Mol Gen Genet (1989) 216:417–421

MGG

Unusual characteristics of *Codium fragile* chloroplast DNA revealed by physical and gene mapping

James R. Manhart[1,*], Kevin Kelly[2], Bernard S. Dudock[2], and Jeffrey D. Palmer[1]

[1] Department of Biology, University of Michigan, Ann Arbor, Michigan 48109-1048; USA
[2] Department of Biochemistry, State University of New York at Stony Brook, Stony Brook, NY 11794, USA

Summary. A complete physical map of the *Codium fragile* chloroplast genome was constructed and the locations of a number of chloroplast genes were determined. Several features of this circular genome are unusual. At 89 kb in size, it is the smallest chloroplast genome known. Unlike most chloroplast genomes it lacks any large repeat elements. The 8 kb spacer region between the 16S and 23S rRNA genes is the largest such spacer characterized to date in chloroplast DNA. This spacer region is also unusual in that it contains the *rps*12 gene or at least a portion thereof. Three regions polymorphic for size are present in the *Codium* chloroplast genome. The *psb*A and *psb*C genes map closely to one of these regions, another region is in the spacer between the 16S and 23S rRNA genes and the third is very close to or possibly within the 16S rRNA gene. The gene order in the *Codium* genome bears no marked resemblance to either the "consensus" vascular plant order or to that of any green algal or bryophyte genome.

Key words: *Codium* chloroplast DNA – Physical – Gene map

Introduction

Most of the chloroplast DNAs (cpDNAs) that have been characterized to date are those of land plants. The cpDNAs of the over 500 land plants that have been examined are unicircular and all but three fall into a size range of 120 to 160 kb (Palmer 1985a, b). With the exceptions of two groups, land plant cpDNAs are segmented into four parts made up of a duplicated sequence present as an inverted repeat that partitions the rest of the molecule into small and large single copy regions. In addition, a complete set of ribosomal RNA (rRNA) genes is present within each repeat unit and the 23S gene is always close to the small single copy region. The linear order of chloroplast genes in most angiosperms also extends to a fern and a gymnosperm, allowing one to infer a consensus gene order for vascular plants (Palmer and Stein 1982, 1986). Two bryophytes have the same gene order as the majority of vascular plants with the exception of one shared inversion in the liverwort *Marchantia* (Ohyama et al. 1986) and the moss *Physcomitrella* (Calie and Hughes 1987) and one unique rearrangement in *Physcomitrella*. Land plant cpDNAs, then, are highly conserved, even among taxa that diverged over 345–375 million years ago (Scagel et al. 1982; Schofield 1985).

The information available on green algal cpDNAs is limited. The chloroplast genome size is known for one or a few species in four genera – *Acetabularia* (Green 1976; Padmanabhan and Green 1978), *Codium* (Hedberg et al. 1981), *Chlorella* (Yamada 1983), and *Chlamydomonas* (Rochaix 1978; Lemieux and Lemieux 1985; Lemieux et al. 1985; Palmer et al. 1985). The genome size of *Acetabularia* is estimated at 2000 kb and that of *Codium* at 85 kb, so in terms of size range alone, green algal cpDNAs are quite different from those of land plants. Structural information is available only for *Chlamydomonas* and *Chlorella*, and detailed gene mapping has been done in *Chlamydomonas* only. While there are large inverted repeats encoding rRNA genes in these two algal genera, the resemblance to land plant cpDNAs stops there. The content and relative position of the genes in the inverted repeats in *Chlamydomonas* are different from land plants. Further, comparative details of gene arrangement reveal many differences even among members of the genus *Chlamydomonas* due to extensive rearrangements (Lemieux et al. 1985). In no case does a *Chlamydomonas* species have a gene order resembling that of any land plant. In summary then, structural features of land plant chloroplast genomes are highly conserved while those of green algae are highly variable. The green algal genomes investigated to date bear no significant resemblance to those of land plants or even to each other except in the case of two pairs of closely related *Chlamydomonas* species (Lemieux et al. 1985; Palmer et al. 1985).

In order to expand our knowledge of algal cpDNAs we have constructed a complete physical map of the *Codium fragile* genome and have mapped a number of chloroplast genes. *Codium* cpDNA has already been shown to be unique by virtue of its small size (Hedberg et al. 1981). In addition, Francis et al. (1987) have shown that the 4.5S RNAs of *Codium* and land plants are unusually divergent in primary sequence, although they retain the potential to form similar secondary structures. Our results show that *Codium* cpDNA is unusual in a number of other respects, including a lack of any large inverted repeats, the presence of three regions of hypervariable size, and a great separation between the 16S and 23S rRNA genes.

Present address: Department of Biology, Texas A&M University, College Station, TX 77843; USA

Offprint requests to: J. Manhart

Table 1. Gene probes[a]

Gene designation	Plant source	Hybridization signal
3′*rps*12-5′*rps*7	Tobacco	+
*rpl*23-5′*rpl*2	Tobacco	−
3′*rpl*2-*rps*19-5′*rpl*22	Tobacco	−
3′*rpl*22-5′*rps*3	Tobacco	−
3′*rps*3-5′*rpl*16	Tobacco	−
3′*rpl*16-*rpl*14-5′*rps*8	Tobacco	−
3′*rps*8-*inf*A	Spinach	−
*rps*11	Spinach	−
*rpo*A	Spinach	−
3′*psb*B	Spinach	+
5′*psb*B	Spinach	+
5′*psb*E	Spinach	−
3′*psb*E-*psb*F	Spinach	−
3′*pet*A	Pea	−
5′*pet*A	Pea	−
3′*rbc*L	Pea	+
*rbc*L	Pea	+
5′*rbc*L	Pea	+
*atp*B	Pea	+
*atp*E	Spinach	+
*psa*A	Spinach	+
*psa*B	Spinach	+
3′*psb*C	Spinach	+
3′*psb*D-5′*psb*C	Pea	+
5′*psb*D	Pea	+
*rpo*B	Tobacco	+
*rpo*C	Pea	−
*rps*2-*atp*I-5′*atp*H	Pea	−
3′*atp*H	Pea	−
*atp*F-5′*atp*A	Spinach	+
3′*atp*A	Spinach	+
*rps*16	Tobacco	−
5′*psb*A	Pea	+
3′*psb*A	Pea	+
*tuf*A[b]	*Chlamydomonas*	+
16S rRNA[c]	Tobacco	+
23S rRNA[d]	Tobacco	+

[a] Sources of gene probes except *tuf*A, 16S rRNA and 23S rRNA are described in detail in Jansen and Palmer (1987)
[b] 374 bp *Pst*I-*Eco*RI internal fragment extending 112–486 bp past the initiation codon (S. Baldauf, personal communication)
[c] 3269 bp *Bam*HI fragment containing the entire 16S rRNA gene and 1226 bp of flanking sequence at its 5′ end and 555 bp of flanking sequence at its 3′ end
[d] 1680 bp *Nhe*I-*Sac*I internal fragment extending from positions 811–2491 past the 5′ end of the 23S rRNA gene

Materials and methods

Plants of *C. fragile* were collected from the intertidal zone at Peconic Bay, Long Island, New York. Three kilogram batches of *Codium* were ground in an Oster Juice Extractor and the liquid extract combined immediately with 1.5 l of isolation buffer (0.32 M sorbitol, 50.0 mM Hepes, pH 7.8, 5.0 mM EDTA, 8.0% polyethylene glycol 400, 75.0 mg/l DTT, 0.5 g/l defatted BSA). The mixture was spun for 10 s in a Sorvall GS3 rotor at 5000 rpm and the chloroplast pellet carefully resuspended in 2.5 l of the isolation buffer. The chloroplast suspension was then passed through a single layer of Miracloth and spun again for 10 s in a Sorvall GS3 rotor at 5000 rpm. The pellets were resuspended in a minimal amount of isolation buffer and the crude chloroplast suspensions combined. This was then layered on top of an equal volume of 1.6 M sucrose, 10 mM Tris, pH 8.0, and spun for 1 h in a Sorvall HB-4 rotor at 10000 rpm. Chloroplasts were recovered from the interface and washed twice by centrifuging in isolation buffer. The resulting pellets were resuspended and lysed in 7.0 ml of TE buffer (10 mM Tris, 1 mM EDTA, pH 7.8) containing 1% sodium sarcosyl. Bisbenzimide H33258 was added to a final concentration of 0.04 mg/ml and the solution was brought to a final refractive index of 1.3895 with cesium chloride. The solution was then spun for 36 h in a Beckman Ti 55.2 rotor at 40000 rpm. The upper cpDNA band was carefully removed and centrifuged again under the above conditions. Following recentrifugation, *Codium* cpDNA was extracted with 1-propanol to remove the bisbenzimide and dialyzed against TE buffer to remove cesium chloride. The DNA was concentrated by butanol extraction, phenol extracted, and collected by ethanol precipitation.

Codium cpDNA was cut with the restriction endonucleases *Bcl*I, *Cla*I, *Eco*RI, and *Xho*I as prescribed by the supplier (BRL). Digests were electrophoresed through 0.8% agarose. Fragment sizes were determined by comparison with known standards.

Restriction site and gene maps were constructed by bidirectional transfer of DNA fragments from agarose gels to Zetabind filters and subsequent Southern hybridization of these filters. Hybridization probes consisted of ^{32}P-labeled gel-isolated *Bcl*I and *Xho*I fragments of *Codium* cpDNA and recombinant plasmids containing the heterologous gene probes listed in Table 1. Nick translations, filter hybridizations (60° C and 4 × SSC for heterologous probes and 65° C and 4 × SSC for homologous probes) were performed as described (Palmer 1982, 1986).

Results and discussion

Physical structure and gene content

Codium cpDNA was digested to completion with the restriction endonucleases *Bcl*I, *Cla*I, *Eco*RI, and *Xho*I (Fig. 1). The summation of fragment sizes for these restriction enzymes indicates a genome size of 89 kb (Table 2). This is in good agreement with the estimate of Hedberg et al. (1981) of 85 kb.

A restriction fragment and gene map, constructed as described in Materials and methods, reveals that the 89 kb *Codium* genome exists as a single circular chromosome (Fig. 2). An unusual feature of *Codium* cpDNA is the absence of any large repeats (Fig. 2). Almost all land plant cpDNAs, as well as those in *Chlamydomonas* (Rochaix 1978; Lemieux and Lemieux 1985; Lemieux et al. 1985; Palmer et al. 1985) and *Chlorella* (Yamada 1983), have an inverted repeat that ranges in size from 10 to 76 kb and contains the rRNA genes. The unusually small size of the *Codium* chloroplast genome can only in part be accounted for by the absence of a large inverted repeat.

In terms of gene content, the major similarities between *Codium* and land plant cpDNAs are the presence of common genes encoding rRNAs and 12 proteins (Table 1). However, a number of gene probes did not give any autoradiographic signal under the experimental conditions used (Table 1). These probes include *rpo*A, *psb*E, *psb*F, *pet*A, *atp*I, *atp*H, and all the ribosomal protein genes tested with one exception (see Table 1). The lack of hybridization of

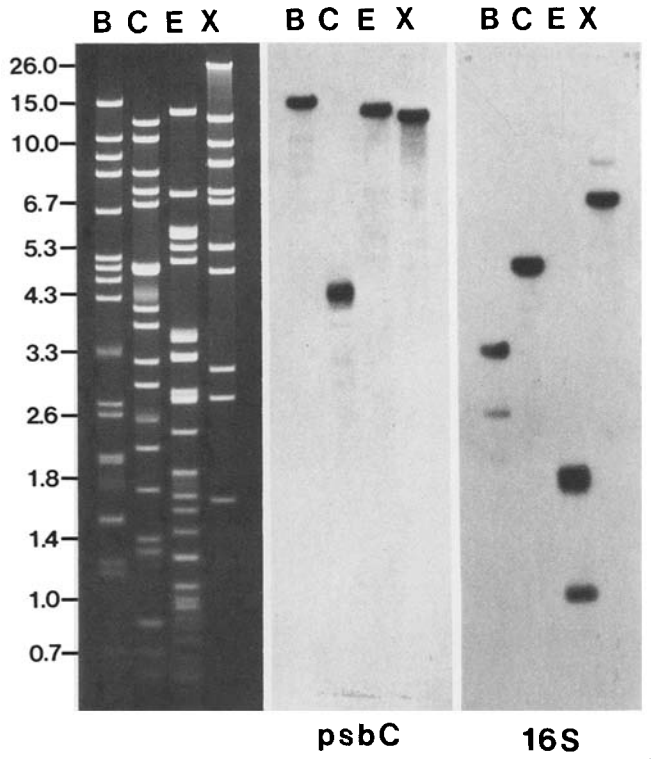

Fig. 1. Digestion products of *Codium fragile* chloroplast DNA and results of Southern hybridizations with radiolabeled *psb*C and 16S rRNA probes. DNA was cut with the following restriction endonucleases: lane B, *Bcl*I; lane C, *Cla*I, lane E, *Eco*RI; lane X, *Xho*I. Fragment sizes were determined by co-electrophoresis with a range of phage lambda DNA fragments produced by single digestions with *Hin*dIII, *Sal*I, and *Sma*I and double digestions with *Eco*RI + *Hin*dIII and *Eco*RI + *Kpn*I. Sizes are indicated in kilobases

Table 2. Restriction endonuclease fragments of *Codium* chloroplast DNA. Sizes are given in Kilobases

*Bcl*I	*Cla*I	*Eco*RI	*Xho*I
15.0	12.0	13.0	26.0
10.3	10.4	7.0	12.3
9.0	8.2	5.8[a]	10.0
8.1	7.3	5.6	8.6
6.5	6.6	5.3	7.2
5.1	4.9[a]	5.0	6.7
4.9	4.8(2)	3.7	5.3
4.7	4.3[a]	3.6	4.8
4.3	4.1	3.4	3.1
3.3[a]	3.8	3.2(2)	2.7
2.7	3.2	2.8	1.7
2.6	2.9	2.7	88.4
2.1	2.5[a]	2.6	
2.0	2.1	2.3	
1.8[a]	1.7	1.9	
1.6 (2)	1.5	1.8[a]	
1.3	1.3	1.7	
1.2	0.8	1.6	
0.7	0.7	1.5	
0.5	0.6	1.2 (2)	
89.3	88.5	1.1 (2)	
		1.05	
		1.00	
		0.97	
		0.95	
		0.90	
		0.68	
		0.60	
		0.58	
		0.50	
		89.5	

[a] Blurred fragment

the other probes may be because these proteins are not under the same constraints with regard to sequence conservation as the majority of the photosynthetic genes and rRNA genes that did give autoradiographic signals. Alternatively, some of these genes may be absent from the *Codium* chloroplast genome. This is a distinct possibility, as the sum of total amino acid, tRNA, and rRNA specifying sequence of land plants, excluding one of the inverted repeats, is 80 kb. That leaves less than 9 kb of noncoding sequence (including all regulatory sequences, introns and spacer sequences) in *Codium*, assuming its chloroplast genome does have the same genic complement as that of land plants.

There is one clear difference between *Codium* cpDNA and land plant cpDNAs with regard to gene content. It has been established from sequence data that the *tuf*A gene, which encodes the polypeptide elongation factor Ef-Tu is absent from the cpDNAs of tobacco and *Marchantia* (Ohyama et al. 1986; Shinozaki et al. 1986). The *tuf*A probe used here, which is from *Chlamydomonas*, hybridizes to *Codium* cpDNA (Fig. 2). Further, a broad survey indicates that *tuf*A is present in the cpDNAs of many green algae, but is absent from all land plant cpDNAs (J. Manhart, S. Baldauf and J. Palmer, unpublished results).

The distance between the 16S and 23S rRNAs genes in *Codium* is estimated as 8 kb. This is the largest separation of these two genes in any cpDNA. The spacer between the 16S and 23S rRNAs normally is 2.5 kb or smaller in size and contains the $tRNA^{Ile}$ and $tRNA^{Ala}$ genes. *Chlorella ellipsoidea* is an exception as the spacer is 4.8 kb and contains the gene for $tRNA^{Ile}$, but not that for $tRNA^{Ala}$ (Yamada and Shimaji 1986). Moreover, there are four open reading frames present in the *Chlorella* spacer that are not found in the spacer of any other organisms characterized to date. With the exception of *Chlorella*, the 16S and 23S rRNA genes in chloroplast genomes are transcribed as a single unit (Whitfield and Bottomley 1983); in *Chlorella* the 23S rRNA gene is coded on the opposite strand in the reverse order (Yamada and Shimaji 1986). In *Codium* cpDNA, the large separation between the 16S and 23S rRNA genes and the presence of at least a portion of the *rps*12 gene in this region (Fig. 2) makes it unlikely that the rRNA genes are cotranscribed. None of the other gene probes mapped to the region between the rRNA genes. If no other genes are present in this region, then the *Codium* chloroplast genome almost certainly must contain fewer genes than the chloroplast genomes of land plants. This would indicate a more extensive transfer of chloroplast genes to the nucleus in *Codium* than in any other photosynthetic organism.

Gene arrangement

The gene order in *Codium* cpDNA does not show any strong similarities to the consensus vascular plant gene order as exemplified by that of tobacco (Fig. 3). One char-

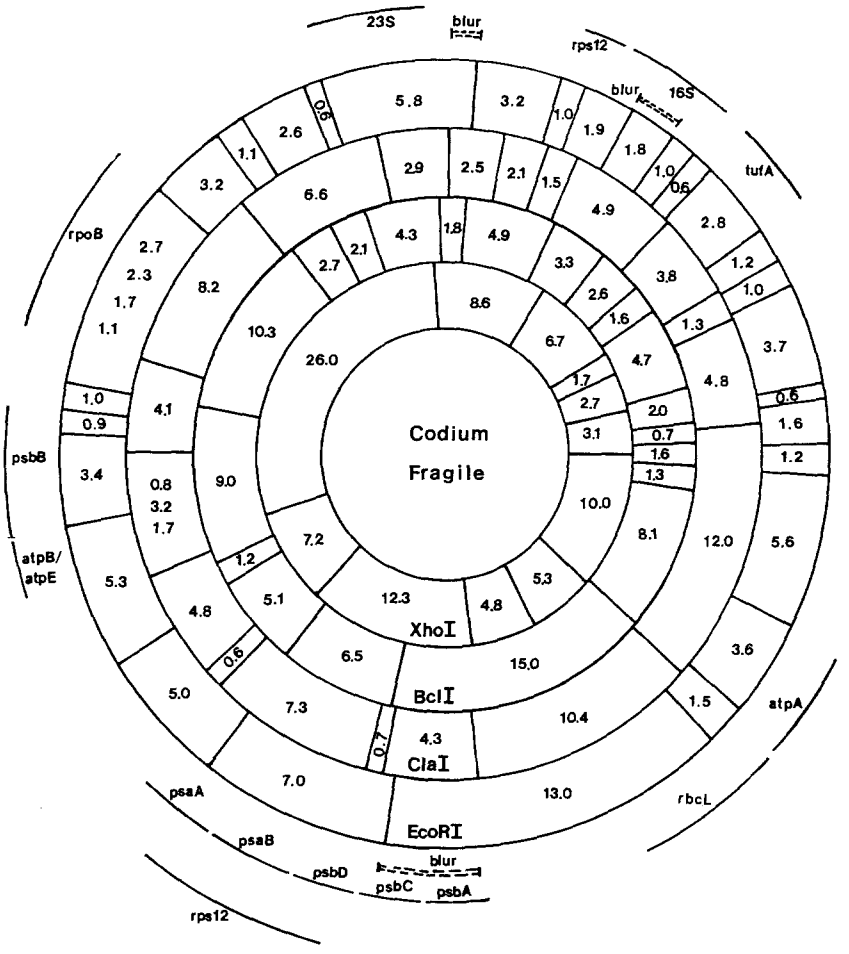

Fig. 2. Physical and gene maps of the *Codium fragile* chloroplast genome. Gene acronyms are as in Table 1. The relative positions of the *atp*B and *atp*E genes have not been determined. The gaps between *psa*A-*psa*B, *psa*B-*psb*D, *psb*D-*psb*C, and *psb*C-*psb*A are approximated but the relative positions are as shown. Sizes are given in kilobases

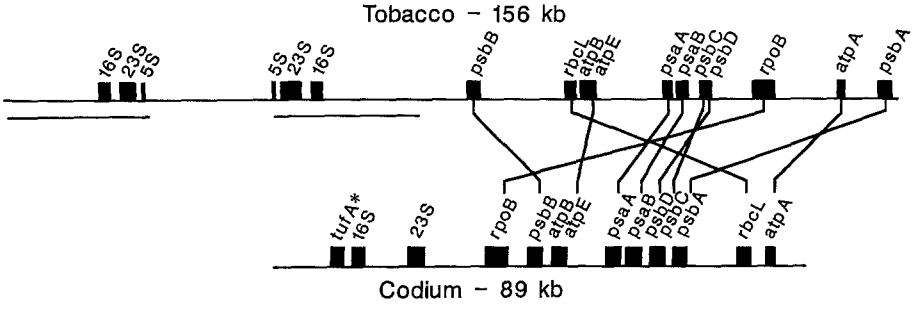

Fig. 3. Comparisons of tobacco and *Codium* chloroplast DNA gene orders. *Codium* gene order is from Fig. 2; tobacco gene order is from Shinozaki et al. (1986). Gene acronyms are as in Table 1. The relative positions of the *atp*B and *atp*E genes have not been determined in *Codium*. **tuf*A is absent from tobacco chloroplast DNA

acteristic that *Codium* and land plant cpDNAs have in common is the conservation of three presumptively cotranscribed pairs of genes – *psb*C-*psb*D, *psa*A-*psa*B, and *atp*B-*atp*E. Although the first two pairs are close in both genomes, their relative orientation is reversed (Fig. 3). The gene order in *Codium* cpDNA does not resemble that of either of the *Chlamydomonas* genomes for which detailed mapping information is available (Lemieux et al. 1985).

It has been demonstrated in tobacco cpDNA that *rps*12 consists of three exons. Exon 1 and exons 2 and 3 are encoded on separate transcripts and a *trans*-splicing event occurs during the formation of mature *rps*12 mRNA (Hildebrand et al. 1988). The 3′*rps*12-5′*rps*7 probe is from tobacco and contains exons 2 and 3 of *rps*12, which are separated by an intron, and 29 bp of *rps*7. This probe mapped to nonadjacent fragments in *Codium* cpDNA (Fig. 2) and hybridized more strongly to the region between the 16S and 23S rRNA genes than to the region on the opposite side of the chromosome which contains the *psa*A, *psa*B, and *psb*D genes. The small amount of *rps*7 sequence present in this probe makes it unlikely that this gene was responsible for hybridization to either of the two regions. This raises the possibility that *trans*-splicing of *rps*12 is also occurring in *Codium*. If this is the case, the hybridization pattern would indicate a different distribution of *rps*12 exons from that found in tobacco. Alternatively, the pattern could be due to gene duplication. Sequence analysis should indicate whether or not *rps*12 is *trans*-spliced in *Codium*.

Hypervariable regions

One unusual characteristic of *Codium* cpDNA is the presence of blurred restriction fragments (Fig. 2). This is especially apparent in the *Cla*I digest. The blurred band at 4.3 kb reflects variation at a single locus since the gel-isolated 15 kb *Bcl*I fragment, the *psb*A probe and the 3′*psb*C probe all hybridize to it (Fig. 1). The size variants appear to extend from 4.1 to 4.5 kb, for a size range of 0.4 kb. These blurred fragments are likely due to size variation since the same blur pattern is seen at a given locus with all restriction enzymes used and the blurring effect is less pronounced in the larger fragments (Fig. 1). This blurring of fragments occurs in two other places and is apparent in the 2.5 kb *Cla*I and 1.8 kb *Eco*RI fragments (Fig. 1). One of these regions maps to the spacer between the 16S and 23S rRNA genes (Fig. 2). The third region is quite close to, or possibly within, the 16S gene (Fig. 2) as the 16S rRNA gene probe hybridizes quite strongly to the 1.8 kb *Eco*RI blurred band (Fig. 1). The range of size variation of each of these variant regions is estimated at 200 bp (Fig. 1).

The presence of a region with size polymorphism has been documented in *Euglena* cpDNA (Schlunegger et al. 1983; Schlunegger and Stutz 1984) and it has been suggested that this region and its immediate vicinity may be involved in the initiation of DNA replication. The total extent of size variation in *Euglena* is 800 bp and is likely due to unequal crossovers within a 55 bp tandem repeat region (Schlunegger and Stutz 1984). Cloning and sequencing of the region between the 16S and 23S rRNA genes, which contains at least one and possibly two of the hypervariable regions in *Codium*, is currently underway and should help determine the cause of the size variation.

Acknowledgements. This study was supported in part by grant BSR-8600167 from NSF to J.R.M., grant GM-25254 from NIH and grant PCM-8201925 from NSF to B.S.D., and grant GM-35087 to J.D.P. Figures were prepared by Nanette Muzzy.

References

Calie PJ, Hughes KW (1987) The consensus land plant chloroplast gene order is present, with two alterations, in the moss *Physcomitrella patens*. Mol Gen Genet 208:335–441

Francis MA, Balint RF, Dudock BS (1987) A novel variety of 4.5S RNA from *Codium fragile* chloroplasts. J Biol Chem 262:1848–1854

Green BR (1976) Covalently closed minicircular DNA associated with *Acetabularia* chloroplasts. Biochim Biophys Acta 447:156–166

Hedberg MF, Huang YS, Hommersand MH (1981) Size of the chloroplast genome in *Codium fragile*. Science 213:445–447

Hildebrand M, Hallick RB, Passavant CW, Bourque DP (1988) Trans-splicing in chloroplasts: The *rps*12 loci of *Nicotiana tabacum*. Proc Natl Acad Sci USA 85:372–376

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

Lemieux B, Lemieux C (1985) Extensive sequence rearrangements in the chloroplast genomes of the green algae *Chlamydomonas eugametos* and *Chlamydomonas reinhardtii*. Curr Genet 10:213–219

Lemieux B, Turmel M, Lemieux C (1985) Chloroplast DNA variation in *Chlamydomonas* and its potential application to the systematics of this genus. Biosystems 18:293–298

Ohyama K, Fukuzawa H, Kochi T, Shirai H, Sano T, Sano S, Umesono K, Shiki Y, Takeuchi M, Chang Z, Aota S, Inokuchi H, Ozeki H (1986) Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. Nature 322:572–574

Padmanabhan U, Green BR (1978) The kinetic complexity of *Acetabularia* chloroplast DNA. Biochim Biophys Acta 521:67–73

Palmer JD (1982) Physical and gene mapping of chloroplast DNA from *Atriplex triangularis* and *Cucumis sativa*. Nucleic Acids Res 10:1593–1605

Palmer JD (1985a) Evolution of chloroplast and mitochondrial DNA in plants and algae. In: MacIntyre RJ (ed) Monographs in evolutionary biology: Molecular evolutionary genetics. Plenum, New York, pp 131–240

Palmer JD (1985b) Comparative organization of chloroplast genomes. Annu Rev Genet 19:325–354

Palmer JD (1986) Conservation of chloroplast genome structure among vascular plants. Curr Genet 11:823–833

Palmer JD, Stein DB (1982) Chloroplast DNA from the fern *Osmunda cinnamomea*: physical organization, gene localization and comparison to angiosperm chloroplast DNA. Curr Genet 5:165–170

Palmer JD, Boynton JE, Gillham NW, Harris EH (1985) Evolution and recombination of the large inverted repeat in *Chlamydomonas* chloroplast DNA. In: Arntzen C, Bogorad L, Bonitz S, Steinback K (eds) Molecular biology of the photosynthetic apparatus. Cold Spring Harbor Press, Cold Spring Harbor, New York, pp 269–278

Rochaix JD (1978) Restriction endonuclease map of the chloroplast DNA of *Chlamydomonas reinhardtii*. J Mol Biol 126:597–617

Scagel RF, Bandoni RJ, Maze JR, Rouse GE, Schofield WB, Stein JR (1982) Nonvascular plants: An evolutionary survey. Wadsworth, Belmont, Calif

Schofield WB (1985) Introduction to bryology, Macmillan, New York

Schlunegger B, Stutz E (1984) The *Euglena gracilis* chloroplast genome: structural features of a DNA region possibly carrying the single origin of DNA replication. Curr Genet 8:629–634

Schlunegger B, Fasnacht M, Stutz E, Koller B, Delius H (1983) Analysis of a polymorphic region of the *Euglena gracilis* chloroplast genome. Biochim Biophys Acta 739:114–121

Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi T, Zaita N, Chunwongse J, Obokata J, Yamaguchi-Shinozaki K, Ohto C, Torazawa K, Meng BY, Sugita M, Deno H, Kamogashira T, Yamada K, Kusuda J, Takaiwa F, Kato A, Tohdoh N, Shimada H, Sugiura M (1986) The complete nucleotide sequence of tobacco chloroplast genome: Its gene organization and expression. EMBO J 5:2043–2049

Whitfield PR, Bottomley W (1983) Organization and structure of chloroplast genes. Annu Rev Plant Physiol 32:279–310

Yamada T (1983) Characterization of inverted repeat sequences and ribosomal RNA genes of chloroplast DNA from *Chlorella ellipsoidea*. Curr Genet 7:481–487

Yamada T, Shimaji M (1986) Peculiar feature of the organization of rRNA genes of the *Chlorella* chloroplast DNA. Nucleic Acids Res 14:3827–3839

Communicated by R.G. Herrmann

Received September 14, 1988