

Evolutionary significance of inversions in legume chloroplast DNAs

Jeffrey D. Palmer¹, Bernardita Osorio², and William F. Thompson³

¹ Department of Biology, University of Michigan, Ann Arbor, MI 48109, USA

² 271 Hill Crest Road, Stafford, PA 19087, USA

³ Department of Botany, North Carolina State University, Raleigh, NC 27695, USA

Summary. Cloned genes from tobacco, spinach, and pea were used as hybridization probes to localize 36 protein genes on the chloroplast chromosomes of four legumes – mung bean, common bean, soybean, and pea. The first three chloroplast DNAs (cpDNAs), all of which retain a large inverted repeat, have an identical gene order with but one exception. A 78 kb segment encompassing nearly the entire large single copy region is inverted in mung bean and common bean relative to soybean and non-legumes. The simplest evolutionary explanation for this difference is a 78 kb inversion, with one endpoint between *rps8* and *infA* and the second between *psbA* and *rpl2*. However, we can not rule out a two-step rearrangement (consisting of successive expansion and contraction of the inverted repeat) leading to the relocation of a block of six ribosomal protein genes (*rps19-rps8*) from one end of the large single copy region to the other. Analysis of gene locations in pea cpDNA, which lacks the large inverted repeat, combined with cross-hybridization studies using 59 clones covering the mung bean genome, leads to a refined picture of the position and nature of the numerous rearrangements previously described in the pea genome. A minimum of eight large inversions are postulated to account for these rearrangements. None of these inversions disrupt groups of genes that are transcriptionally linked in angiosperm cpDNA. Rather, the endpoints of inversions are associated with relatively spacer-rich segments of the genome, many of which contain tRNA genes. All of the pea-specific inversions are shown to be positionally distinct from those recently described in a closely related legume, broad bean.

Key words: Gene mapping – Rearrangements – Chloroplast DNA evolution – Inverted repeat

Introduction

Legumes represent an excellent group in which to study contrasting patterns of cpDNA evolution. Those legume cpDNAs that retain an ancestral large inverted repeat also share a highly conserved sequence order that differs by only a single inversion from the arrangement typical of most vascular plants (Palmer and Thompson 1982; Palmer et al. 1983, 1987a; Palmer and Stein 1986; reviewed in Palmer 1985). In contrast, another group of legumes, whose genomes lack the inverted repeat, undergo more frequent rearrangement. One of the most divergent of these rapidly rearranging legume genomes is that of Mendel's garden pea, *Pisum sativum* (Palmer and Thompson 1981a, 1982). A lesser degree of rearrangement has occurred in the related species broad bean (*Vicia faba*); this has facilitated detailed mapping of rearrangement endpoints and resulted in three recent, independent models for the evolutionary derivation of its genome structure (Ko et al. 1987; Michalowski et al. 1987; Palmer et al. 1987a).

We have mapped 36 protein genes in four legume cpDNAs and also performed extensive cross-hybridization experiments between two of the four in order to address two related questions of chloroplast genome evolution. First, where are the precise boundaries of the many rearrangements in pea cpDNA? Such information should allow us to infer the nature and number of these events. In addition, since a detailed transcriptional map is available for pea cpDNA (Woodbury et al. 1988), we should be able to examine the relationship of rearrangements to the transcriptional organization of the chloroplast genome. Second, how stable in structure are three legume cpDNAs that retain the large inverted repeat? We show that two of the three genomes studied are invariant in order for the 36 genes mapped and that these two differ from the third by a single large inversion.

Materials and methods

Chloroplast DNAs were purified from mung bean (*Vigna radiata* cv. Berken) and common bean (*Phaseolus vulgaris* cv. Hawkesbury Wonder) by the sucrose gradient method (Palmer 1986). cpDNAs were purified from soybean (*Glycine max* cv. Harcor) and pea (*Pisum sativum* cv. Alaska) by DNAase I-treatment of isolated chloroplasts (Kolodner and Tewari 1975). Plasmid DNAs were purified from *E. coli* by the alkaline miniprep procedure of Birnboim and Doly (1979).

Restriction endonuclease digestions, agarose gel electrophoresis, nick-translations, and filter hybridizations were performed as described (Palmer 1986). Southern blotting was by a combination of the bidirectional blotting procedure of Smith and Summers (1980) and the alkaline transfer procedure of Reed and Mann (1985). Zetabind nylon filters were purchased from AMF Cuno, Inc. All filters were washed in 2 × SSC (0.3 M NaCl, 30 mM trisodium citrate) and 0.5% SDS at 65° prior to autoradiography. Prior to reuse, filters were stripped of hybridized probe with 0.4 N NaOH at 42° for 1 h and then washed in 0.1 × SSC, 0.5% SDS and 0.2 M Tris (pH 7.5) for 30 min.

The 39 gene probes listed in Table 1 consist of plasmid clones containing the indicated fragments as single inserts. These were subcloned from clones containing larger fragments of cpDNAs of spinach and pea (Palmer and Thompson 1981b) and tobacco (Sugiura et al. 1986) as described in Jansen and Palmer (1987). A bank of plasmid clones containing 59 cpDNA fragments (0.3–4.8 kb in size) that cover the mung bean genome were constructed from a set of 15 larger clones described previously (Palmer and Thompson 1981b). The positions of these fragments in the mung bean genome and the identities of the nine enzymes used for subcloning are shown in Fig. 5. The indicated fragments were isolated from parent clones by gel electrophoresis and electroelution, and were then cloned into either pUC8 or pUC12.

Results and discussion

Gene maps of four legume chloroplast DNAs

A total of 39 plasmids, containing portions of 36 proteins genes from tobacco, spinach, and pea (Table 1), were nick-translated and hybridized to Zetabind filters containing digests of cpDNAs from mung-bean, common bean, soybean, and pea produced by the restriction enzymes whose maps are shown in Fig. 1. All of the gene probes hybridized strongly to each of the four cpDNAs under standard hybridization conditions (65 °C, 2 × SSC wash). This result suggests that each of the 36 genes is likely to be present in these legume cpDNAs, although ultimate proof requires complete DNA sequence determination. Positive hybridization with gene probes located either strictly (e.g. *rbcL* middle) or nearly (e.g. *rpoA*) internal to a single gene is evidence for the presence of at least a portion of the gene in question in the legume cpDNAs. Assignment is less certain for those genes mapped with probes containing large blocks of adjacent noncoding sequences (e.g. *infA*) or parts of two or more genes (e.g. *psbE-psbF*).

Our hybridization data have been in large part validated by the extensive independent hybridization mapping and sequence analysis performed in the case of pea cpDNA. No fewer than 17 of the 36 genes mapped in Fig. 1 have been shown to be fully present in pea cpDNA by DNA sequencing (for genes and references, see legend to Fig. 5). More critically, all regions to which we have assigned a particular gene have, when sequenced, been shown to contain that gene. Also, two other groups have used hybridization mapping to reach conclusions identical to ours regarding the positions of several of the unsequenced pea genes (Courtice et al. 1985; Berends et al. 1986). The only one of the 36 genes whose complete presence in pea cpDNA we have reason to doubt is *rpl22* (indicated in Fig. 1 with a filled circle). In recent experiments, we find that a 209 bp tobacco fragment containing the 5'-178 bp of *rpl22* coding sequences fails completely to hybridize to all legume cpDNAs tested under conditions in which it hybridizes strongly to 13 nonleguminous angiosperm cpDNAs (J. Palmer, unpublished data).

The probable direction of transcription in the four legume genomes is shown for all 36 genes in Fig. 1. For a number of genes, the differential hybridization of 5' and 3' probes (Table 1) allowed direct assignment of gene orientation. Indirect assignments for the remaining genes were based on their known orientations in tobacco, spinach, and pea (from which the gene probes were constructed), their tight clustering in these species, often as part of polycistronic transcription units (see references in Shinozaki et al. 1986, Westhoff et al. 1988, and the legend to Fig. 5), and their highly similar clustering in these legumes.

A large inversion among legume cpDNAs containing the inverted repeat

The inverted repeat-containing cpDNAs of mung bean and common bean are identical in gene order (Fig. 1). However, they differ from soybean by the inversion of a 78 kb segment constituting nearly the entire large single copy region (Fig. 2). The left endpoint of this inversion maps roughly to the spacer region between *rps8* and *infA*, a segment that is 135 bp in length in tobacco (Shinozaki et al. 1986). The right endpoint maps to the spacer between *psbA* and *rpl2*, a region that is 600 bp long in tobacco (Shinozaki et al. 1986). Within this 600 bp region lies a gene for tRNA^{His}, which is also contained (Shapiro and Tewari 1986) on the 3' *psbA* probe (Table 1) that hybridized solely to the *rps8* end of the large single copy region in mung bean and common bean. Hence, this inversion endpoint may well be located within the spacer that normally separates *trnH* and *rpl2*.

Table 1. Sources of gene probes

Gene name	Species	Fragment ^a	Gene location ^b	Reference
1. 3' <i>rps12</i>	Tobacco	969 bp PstI-BamHI	-119/+19	c
2. <i>rps7</i>	Tobacco	1,833 bp PstI-SalI	+319/++1,379	c
3. <i>rpl23-5'rpl2</i>	Tobacco	1,250 bp BamHI-PstI	-462/+486	d
4. 3' <i>rpl2-rps19-5'rpl22</i>	Tobacco	1,573 bp PstI	+486/+178	d
5. 3' <i>rpl22-5'rps3</i>	Tobacco	536 bp PstI-SalI	+178/+262	d
6. 3' <i>rps3-5'rpl16</i>	Tobacco	1,825 bp SalI-BamHI	+262/+297	d
7. 3' <i>rpl16-rpl14-5'rps8</i>	Tobacco	1,055 bp BamHI-SalI	+297/+250	d
8. <i>infA</i>	Spinach	670 bp SalI	-279/++160	d, e
9. <i>rps11</i>	Spinach	635 bp SalI-XbaI	-222/+413	e
10. <i>rpoA</i>	Spinach	1,040 bp XbaI	-78/+962	e
11. 3' <i>petD</i>	Spinach	416 bp BamHI-XbaI	+134/++132	f
12. 5' <i>petD</i>	Spinach	296 bp BamHI	-162/+134	f
13. <i>psbH-petB</i>	Spinach	2,806 bp SalI-BamHI	-238/++846	f, g
14. 3' <i>psbB</i>	Spinach	1,597 bp BamHI-SalI	+258/++327	h
15. 5' <i>psbB</i>	Spinach	338 bp BamHI	-80/+258	h
16. 5' <i>psbE</i>	Spinach	0.65 kb EcoRI	-465/+185	i
17. 3' <i>psbE-psbF</i>	Spinach	0.50 kb EcoRI	+185/++433	i
18. 3' <i>petA</i>	Pea	1.1 kb BamHI	+569/++500	j
19. 5' <i>petA</i>	Pea	0.9 kb HindIII-BamHI	-300/+569	j
20. 3' <i>rbcL</i>	Pea	0.8 kb HindIII-BamHI	+1,340/++700	k
21. <i>rbcL</i> middle	Pea	1,167 bp PstI-HindIII	+173/+1,340	k
22. 5' <i>rbcL</i>	Pea	685 bp XbaI-PstI	-512/+173	k
23. <i>atpB</i>	Pea	1,171 bp PstI-XbaI	-182/+989	k
24. <i>atpE</i>	Spinach	420 bp EcoRI-XbaI	+41/++79	l
25. <i>psaA-5'psaB</i>	Spinach	2,505 bp BamHI	+50/++277	m
26. 3' <i>psaB</i>	Spinach	1,683 bp BamHI	+598/++80	m
27. 3' <i>psbC</i>	Spinach	367 bp BamHI-PstI	+983/+1,350	n
28. 3' <i>psbD-5'psbC</i>	Pea	1,150 bp PstI	+227/+369	o
29. 5' <i>psbD</i>	Pea	707 bp BamHI-PstI	-480/+227	o
30. <i>rpoB</i> middle	Tobacco	1,063 bp BamHI	+1,306/+2,369	p
31. 3' <i>rpoB-5'rpoC1</i>	Tobacco	3,182 bp BamHI	+2,369/+1,180	q, r
32. 3' <i>rpoC1</i>	Tobacco	459 bp BamHI-KpnI	+1,180/+1,639	q, r
33. <i>rpoC2</i>	Pea	3,292 bp KpnI	+868/+4,160	r, s
34. <i>rps2-atpI-5'atpH</i>	Pea	2,887 bp KpnI-PstI	-405/+92	t
35. 3' <i>atpH-5'atpF</i>	Pea	767 bp PstI-BamHI	+92/+111	u
36. 3' <i>atpF-5'atpA</i>	Spinach	1,452 bp SalI-HindIII	+144/+715	u
37. 3' <i>atpA</i>	Spinach	1,024 bp HindIII-SalI	+715/++309	u
38. 5' <i>psbA</i>	Pea	532 bp EcoRI-PstI	-58/+464	v
39. 3' <i>psbA</i>	Pea	1.2 kb PstI-EcoRI	+464/++700	v

^a Sizes in bp are based on complete sequence data; sizes in kb are based on electrophoretic mobility of incompletely sequenced fragments

^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	Fromm et al. (1986)	j	Willey et al. (1984)	q	Shinozaki et al. (1986)
d	Tanaka et al. (1986)	k	Zurawski et al. (1986a, b)	r	Ohyama et al. (1986)
e	Sijben-Muller et al. (1986)	l	Zurawski et al. (1982)	s	Cozens and Walker (1986)
f	Heinemeyer et al. (1984)	m	Kirsch et al. (1986)	t	Cozens et al. (1986)
g	Westhoff et al. (1986)	n	Alt et al. (1984)	u	Hudson et al. (1987)
h	Morris and Herrmann (1984)	o	Rasmussen et al. (1984a)	v	Oishi et al. (1984)
i	Herrmann et al. (1984)	p	Ohme et al. (1986)		

The 78 kb inversion described here was not detected in an earlier study (Palmer et al. 1983) for the following reason. The inversion endpoints map at or near the large single copy boundary of two mung bean fragments (Pst3, of 17.2 kb, and Pst4, of 16.2 kb; cf. Figs. 1 and 2) that

are largely identical, containing over half of the 25 kb inverted repeat (Fig. 1). In the 1983 study, these two mung bean fragments gave essentially identical patterns of hybridization to soybean cpDNA (which was lacking in restriction sites in the hybridizing regions), making it

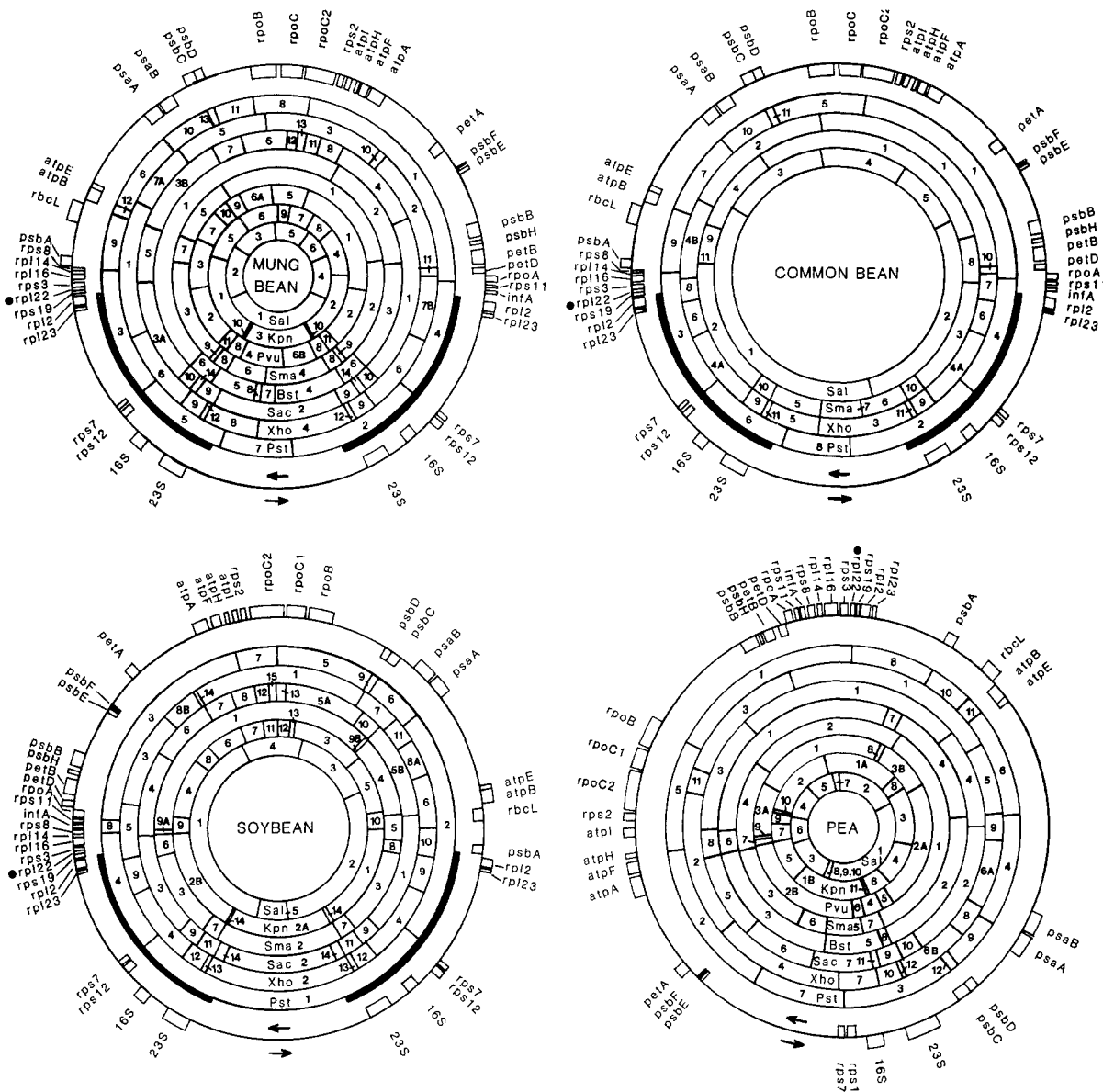


Fig. 1. Physical and gene maps of four legume cpDNAs. Restriction fragments are numbered according to size, starting with the largest fragment. The mung bean and pea restriction maps are from Palmer and Thompson (1981a), the common bean map is from Mubumbila et al. (1983) and Palmer et al. (1983), and the soybean map is from Palmer et al. (1983) and Spielmann et al. (1983). *The two long, heavy black arcs* represent the large inverted repeats present in three of the four genomes. The single copy regions in these three genomes are shown in only one of two possible relative orientations (reviewed in Palmer 1985). *Arrows* at bottom of maps indicate direction of transcription of genes shown on the inside (transcribed clock-wise) and outside (transcribed counterclock-wise) of the outermost circle. Genes that are at least partially deleted are indicated with *filled circles*. Most of the gene assignments presented here are new, except for a few genes previously mapped in mung bean (see Palmer et al. 1987a and references therein) and soybean (see Singh et al. 1985, von Allmen and Stutz 1987 and references therein), and a large number of genes previously mapped in pea (see legend to Fig. 5)

impossible to detect such an inversion having both endpoints near the ends of the large single copy region. The use of much smaller gene probes has now permitted detection of the 78 kb inversion and leads us to suggest that the "two regions of deletions/additions ... at the ends of the large single copy DNA region" (Palmer et al. 1983) are probably artefacts of the previously improper alignment of the mung bean and soybean genomes.

All legumes share a derived 50 kb inversion within their large single copy region relative to all nonlegumes (Figs. 2 and 3; Palmer and Thompson 1982; Palmer et al. 1982; Spielmann et al. 1983; Palmer et al. 1987a). Aside from this one difference, gene order in soybean is identical to that of tobacco (Fig. 2) and most other vascular plants (Palmer and Stein 1986). Therefore the 78 kb inversion must have occurred in the common ancestor (Fig. 3) of

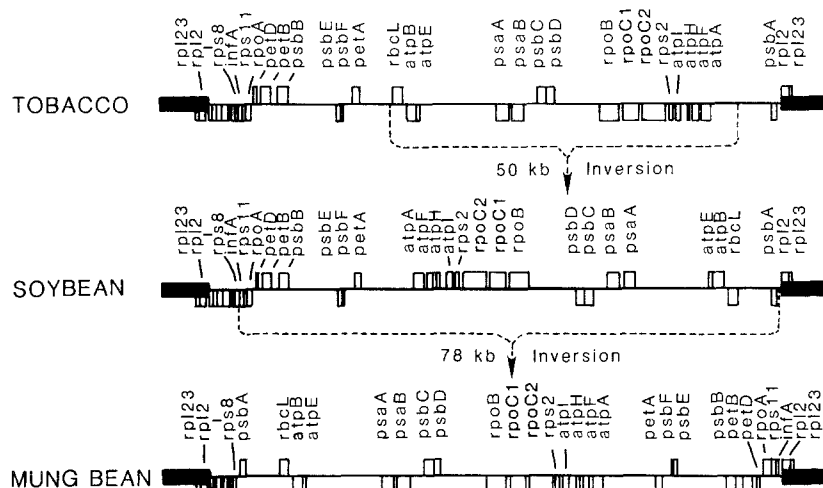


Fig. 2. A model for the derivation of the soybean and mung bean gene orders in the large single copy region by successive inversions in a tobacco-like genome. Soybean and mung bean maps are from Fig. 1 and tobacco map is from Shinozaki et al. (1986)

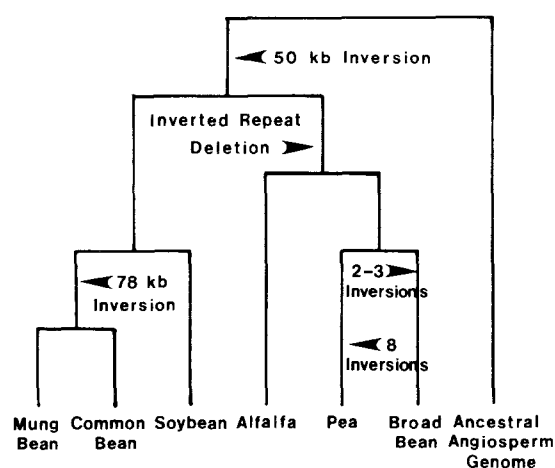


Fig. 3. Phylogeny of legumes based on cpDNA rearrangements. References for each set of rearrangements: 50 kb inversion (Palmer and Thompson 1982; Palmer et al. 1982); inverted repeat deletion (Koller and Delius, 1980; Palmer and Thompson 1981a, 1982; Palmer et al. 1987a); 2-3 inversions in broad bean (Ko et al. 1987; Michalowski et al. 1987; Palmer et al. 1987); 8 inversions in pea (this paper); 78 kb inversion (this paper)

mung bean and common bean (which both belong to the subtribe Phaseolinae of the tribe Phaseoleae), after the divergence of this lineage from that leading to soybean (which belongs to subtribe Glycinae of Phaseoleae).

While a single inversion is the simplest hypothesis to explain the difference in gene order among these three legumes, an alternative is suggested by the proximity of the rearrangement breakpoints to the boundaries of the inverted repeat and large single copy regions. Figure 4 diagrams a two-step process, of expansion of the inverted repeat and duplication of the approximately 4 kb *rps19-rps8* gene cluster, followed by contraction to the same starting point but with deletion of the original copy of the gene cluster. Since the two segments of the inverted repeat are completely identical (Ohyama et al. 1986;

Shinozaki et al. 1986) and since frequent "flip-flop" recombination between repeat segments randomizes polarity between large and small single copy regions (Palmer 1983; Palmer et al. 1984), this expansion-contraction cycle (Fig. 4) leads to the same relative arrangement of mung bean and soybean as the inversion shown in Fig. 2.

Although we favor the inversion model over the expansion-contraction one for reasons of both parsimony and simplicity, the latter model is by no means unreasonable. Precedents for evolutionary expansions and contractions of the inverted repeat are known in *Nicotiana acuminata* (3.4 kb expansion; Shen et al. 1982), geranium (50 kb expansion; Palmer et al. 1987b), and coriander (15 kb contraction Palmer 1985). Moreover, many laboratory mutants of *Chlamydomonas reinhardtii* have inverted repeats of varying degrees of expansion and contraction (Palmer et al. 1985a). The likelihood of an expansion-contraction cycle restoring the original (within the limits of resolution of Southern hybridization) boundary between inverted repeat and large single copy region might seem small. However, the fact that this boundary is rather precisely conserved in a wide range of dicots and monocots (Spielmann and Stutz 1983; Zurawski et al. 1984; Posno et al. 1985; Sugita et al. 1984) suggest that there may be factors tending to constrain and, indeed, even restore this junction.

Nature and timing of rearrangements in pea cpDNA

The existence of multiple rearrangements in pea cpDNA was inferred several years ago based on the complex hybridization patterns obtained using cloned fragments of mung bean cpDNA as probes (Palmer and Thompson 1981a, 1982). Figure 5 summarizes our current understanding of the relative organization of the pea and mung bean chloroplast genomes as derived from three sources

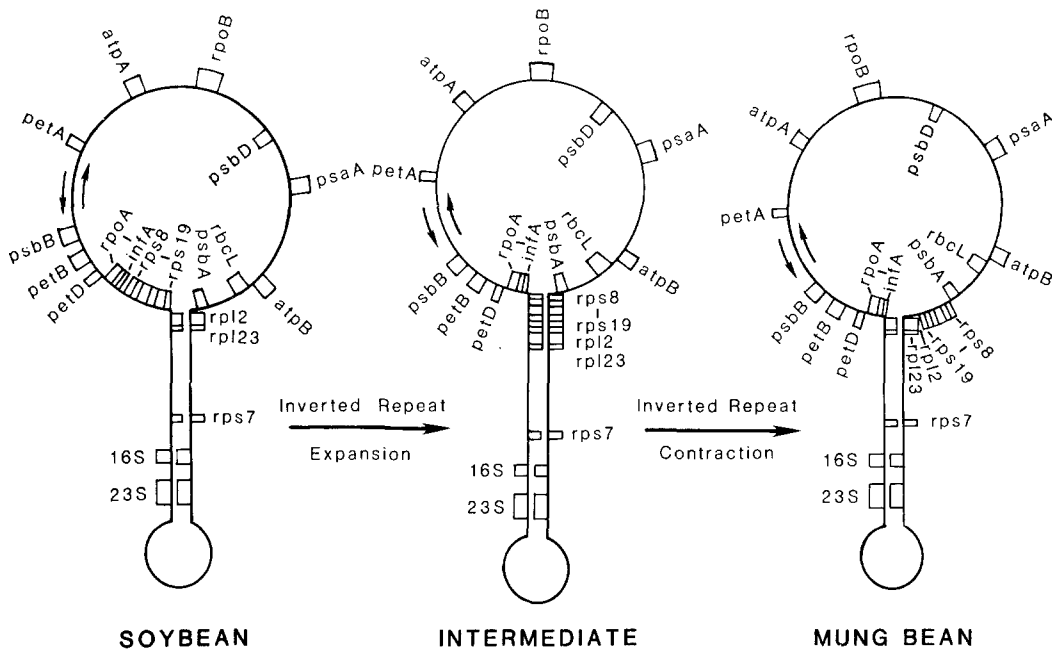


Fig. 4. An alternative model compared to that in Fig. 2 for the derivation of the mung bean gene order in the large single copy region by successive expansion and contraction of the inverted repeat of a soybean-like genome. Simplified gene maps from Fig. 1 are shown

of information: 1) The hybridization experiments that map the 36 genes shown in Fig. 1; 2) The DNA sequence studies that have led to the precise positioning in pea of 20 protein genes, 19 tRNA genes, and four rRNA genes (for references, see legend to Fig. 5); 3) A high resolution series of hybridizations between mung bean and pea cpDNAs. In our previous studies, a small number (14) of relatively large (average size 10.6 kb) mung bean fragments were used to map rearrangements in pea cpDNA (Palmer and Thompson 1981a, 1982). In the present study, we have subdivided these 14 original clones into 59 subclones averaging 2.1 kb in size, each of which was hybridized to Zetabind filters containing digests of pea cpDNA with KpnI, PstI, SacI, SalI, SmaI, and XhoI. Certain mung bean clones were also hybridized to digests of pea cpDNA with BamHI, BglII, EcoRI, and HindIII to provide resolution in regions deficient in sites for the first six enzymes.

The mung bean genome in Fig. 5 is represented in a soybean-like form — without the 78 kb inversion; cf. Figs. 2 and 5 — to focus attention on the rearrangements that occurred in a pea-specific lineage. The pea rearrangements shown in Fig. 5 approximate in number and location those described in our earlier, low-resolution studies (Palmer and Thompson 1981a, 1982), but are here positioned more precisely with respect to their endpoints and gene content. This finer level of analysis reveals that, although rearrangement is indeed quite extensive in the pea genome, the average block of rearranged sequences is large (only blocks 2, 5, and 9B are less than 5 kb) and generally contains multiple genes.

At our current level of resolution, the pea and mung bean genomes contain largely, if not entirely, the same sequence content. Except for a 2.8 kb XbaI fragment from the small single copy region, every mung bean subclone hybridized strongly to pea cpDNA (Fig. 5). Interestingly, the one region in pea cpDNA that is not hybridized to by any mung bean clones has the same flanking sequences (blocks 4A and 4B) as the mung bean fragment that failed to hybridize to pea cpDNA (Fig. 5). It is possible that this is a truly homologous region, but one in which sequence change has been so rapid that pea and mung bean sequences fail to form stable hybrids under our experimental conditions. Alternatively, the pea- and mung bean-specific sequences could simply be nonhomologous, having different histories of insertion and deletion in the two genomes.

Knowledge of the precise number and positions of rearrangement breakpoints in pea cpDNA (Fig. 5) allows us to put forward a model for the evolution of the pea genome from an unrearranged legume genome (Fig. 6). The nature of the first and last of these alterations is unambiguous. As described in detail in an earlier paper (Palmer et al. 1987a), the deletion of one segment of the large inverted repeat preceded all other known rearrangements in such rearranged genomes as broad bean, subclover, and also pea. Figure 5 presents evidence in support of this conclusion, i.e. pea cpDNA has lost the same one of the four possible inverted repeat configurations as shown previously (Palmer et al. 1987a) for these and other legumes. The most recent of pea rearrangements is a small inversion (of region 9B in Figs. 5 and 6) occur-

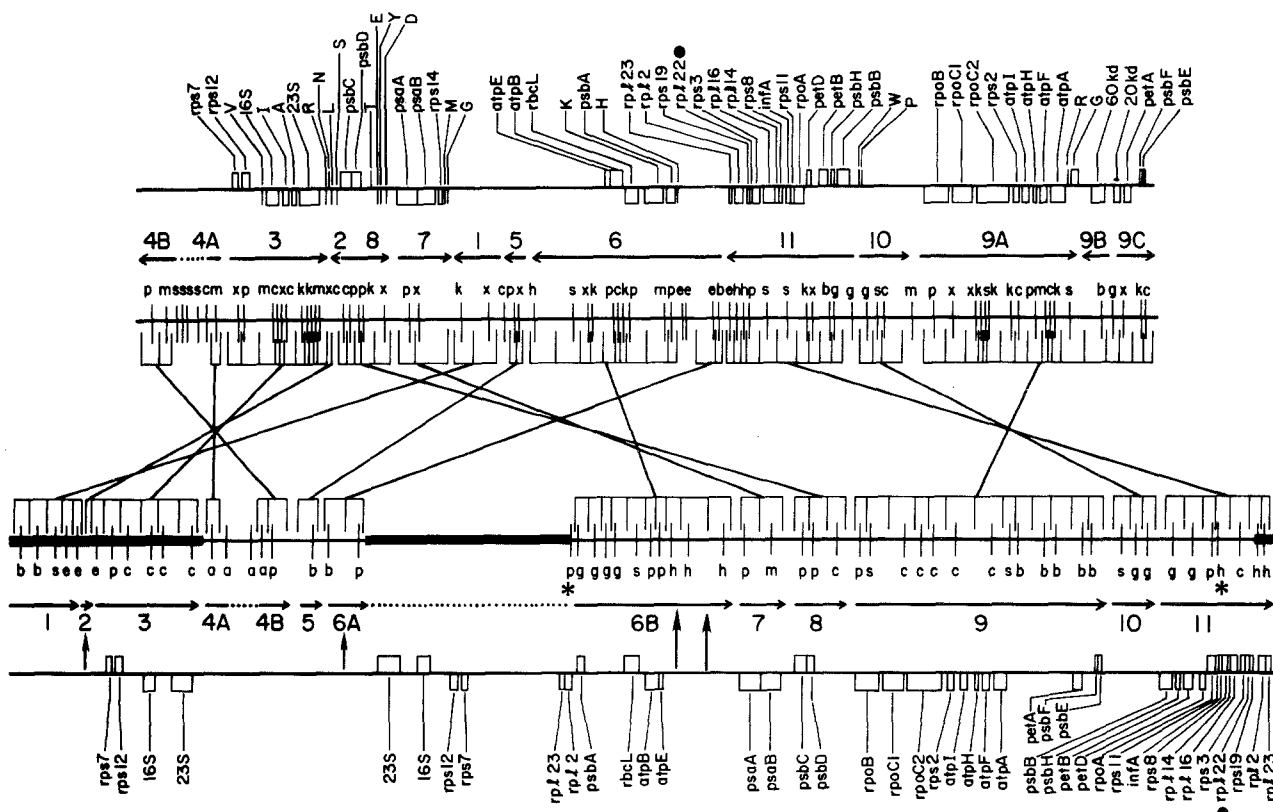


Fig. 5. Arrangement of homologous sequences in the mung bean (*bottom maps*) and pea (*top maps*) chloroplast genomes. A total of 59 cloned restriction fragments covering the mung bean genome were hybridized to Southern blots containing KpnI (k), PstI (p), SacI (c), SalI (s), SmaI (m), and XhoI (x) digests of pea cpDNA. In addition, to achieve adequate resolution in regions of the genome depauperate in sites for these six enzymes, selected mung bean clones were hybridized to digests of pea cpDNA with BamHI (b), BglII (g), EcoRI (e), and HindIII (h). Sites in the mung bean map are indicated with the same letters as in pea, with XbaI sites indicated with “a”. *Numbered horizontal arrows* indicate the position and relative orientation of blocks of sequences that cross-hybridize and whose arrangement has been conserved between the two genomes. All *arrows* are drawn to scale, except those representing block 2, which consists of but a single tRNA gene (for Leu-CAA). The inversion in the pea genome of block 9B relative to 9A and 9C represents a previously described inversion in *Pisum humile*, a wild ancestor of the garden pea, *P. sativum* (Palmer et al. 1985b). The *crossing lines* connect homologous sequence blocks. *Dotted lines* within the horizontal arrows indicate sequences present in only one of the two genomes. The *vertical arrows* in the mung bean map indicate the positions of the four endpoints of two inversions mapped in broad bean cpDNA (Palmer et al. 1987a). Note that the mung bean genome is shown in a soybean-like arrangement (the asterisks indicate the region that has been inverted; cf. Fig. 2) in order to represent the ancestral legume gene order and facilitate presentation of the *pea-specific* rearrangements. Genes shown above the top and bottom lines are transcribed from right-to-left, while those shown below the lines are transcribed from left-to-right. Mung bean gene positions are from Fig. 1, while pea data are from Fig. 1 and the following sources of sequenced pea genes (genes are given in order from left-to-right in the genome; tRNA genes are indicated with the single letter amino acid code): *trnV-5S* rRNA (Stumman et al. 1987); *trnR-trnL* (Shapiro and Tewari 1986); *trnS* and *psbC* (Brookjans et al. 1986); *psbD* (Rasmussen et al. 1987); *trnE-trnD* (Rasmussen et al. 1984b); *psaA* and *psaB* (Lehmbeck et al. 1986); *rps14-trnG* (Lehmbeck et al. 1987); *atpE* and *atpB* (Zurawski et al. 1986a); *rbcL* (Zurawski et al. 1986b); *psbA* (Oishi et al. 1984); *trnH* (Shapiro and Tewari 1986); *rps11* (Purton and Gray 1987); *petD* (Phillips and Gray 1984); *trnW* and *trnP* (Lehmbeck et al. 1987); *rpoC2* and *rps2* (Cozens and Walker 1986); *atpI* (Cozens et al. 1986); *atpH-atpA* (Hudson et al. 1987); *trnR* and *trnG* (Shapiro and Tewari 1986); “60 kd” (Smith and Gray 1986; J. Gray, personal communication); “20 kd” and *petA* (Willey et al. 1984). In addition, *trnK* and its intron-encoded maturase-related gene (Neuhaus and Link 1987; Sugita et al. 1985) were mapped in pea using as hybridization probe a 961 bp BamHI-BglII fragment cloned from tobacco. *Filled circles* indicate genes that are partially deleted

ing in a single population of *Pisum humile*, a close relative of the domesticated pea, *P. sativum* (Palmer et al. 1985b).

Figure 6 presents one of several equally parsimonious (in number of steps) pathways for interconverting genomes containing the first (an “alfalfa” type; Palmer et al. 1987a), but lacking the last (a “*P. sativum*” type) of

these pea rearrangements. This scheme postulates seven rearrangements, all of them inversions, and is neutral with respect to the temporal order of the seven inversions shown, i.e. the same seven inversions could have been diagrammed in a different order to produce the same end-product genome. Other seven-inversion models that also account for the present day arrangement of the pea ge-

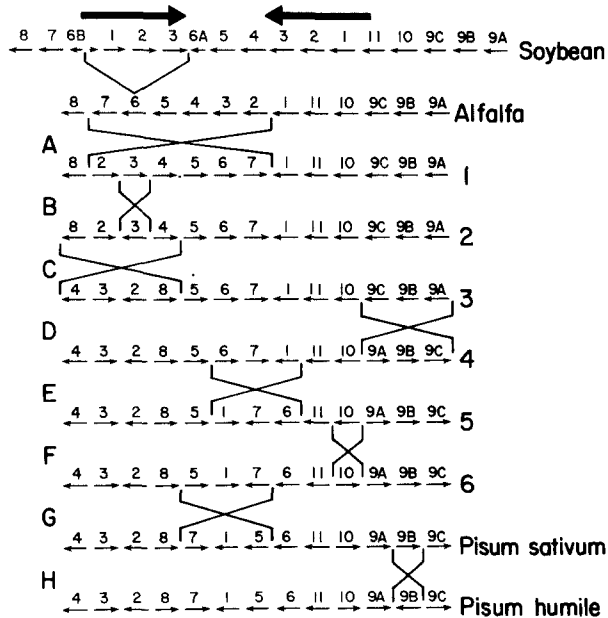


Fig. 6. Model for the evolution of the pea chloroplast genome by a step-wise series of rearrangements of a soybean-like ancestral legume cpDNA type. The *horizontal arrows* represent conserved sequence blocks; these are numbered as in Fig. 5, but are not drawn to scale

nome predict inversions between different pairs of mapped rearrangement breakpoints. This modeling exercise serves to transform a static picture of great complexity (Fig. 5) into a dynamic one of lower complexity, i.e. seven specific rearrangements (Fig. 6). Further, it suggests that a single type of rearrangement – inversion – is the major historical force responsible for the reshaping of the pea chloroplast genome. This feature unites pea cpDNA with most other angiosperm cpDNAs, where inversions are the most common internal rearrangements found, albeit usually more rarely than in pea (reviewed in Palmer 1985). In contrast, in subclover and geranium transposition may be a major force leading to cpDNA reorganization (Palmer et al. 1987a, 1987b).

Three of four inversion endpoints previously mapped in broad bean cpDNA (Palmer et al. 1987a) clearly do not align with any of the pea inversion boundaries; in fact all three are located within a single block (no. 6; Fig. 5) of sequences that are unrearranged in pea. Since only one of the four does align with a pea breakpoint, we conclude that each of the 2–3 inversions mapped in broad bean (Ko et al. 1987; Michalowski et al. 1987; Palmer et al. 1987a) probably occurred independently of the 8 inversions described in pea (Fig. 3). That pea and broad bean have independent histories of cpDNA inversions is striking in view of their close taxonomic relationship, i.e. both belong to the same tribe (Vicieae) of legumes (Polhill et al. 1981).

Mechanisms and functional significance of pea cpDNA inversions

A major constraint on rearrangements in pea appears to be the organization of the chloroplast genome into large clusters of polycistronically transcribed genes. Five of the larger rearrangement blocks (numbers 3, 8, 7, 11, and 9A) are nearly saturated with genes and in all cases appear to be occupied primarily by one or a few large clusters of genes thought to be polycistronic operons in such plants as spinach and tobacco (Fig. 5). Block 3 contains primarily the rRNA operon (Stumman et al. 1987) and the 3' *rps12-rps7* operon (Fromm et al. 1986; Koller et al. 1987; the location of the trans-spliced 5' exon of *rps12* is not known in pea). Blocks 8 and 7 have been essentially entirely sequenced by Henningsen's group (Rasmussen et al. 1984a, 1984b, 1987; Bookjans et al. 1986; Lehmbeck et al. 1986, 1987). Each is largely occupied by a single putative operon of protein genes [*psbD-psbC* in the case of block 8 (Alt et al. 1984) and *psaA-rps14* in block 7 (Kirsch et al. 1986)] and also includes a few tRNA genes whose transcriptional linkage, if any, is unclear. Block 11 is 14 kb in size and contains 15 protein genes, yet may be organized into as few as two operons *bB-petD* (Heinemeyer et al. 1984; Morris and Herrmann 1984; Westhoff et al. 1986) and *rpl23-rpoA* (H. Grune and P. Westhoff, personal communication)]. Block 9A is 19 kb in size, yet may contain but two operons of protein genes [*rpoB-rpoC2* (Hudson et al. 1988) and *rps2-atpA* (Hudson et al. 1987; Westhoff et al. 1988)].

Although the strongest evidence for these evolutionarily conserved sequence blocks being operons comes from non-legumes, the accompanying paper (Woodbury et al. 1988) presents Northern data consistent with some also being operons in pea. For example, large transcripts spanning multiple protein genes are found in pea in the case of the putative operons for 3' *rps12-rps7*, *psbD-psbC*, *psaA-rps14*, and *psbB-petD* (Woodbury et al. 1988). In addition, large, putatively polycistronic transcripts that map to these same regions are also found in chloroplasts of mung bean (Palmer et al. 1984; data not shown).

Operon-spanning transcripts are not found in pea in the regions of *rpoB-rpoC2* and *rpl23-rpoA* (Woodbury et al. 1988). Their absence may reflect a very low level of primary transcript from these putative operons, as no transcripts of any size were detected in pea for large parts of both regions (Woodbury et al. 1988). Additionally, there may be a fundamentally different organization of these regions in pea, and perhaps other legumes, compared to such plants as spinach and tobacco. One possibility is that certain genes are no longer actively transcribed, perhaps existing as pseudogenes. This appears to be the case for *rpl22*, at least part of which has been deleted from all legume cpDNAs (see section on Gene Maps), and may hold also for *rps19*, which is adjacent to *rpl22*.

Neither of these genes is represented by detectable transcripts in pea (Woodbury et al. 1988). Another possibility is that transcription units are smaller and more numerous in legume cpDNAs. Indirect evidence for this idea comes from consideration of two rearrangements found in legumes other than pea. First, the 78 kb inversion present in mung bean and common bean (Fig. 2) disrupts the physical and transcriptional (H. Grune and P. Westhoff, personal communication) linkage between the eight gene *rpl23-rps8* cluster and the three gene *infA-rpoA* cluster that is conserved not only in pea and most angiosperms but also the liverwort *Marchantia polymorpha* (Ohyama et al. 1986). Second, a rearrangement in subclover cpDNA separates *rpoB-rpoC1* from *rpoC2* (B. Milligan, J. Hampton, J. Palmer, unpublished data), all three of which are cotranscribed in spinach (Hudson et al. 1988).

Transfer RNA genes are located at or near the endpoints of many of the rearranged areas (blocks 3, 8, 7, 10, 9A) in pea cpDNA (Fig. 5). In addition, block 2 appears to consist of but a single tRNA gene (for Leu-CAA). An association between tRNA genes and rearrangement borders has also been noted for three inversions in wheat cpDNA (Howe 1985; Quigley and Weil 1985; Howe et al. 1988) and for several rearrangements in animal mtDNA (Cantatore et al. 1987; Moritz and Brown 1987; Turker et al. 1987), prompting the suggestion that similarities in tRNA gene primary and/or secondary structures might facilitate recombination and rearrangement. However, in chloroplasts the proximity of tRNA genes and inversion boundaries may be coincidental, reflecting the fact that chloroplast tRNA genes are often found as small islands within relatively large spacer regions (Shinozaki et al. 1986). Sequence analysis of the endpoints of a number of cpDNA rearrangements is needed to evaluate further the relative contributions of tRNA genes and spacer sequence elements (e.g., short dispersed repeats) in generating these alterations. To be most informative, such studies should focus on recent rearrangements and, for proper comparison, include the most closely related unrearranged genome for each event analyzed.

Acknowledgements. We thank L. Herbon and I. Apel for providing technical assistance, M. Sugiura for generously providing tobacco clones used in making gene probes, C. Howe, P. Westhoff, and P. Whitfeld for making their unpublished data available to us, G. P. Singh for providing a sample of soybean cpDNA, P. Calie for critical reading of the manuscript, and A. Srinivasen for help with the figures. This research was supported by a grant to J. D. P. from the United States Department of Agriculture (85-CRCR-1-1778) and to W. F. T. from the National Science Foundation (PCM-81-09795).

References

- Alt J, Morris J, Westhoff P, Herrmann RG (1984) *Curr Genet* 8:597–606
- Berends T, Kubicek Q, Mullet JE (1986) *Plant Mol Biol* 6:125–134
- Birnboim HC, Doly J (1979) *Nucleic Acids Res* 7:1513–1523
- Bookjans G, Stummann BM, Rasmussen OF, Henningsen (1986) *Plant Mol Biol* 6:359–366
- Cantatore P, Gadaletta MN, Roberti R, Saccone C, Wilson AC (1987) *Nature* 329:853–855
- Courtice GRM, Bowman CM, Dyer TA, Gray JC (1985) *Curr Genet* 10:329–333
- Cozens AL, Walker JE (1986) *Biochem J* 236:453–460
- Cozens AL, Walker JE, Phillips AL, Huttly AK, Gray JC (1986) *EMBO* 5:217–222
- Fromm H, Edelman M, Koller B, Goloubinoff P, Galun E (1986) *Nucleic Acids Res* 14:883–898
- Heinemeyer W, Alt J, Herrmann RB (1984) *Curr Genet* 8:543–549
- Herrmann RG, Alt J, Schiller B, Widger WR, Cramer WA (1984) *FEBS Lett* 176:239–244
- Howe CJ (1985) *Curr Genet* 10:139–145
- Howe CJ, Barker RF, Bowman CM, Dyer TA (1988) *Curr Genet* 13:343–349
- Hudson GS, Mason JG, Holton TA, Koller B, Cox GB, Whitfeld PR, Bottomley W (1987) *J Mol Biol* 196:283–298
- Hudson GS, Holton TA, Whitfeld PR, Bottomley W (1988) *J Mol Biol* 200:639–654
- Jansen RK, Palmer JD (1987) *Curr Genet* 11:553–564
- Kirsch W, Seyer P, Herrmann RG (1986) *Curr Genet* 10:843–855
- Ko K, Orfanides AG, Straus NA (1987) *Theor Appl Genet* 74:125–139
- Koller B, Delius H (1980) *Mol Gen Genet* 178:261–269
- Koller B, Fromm H, Galun E, Edelman (1987) *Cell* 48:111–119
- Kolodner R, Tewari KK (1975) *Biochem Biophys Acta* 401:372–390
- Lehmbeck J, Rasmussen OF, Bookjans BG, Jepsen BR, Stummann BM, Henningsen KW (1986) *Plant Mol Biol* 7:3–10
- Lehmbeck J, Stummann BM, Henningsen KW (1987) *Nucleic Acids Res* 15:3630
- Michalowski C, Breunig KD, Bohnert HJ (1987) *Curr Genet* 11:265–274
- Moritz C, Brown WM (1987) *Proc Natl Acad Sci USA* 84:7183–7187
- Morris J, Herrmann RG (1984) *Nucleic Acids Res* 12:2837–2850
- Mubumbila M, Gordon KHJ, Crouse EJ, Burkard G, Weil JH (1983) *Gene* 21:257–266
- Neuhaus H, Link G (1987) *Curr Genet* 11:251–257
- Ohme M, Tanaka M, Chunwongse J, Shinozaki K, Sugiura M (1986) *FEBS Lett* 200:87–90
- Ohyama K, Fukuzawa H, Kohchi T, Shirai H, Sano T, Sano S, Umesono K, Shiki Y, Takeuchi M, Chang Z, Aota S, Inokuchi H, Ozeki H (1986) *Nature* 322:572–574
- Oishi KK, Shapiro DR, Tewari KK (1984) *Mol Cell Biol* 4:2556–2563
- Palmer JD (1983) *Nature* 301:92–93
- Palmer JD (1985) *Annu Rev Genet* 19:325–354
- Palmer JD (1986) *Methods Enzymol* 118:167–186
- 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 (1983) *Mol Gen Genet* 190:13–19
- 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*. New York, pp 273–283
- Palmer JD, Boynton JE, Gillham NW, Harris EH (1985a) 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
- Palmer JD, Jorgensen RA, Thompson WF (1985b) *Genetics* 109:195–213
- Palmer JD, Osorio B, Aldrich J, Thompson WF (1987a) *Curr Genet* 11:275–286
- Palmer JD, Nugent JM, Herbon LA (1987b) *Proc Natl Acad Sci USA* 84:769–773
- Phillips AL, Gray JC (1984) *Mol Gen Genet* 194:477–484
- Polhill RM, Raven PH (1981) *Advances in legume systematics, Part 1*. Royal Botanic Gardens, Kew
- Posno M, Torenlvliet DJ, Lustig H, van Noort M, Groot GSP (1985) *Curr Genet* 9:211–219
- Purton S, Gray JC (1987) *Nucleic Acids Res* 15:1873
- Quigley F, Weil JH (1985) *Curr Genet* 9:495–503
- Rasmussen OF, Bookjans G, Stummam BM, Henningsen KW (1984a) *Plant Mol Biol* 3:191–199
- Rasmussen OF, Stummam BM, Henningsen KW (1984b) *Nucleic Acids Res* 12:9143–9153
- Rasmussen O, Jepsen B, Stummam B, Henningsen KW (1987) *Nucleic Acids Res* 15:854
- Reed KC, Mann DA (1985) *Nucleic Acids Res* 13:7207–7221
- Shapiro DR, Tewari KK (1986) *Plant Mol Biol* 6:1–12
- Shen GF, Chen K, Wu M, Kung SD (1982) *Mol Gen Genet* 187:12–18
- Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi T, Zaita N, Chunwongse K, Obokata J, Yamaguchi-Shiozaki K, Ohto C, Torazawa K, Meng BY, Sugita M, Deno H, Kamogashira T, Yamada K, Kusuda J, Takaiwa F, Kato A, Tohdoh N, Shimada H, Sugiura M (1986) *EMBO* 5:2043–2049
- Sijben-Muller G, Hallick RB, Alt J, Westhoff P, Herrmann RG (1986) *Nucleic Acids Res* 14:1029–1044
- Singh GP, Wallen DG, Pillay DTN (1985) *Plant Mol Biol* 4:87–93
- Smith AG, Gray JC (1984) *Mol Gen Genet* 194:471–476
- Smith GE, Summers MD (1980) *Anal Biochem* 109:123–129
- Spielmann A, Ortiz W, Stutz E (1983) *Mol Gen Genet* 190:5–12
- Spielmann A, Stutz E (1983) *Nucleic Acids Res* 11:7157–7167
- Stummam BM, Lehmbeck J, Bookjans G, Henningsen KW (1988) *Physiol Plant* 72:139–146
- Sugita M, Kato A, Shimada H, Sugiura M (1984) *Mol Gen Genet* 19:200–205
- Sugita M, Shinozaki K, Sugiura M (1985) *Proc Natl Acad Sci USA* 82:3557–3561
- Sugiura M, Shinozaki K, Zaita N, Kusuda M, Kumano M (1986) *Plant Sci* 44:211–216
- Tanaka M, Wakasugi T, Sugita M, Shinozaki K, Sugiura M (1986) *Proc Natl Acad Sci USA* 83:6030–6034
- Turker MS, Domenico JM, Cummings DJ (1987) *J Mol Biol* 198:171–185
- von Allmen JM, Stutz E (1987) *Nucleic Acids Res* 15:2387
- Westhoff P, Farchaus JW, Herrmann RG (1986) *Curr Genet* 11:165–169
- Westhoff P, Grune H, Schrubar H, Oswald A, Streubel M, Ljungberg U, Herrmann RG (1988) In: Scheer H, Schneider S (eds) *Photosynthetic light-harvesting systems: structure and function*. de Gruyter, Berlin (in press)
- Willey DL, Auffret AD, Gray JC (1984) *Cell* 36:555–562
- Woodbury NW, Roberts LL, Palmer JD, Thompson (1988) *Curr Genet* 14:75–89
- Zurawski G, Bottomley W, Whitfeld PR (1982) *Proc Natl Acad Sci USA* 79:6260–6264
- Zurawski G, Bottomley W, Whitfeld PR (1984) *Nucleic Acids Res* 12:6547–6558
- Zurawski G, Bottomley W, Whitfeld PR (1986a) *Nucleic Acids Res* 14:3974
- Zurawski G, Bottomley W, Whitfeld PR (1986b) *Nucleic Acids Res* 14:3975

Communicated by R. W. Lee

Received November 11, 1987 / February 3, 1988