Curr Genet (1986) 10:823–833

Current Genetics

Conservation of chloroplast genome structure among vascular plants

Jeffrey D. Palmer[1] **and Diana B. Stein**[2]

[1] Department of Biology, University of Michigan, Ann Arbor, MI 48109, USA
[2] Department of Biological Sciences, Mount Holyoke College, South Hadley, MA 01075, USA

Summary. We have constructed the first physical map of a gymnosperm chloroplast genome and compared its organization with those of a fern and several angiosperms by heterologous filter hybridization. The chloroplast genome of the gymnosperm *Ginkgo biloba* consists of a 158 kb circular chromosome that contains a ribosomal RNA-encoding inverted repeat approximately 17 kb in size. Gene mapping experiments demonstrate a remarkable similarity in the linear order and absolute positions of the ribosomal RNA genes and of 17 protein genes in the cpDNAs of *Ginkgo biloba,* the fern *Osmunda cinnamomea* and the angiosperm *Spinacia oleracea.* Moreover, filter hybridizations using as probes cloned fragments that cover the entirety of the angiosperm chloroplast genome reveal a virtually colinear arrangement of homologous sequence elements in these genomes representing three divisions of vascular plants that diverged some 200–400 million years ago. The only major difference in chloroplast genome structure among these vascular plants involves the size of the rRNA-encoding inverted repeat, which is only 10 kb in *Osmunda,* 17 kb in *Ginkgo,* and about 25 kb in most angiosperms. This size variation appears to be the result of spreading of the repeat through previously single copy sequences, or the reverse process of shrinkage, unaccompanied by any overall change in genome complexity.

Key words: Inverted repeat – Gene order – Chloroplast genome arrangement – Vascular plant

Introduction

Recent studies have revealed a high degree of conservation of the basic size and sequence arrangement of chloroplast DNA (cpDNA) among over 200 species of flowering plants representing some 150 million years of angiosperm evolution (reviewed in Whitfeld and Bottomley 1983; Crouse et al. 1985; Gillham et al. 1985; Palmer 1985a, 1985b). All angiosperm cpDNAs are circular molecules, the great majority of which cluster in a rather narrow size range of between 135 kb and 160 kb. With the exception of one group of leguminous genomes (Kolodner and Tewari 1979; Koller and Delius 1980; Palmer and Thompson 1981a, 1982; Chu and Tewari 1982), all angiosperm cpDNAs contain a large inverted repeat some 20–30 kb in size that encodes a complete set of ribosomal RNA genes. Most angiosperm cpDNAs have the same linear order of genes and other sequences (Fluhr and Edelman 1981; Palmer and Thompson 1982; Palmer et al. 1983a, 1983b; De Heij et al. 1983). Where gene order differences are found, these can often be attributed to single inversions (Palmer and Thompson 1982; Herrmann et al. 1983; Howe et al. 1983), although more extensive rearrangement has been noted in certain of those leguminous cpDNAs that lack the large inverted repeat (Palmer and Thompson 1981a, 1982; Palmer et al. 1985a; Mubumbila et al. 1984).

Much less is known about structural variation in cpDNAs from non-angiosperms. CpDNAs from the few non-angiospermous land plants and most of the algae examined fall into the same narrow size range as angiosperm genomes (reviewed in Crouse et al. 1985; Palmer 1985a, 1985b). Green algae cpDNAs, however, are exceptionally variable in size, ranging from 85 kb in *Codium* (Hedberg et al. 1981) to over 2,000 kb in *Acetabularia* (Padmanabham and Green 1978; Tymms and Schweiger 1985). Moreover, cpDNAs from within a single genus of green algae, *Chlamydomonas*, vary extensively both in size and in linear sequence arrangement (Lemieux and Lemieux 1985; Lemieux et al. 1985a, 1985b; Palmer et al. 1985b).

Offprint requests to. J. D. Palmer

Table 1. Sources of gene probes

Gene name	Species	Fragment[a]	Gene location[b]	Reference
rpl2-rps19	Spinach	1.5 kb XhoI-PstI	+30 *rpl2*/++300 *rps19*	c
*pet*D	Spinach	296 bp BamHI	−162/+134	d
*pet*B	Spinach	2.4 kb SalI-BamHI	−900/++864	d
3′ *psb*B	Spinach	1,597 bp BamHI-SalI	+258/++327	e
5′ *psb*B	Spinach	338 kp BamHI	−80/+258	e
5′ *psb*E	Spinach	0.65 kb EcoRI	−465/+185	f
3′ *psb*E	Spinach	0.50 kb EcoRI	+185/++433	f
3′ *pet*A	Pea	1.1 kb BamHI	+569/++600	g
5′ *pet*A	Pea	3.2 kb BamHI	−2,600/+569	g
3′ *rbc*L	Pea	1,167 bp PstI-HindIII	+173/+1,340	h
5′ *rbc*L	Pea	0.7 kb XbaI-PstI	−500/+173	h
*atp*B	Spinach	1,980 bp EcoRI	−440/++46	i
*psa*A1	Spinach	2.4 kb BamHI	+100/++300	j
*psa*A2	Spinach	1.6 kb BamHI	+600/++100	j
*psb*C	Spinach	367 bp BamHI-PstI	+983/+1350	j
*psb*D	Pea	1,150 bp PstI	+227/++315	k
*atp*H	Pea	0.8 kb PstI-BamHI	+92/++600	l
*atp*F-5′ *atp*A	Spinach	1.6 kb SalI-HindIII	+50 *apt*F/+800 *atp*A	m
*atp*F-*atp*A	Spinach	2.4 kb SalI	+50 *atp*F/++150 *atp*A	m
5′ *psb*A	Pea	532 bp EcoRI-PstI	−58/+464	n
3′ *psb*A	Pea	1.2 kb PstI-EcoRI	+464/++700	n
16S rRNA	Tobacco	16S rRNA		o
23S rRNA	Tobacco	23S rRNA		o

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

[b] "−x" indicates gene probe starts x bp before the initiation codon; "+x" indicates gene probe either starts or ends x bp following the initiation codon; "++x" indicates gene probe ends x bp following the termination codon

[c] Zurawski et al. (1984)

[d] Heinemeyer et al. (1984)

[e] Morris and Herrmann (1984)

[f] Herrmann et al. (1984)

[g] Willey et al. (1984)

[h] Zurawski et al. (1986)

[i] Zurawski et al. (1982)

[j] Alt et al. (1984)

[k] Rasmussen et al. (1984)

[l] J. Gray (unpublished data)

[m] Westhoff et al. (1985)

[n] Oishi et al. (1984)

[o] Gift of D. Bourque

The only two non-angiospermous land plants whose cpDNAs have been studied in any detail are the fern *Osmunda cinnamomea* (Palmer and Stein 1982) and the bryophyte *Marchantia polymorpha* (Ohyama et al. 1983). At 144 kb, the fern genome is of a comparable size to most angiosperm genomes, while the bryophyte genome, at 121 kb, is somewhat smaller. Both genomes contain a rRNA-encoding inverted repeat located in the same relative position and orientation as in angiosperm genomes. However, at 10–11 kb in size, the fern and bryophyte repeats are considerably smaller than the 20–30 kb repeats found in most angiosperms. Only a single protein gene has been mapped in *Marchantia* and four in *Osmunda;* these are present in approximately the same locations as in angiosperms.

In order to gain a deeper understanding of the comparative structure of chloroplast genomes among land plants we have conducted detailed comparisons of cpDNAs from several angiosperms with those of the fern *Osmunda cinnamomea* and the gymnosperm *Ginkgo biloba.* We report the first physical map for any gymnosperm and show that the *Ginkgo* chloroplast genome is of standard size and contains the usual rRNA-encoding inverted repeat. We then present extensive gene mapping and cross-hybridization experiments showing that, except for one or two alterations relating to differences in size of the inverted repeat, the chloroplast genomes of these three major groups of vascular plants are colinear in sequence order. This allows us to derive a consensus structure for the ancestral chloroplast genome of all vascular plants.

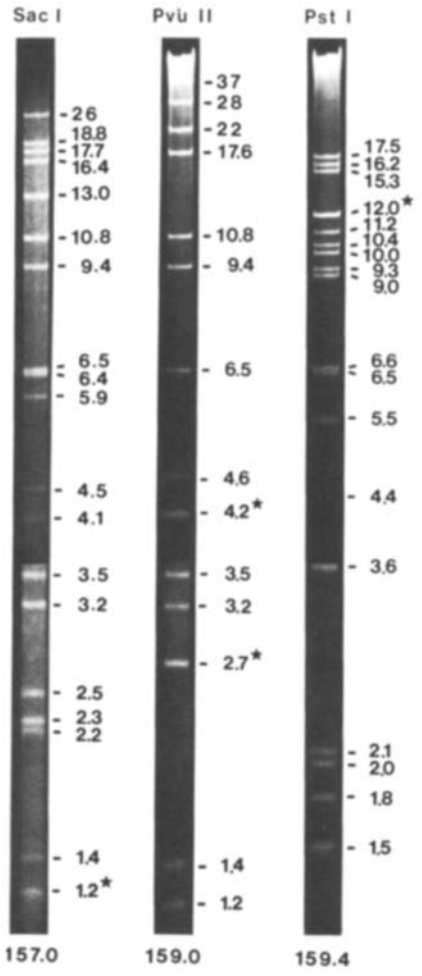

Fig. 1. Separation of *Ginkgo* cpDNA SacI, PvuII and PstI fragments on a 0.7% agarose gel. Fragment sizes are given in kb. Doublet intensity bands are marked with a *star*. Additional PstI fragments of 1.0, 0.9 and 0.6 kb were observed on other gels. Numbers at bottom indicate summation of restriction fragment sizes, taking into account fragment stoichiometries. The bottom portion of the gel (below 3.7 kb) is shown at a longer exposure than the top portion

Materials and methods

Chloroplast DNA was isolated from *Osmunda cinnamomea* as described (Palmer and Stein 1982). CpDNA was purified from *Ginkgo biloba* by the sucrose gradient technique (Palmer 1982, 1985c), with the following modifications. The initial low-speed pellet (i.e. the product of blendor homogenization, cheesecloth and miracloth filtration, and centrifugation at 1,000 g for 15 min) was observed by microscopy to consist primarily of unbroken cells and clumps of cells. This material was resuspended in homogenization buffer and cells were disrupted using a Brinkman homogenizer (model PT 10-35) for 30 s at speed setting of 7–8. This secondary homogenate was then filtered through miracloth and further treated exactly as described (Palmer 1982, 1985c).

Recombinant plasmids containing the second through the seventh gene probes listed in Table 1 were constructed in collaboration with K. Ko, who subcloned the indicated restriction fragments (Table 1) from parental spinach PstI clones (Palmer and Thompson 1981b). Each of these six gene probes was cloned into pBR322, except for the 5′ and 3′ *psb*E probes, which are in pDPL13. The other 15 gene probes were constructed by isolating the indicated restriction fragments from parental PstI clones (Palmer and Thompson 1981b) and ligating them with appropriately digested pUC8 or pUC12 (except for the *rpl2-rps19* gene clone, for which pIC20H is the vector). The constructs were used to transform *E. coli* strain JM83, recombinant white colonies were selected on ampicillin/X-gal plates, and the desired clones were identified by restriction analysis of purified plasmid DNA. All plasmid DNAs were isolated from bacteria by the alkaline extraction procedure of Birnboim and Doly (1979).

Restriction endonuclease digestions, agarose gel electrophoresis, bidirectional transfers of DNA fragments from agarose gels to nitrocellulose filters, labeling of recombinant plasmids by nick-translation, and filter hybridizations were performed as described (Palmer 1982, 1985c). All filters were washed in 2 × SSC (0.3 M NaCl/30 mM trisodium citrate) and 0.5% SDS at 65 °C prior to autoradiography.

Results

Physical mapping of Ginkgo chloroplast DNA

We wished to compare the relative arrangement of cpDNA sequences in the three major lineages of vascular plants – ferns, gymnosperms and angiosperms. Complete restriction maps to serve as a framework for such comparisons were available for a single fern (Palmer and Stein 1982) and for numerous angiosperms (Whitfeld and Bottomley 1983; Crouse et al. 1985; Gillham et al. 1985), but not for any gymnosperms. We therefore constructed a restriction map for a gymnosperm, the broad-leaved deciduous tree *Ginkgo biloba*.

A DNA preparation substantially enriched in cpDNA sequences (herein termed "cpDNA") was purified from one month old green leaves of *Ginkgo biloba* as described in Materials and methods. This DNA consisted of more or less equal proportions of cpDNA and nuclear DNA (Fig. 1). The cpDNA sequences were present as discrete fragments ranging from 1.2 kb to 37 kb, and the nuclear DNA sequences as a diffuse smear in the SacI lanes and as much tighter, high molecular weight smears in the PvuII and PstI lanes. The failure of these two enzymes to cut *Ginkgo* nuclear DNA suggests a high degree of methylation at cytosine-containing dinucleotides that are part of the basic trinucleotide C-X-G, as is the case for nuclear DNA of angiosperms (Gruenbaum et al. 1981; Palmer 1985c).

A map of the cpDNA restriction fragments produced by the three enzymes whose digest patterns are shown in Fig. 1 was constructed according to the mapping strategy previously described (Palmer 1982). Total *Ginkgo* cpDNA was labeled with ^{32}P by nick-translation, digested with PstI, and the resulting radioactive fragments were

Table 2. Summary of *Ginkgo* restriction mapping hybridizations. Cross-hybridization between PstI probe fragments was observed only between the 15.3 kb and 6.5 kb fragments and between the 16.2 kb and 3.6 kb fragments. PstI fragments of 1.0 kb and 0.9 kb were not used as probes

Probe DNA	Filter-bound DNA hybridized			
PstI	SacI	SacI-PstI	PvuII	PvuII-PstI
17.5	17.7, 6.5, 4.5, 4.1, 3.5	4.5, 4.1, 3.8, 3.5, 1.5	22, 17.6, 1.2	10.8, 5.1, 1.2
16.2	26, 9.4	11.8, 4.0, 3.6	28	16.2, 3.6
15.3	18.8, 16.4, 3.2, 1.4	12.8, 6.5, 1.4, 1.2	37, 9.4	15.3, 6.3
12.0	26, 18.8, 16.4, 9.4	6.0, 5.5	37, 28, 9.4, 4.2, 2.7	4.2, 3.1, 2.7, 1.7
11.2	13.0, 6.4, 5.9, 2.5, 1.2	5.9, 2.5, 1.2	10.8, 6.5, 3.5, 1.4	5.0, 3.4, 1.3, 1.2
10.4	10.8, 2.3, 1.2	6.5, 2.3, 1.2	37	10.4
10.0	13.0	10.0	10.8	9.6
9.3	17.7, 6.4	5.7, 3.7	6.5, 4.6, 3.2	4.6, 3.2, 1.4
9.0	17.7	9.0	22	9.0
6.6	16.4, 2.2	4.1, 2.2	17.6	6.6
6.5	18.8, 16.4	12.8, 6.5	37, 9.4	15.3, 6.3
5.5	6.5	5.5	17.6	5.5
4.4	26	4.4	28	4.4
3.6	26, 9.4	4.0, 3.6	28	16.2, 3.6
2.1	3.2	2.0	37	2.1
2.0	10.8	2.0	37	2.0
1.8	13.0	1.8	37	1.8
1.5	10.8	1.5	37	1.5

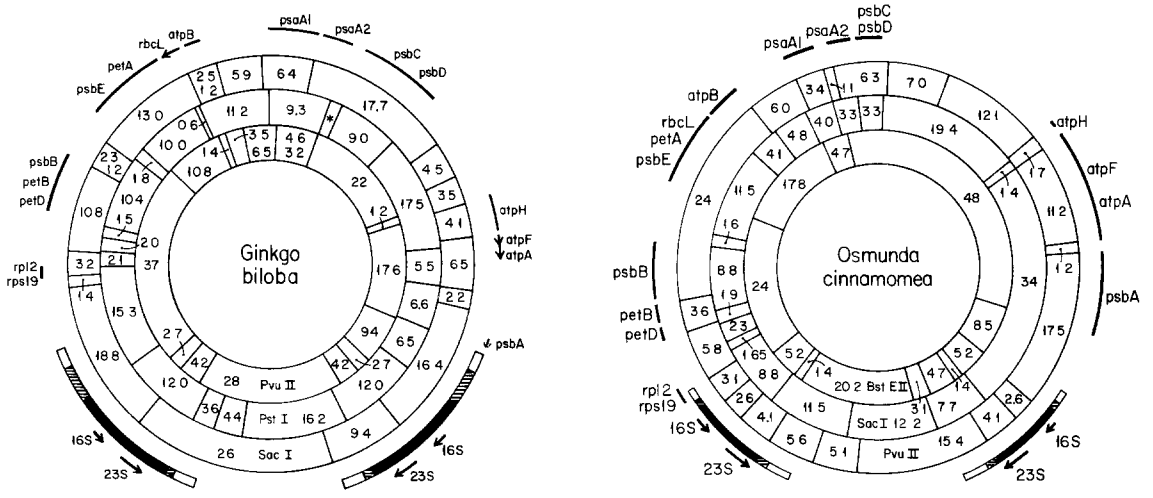

Fig. 2. Physical and gene maps of the *Ginkgo biloba* and *Osmunda cinnamomea* chloroplast chromosomes. Cleavage sites and gene locations are based on the data presented in Tables 2 and 3 for *Ginkgo* and on Palmer and Stein (1982) and Table 3 for *Osmunda*. Gene orientations are assigned in those cases where 5′ and 3′ probes for a given gene hybridized differentially across a restriction site (Table 3). The orientation of the rRNA genes was assigned as described in the text. In each map, the two long, filled lines represent the minimum mapped extent of the inverted repeat, the hatched extensions of these lines represent the estimated extent of the repeat, and the open extensions of these lines represent its maximum possible extent (see text for explanation; also see Palmer and Stein 1982). The asterisk in the *Ginkgo* PstI map indicates a "hole" which is probably occupied by one or both of the two unmapped PstI fragments of 1.0 kb and 0.9 kb. Note that each map shows only one of two possible genome orientations (cf. Palmer 1983 and Stein et al. 1986)

separated by electrophoresis in a 0.7% agarose gel. Gel slices containing each fragment were diluted with hybridization buffer, boiled, and each used as a hybridization probe against replica nitrocellulose filter strips containing *Ginkgo* cpDNA digested with PstI, SacI, SacI-PstI, PvuII, and PvuII-PstI. These mapping hybridizations (Table 2) establish a single circular linkage map (Fig. 2), 158 kb in size, that incorporates all of the *Ginkgo* cpDNA fragments seen in Fig. 1. The major structural feature of this map is a large inverted repeat sequence of

Table 3. Summary of gene mapping hybridizations

Probe DNA	Filter-bound DNA hybridized				
	Osmunda		*Ginkgo*		
	PvuII	SacI	SacI	PstI	PvuII
rpl2-rps19	5.8, 3.1	8.8	3.2	15.3	[a]
*pet*D	3.6	1.9	10.8	10.4	[a]
*pet*B	3.6	8.8	10.8	10.4	[a]
3′ *psb*B	24	8.8	10.8	10.4	[a]
5′ *psb*B	24	8.8	10.8	10.4	[a]
5′ *psb*E	24	11.5	13.0	10.0	[a]
3′ *psb*E	24	11.5	13.0	10.0	[a]
3′ *pet*A	24	11.5	13.0	10.0	[a]
5′ *pet*A	24	11.5	13.0	10.0	[a]
3′ *rbc*L	[b]	[b]	[a]	10.0	10.8
5′ *rbc*L	b	b	a	0.6	10.8
*atp*B	b	b	a	11.2	1.4
*psa*A1	3.4	4.0	6.4	9.3	[a]
*psa*A2	6.3	3.3	17.7	9.3	[a]
*psb*C	6.3	3.3	17.7	9.0	[a]
*psb*D	6.3	3.3	17.7	9.0	[a]
*atp*H	1.7	19.4	4.1	17.5	[a]
*atp*F-5′ *atp*A	11.2	34	6.5	17.5	[a]
*atp*F-*atp*A	11.2	34	6.5	17.5, 5.5	[a]
5′ *psb*A	[b]	[b]	[a]	6.6, 6.5	17.6, 9.4
3′ *psb*A	b	b	a	6.5	9.4
16S rRNA	b	b	18.8, 16.4	12.0	37, 9.4, 2.7
23S rRNA	b	b	26, 9.4	12.0	4.2

[a] Digest not tested
[b] Gene already mapped (Palmer and Stein 1982)

minimum size 12.0 kb (12.0 kb doublet PstI fragment) and maximum size 21.9 kb (12.0 kb and 3.6 kb PstI fragments, plus the 6.3 kb PstI-PvuII fragment contained within the 6.5 kb PstI fragment and hybridized to by the 15.3 kb PstI fragment). The actual size of the inverted repeat was estimated by comparing cross- and self-hybridization intensities of fragments carrying the ends of the repeat. The 6.5 kb PstI fragment carrying the upper right end of the repeat hybridized about two-thirds as strongly to the 15.3 kb PstI fragment carrying the upper left end of the repeat as to itself (data not shown); therefore we estimate that the repeat extends about 4 kb into these two fragments. By similar reasoning, the repeat was estimated to extend about 1 kb into the cross-hybridizing 3.6 kb and 16.2 kb PstI fragments. The overall estimated size of the repeat is therefore 17 kb.

Gene mapping of Ginkgo and Osmunda chloroplast DNAs

We first compared the linear arrangement of cpDNA sequences in *Ginkgo*, *Osmunda* and angiosperms by mapping homologous genes. We used as probes small cloned restriction fragments containing segments of well-characterized chloroplast genes from spinach and pea (Table 1). Construction of these gene clones was based on complete sequence data for 13 of the 17 genes and on fine structure restriction site mapping for the other 4 (*psa*A1, *psa*A2, *atp*F, *atp*A; Table 1). Each of the angiosperm gene probes hybridized strongly and specifically (at hybridization wash conditions of 65 °C and 2 x SSC) to one or two restriction fragments in each of the digests of *Ginkgo* and *Osmunda* cpDNA tested (Table 3). The approximate locations of these genes on the *Ginkgo* and *Osmunda* chromosomes are diagrammed in Fig. 2. In several cases, the differential hybridization of 5′ and 3′ probes for a given gene allowed us to assign tentatively the gene's orientation in the *Ginkgo* genome (Table 3; Fig. 2). We have assumed in all cases that the observed hybridizations are the result of homologies between the identified genes whose names are given to the clones, even though many of the gene probe fragments are not completely internal to the indicated gene.

In addition to mapping 17 protein genes in each of the two genomes, we also located the 16S and 23S rRNA genes in *Ginkgo* [the *Osmunda* rRNA genes were

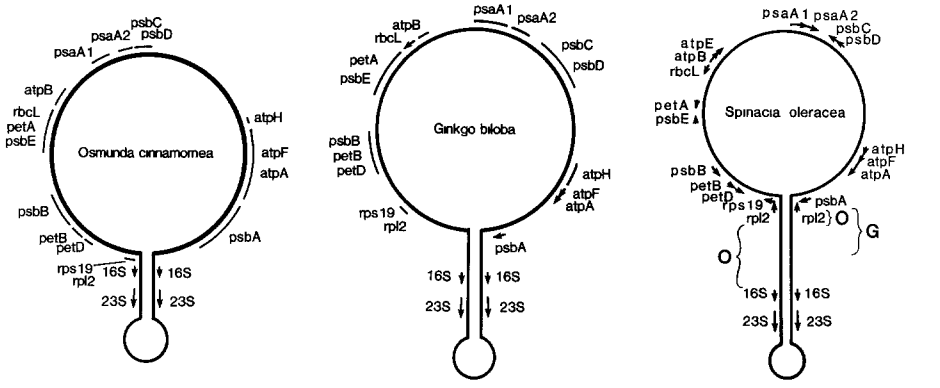

Fig. 3. Comparison of gene order and inverted repeat sizes among vascular plant cpDNAs. The *Ginkgo* and *Osmunda* maps are from Fig. 2, while the spinach mapping data are taken from Alt et al. (1984), Westhoff et al. (1985), and Zurawski et al. (1984). Brackets in the spinach map indicate regions of the genome that are part of the inverted repeat in spinach, but are single copy in *Ginkgo* and *Osmunda*

mapped previously (Palmer and Stein 1982)] by hybridizing with 16S and 23S rRNA purified from tobacco (Table 3). We assigned the orientation of the rRNA genes in both genomes (Fig. 2) by analogy to the known transcriptional orientation of the rRNA operon in all characterized chloroplast and eubacterial genomes (reviewed in Bohnert et al. 1982; Whitfeld and Bottomley 1983; Gillham et al. 1985; Palmer 1985a, 1985b). Furthermore, we note the presence of the two almost universally conserved (among vascular plants) PvuII fragments of 4.2 kb and 2.7 kb in *Ginkgo* (Fig. 2). These fragments frame the *Ginkgo* rRNA genes in a manner already described in detail in the case of the *Osmunda* genome (Palmer and Stein 1982; see also Fig. 2). This conservation suggests a similar organization of the rRNA operon – for example, containing a large transcribed spacer with two tRNA genes split by large introns – among all vascular plants.

The gene mapping studies revealed precisely the same order of genes in *Ginkgo* and *Osmunda* (Fig. 2), and only one structural difference relative to spinach, an angiosperm whose sequence arrangement approximates that of the ancestral angiosperm (Fig. 3; Fluhr and Edelman 1981; Palmer and Thompson 1982; De Heij et al. 1983). In all three genomes the rRNA genes are located within the inverted repeat and are oriented towards the small single copy region (Figs. 2, 3). The 17 mapped protein genes all lie within the large single copy region and in the same order in *Ginkgo* and *Osmunda.* The major difference is that the entire spectrum of protein genes is shifted about 25 °C clockwise in *Ginkgo* relative to *Osmunda* (Figs. 2, 3). For example, *rpl*2 and *rps*19 are only about 1 kb from the end of the inverted repeat in *Osmunda,* but are roughly 11 kb away in *Ginkgo,* whereas the reverse is true for *psb*A, located on the other side of the genome. The 17 protein genes that are colinear in order in *Ginkgo* and *Osmunda* are present, with one major exception, in the same order in spinach. *Rpl*2, a single copy gene in *Ginkgo* and *Osmunda,* is located at the terminus of the inverted repeat in spinach (Fig. 3; Zurawski et al. 1984). With respect to this duplication, it should be noted that the inverted repeat is about 25 kb in size in spinach (Kolodner and Tewari 1979), 17 kb in *Ginkgo* (Fig. 2), and only 10 kb in *Osmunda* (Fig. 2). Thus in spinach the inverted repeat occupies a region of the genome that is single copy in *Ginkgo* and *Osmunda.*

Rearrangement hybridizations of Ginkgo and Osmunda chloroplast DNAs

The gene mapping hybridizations described above reveal a highly conserved arrangement of genes in *Ginkgo, Osmunda* and spinach. However, the strength of this conclusion is limited by the small size of the gene probes used, which in sum cover not much more than 20% of a typical vascular plant genome of 150 kb. To provide a more complete coverage of the genome, although at a diminshed level of resolution, we examined the arrangement of homologous sequences using as hybridization probes cloned fragments covering an entire angiosperm chloroplast genome. These clones contained fragments from two chloroplast genomes. A set of petunia clones was used that cover the entire large single copy region and the outer half of the inverted repeat (Palmer et al. 1983a; E. Clark and M. Hanson, unpublished data). Petunia was chosen as a representative angiosperm since similar cross-hybridization studies performed among angiosperms cpDNAs have revealed that, as in the case of spinach, its arrangement of cpDNA sequences typifies that of the ancestral angiosperm (Fluhr and Edelman 1981; Palmer and Thompson 1982; De Heij et al. 1983). The available petunia clones (Palmer et al. 1983a) provided rather limited resolution of the inner half of the inverted repeat and the small single copy region. Therefore, to obtain better resolution in this region, clones containing this segment of the mung bean genome (Palmer and Thompson 1981a, 1981b) were used as probes. The mung bean and petunia probes extend to virtually the same site within the inverted repeat, about 3 kb upstream from the 16S rRNA gene (Palmer and Thompson 1981a; Palmer et al. 1983a). We have therefore re-

Table 4. Summary of rearrangement hybridizations. Cloned mung bean and petunia cpDNA restriction fragments used as hybridization probes are designated as "MB" or "PET", respectively, followed by the size in kb of the insert. All of the cloned fragments are PstI fragments, except for PET13.1, PET8.0, and PET8.9, which are PstI-SalI fragments, and PET11.4, which is a SalI fragment (see Fig. 5)

Probe DNA	Filter-bound DNA hybridized			
	Osmunda		*Ginkgo*	
	PvuII	SacI	SacI	PstI
MB9.7	15.4, 5.6, 5.1	12.2, 11.5	26	16.2, 4.4, 3.6
MB18.8	17.5, 15.4, 5.6, 4.1, 3.1, 2.6	34, 12.2, 11.5, 8.8, 7.7	26, 18.8, 16.4, 9.4	16.2, 12.0
MB12.8	17.5, 15.4, 5.6, 4.1, 3.1, 2.6	34, 11.5, 8.8, 7.7	26, 18.8, 16.4, 9.4	12.0
PET4.6	17.5	34	18.8, 16.4	15.3, 6.5
PET7.6	17.5, 3.1	34, 8.8	18.8	15.3
PET1.4	17.5, 5.8, 3.1	34, 8.8	16.4, 3.2	15.3, 6.5
PET9.0	11.2	34	16.4, 6.5, 2.2	6.6, 5.5
PET9.2	12.1, 11.2, 1.7	34, 19.4, 1.4	6.5, 4.1, 3.5	17.5, 5.5
PET15.3	12.1, 7.0, 6.3	19.4, 3.3	17.7, 4.5	17.5, 9.0
PET13.1	6.3, 6.0, 3.4	4.8, 4.0, 3.3	17.7, 6.4, 5.9	11.2, 9.3
PET8.0	24, 6.0	11.5, 4.8, 4.1	13.0, 5.9, 2.5	11.2
PET4.1	24	11.5	13.0	10.0
PET8.9	24	11.5, 1.6	13.0	10.0, 1.8
PET11.4	24, 5.8, 3.6	8.8, 2.3, 1.9	10.8, 2.3	10.4, 2.0, 1.5
PET2.6	5.8	2.3, 1.65	10.8, 3.2	2.1, 2.0
PET1.5	5.8, 3.1	8.8	3.2	15.3

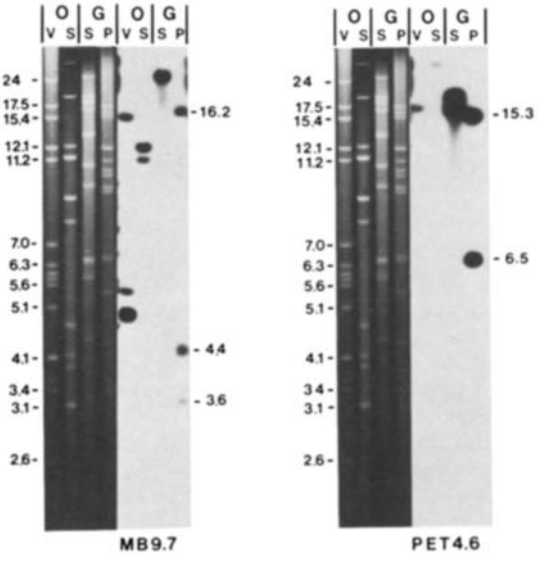

Fig. 4. Representative hybridizations of angiosperm cpDNA restriction fragments to filter blots of *Osmunda* and *Ginkgo* cpDNAs. MB9.7, a clone containing a small single copy fragment of mung bean cpDNA, and PET4.6, a clone containing an inverted repeat fragment of petunia cpDNA (see Fig. 2 for genomic locations of these fragments), were labeled with ^{32}P by nick-translation and hybridized to nitrocellulose filters containing *Osmunda* (O) PvuII (V) and SacI (S) fragments and *Ginkgo* (G) SacI (S) and PstI (P) fragments separated on a 0.7% agarose gel. Sizes of most of the *Osmunda* PvuII fragments are indicated on the left of each panel, while sizes of hybridizing *Ginkgo* PstI fragments are indicated on the right

presented these clones as abutting to form a perfect seam within the map of the "hybrid" angiosperm chloroplast genome shown in Fig. 5.

Each of the 16 petunia and mung bean clones tested hybridized to one or more restriction fragments from *Ginkgo* and *Osmunda* (Table 4). In almost all cases the angiosperm probes gave equally strong signals with each of the two genomes (e.g. MB9.7 hybridization in Fig. 4). Yet angiosperms and gymnosperms last shared a common ancestor some 200–250 million years ago, at least 130 million years more recently than their last common ancestry with the fern lineage (Banks 1970; P. G. Gensel, personal communication). One explanation for this apparent anomaly is that a saturation or near-saturation of nucleotide substitutions has occurred in that limited fraction of positions that is readily substitutable. What is perhaps more interesting is the difference in strength of hybridization of a few of the angiosperm clones to one genome relative to the other. By far the most extreme example of this is shown in Fig. 4 for clone PET4.6. This clone hybridized about 10 times more strongly to *Ginkgo* cpDNA than to *Osmunda* cpDNA. At one extreme, this could be due to the insertion in a gymnosperm-angiosperm-specific lineage of exogenous sequences that comprise most of the PET4.6 fragment, or conversely, of the deletion of such sequences in a fern-specific lineage. At the other extreme, this hybridization difference could reflect a rapid acceleration in the rate of sequence evolution in this portion of the genome in a fern-specific

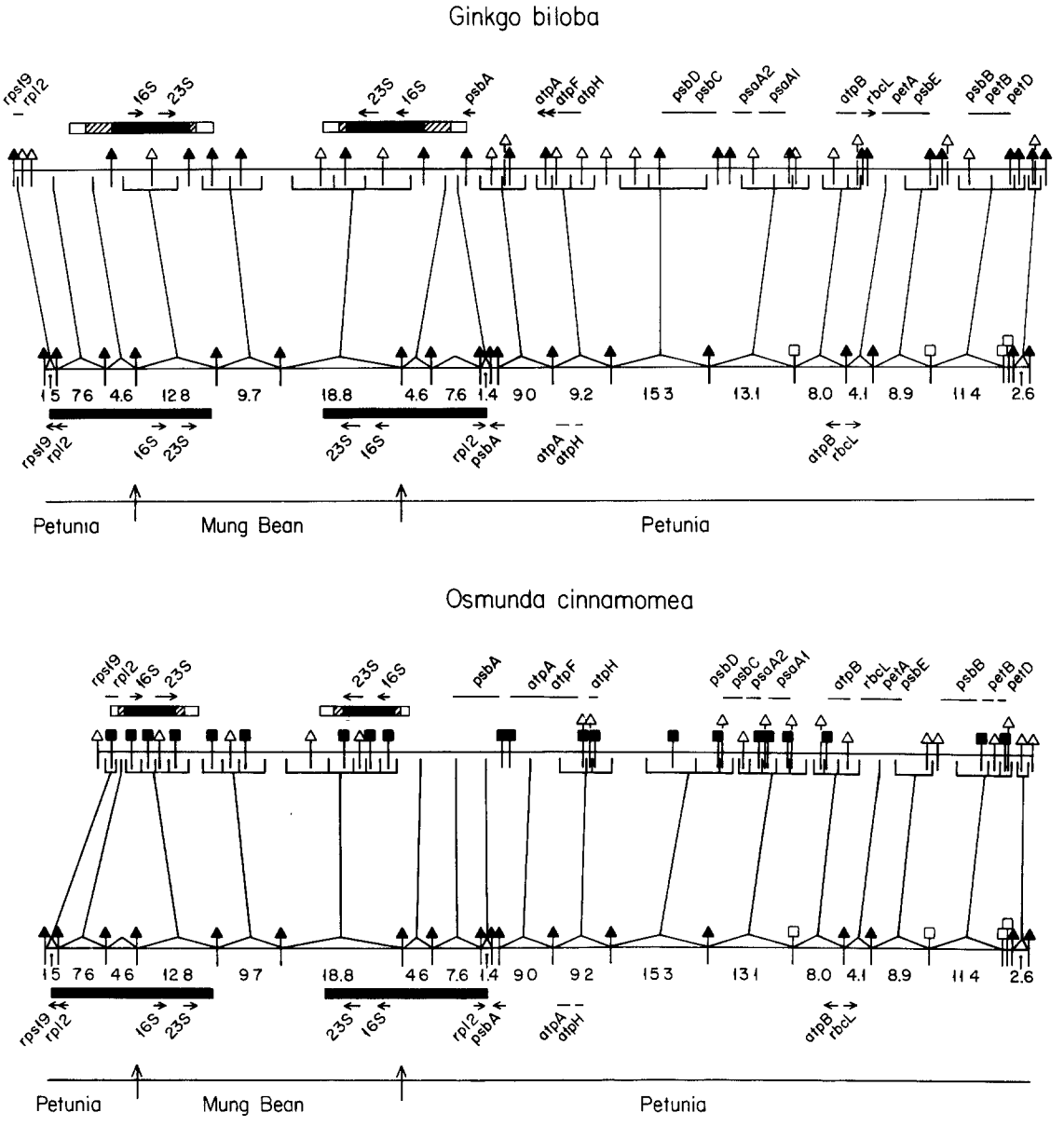

Fig. 5. Colinearity of *Ginkgo, Osmunda,* and angiosperm cpDNAs. The hybridization data summarized in Table 4 are diagrammed in terms of the *Ginkgo* and *Osmunda* (Fig. 2), mung bean (Palmer and Thompson 1981a), and petunia (Palmer et al. 1983a; Bovenberg et al. 1984a, 1984b; J. D. Palmer, unpublished data) physical and gene maps. The extent of the mung bean and petunia fragments used as hybridization probes is indicated by the two lines that converge above the fragments, while the size of each fragment is given below in kb. The *Ginkgo* (or *Osmunda*) fragments to which the probes hybridize are indicated by the lines leading from the probe fragments to the *Ginkgo* fragments. Wherever two or more probe fragments hybridized to the same filter-bound fragments the two following conventions were adopted: (1) If a probe fragment hybridized to two adjacent *Ginkgo* fragments it was assumed that the regions of homology are continguous and thus a horizontal line segment was drawn connecting the two regions of hybridization. (2) If a given probe fragment hybridized to only a single filter-bound fragment but an adjacent probe fragment also hybridized to one end of the same *Ginkgo* fragment, then a line was drawn from the first probe fragment to a region of the *Ginkgo* fragment adjacent to that region where the second probe fragment was found to hybridize. Restriction sites shown: (▲), PstI; (▽), SalI; (△), SacI; (▼), PvuII

lineage unaccompanied by any major deletions/insertions, or its converse, an abrupt slow-down in the rate of sequence evolution in a gymnosperm-angiosperm lineage. The former pair of explanations is particularly tantalizing, given that there are as yet no clearly proven differences in chloroplast gene content among all examined green algae and land plants (reviewed in Bottomley and Bohnert 1982; Gillham et al. 1985; Palmer 1985b).

Overall, these cross-hybridization experiments reveal an amazing conservation of sequence order in the three major lineages of vascular plants (Fig. 5). As an example, Fig. 4 shows that MB9.7, a fragment from the small single

copy region of mung bean, hybridizes exclusively to three *Osmunda* PvuII fragments (of 15.4 kb, 5.6 kb, and 5.1 kb) and three *Ginkgo* PstI fragments (of 16.2 kb, 4.4 kb, and 3.6 kb) that cover the entire small single copy regions of each genome, extending slightly into the inverted repeat (Fig. 2). The major differences in genome organization among these species occur in the outer portion of the inverted repeat and flanking large single copy regions. PET4.6, a fragment from near the middle of the petunia inverted repeat, hybridizes to the outer portion of the *Ginkgo* inverted repeat, but to the right end of the large single copy region in *Osmunda* (Figs. 4, 5). This difference can be imagined most easily as the consequence of the inverted repeat having spread through a region in *Ginkgo* and petunia that is single copy in *Osmunda.* PET7.6, the next outermost inverted repeat fragment in petunia, hybridizes to the left end of the large single region in *Ginkgo* (Table 4; Fig. 5). This difference may be the result of still further spreading of the inverted repeat in the angiosperm. We again point out the much larger size of the angiosperm inverted repeat, 25 kb in petunia (Bovenberg et al. 1981; Palmer et al. 1983) and also spinach (see preceding section), compared to the 17 kb and 10 kb repeats in *Ginkgo* and *Osmunda,* respectively.

Discussion

The major finding of this study is that the structure and linear arrangement of cpDNA sequences is extraordinarily conserved among a fern, a gymnosperm, and several representative angiosperms – species which last shared a common ancestor about 380 million years ago (P. G. Gensel, personal communication). Chloroplast genomes in these three fundamental divisions of vascular plants are overwhelmingly similar in size; in presence, location and orientation of a large, rRNA-encoding inverted repeat; and in the linear order and linkage relationships of sequences comprising virtually the entire cpDNA molecule. This last conclusion is based on detailed filter hybridizations of two types. First, by using small, nearly gene-specific fragments as hybridization probes, we have shown that the linear placement of 17 protein genes is, with one exception, identical in cpDNAs from the fern *Osmunda cinnamomea,* the gymnosperm *Ginkgo biloba,* and the angiosperm spinach (*Spinacia oleracea*). Second, by hybridizing with large fragments that in sum cover the entire chloroplast genome, we have shown that the *overall* order of cpDNA sequences is highly conserved in these three groups of vascular plants. These results thus extend earlier findings demonstrating that chloroplast genome arrangement is very stable among most angiosperms (see Introduction) by showing that this same arrangement is also characteristic of a fern and a gymnosperm. Moreover, as described in the accompanying paper, this conserved arrangement of cpDNA sequences is not unique to a single fern species, but characterizes a lineage of species representing some 70 million years of fern evolution (Stein et al. 1986).

We have already speculated about factors that might be responsible for the conserved arrangement of chloroplast genomes among most angiosperms (Palmer and Thompson 1982; Palmer et al. 1984; Palmer 1985a, 1985b). Most prominent among these are the compact gene organization of cpDNA (i.e. there is a relative lack of large spacer sequences that could accept inversions and other rearrangements without disruption of gene function), the lack of dispersed repeated sequences within these spacers that could serve as substrates for recombination, and structural and recombinational constraints imposed by the large inverted repeat. In the present context, we emphasize that all of these possible factors appear to be equally operative for the *Osmunda* and *Ginkgo* genomes. These have a similar large inverted repeat and are of the same approximate size and gene composition as angiosperm cpDNAs. In the absence of any direct sequence information on the *Osmunda* and *Ginkgo* genomes, we speculate that their genes are as compactly arranged as those in angiosperm genomes. The restriction mapping hybridizations reported for *Ginkgo* in this paper and for *Osmunda* previously (Palmer and Stein 1982) did not reveal any major dispersed repeated elements in these genomes, although the sensitivity of these experiments is limited by the use of uncloned fragments isolated from digests of total cpDNA as hybridization probes. However, in the case of *Osmunda,* self-hybridization experiments using cloned fragments representing half the genome still failed to detect any dispersed repeats (Stein et al. 1986).

Angiosperm cpDNA is highly conserved not only in structure (see Introduction), but in primary sequence (reviewed in Whitfeld and Bottomley 1983; Curtis and Clegg 1984, Palmer 1985b). While the studies reported here and in the accompanying paper (Stein et al. 1986) indicate a similarly low rate of cpDNA structural change in other major lineages of vascular plants, there are as yet no sequence data that bear directly on the question of rates of sequence change among fern and gymnosperm cpDNAs. However, our ability to detect positive hybridization under stringent conditions between fern and gymnosperm cpDNAs and each of 39 angiosperm cpDNA probe fragments implies a strong conservation in primary sequence. We therefore predict that the rate of primary sequence evolution will be found to be equally slow along among all vascular plant cpDNAs.

The major difference in genome organization among the species investigated in this study is the size of the inverted repeat, which is about 25 kb in most angiosperms,

17 kb in *Ginkgo*, and 10 kb in *Osmunda*. Associated with this difference is variation in the position and copy number of sequences and genes (such as *rpl*2, *rps*19, *psb*A) located either at the ends of the inverted repeat (in the larger inverted repeat genomes) or in flanking portions of the large single copy region (in the smaller inverted repeat genomes) (Figs. 3, 5). As diagrammed in Fig. 3, these changes can be envisioned as occurring by spreading of the inverted repeat in angiosperms (such as spinach) through adjacent single copy regions. Relative to *Ginkgo* this duplication can be imagined as having occurred through the lower left portion of its large single copy region, while relative to *Osmunda* one must postulate spreading of the repeat through both the right and left lower portions of the large single copy region (Fig. 3).

We present the above speculations without intending to imply any knowledge of the phylogenetic direction of the inverted repeat size changes. One could equally well imagine a larger ancestral inverted repeat that has been diminished by deletion (shrinkage) in the fern and gymnosperm lineages. However, the best available outgroup for this comparison, the bryophyte *Marchantia polymorpha* (Ohyama et al. 1983), has a cpDNA inverted repeat of about the same size as that of *Osmunda*. While much more extensive phylogenetic comparisons are obviously desirable, this observation nonetheless suggests that the larger inverted repeats in angiosperms and the intermediate-sized repeat in the gymnosperm *Ginkgo* may represent derived conditions.

It is now apparent that the linear order of genes and other sequences in the chloroplast genome has been highly conserved over the course of several hundred million years of vascular plant evolution, with the only major changes evident being shifts in size of the large inverted repeat. One now wonders whether this same genome arrangement can be found among extant non-vascular land plants and perhaps even among the likely green algal ancestors of land plants.

Acknowledgements. We are grateful to W. F. Thompson, in whose laboratory the mapping studies on *Ginkgo* cpDNA were initiated, to J. Nugent for technical assistance, to K. Ko for collaboration in construction of some of the gene probes, to E. Clark and M. Hanson for a gift of petunia SalI clones, to D. Bourque for a gift of tobacco chloroplast ribosomal RNA, to J. Gray and G. Zurawski for supplying unpublished data on the sequences of pea *atp*H and *rbc*L, to L. Herbon for assistance in preparing the figures, and to O. L. Stein and M. Zolan for critical reading of the manuscript. This research was partially funded by a William and Flora Hewlett Foundation grant of Research Corporation to D. B. Stein and by NSF grant BSR-8415934 to J. D. Palmer.

References

Alt J, Morris J, Westhoff P, Herrmann RG (1984) Curr Genet 8:597–606

Banks HP (1970) Evolution and plants of the past. Wadsworth, Belmont, CA, pp 1–170

Birnboim HC, Doly JC (1979) Nucleic Acids Res 7:1513–1523

Bohnert HJ, Crouse EJ, Schmitt JM (1982) Encycl Plant Physiol 14B:475–530

Bottomley W, Bohnert HJ (1982) Encycl Plant Physiol 14B:531 to 596

Bovenberg WA, Kool AJ, Nijkamp HJJ (1981) Nucleic Acids Res 9:503–517

Bovenberg WA, Howe CH, Kool AJ, Nijkamp HJJ (1984a) Curr Genet 8:283–290

Bovenberg WA, Koes RE, Kool AJ, Nijkamp HJJ (1984b) Curr Genet 8:231–241

Chu NM, Tewari KK (1982) Mol Gen Genet 186:23–32

Crouse EJ, Schmitt JM, Bohnert HJ (1985) Plant Mol Biol Rep 3:43–89

Curtis SE, Clegg MT (1984) Mol Biol Evol 1:291–301

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

Gillham NW, Boynton JE, Harris EH (1985) In: Cavalier-Smith T (ed) DNA and evolution: natural selection and genome size. Wiley, New York, pp 299–351

Gruenbaum Y, Nevah-Many T, Cedar H, Razin A (1981) Nature (London) 292:860–862

Hedberg MF, Huang YS, Hommersand MH (1981) Science 213: 445–447

Heinemeyer W, Alt J, Herrmann RG (1984) Curr Genet 8:543–549

Herrmann RG, Westhoff P, Alt J, Winter P, Tittgen J, Bisanz C, Sears BB, Nelson N, Hurt E, Hauska G, Viebrock A, Sebald W (1983) In: Ciferri O, Dure L III (eds) Structure and function of plant genomes. Plenum Press, New York, pp 143–153

Herrmann RG, Alt J, Schiller B, Widger WR, Cramer WA (1984) FEBS Lett 176:239–244

Howe CJ, Bowman CJ, Dyer TA, Gray JC (1983) Mol Gen Genet 190:51–55

Koller B, Delius H (1980) Mol Gen Genet 178:261–269

Kolodner R, Tewari KK (1979) Proc Natl Acad Sci USA 76:41–45

Lemieux B, Lemieux C (1985) Curr Genet 10:213–219

Lemieux C, Turmel M, Lee RW, Bellemare G (1985a) Plant Mol Biol 5:77–84

Lemieux C, Turmel M, Seligy VL, Lee RW (1985b) Curr Genet 9:139–145

Morris J, Herrmann RG (1984) Nucleic Acids Res 12:2837–2850

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

Oishi KK, Shapiro DR, Tewari KK (1984) Mol Cell Biol 4:2556 to 2563

Padmanabhan H, Green BR (1978) Biochim Biophys Acta 521: 67–73

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 Press, New York, pp 131–240

Palmer JD (1985c) Methods Enzymol 118:167–186

Palmer JD, Stein DB (1982) Curr Genet 5.165–170

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, Shields CR, Cohen DB, Orton TJ (1983a) Theor Appl Genet 65:181–189

Palmer JD, Singh GP, Pillary DTN (1983b) 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. 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

Rasmussen OF, Bookjans G, Stummann BM, Henningsen KW (1984) Plant Mol Biol 3:191–199

Stein DB, Palmer JD, Thompson WF (1986) Curr Genet 10: 835–841

Tymns MJ, Schweiger HG (1985) Proc Natl Acad Sci USA 82: 1706–1710

Westhoff P, Alt J, Nelson N, Herrmann RG (1985) Mol Gen Genet 199.290–299

Whitfeld PR, Bottomley W (1983) Annu Rev Plant Physiol 34: 279–310

Willey DL, Auffret AD, Gray JC (1984) Cell 36:555–562

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, Whitfeld PR, Bottomley W (1986) Nucleic Acids Res 14:3975

Communicated by K. P. van Winkle-Swift

Received December 3, 1985/March 3, 1986