	35, 5.0
16.2	15.5, 4.4	35, 12.7	35, 5.0
13.3	18.2, 13.5	35	35, 20.5
12.8	27, 4.4, 2.6	18.0, 7.0	17.8, 15.1, 5.0, 3.6
11.1	14.8	12.5	17.8, 5.5, 3.7, 0.8
10.6	15.5, 13.5	35	35
9.8	18.2, 9.8	35, 21.0	20.5
9.7	27	7.0, 5.1	17.8, 15.1
7.8	11.7, 9.8	21.0	20.5, 7.2, 3.0
7.5	27, 14.8, 3.4	7.1, 7.0	17.8
7.0	14.8, 11.7, 5.2	12.5, 6.2	9.1, 5.5
5.6	11.7	21.0	9.1, 7.2
1.2	27	7.1	17.8
Alfalfa	Broad bean	Mung bean	
PstI	KpnI-SalI	PstI	SalI
21.0	17.2, 12.3, 4.9, 3.1	34, 7.8, 5.6, 1.1	21.5, 16.5
18.0	14.0, 6.6, 5.0	18.8, 12.8	54
12.7	17.2, 9.3, 3.2	17.2, 16.2	54, 24.4, 20.5
12.5	17.2, 14.0, 3.9, 3.2, 1.5	11.1, 7.0	24.4, 21.5
7.1	10.2, 3.9	7.5, 1.2	24.4
7.0	14.0, 10.2, 1.5	12.8, 9.7, 7.5	54, 24.4
6.2	17.2	7.0	21.5
5.1	14.0	9.7	54

[a] The mung bean fragments are all PstI fragments, except for one SalI fragment (13.3 kb) and two SalI-PstI fragments (10.6 kb and 9.8 kb)

Comparative organization of the alfalfa, mung bean, and broad bean genomes

Given that alfalfa chloroplast DNA and those from broad bean (Koller and Delius 1980) and pea (Palmer and Thompson 1981a) are similar in terms of size, lack of any detectable repeats and the presence of only one set of rRNA genes, it was of interest to determine whether alfalfa chloroplast DNA has undergone rearrangements similar to those found in broad bean and pea (Palmer and Thompson 1982). In order to compare the linear sequence organization of legume chloroplast DNAs we initially hybridized each of the eight cloned alfalfa PstI restriction fragments to filter-bound chloroplast DNAs from mung bean (which contains the large inverted repeat missing in the other three legumes) and broad bean (Table 2). The three representative hybridizations shown in Fig. 2 demonstrate that each alfalfa PstI probe fragment has specific homologies to between two and five different mung bean or broad bean fragments in any given enzyme digest.

Figure 3 summarizes these cross-hybridization experiments in terms of the physical maps of the alfalfa, mung bean and broad bean genomes. The alfalfa-mung bean comparison is simplified by showing alfalfa hybridization to only one of the two mung bean inverted repeat segments (cf. Table 2, Fig. 2). This simplification dramatizes what we believe is the most likely pathway for derivation of the alfalfa genome from a mung bean-like ancestral genome – that is, deletion of the mung bean inverted repeat segment which lies between the sequences in the small and large single copy regions that hybrid-

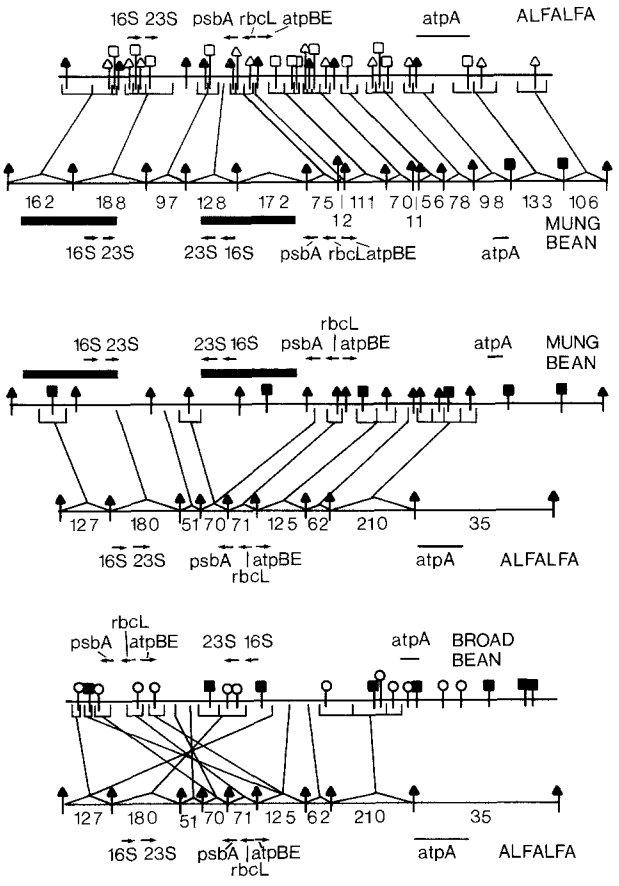

Fig. 3. Arrangement of homologous sequences in the alfalfa, mung bean and broad bean genomes. The hybridization data summarized in Table 2 are diagrammed in terms of the alfalfa (Fig. 2), mung bean (Palmer and Thompson 1981a) and broad bean (Koller and Delius 1980) physical maps. Gene mapping data for alfalfa and mung bean are from Table 1 and for broad bean are from Ko et al. (1984) and Shinozaki et al. (1984). The extent of the alfalfa or mung bean fragments used as probes is indicated by the two lines that converge above the fragments, while the size of each fragment is given below in kb. The filter-bound alfalfa, mung bean or broad bean fragments to which the probe fragments hybridize are indicated by the lines leading from the probe fragments to the filter-bound fragments. The two alfalfa-mung bean comparisons have been simplified by omitting the observed hybridization between MB17.2 and A12.7 (see text for rationale). Restriction sites shown: (▲), PstI; (▼), SalI; (⚲), KpnI; (⚲), XhoI; (⚲), SphI

ize to A7.0. If this interpretation is correct, then the entirety of this inverted repeat segment has been lost from the alfalfa genome. A7.0 does not hybridize to either MB16.2 or MB17.2, which carry approximately 60% of the inverted repeat, and even at autoradiographic exposures 20 times longer that shown in Fig. 2, it hybridizes to only one (MB12.8) of the two fragments, MB12.8 and MB18.8, which carry the remainder of the mung bean inverted repeat (Fig. 3, Table 2). We interpret these results to indicate that A7.0 hybridizes only

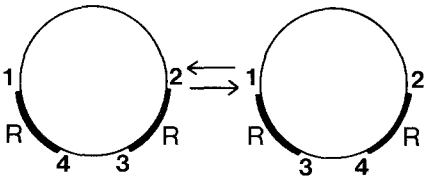

Fig. 4. Orientation heterogeneity of inverted repeat-containing chloroplast genomes resulting from intramolecular recombination between segments of the inverted repeat. Sequences 1–4 are single copy regions flanking the inverted repeat (*R*). In the context of the mung bean-alfalfa comparisons presented herein, the 1R4 sequence combination might be MB16.2 and MB18.8, 2R3 be MB17.2 and MB12.8, 1R3 be MB16.2 and MB12.8, and 2R4 be MB17.2 and MB18.8

to the small single copy portion of MB12.8 and that the entirety of the inverted repeat segment corresponding to MB12.8–MB17.2 has been deleted from alfalfa.

A further consequence of these comparisons is that one can determine which of the four inverted repeat configurations has been lost in alfalfa. It now appears that all inverted repeat-containing chloroplast genomes exist as two equimolar populations of molecules differing only in the relative orientation of their single copy sequences (Bohnert and Loffelhardt 1982; Palmer 1983; Mubumbila et al. 1983; Palmer et al. 1984, 1985b; Aldrich et al. 1985; Brears et al. 1986; Stein et al. 1986). The consequence of this inversion heterogeneity for producing four different combinations of the inverted repeat and flanking single copy sequences is diagrammed schematically in Fig. 4. Using mung bean as the reference genome, one can conclude that the inverted repeat deletion must have occurred in a molecule in which MB12.8 and MB17.2, and also MB18.8 and MB16.2, were adjacent (as opposed to the other 50% of the molecules with MB12.8 adjacent to MB16.2 and MB18.8 to MB17.2) and, as discussed above, within the MB12.8–MB17.2 kb repeat segment.

To confirm these inferences, and to complete the comparison of the alfalfa and mung bean genomes in the region of the uncloned 35 kb alfalfa PstI fragment, we hybridized cloned mung bean restriction fragments to filter-bound alfalfa chloroplast DNA (Table 2). Alfalfa and mung bean chloroplast DNAs are colinear over their entire lengths except for the deletion of one inverted repeat segment in alfalfa (Fig. 3). This colinearity extends on a fine-scale level to the precise positions and orientations of the various genes whose map positions are shown in Fig. 3.

The alfalfa-broad bean hybridizations reveal that certain sequences common to the two DNAs have undergone rearrangement (Figs. 2, 3, Table 2). In particular, A12.7 and A7.0 each hybridize to two distinct regions of the braod bean genome and A12.5 to three different regions. Figure 5 presents a model for the evolution of the broad

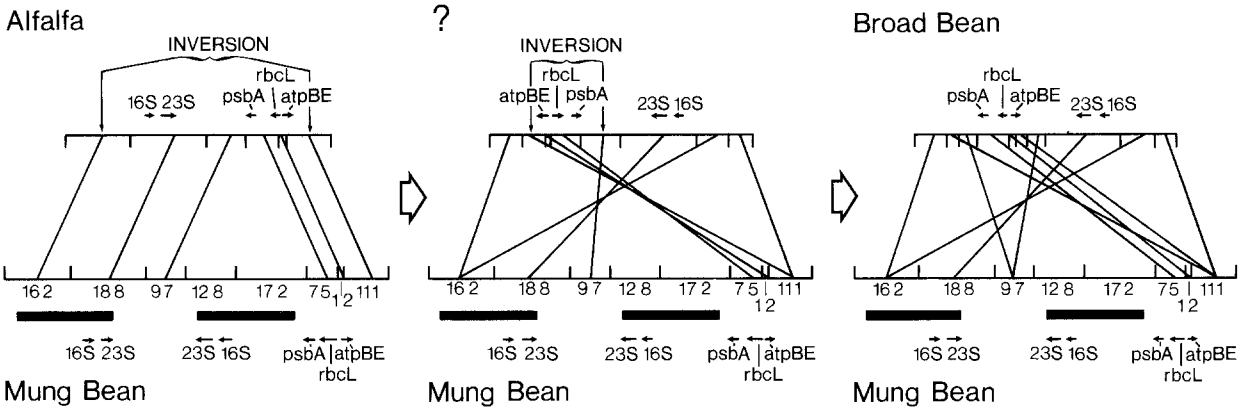

Fig. 5. Derivation of a portion of the broad bean chloroplast genome from a portion of an alfalfa-like ancestral chloroplast genome. *Left*, mung bean hybridization pattern to a portion of the alfalfa genome (Fig. 3); *middle*, to a portion of a hypothetical legume chloroplast genome which is derived from an alfalfa-like genome by a single inversion with endpoints within MB16.2 and MB11.1, and *right*, to a portion of the broad bean genome (Fig. 3 and also Palmer and Thompson 1982). PstI restriction sites are shown for the mung bean maps, while the other three maps are idealized so that the sites shown are the same distance apart as for mung bean. The two inversions are represented as occuring exactly in the middle of a given PstI fragment. For clarity of presentation we have omitted the hybridization of the small single copy-specific portion of MB12.8 to alfalfa DNA and its derivatives

Table 3. Hybridization of mung bean fragments to broad bean inversion endpoint regions

Probe fragment		Filter-bound fragment hybridized (broad bean SalI-KpnI)
Primary clone	Subclone used in hybridization	
16.2 kb PstI	0.7 kb HindIII	9.3 kb
16.2 kb PstI	2.3 kb HindIII-BamHI	9.3
16.2 kb PstI	1.1 kb BamHI	9.3
16.2 kb PstI	0.8 kb BamHI	[a]
16.2 kb PstI	2.2 kb BamHI-SalI	9.3
16.2 kb PstI	1.3 kb SalI-EcoRI	9.3, 3.2
16.2 kb PstI	1.2 kb EcoRI	3.2
16.2 kb PstI	2.2 kb EcoRI	17.2, 3.2
16.2 kb PstI	1.8 kb EcoRI-PstI	17.2
9.7 kb PstI	3.1 kb PstI-BamHI	14.0, 10.2, 1.5
9.7 kb PstI	1.7 kb BamHI	14.0
9.7 kb PstI	4.8 kb BamHI-PstI	14.0
11.1 kb PstI	1.0 kb PstI-HindIII	3.9
11.1 kb PstI	1.5 kb HindIII	3.9
11.1 kb PstI	1.9 kb HindIII	14.0, 3.9, 1.5
11.1 kb PstI	4.2 kb HindIII	17.2, 3.2, 1.5
11.1 kb PstI	2.4 kb HindIII-PstI	17.2

[a] No detectable hybridization

bean genome from an ancestral, alfalfa-like genome by two specific inversions – each with one end in the region homologous to MB11.1 (a nearly equivalent fragment to A12.5) and with a second end in either MB16.2 or in MB9.7. Clearly, the two inversions may have occurred with an opposite temporal order from that shown in Fig. 5. Further examination of additional legume species that lack the inverted repeat may be expected to reveal an intermediate genome (designated in Fig. 5 by a question mark) which has sustained only one of the two inversions. Note that essentially the same pattern of rearrangements found here between alfalfa and broad bean (Fig. 3) was previously observed using large cloned fragments from mung bean to map broad bean homologies (Palmer and Thompson 1982).

To locate the endpoints of the two braod bean inversions more precisely we have carried out additional heterologous mapping hybridizations using subclones of the original mung bean clones as probes. These hybridizations indicate that the inversion breakpoints are located within regions homologous to a 2.2 kb EcoRI fragment from the mung bean inverted repeat, a 3.1 kb PstI-BamHI fragment from the small single copy region, and HindIII fragments of 1.9 kb and 4.2 kb from the large single copy region (Table 3, Fig. 6). Relative to broad bean, the breakpoints are located in SalI-KpNI fragments of 3.2 kb, 1.5 kb, 14.0 kb, and 17.2 kb. We conclude that the broad bean genome was derived from a mung bean-like ancestral genome by three major sequence rearrangements: (1) A deletion of the entire 26 kb MB12.8–MB17.2 inverted repeat segment; (2) an inversion, approximately 49 kb in size, with endpoints located about 7 kb upstream of the rRNA operon and 6 kb downstream of the *atp*BE operon; and (3) a smaller inversion, approximately 17 kb in size, with endpoints located about 3 kb downstream from *psb*A and also 3 kb downstream from *atp*BE. As indicated in the preceding paragraph, we have no evidence bearing on the temporal order of the two inversions. However, we feel it likely that the inverted repeat deletion occurred prior to both inversion, in the common ancestor of broad bean and alfalfa (see Discussion).

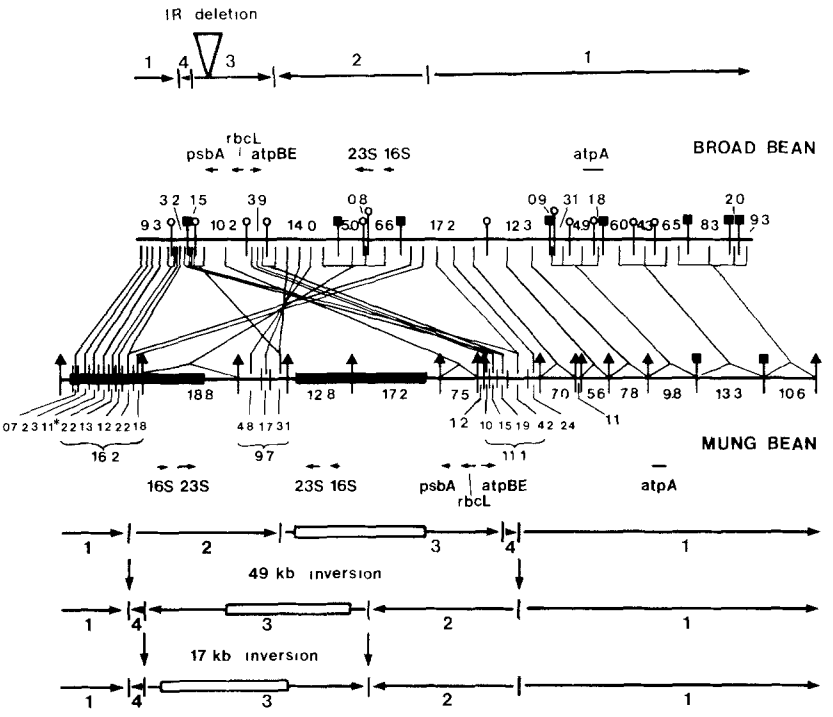

Fig. 6. Location of inversion endpoints in broad bean chloroplast DNA. Crossing diagram shows the hybridization pattern of two sets of mung bean chloroplast clones – a set of 13 primary clones containing PstI, SalI or SalI-PstI fragments that together cover 99% of the genome and whose hybridization was previously reported (Palmer and Thompson 1982), and a set of secondary subclones derived from the primary PstI clones of 16.2 kb, 9.7 kb, and 11.1 kb and whose hybridization is reported in Table 3 – to SalI-KpnI fragments of broad bean. Gene mapping sources and conventions used in constructing this diagram are described in the legend to Fig. 3. The *horizontal arrows* (labeled 1, 2, 3 and 4) above and below the crossing diagram indicate blocks of sequences which retain colinearity between the two genomes. *Vertical slashes* between arrows indicate inversion breakpoints. The *open box* within region 3 indicates the inverted repeat segment that is absent in broad bean. Restriction sites shown: (↑), PstI; (▼), SalI; (♀), KpnI

Comparative organization of the lupine, wisteria, subclover and mung bean genomes

Cloned restriction fragments covering the entire mung bean chloroplast genome were hybridized to Southern blots of three previously uncharacterized legume chloroplast DNAs in order to compare sequence organization in the four genomes. We reasoned that heterologous mapping hybridizations should allow us to construct fairly complete restriction maps and also to determine the nature and extent of simple rearrangements for genomes whose basic organization is similar to that of mung bean. Where extensive rearrangement has occurred, heterologous hybridizations are not expected to yield an unambiguous map, but should at least give an initial estimate of the amount, and perhaps also the type, of rearrangement. We chose for analysis chloroplast DNAs from subclover, a member of the same tribe as alfalfa, and lupine and wisteria, members of two unrelated tribes of legumes so far unexamined in terms of their chloroplast genomes.

Hybridization of mung bean clones to Southern blots containing lupine and wisteria chloroplast DNA digests allowed construction of complete fragment maps for PvuII and SacI (Fig. 7). The lupine genome, estimated to be 147 kb in size, features a large inverted repeat at least 21.6 kb long, which is homologous in sequence and position to the mung bean inverted repeat. The lupine and mung bean genomes are colinear throughout the two single copy regions, with the exception of slight misalignments at the two ends of the large single copy region (Fig. 7). Similar misalignments were found in comparing the mung bean and soybean chloroplast genomes (Palmer et al. 1983a), and probably reflect the accumulation of small deletions and insertions in these two regions.

The 130 kb wisteria genome lacks one segment of the inverted repeat, but is otherwise colinear with the mung bean genome (Fig. 7). It is thus extremely similar to the alfalfa genome in overall sequence arrangement (cf. Figs. 3, 7). In order to determine whether the same inverted repeat configuration was lost in wisteria as in alfalfa, additional experiments were performed using smaller mung bean fragments as hybridization probes against wisteria DNA digested with XhoI, an enzyme found in preliminary analysis to cleave within the region of MB9.7 homology. These hybridizations (Fig. 8) establish a linkage in wisteria between small and large single copy sequences immediately flanking the MB12.8–MB17.2 inverted repeat configuration, supporting the hypothesis that this is probably the segment which was deleted in an ancestral wisteria genome. Thus, it appears that the same one of the four inverted repeat orientations was lost in both wisteria and alfalfa (cf. Figs. 3, 8).

In contrast to the situation for lupine and wisteria, we were unable to construct an unambiguous restriction map for subclover based solely on heterologous hybridizations. Table 4 summarizes the hybridization results obtained using mung bean fragments to probe subclover BalI fragments; similarly complex results were also obtained with subclover SacI fragments. One can at best

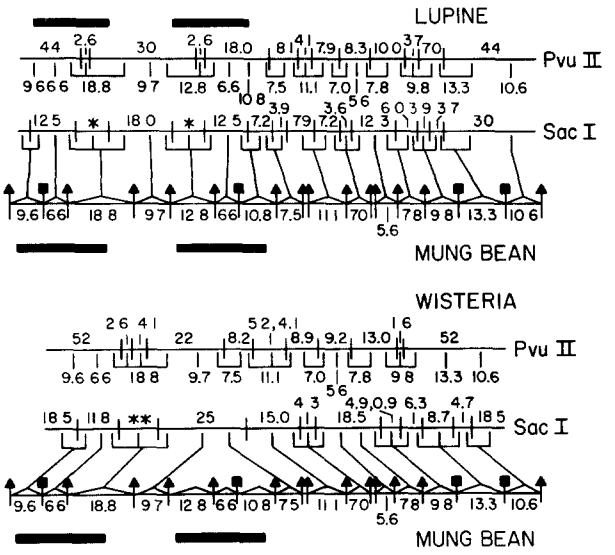

Fig. 7. Arrangement of homologous sequences in the lupine, wisteria and mung bean genomes. The fourteen cloned fragments whose sizes are given in the mung bean maps were each hybridized to replica nitrocellulose filters containing PvuII and SacI restriction fragments from lupine and wisteria chloroplast DNA. The size and inferred order of the lupine and wisteria fragments are shown relative to the order of the mung bean fragments to which they hybridize. The diagrams have been simplified by showing only one of the regions of hybridization of each of the mung bean inverted repeats (i.e. mung bean fragments of 12.8 kb, 6.6 kb and 10.8 kb hybridize to the same lupine and wisteria fragments as do those probe fragments of 18.8 kb, 6.6 kb and 9.6 kb, respectively). The *single asterisks* denote a 9.1 kb region within the lupine inverted repeat which is composed of five separate, but unordered SacI fragments, 2.8, 2.6, 1.9, 1.2 and 0.65 kb in size. The *double asterisk* denotes an 11.2 kb region in wisteria which is composed of four separate, but unordered SacI fragments, 5.4, 2.8, 1.8 and 1.2 kb in size. The *long, heavy black lines* indicate the approximate extent of the inverted repeat in the mung bean (Palmer and Thompson 1981a; Chu and Tewari 1982) and its minimum extent in lupine, as defined by a doublet SacI fragment of 12.5 kb which maps adjacent to the cluster of five doublet SacI fragments of 2.8 through 0.65 kb. Mung bean restriction sites shown: (↑), PstI; (↟), SalI

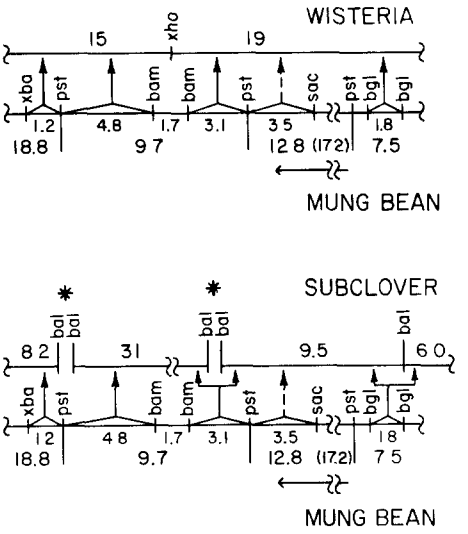

Fig. 8. Linkage relationships in the region of the deleted inverted repeat segment in wisteria and subclover relative to mung bean. Five mung bean fragments – from left to right, a 1.2 kb XbaI-PstI fragment subcloned from MB18.8, a 4.8 kb PstI-BamHI fragment and a 3.1 kb BamHI-PstI fragment, both subcloned from MB9.7, a 3.5 kb PstI-SacI fragment subcloned from MB12.8, and a 1.8 kb BglII fragment isolated from MB7.5 – were each labeled with ^{32}P and hybridized to replica nitrocellulose filters containging wisteria XhoI fragments and subclover BalI fragments. The *vertical arrows* indicate regions of cross-hybridization. The *horizontal arrows* beneath the mung bean map indicate the extent of the inverted repeat. The *dashed arrow* indicates the region of hybridization attributed to the small single copy portion of the 3.5 kb probe. The majority of the hybridization of this probe was to the same region (i.e., the 15 kb XhoI fragment in wisteria and the 8.3 kb BalI fragment in subclover) hybridized to by the 1.9 kb probe; this cross-hybridization is not shown in order to simplify the diagrams and dramatize linkage relationships in the region of the deleted inverted repeat segment. The *asterisks* indicate the positions of possible rearrangement endpoints in subclover

deduce only relatively small subclover linkage groups from these data. For example, MB5.6 and MB7.0 each hybridize solely to a BalI fragment of 16.5 kb, which is one of three BalI fragments to which MB11.1 hybridizes (Table 4). Since these three mung bean fragments are adjacent to one another in the genome, it is quite possible that this region has a conserved organization in subclover. Note that even this region of conservation cannot extend over the entire 16.5 kb BalI fragment, since MB13.3, which is located 18 kb away from these three mung bean fragments, also hybridizes to this subclover fragment.

We interpret these complex cross-hybridization patterns as indicating that the subclover and mung bean genomes are extensively rearranged relative to one another. We base this conclusion on the following observations: (1) In a number of cases, for example, as described above, non-linked mung bean fragments hybridize to the same linkage group in subclover. (2) Several relatively small subclover BalI fragments (e.g. of 5.4 kb and 7.2 kb) each hybridize with a number of relatively large mung bean fragments in a manner totally at odds with a colinear arrangement of the cross-hybridizing fragments. (3) Similarly, several mung bean fragments (e.g. of 7.5 kb, 11.1 kb, 9.8 kb and 10.6 kb) each hybridize to an excessively large number of subclover fragments in a manner inconsistent with their colinearity.

In studies to be published elsewhere, we have recently constructed detailed restriction site and gene maps for the subclover genome and shown conclusively that it

Table 4. Summary of mung bean and subclover hybridizations. A "+" indicates positive hybridization between the indicated mung bean and subclover fragments. Mung bean probe fragments are given according to their map order (see Fig. 7).

Filter-bound subclover BalI fragment hybridized	Mung bean probe fragment												
	9.6	6.6	18.8	9.7	12.8[a]	7.5	11.1	7.0	5.6	7.8	9.8	13.3	10.6
31	+	+	+	+		+	+						+
20.2											+	+	+
16.5							+	+	+			+	
13.4	+	+				+	+						
9.5				+	+	+							
8.2			+							+			
7.2										+	+		+
6.0						+							
5.4										+	+	+	+
4.6	+												
4.4													
3.5			+										
3.3											+		
2.3											+		
1.7													
1.1 (2x)	+										+		
1.0											+		

[a] Only the single copy specific hybridization is given (see Fig. 8)

is indeed highly rearranged relative to mung bean and that one entire segment of the inverted repeat has been deleted in subclover (B. Milligan and J. Palmer, unpublished data). In the context of the present study, it is of interest to examine whether subclover lacks the same inverted repeat segment configuration as is missing in alfalfa and wisteria. Three relatively small, single copy mung bean fragments immediately flanking the ends of the MB12.8–MB17.2 kb repeat segment each hybridize to the 9.5 kb subclover BalI fragment (Fig. 8). Precisely the same linkage relationships in this region were previously observed for alfalfa (Fig. 3) and wisteria (Fig. 8), leading us to conclude that all three legume genomes lack the same (out of four possible ones) inverted repeat segment.

An unusual, for land plant chloroplast genomes, aspect of the rearrangements in subclover is the presence of at least one major family of dispersed repeated sequences. These were found in the course of initial subclover cloning and mapping experiments and are illustrated in Fig. 9. Each of two cloned subclover PstI fragments, 3.5 kb and 7.2 kb in size, hybridizes strongly to the same five BalI fragments, of 31, 7.2, 5.4, 4.6, and 1.1 kb in size, while the 3.5 kb PstI clone also hybridzes to a 1.7 kb BalI fragment. The 7.2 kb probe lacks any BalI sites and is contained within the 31 kb fragment (data not shown); thus, it must contain repeated sequences which are also present on each of the four other cross-hybridizing BalI fragments. The 3.5 kb probe has an internal BalI fragment of 1.1 kb and overlaps with BalI fragments of 7.2 kb and 4.6 kb; thus it must contain repeats present at a minimum on BalI fragments of 31 kb, 5.4 kb and 1.7 kb. The nearly identical qualitative hybridization patterns of the two PstI clones suggests that the cross-hybridizing repeats are probably part of one or two higher copy number repeat families rather than a larger number of lower copy number families. That is, it seems more likely that the repeats responsible for the 7.2 kb PstI fragment's cross-hybridizations are part of a single family of at least five different members than that they belong to four different two-member families. Based on the strength of hybridization signals, we roughly estimate the repeats to be several hundred base pairs in size.

Discussion

Alfalfa and wisteria chloroplast DNAs occupy an intermediate evolutionary position among angiosperm chloroplast DNAs. They share with pea (Palmer and Thompson 1981a), broad bean (Koller and Delius 1980), chickpea (Chu and Tewari 1982) and subclover (this report) the loss of a prominent inverted repeat structure which is otherwise universally present among land plants. However, unlike subclover, pea, and broad bean, which are considerably rearranged relative to one another and also relative to the inverted repeat-containing mung bean

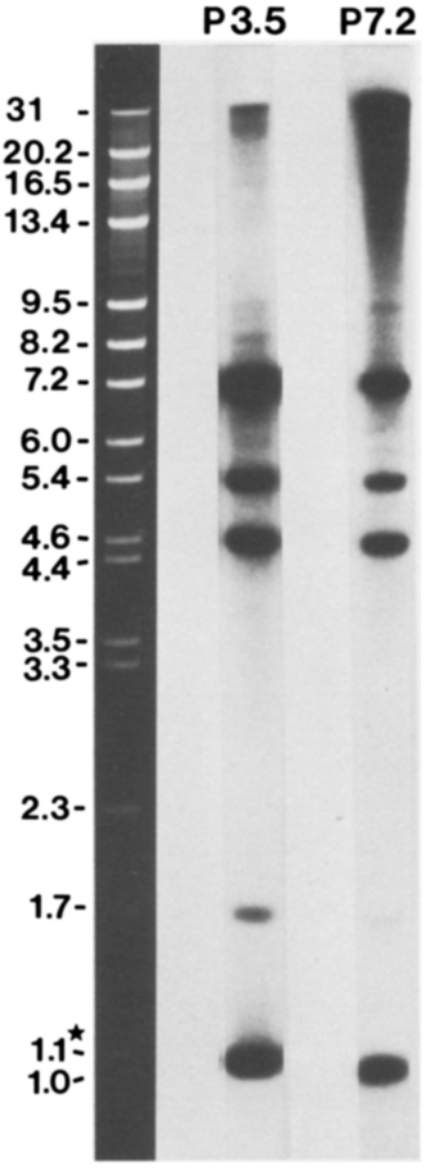

Fig. 9. Repeated sequences in the subclover chloroplast genome. Two cloned subclover chloroplast DNA PstI fragments, P3.5 and P7.2, were each labeled with ^{32}P by nick-translation and hybridized to replica nitrocellulose filters containing subclover BalI fragments separated on a 0.7% agarose gel. Fragment sizes are given in kb. A doublet intensity band of 1.1 kb is marked with an *asterisk*

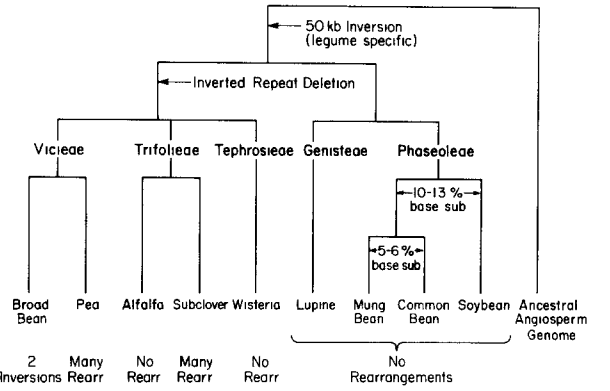

Fig. 10. Legume phylogeny based on chloroplast DNA rearrangements. Data are from Palmer and Thompson (1982), Palmer et al. (1983a) and this report. Tribal names are written across the middle of the tree

genome (Palmer and Thompson 1982), alfalfa and wisteria chloroplast DNAs differ from the mung bean genome only by the simple deletion of one entire segment of the inverted repeat. (Chickpea has not been examined in this regard.) Thus, despite the loss of the inverted repeat, alfalfa and wisteria resemble inverted repeat-containing chloroplast DNAs (Fluhr and Edelman 1981; Palmer and Thompson 1982; Palmer et al. 1983a, b; de Heij et al. 1983; Palmer and Stein 1986) in being stable in sequence organization over quite long periods of evolutionary time. Of the four legumes newly studied herein, lupine most closely resembles the inverted repeat-containing class of legumes (mung bean, common bean, and soybean; Palmer et al. 1983a), not only in retaining the primitive inverted repeat structure, but also in sharing overall sequence colinearity (Figs. 7, 10).

Has the inverted repeat been lost only once among legumes, or has it been lost on multiple, independent occasions, for example, once each within the tribes Tephrosieae (wisteria), Trifolieae (alfalfa and subclover), Vicieae (pea and broad bean) and Cicereae (chickpea)? We favor the first hypothesis – that the inverted repeat deletion occurred as a single, common event among these legumes – for three reasons. First, it is more parsimonious to postulate one event, rather than multiple independent events, particularly when the event in question is such a rare one in the plant kingom as a whole. Second, there are no incongruities between a dichotomous classification (Fig. 10) of legume species according to whether they retain or have lost the inverted repeat and standard phylogenetic schemes for the legume family (Polhill and Raven 1981). Third, it appears that the same one of the four inverted repeat configurations was lost in alfalfa, wisteria, subclover, broad bean and pea. The hybridization data presented in this paper, using cloned mung bean fragments as probes, provide support for this conclusion in the case of alfalfa, wisteria, subclover and broad bean (Figs. 2, 3, 6, 8). Furthermore, the ability to model broad bean evolution from alfalfa by only two specific inversions is also consistent with the two species having lost the same inverted repeat segment. Finally, cross-hybridizations between cloned mung bean fragments and pea chloroplast DNA indicate that the same linkage relationships exist in the region surrounding the deleted inverted repeat segment in pea as we have inferred in this paper for the other four legumes (J. Palmer and W. Thompson, unpublished data).

If, indeed, there has been only a single loss of the inverted repeat among legumes, then, since two of the species (alfalfa and wisteria) which have sustained this loss are otherwise unrearranged relative to legumes which retain the inverted repeat (Figs. 3, 7), it follows that all the other sequence rearrangements observed in subclover, broad bean and pea must have occurred subsequent to the inverted repeat deletion. Furthermore, we can distinguish among these three rearranged legume chloroplast genomes in terms of the nature, frequency and phylogenetic independence of their specific rearrangements. The taxonomic placement of subclover in the same tribe as alfalfa implies that the subclover rearrangements occurred during intratribal divergence and therefore independently from the rearrangements in pea and broad bean, members of a separate tribe (Fig. 10). While the pea and broad bean rearrangements cannot be distinguished from one another by such phylogenetic arguments, hybridization analysis indicates that there are no shared rearrangements between the two species. That is, both broad bean inversions have one endpoint within the MB11.1-homologous region, one occurring within the 1.9 kb HindIII subfragment of MB11.1 and one within the 4.2 kb HindIII subfragment (Figs. 5, 6, Table 3). In pea, however, there is only a single inversion within the MB11.1 region (Palmer and Thompson 1981, 1982), and, most importantly, this maps to a different locale (the 2.4 kb HindIII-PstI subfragment) than either of the broad bean inversions (Palmer et al. 1984).

A striking feature of the subclover genome is the presence of dispersed repetitive sequence elements and the transposition events such repeats imply. In contrast, there is no evidence for the existence of any repeats of this size in either pea (Palmer and Thompson 1981a, 1982) or broad bean (Ko et al. 1983) and there is no need to invoke transposition in modeling the evolution of their chloroplast genomes (Figs. 5, 6; Palmer and Thompson 1982, unpublished data). While all of the sequence rearrangements in pea and broad bean may well be inversions, it is nonetheless clear that the frequency of these events has been significantly higher in the pea lineage than in the broad bean lineage (Figs. 5, 6; Palmer and Thompson 1982), and moreover, that inversion is still occurring even among very closely related species and populations of peas (Palmer et al. 1985a). In summary, a wide variety of evolutionary pathways have been taken by those legume chloroplast genomes that lack the inverted repeat. Some, such as those of alfalfa and wisteria, have otherwise retained the primitive and highly conserved arrangement of chloroplast sequences which characterizes all legume genomes that retain the inverted repeat. In contrast, three distinct, both qualitatively and quantitatively, and mutually independent patterns of sequence rearrangement are found in subclover, pea and broad bean. Whether or not the degree of instability evidenced by these three chloroplast lineages is in any way causally related to the absence of the inverted repeat structure, as suggested in our earlier paper (Palmer and Thompson 1982), remains an open question.

Subclover is unique among characterized land plants in possessing a family of rather large (several hundred bp) dispersed repeats in its chloroplast genome. These repeats were readily detected using two cloned subclover fragments as hybridization probes, whereas similar hybridization experiments using cloned probes failed to detect repeated sequences of this size in chloroplast DNAs from pea and mung bean (Palmer and Thompson 1981a, 1982), broad bean (Ko et al. 1983), soybean (Spielmann et al. 1983), alfalfa (this report), *Spirodela* (de Heij et al. 1983), flax (Coates and Cullis 1982), tobacco (Fluhr et al. 1983), pearl millet (Thomas et al. 1984) rice (Hirai et al. 1985), *Osmunda* (Stein et al. 1986), and petunia, spinach, lettuce and turnip (J. Palmer, R. Jansen, and J. Nugent, unpublished data). Interestingly, however, substantially shorter repeats, a duplication of 119 bp and a triplication of 70 bp, have recently been found in wheat chloroplast DNA and shown to be associated with the endpoints of evolutionary inversions (Quigley and Weil 1985; Howe 1985). In addition, wheat and other cereals also contain a short duplication of sequences adjacent to *rbc*L and at the end of the inverted repeat, although this repeat has not been implicated in any evolutionary rearrangements (Day and Ellis 1984; Dang and Pring 1986). In contrast to their relative absence from land plant genomes, dispersed repeats are a prominent feature of chloroplast DNA of the green alga *Chlamydomonas reinhardtii*, where 25–40 copies of a short inverted repeat sequence 100–300 bp in length are found scattered throughout the genome (Rochaix 1978; Rochaix and Malnoe 1978; Gelvin and Howell 1979; Palmer et al. 1985b).

Since alfalfa, a member of the same tribe as subclover, lacks any detectable repeats, it is likely that these repeats have originated fairly recently in the subclover genome. In this regard, it should be interesting to examine other taxa in the genus *Trifolium* to see whether they also possess dispersed repeats, and if so, whether the repeats are present at the same copy number and positions as those in *T. subterraneum*. Such evolutionary comparisons, as well as direct sequence analysis of the repeats in *T. subterraneum*, should indicate whether these elements have the general behavioral and structural properties of transposable elements (Calos and Miller 1980). In addition, such studies might be expected to reveal whether the repeats, even if they themselves are not in any way actively transposable, generate evolutionary rearrangement (such as inversions) by serving as sites for homologous intragenomic recombination.

Acknowledgements. We thank H. Bohnert for the gift of the *psb*A 3′ probe, L. Bogorad for the gift of the *rbc*L 5′ probe, D. Bourque for the gift of the tobacco 16S and 23S rRNAs, N. W. Gillham, M. Zolan and J. Watson for critical reading of the manuscript, and B. Milligan for the gift of the subclover chloroplast DNA used in the experiment shown in Fig. 9. This research was supported in part by grants to J. D. P. from the United States Department of Agriculture (85-CRCR-1-1778) and the University of Michigan Rackham Graduate School, and to W. F. T. from the National Science Foundation (PCM-81-09795).

References

Aldrich J, Cherny B, Merlin E, Williams C, Mets L (1985) Curr Genet 9:233–238
Birnboim HC, Doly J (1979) Nucleic Acids Res 7:1513–1523
Bohnert HJ, Loffelhardt W (1982) FEBS Lett 150:403–406
Brears T, Schardl CL, Lonsdale DM (1986) Plant Mol Biol 6: 171–177
Calos MP, Miller JH (1980) Cell 20:579–595
Chu NM, Tewari KK (1982) Mol Gen Genet 186:23–32
Coates D, Cullis CA (1982) Plant Mol Biol 1:183–189
Dagert M, Ehrlich SD (1979) Gene 6:23–38
Dang LH, Pring DR (1986) Plant Mol Biol 6:119–123
Day A, Ellis THN (1984) Cell 39:359–368
de Heij HT, Lustig H, Moeskops DJM, Bovenberg WA, Bisanz C, Groot GSP (1983) Curr Genet 7:1–6
Fluhr R, Edelman M (1981) Nucleic Acids Res 9:6841–6853
Fluhr R, Fromm H, Edelman M (1983) Gene 25:271–280
Gelvin SB, Howell SH (1979) Mol Gen Genet 173:315–322
Gillham NW, Boynton JE, Harris EH (1985) In: Cavalier-Smith T (ed) DNA evolution: natural selection and genome size. Wiley, New York, pp 299–351
Hirai A, Ishibashi T, Morikami A, Iwatsuki N, Shinozaki K, Sugiura M (1985) Theor Appl Genet 70:117–122
Howe CJ (1985) Curr Genet 10:139–145
Ko K, Strauss NA, Williams JP (1983) Curr Genet 7:255–263
Ko K, Strauss NA, Williams JP (1984) Curr Genet 8:359–367
Koller B, Delius H (1980) Mol Gen Genet 178:261–269
Kolodner R, Tewari KK (1975) Biochim Biophys Acta 402: 372–390
Maizels N (1976) Cell 9:431–438
McIntosh L, Poulsen C, Bogorad L (1980) Nature (London) 288:556–560
Mubumbila M, Gordon KHJ, Crouse EJ, Burkard G, Weil JH (1983) Gene 21:257–266
Mubumbila M, Crouse EJ, Weil JH (1984) Curr Genet 8:379–385
Ohyama K, Yamano Y, Fukuzawa H, Komano T, Yamagishi H, Fujimoto S, Sugiura M (1983) Mol Gen Genet 189:1–9
Palmer JD (1982) Nucleic Acids Res 10:1593–1605
Palmer JD (1983) Nature (London) 301:92–93
Palmer JD (1985a) Annu Rev Genet 19:325–354
Palmer JD (1985b) In: MacIntyre RJ (ed) Monographs in evolutionary biology: molecular evolutionary genetics. Plenum, New York, pp 131–240
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 29:537–550
Palmer JD, Edwards H, Jorgensen RA, Thompson WF (1982) Nucleic Acids Res 10:6819–6832
Palmer JD, Singh GP, Pillay DTN (1983a) Mol Gen Genet 190: 13–19
Palmer JD, Shields CR, Cohen DB, Orton TJ (1983b) Theor Appl Genet 65:181–189
Palmer JD, Osorio B, Watson JC, Edwards H, Dodd J, Thompson WF (1984) In: Thornber JP, Staehelin LA, Hallick RB (eds) Biosynthesis of the photosynthetic apparatus: molecular biology, development and regulation. Liss, New York, pp 273–283 (UCLA Symposia on Molecular and Cellular Biology, new series, vol 14)
Palmer JD, Jorgensen RA, Thompson WF (1985a) Genetics 109:195–213
Palmer JD, Boynton JE, Gillham NW, Harris EH (1985b) In: Steinback KE, Bonitz S, Arntzen CJ, Bogorad L (eds) Molecular biology of the photosynthetic apparatus. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, pp 269–278
Polhill RM, Raven PH (1981) Advances in legume systematics, part 1. Royal Botanic Gardens, Kew
Quigley F, Weil JH (1981) Curr Genet 9:495–503
Rochaix JD (1978) J Mol Biol 126:597–617
Rochaix JD, Malnoe P (1978) Cell 15:661–670
Shinozaki K, Sun CR, Sugiura M (1984) Mol Gen Genet 197: 363–367
Spielmann A, Ortiz W, Stutz E (1983) Mol Gen Genet 190:5–12
Stein DB, Palmer JD, Thompson WF (1986) Curr Genet 10: 835–841
Takaiwa F, Sugiura M (1982) Eur J Biochem 124:13–19
Thomas KM, Wood BJ, Bassett CL, Rawson JRY (1984) Curr Genet 8:291–297
Tohdoh N, Sugiura M (1982) Gene 17:213–218
Vieira J, Messing J (1982) Gene 19:259–268
Westhoff P, Nelson N, Bunemann H, Herrmann RG (1981) Curr Genet 4:109–120
Whitfeld PR, Bottomley W (1983) Annu Rev Plant Physiol 34: 279–310
Zurawski G, Perrot B, Bottomley W, Whitfeld PR (1981) Nucleic Acids Res 9:3251–3270
Zurawski G, Bohnert HJ, Whitfeld PR, Bottomley W (1982a) Proc Natl Acad Sci USA 79:7699–7703
Zurawski G, Bottomley W, Whitfeld PR (1982b) Proc Natl Acad Sci USA 79:6260–6264

Communicated by C. W. Birky, Jr.

Received March 12 / July 29, 1986