The Plant Cell, Vol. 5, 465–475, April 1993 © 1993 American Society of Plant Physiologists

A High-Resolution Gene Map of the Chloroplast Genome of the Red Alga *Porphyra purpurea*

Michael Reith[1] and Janet Munholland

National Research Council of Canada, Institute for Marine Biosciences, 1411 Oxford Street, Halifax, Nova Scotia, B3H 3Z1 Canada

Extensive DNA sequencing of the chloroplast genome of the red alga *Porphyra purpurea* has resulted in the detection of more than 125 genes. Fifty-eight (approximately 46%) of these genes are not found on the chloroplast genomes of land plants. These include genes encoding 17 photosynthetic proteins, three tRNAs, and nine ribosomal proteins. In addition, nine genes encoding proteins related to biosynthetic functions, six genes encoding proteins involved in gene expression, and at least five genes encoding miscellaneous proteins are among those not known to be located on land plant chloroplast genomes. The increased coding capacity of the *P. purpurea* chloroplast genome, along with other characteristics such as the absence of introns and the conservation of ancestral operons, demonstrate the primitive nature of the *P. purpurea* chloroplast genome. In addition, evidence for a monophyletic origin of chloroplasts is suggested by the identification of two groups of genes that are clustered in chloroplast genomes but not in cyanobacteria.

INTRODUCTION

The complete DNA sequences of three photosynthetic land plant chloroplast genomes have revealed an enormous amount of functional and evolutionary information. The chloroplast genomes of tobacco (Shinozaki et al., 1986), rice (Hiratsuka et al., 1989), and the liverwort *Marchantia polymorpha* (Ohyama et al., 1986) have each been shown to contain 110 to 118 genes. The products of these genes are primarily involved in two processes: gene expresssion (59 to 60 genes) and photosynthesis (29 genes). In addition, 11 genes encoding subunits of NADPH dehydrogenase (Arizmendi et al., 1992) as well as a number of conserved open reading frames (ORFs) are contained on these genomes. The gene content of all land plant chloroplast genomes investigated is surprisingly conserved (for review, see Palmer, 1991).

Algal chloroplast genomes, on the other hand, have not been so extensively characterized. Current information on the chloroplast genomes of chlorophyll *a/b*–containing algae suggest that their gene content is not too different from that of land plants, although a few genes that are absent from land plant chloroplast genomes (e.g., *tufA* in several species [Baldauf et al., 1990], *rpl5* [Christopher and Hallick, 1989], and *ccsA* [Orsat et al., 1992] in *Euglena gracilis*) have been identified. Of greater significance, however, may be the observation of substantial genome rearrangements in green algal chloroplast genomes in the form of the splitting up of ancestral operons in several species of *Chlamydomonas* (e.g., Woessner et al., 1987) and the large number of introns, including introns within introns, in *E. gracilis* (Christopher and Hallick, 1989; Copertino and Hallick, 1991). These observations are indicative of chloroplast genomes that are evolving under more relaxed constraints than land plant chloroplast genomes.

Information about the gene content of chloroplast genomes from chromophyte (chlorophyll *a/c*–containing) and rhodophyte/glaucophyte (chlorophyll *a*/phycobilisome–containing) algae is even more fragmentary. However, the chloroplast genomes of both groups of algae have been shown to contain many genes that are not located on the chloroplast genomes of land plants. Genes localized to the chloroplast genomes of both groups include *rbcS*, *secY*, *dnaK*, *petF*, *acpA*, *ompR*, *ilvB*, *atpD*, *atpG*, and those for several ribosomal proteins (for review, see Douglas, 1992; also Kostrzewa and Zetsche, 1992; Pancic et al., 1992). Additional examples include *hlpA* (Wang and Liu, 1991) and *secA* (Scaramuzzi et al., 1992) in chromophytes, *psaE* (Reith, 1992) and *fabH* (Reith, 1993) in rhodophytes, and *nadA* (Michalowski et al., 1991b) and *crtE* (Michalowski et al., 1991a) in glaucophytes. These observations suggest either of two possibilities: chromophyte and rhodophyte chloroplast genomes contain a different set of genes than land plant chloroplast genomes, or they contain an increased number of genes relative to land plant chloroplast genomes.

As the result of encountering a number of genes not expected to be localized on the chloroplast genome of the red alga *Porphyra purpurea* (Reith and Munholland, 1991, 1993; Reith, 1992, 1993), we have undertaken an extensive characterization of this genome. This alga was originally referred to as *P. umbilicalis*, but a reinvestigation of its taxonomy (C. Bird,

[1] To whom correspondence should be addressed.

J. Munholland, and M. Reith, unpublished results) suggests that *P. purpurea* is more appropriate. In addition to the unusual gene complement, the *P. purpurea* chloroplast genome is shown elsewhere (M. Reith and J. Munholland, manuscript submitted) to have an unusual genome organization in that it contains nonidentical, nontandem, direct rRNA repeats. In this study, we have used DNA sequencing to localize more than 125 genes on the *P. purpurea* chloroplast genome and thus provide evidence that rhodophyte chloroplast genomes contain an increased coding capacity relative to land plant chloroplast genomes.

RESULTS

Gene Mapping

Two distinct approaches were used to locate genes on the chloroplast genome of *P. purpurea.* For a limited number of genes, polymerase chain reaction (PCR) experiments were conducted to generate homologous probes. These experiments relied on redundant oligonucleotide primers based on highly conserved amino acid sequences. The PCR products were directly sequenced to confirm the identity of the probes, and DNA gel blot hybridization and additional PCR experiments were done to place these genes precisely on the map. When appropriate clones had been isolated, further sequencing was done based on the PCR-derived sequence. Genes mapped in this manner include *dnaK*, *ilvB*, *petF*, *petJ*, *cpcB/A*, *cpeB/A*/ORF 301, *apcE/A/B*, *psbA*, *tufA*, *psaA/B*, *trnL*(UAA), and *groEL.* The generation of PCR probes, cloning, and sequencing of the *dnaK* and *ilvB* genes have been presented previously (Reith and Munholland, 1991, 1993).

All other genes have been identified through the random sequencing of EcoRI clones followed by FASTA searches (Pearson and Lipman, 1988) of the Swiss-Prot or GenPept protein sequence data bases. Approximately 65 kb of the *P. purpurea* chloroplast genome have now been sequenced and more than 125 genes have been identified, as shown in Figure 1. The genes identified so far are shown in Table 1. Many of these genes are found on all chloroplast genomes (e.g., *rbcL*, *psbA*, or *psaA/B*). However, 58 of these genes (~46%) are not found on the chloroplast genomes of land plants.

Photosynthetic Genes

Among the genes encoding proteins involved in photosynthesis, we have detected genes for seven photosystem I proteins, 11 photosystem II proteins, four photosynthetic electron transport proteins, both ribulose-1,5-bisphosphate carboxylase (Rubisco) subunits, eight ATPase proteins, and eight phycobilisome polypeptides. Three of the photosystem I protein genes are not located on the chloroplast genomes of land plants: *psaE* (Reith, 1992), *psaF*, and *psaL*, which encode subunits IV, III, and XI, respectively. *psaF* has been found on the cyanelle genome of *Cyanophora paradoxa* (V. L. Stirewalt and D. A. Bryant, unpublished results), whereas *psaL* has been detected on the chloroplast genome of *Cryptomonas* Φ (Douglas, 1992). Three of the four photosynthetic electron transport proteins so far detected on the *P. purpurea* chloroplast genome are also not found on the chloroplast genomes of land plants. Ferredoxin (*petF*) is encoded in the nucleus in land plants but has been found on the chloroplast genomes of *Cyanophora paradoxa* (Neumann-Spallart et al., 1990; Bryant et al., 1991) and *Cryptomonas* (Douglas, 1992). Neither cytochrome c_{553} (*petJ*) nor cytochrome c_{550} (*petK*) is present in land plant chloroplasts. Cytochrome c_{553} is functionally replaced by plastocyanin in land plants, whereas the function of cytochrome c_{550} is unclear. This protein is stoichiometrically associated with purified, oxygen-evolving photosystem II core complexes from the cyanobacterium *Synechococcus vulcanus* (Shen et al., 1992). As has been found for all rhodophyte and chromophyte algae investigated to date, both subunits of Rubisco are encoded on the chloroplast genome of *P. purpurea.* Also, the genes for all but one protein of the ATPase complex (the γ subunit encoded by *atpC*) have been found on the chloroplast genome of *P. purpurea.* The *atpD* and *atpG* genes, which are nuclear in green plants, are also located in the chloroplasts of the diatom *Odontella sinensis* (Pancic et al., 1992), *Cyanidium caldarium* (Kostrzewa and Zetsche, 1992), and *Cyanophora paradoxa* (M. Annarella, V. L. Stirewalt, and D. A. Bryant, unpublished results).

Transcriptional and Translational Genes

In addition to the three rRNA genes on the chloroplast genome of *P. purpurea*, 20 tRNA genes have been identified including three (*trnA*[GGC], *trnL*[GAG], and *trnS*[CGA]) not found in land plant chloroplast genomes. In addition, 21 ribosomal protein genes have been located, nine of which are absent from the chloroplast genomes of land plants. Many of the ribosomal protein genes are arranged in a large operon that is organized much like the S10, *spc*, α, and *str* ribosomal protein operons of *Escherichia coli*, as has already been noted in *Cryptomonas* Φ (Douglas, 1991, 1992). Genes for the four subunits of RNA polymerase as well as several genes encoding proteins involved in translation, DNA replication, or control of transcription have been identified. Genes for both subunits of translation elongation factor T (encoded by *tufA* and *tsf*) are present on the *P. purpurea* chloroplast genome as well as initiation factor β (*infB*). *dnaB*, encoding a protein involved in the strand separation step of DNA replication, has also been detected. Among the more intriguing genes identified are *trsA* and *trsB* (for *t*ranscriptional *r*egulatory *s*ystem), which are similiar to the genes for the membrane kinase (e.g., *envZ*) and DNA binding proteins (e.g., *ompR*), respectively, of the so-called bacterial two-component regulatory systems (Stock et al., 1989). *trsB*

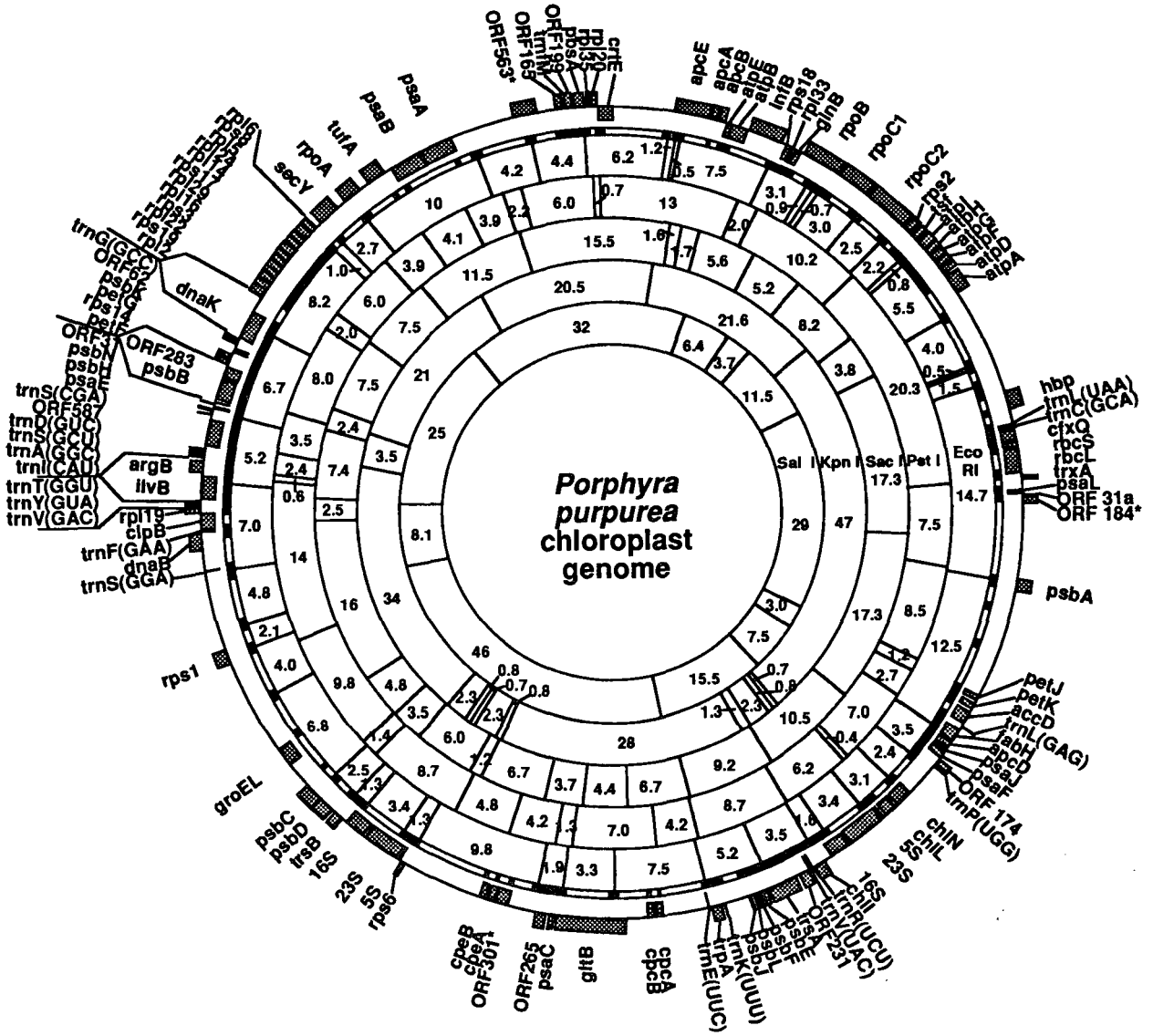

Figure 1. Physical and Gene Map of the *P. purpurea* Chloroplast Genome.

Gene names are as given in Table 1. Genes located on the outside of the outermost circle are transcribed in a clockwise direction, while those on the inside of this circle are transcribed counterclockwise. The narrow circle outside the EcoRI restriction fragment map indicates the regions of the genome for which sequence data are available (black boxes). Information on the restriction enzyme map will be presented elsewhere (M. Reith and J. Munholland, manuscript submitted).

is also present on the cyanelle genome of *Cyanophora paradoxa* (V. L. Stirewalt and D. A. Bryant, unpublished results) and on the chloroplast genomes of *Cryptomonas* Φ (Douglas, 1992), *Cyanidium caldarium*, and *Antithamnion* sp (Kessler et al., 1992). Currently, it is unclear which genes are regulated by this system as well as what external factor activates the regulatory system.

Biosynthesis Genes

Another group of genes encodes proteins involved in several biosynthetic processes. These include amino acid biosynthesis genes (*argB*, *ilvB*, and *trpA*), a carotenoid biosynthesis gene (*crtE*), chlorophyll biosynthesis genes (*chlI*, *chlL*, and *chlN*), and fatty acid biosynthesis genes (*fabH* [Reith, 1993] and

Table 1. Genes Identified on the Chloroplast Genome of *P. purpurea*

Gene	Gene Product
Photosynthesis	
psaA	Photosystem I P700 apoprotein A1
psaB	Photosystem I P700 apoprotein A2
psaC	Photosystem I subunit VII (F_A/F_B containing)
psaE[a]	Photosystem I subunit IV
psaF[a]	Photosystem I subunit III
psaJ	Photosystem I 5-kD protein
psaL[a]	Photosystem I subunit XI
psbA	Photosystem II core 32-kD protein
psbB	Photosystem II CP47 chlorophyll apoprotein
psbC	Photosystem II CP43 chlorophyll apoprotein
psbD	Photosystem II core 34-kD protein
psbE	Photosystem II cytochrome b_{559} α subunit
psbF	Photosystem II cytochrome b_{559} β subunit
psbH	Photosystem II 10-kD protein
psbJ	Photosystem II J protein
psbK	Photosystem II 3.9-kD protein
psbL	Photosystem II L protein
psbN	Photosystem II N protein
petF[a]	Ferredoxin
petG	Cytochrome b_6/f complex subunit V
petJ[a]	Cytochrome c_{553}
petK[a]	Cytochrome c_{550}
rbcL	Rubisco large subunit
rbcS[a]	Rubisco small subunit
atpA	ATPase α subunit
atpB	ATPase β subunit
atpD[a]	ATPase δ subunit
atpE	ATPase ε subunit
atpF	ATPase subunit I
atpG[a]	ATPase subunit II
atpH	ATPase subunit III
atpI	ATPase subunit IV
apcA[a]	Allophycocyanin α subunit
apcB[a]	Allophycocyanin β subunit
apcD[a]	Allophycocyanin B α subunit
apcE[a]	Phycobilisome anchor protein
cpcA[a]	Phycocyanin α subunit
cpcB[a]	Phycocyanin β subunit
cpeA[a]	Phycoerythrin α subunit
cpeB[a]	Phycoerythrin β subunit
Biosynthesis	
argB[a]	Acetylglutamate kinase
ilvB[a]	Acetohydroxyacid synthase (acetolactate synthase)
trpA[a]	Tryptophan synthase α subunit
gltB[a]	Glutamate synthase (GOGAT)
chlI[a]	Chlorophyll biosynthesis
chlL	Chlorophyll biosynthesis (= FrxC)
chlN	Chlorophyll biosynthesis (= ORF 465)
crtE[a]	Carotenoid biosynthesis
pbsA[a]	Heme oxygenase (phycobilin synthesis)
fabH[a]	3-Ketoacyl-ACP synthase III
glnB[a]	Regulator of glutamine synthetase
accD	Acetyl-CoA carboxylase carboxytransferase β subunit (= ZpfA)
Transcription/Translation/Replication	
rpoA	RNA polymerase α subunit
rpoB	RNA polymerase β subunit
rpoC1	RNA polymerase β′ subunit
rpoC2	RNA polymerase β″ subunit
infB[a]	Translation initiation factor β
tufA[a]	Translation elongation factor Tu
tsf[a]	Translation elongation factor Ts
trsA[a]	Transcriptional regulatory protein modulator (EnvZ-like)
trsB[a]	Transcriptional regulatory protein (OmpR-like)
dnaB[a]	DNA replication helicase
Ribosomes	
23S	23S ribosomal RNA
Ribosomes (*continued*)	
16S	16S ribosomal RNA
5S	5S ribosomal RNA
rpl2	Ribosomal protein L2
rpl5[a]	Ribosomal protein L5
rpl6[a]	Ribosomal protein L6
rpl14	Ribosomal protein L14
rpl16	Ribosomal protein L16
rpl19[a]	Ribosomal protein L19
rpl20	Ribosomal protein L20
rpl22	Ribosomal protein L22
rpl24[a]	Ribosomal protein L24
rpl29[a]	Ribosomal protein L29
rpl33	Ribosomal protein L33
rpl35[a]	Ribosomal protein L35
rps1[a]	Ribosomal protein S1
rps2	Ribosomal protein S2
rps3	Ribosomal protein S3
rps6[a]	Ribosomal protein S6
rps8	Ribosomal protein S8
rps14	Ribosomal protein S14
rps17[a]	Ribosomal protein S17
rps18	Ribosomal protein S18
rps19	Ribosomal protein S19
Transfer RNAs	
trnA(GGC)[a]	Alanine tRNA
trnC(GCA)	Cysteine tRNA
trnD(GUC)	Aspartic acid tRNA
trnE(UUC)	Glutamic acid tRNA
trnF(GAA)	Phenylalanine tRNA
trnG(GCC)	Glycine tRNA
trnI(CAU)	Isoleucine tRNA
trnK(UUU)	Lysine tRNA
trnL(GAG)[a]	Leucine tRNA
trnL(UAA)	Leucine tRNA
trnfM(CAU)	Initiator methionine tRNA
trnP(UGG)	Proline tRNA
trnR(UCU)	Arginine tRNA
trnS(CGA)[a]	Serine tRNA
trnS(GCU)	Serine tRNA
trnS(GGA)	Serine tRNA
trnT(GGU)	Threonine tRNA
trnV(UAC)	Valine tRNA
trnV(GAC)	Valine tRNA
trnY(GUA)	Tyrosine tRNA
Miscellaneous Proteins and ORFs[b]	
dnaK[a]	Hsp70-type protein—presumptive chaperonin
groEL[a]	Chaperonin 60
trxA[a]	Thioredoxin
secY[a]	Similar to *E. coli* SecY
clpB[a]	ATP binding subunit of Clp protease
cfxQ[a]	ORF downstream of *rbcS* in *X. flavus* and *A. eutrophus*
hbp	Putative heme binding protein
ORF 31	Similar to land plant ORF 35 near *psbB*
ORF 31a	Similar to land plant ORF 31 near *petG*
ORF 62	Similar to *C. paradoxa* ORF 65, land plant *ORF 62*
ORF 165[a]	Unknown
ORF 174[a]	Similar to *C. paradoxa* ORF 173
ORF 184[c]	Similar to land plant ORF 184
ORF 199[a]	Similar to mouse MER5, *E. histolytica* surface antigen, *H. pylori* antigen, and *S. typhimurium* alkyl hydroperoxide reductase
ORF 231[a]	Unknown
ORF 265[a]	Similar to *C. paradoxa* ORF 243
ORF 283[a]	Similar to *C. paradoxa* ORF 290
ORF 301[c]	Similar to *Cryptomonas* ORF 301, land plant ORF 313/320
ORF 563[a,c]	Similar to *C. eugametos* ORF 563
ORF 587[a]	Similar to several eukaryotic ATP binding proteins: Sec18p, NSF, Cdc48p, VCP, Pas1p

[a] Not found in land plant chloroplast genomes.
[b] Only ORFs longer than 100 amino acids or similar to known genes are included.
[c] The length of these ORFs in *P. purpurea* is unknown because they have not been completely sequenced. The number refers to the size of the homolog in other chloroplast genomes.

accD). *crtE* is also located on the chloroplast genome of *Cyanophora paradoxa* (Michalowski et al., 1991), whereas *chlI* has also been detected on the chloroplast genomes of *Cryptomonas* Φ (Douglas, 1992) and *E. gracilis* (as *ccsA* [Orsat et al., 1992]). *chlL* (*frxC*) and *chlN* are apparently located on the chloroplast genome of all plants except angiosperms, although *chlL* also appears to be absent from the chloroplast genome of the fern *Psilotum nudum* (Suzuki and Bauer, 1992). *accD*, which was previously identified in pea chloroplast DNA as *zpfA* (Sasaki et al., 1989, but see Smith et al., 1991) on the basis of a single zinc finger motif (also as *M. polymorpha* ORF 316 [Ohyama et al., 1986] and tobacco ORF 512 [Shinozaki et al., 1986]), is homologous to *E. coli dedB* (*usg*), which has recently been shown to encode the β subunit of acetyl-coenzyme A carboxylase carboxytransferase (Li et al., 1992). In addition, two proteins involved in nitrogen assimilation, glutamate synthase (GOGAT), encoded by *gltB*, and a protein involved in the regulation of both the activity and transcription of glutamine synthetase (encoded by *glnB*), are encoded on the *P. purpurea* chloroplast genome. The most intriguing of this group of genes is one encoding heme oxygenase (*pbsA* for *p*hyco*b*ilin *s*ynthesis), which was detected by its strong similarity to mammalian heme oxygenases. *gltB*, *glnB*, and *pbsA* have not been previously detected on any chloroplast genome.

Miscellaneous Genes

Finally, several miscellaneous protein-encoding genes and ORFs are found on the *P. purpurea* chloroplast genome. Only ORFs longer than 100 codons have been included in Table 1 or Figure 1, unless they are similar to ORFs known from other chloroplast genomes. Among the identified genes in this group are those for chaperonin proteins (*dnaK* and *groEL*), thioredoxin (*trxA*), and a protein probably involved in protein transport through the thylakoid membrane (*secY*). A homolog of *E. coli clpB* and tomato nuclear genes CD4A and CD4B, which encode the regulatory subunit of the Clp protease (Gottesman et al., 1990), has also been localized on the *P. purpurea* chloroplast genome. Downstream of *rbcS*, we detected a gene (*cfxQ*) that is located in the same position in *Xanthobacter flavus* (Meijer et al., 1991) and *Alcaligenes eutrophus* (Kusian et al., 1992) and that encodes an ATP binding protein of unknown function. A putative heme binding protein gene (*hbp*) (Willey and Gray, 1990) and three small ORFs (ORFs 31, 31a, and 62) similar to land plant ORF 35, ORF 31, and ORF 62, respectively, have also been located. Among the larger ORFs that have been completely sequenced or otherwise identified by homology to known ORFs, two (ORFs 165 and 231) have no known homologs, three (ORFs 174, 265, and 283) are similar to *C. paradoxa* ORFs (V. L. Stirewalt and D. A. Bryant, unpublished results) and are located near similar genes, and one (ORF 563) is homologous to a *Chlamydomonas moewusii* ORF (Richard and Bellemare, 1990). Two other ORFs are each similar to several proteins in the Swiss-Prot and GenPept data bases, but still defy identification. ORF 199 is similar to a mouse erythroleukemia cell line–specific protein (MER5)(Yamamoto et al., 1989), a *Helicobacter pylori* antigen (O'Toole et al., 1991), an *Entamoeba histolytica* surface antigen (Torian et al., 1990), and *Salmonella typhimurium* alkyl hydroperoxide reductase (Tartaglia et al., 1990). ORF 587 is similar to a group of eukaryotic proteins, Pas1p, Sec18p, NSF, Cdc48p, and VCP, that have a conserved domain involved in ATP hydrolysis (Erdmann et al., 1991). Obviously, further work will be necessary to identify the function of the products of these ORFs.

DISCUSSION

Gene Content

The chloroplast genome of *P. purpurea* has been shown to contain a significantly greater number of genes than are found on other well-studied chloroplast genomes. Having identified more than 125 genes that cover ~60% of the genome, we estimate a potential coding capacity for the *P. purpurea* chloroplast genome of 200 to 220 genes. These calculations suggest that at least 75 more genes remain to be identified on the *P. purpurea* chloroplast genome. If rhodophyte chloroplast genomes contain all the photosynthetic and gene expression-related genes found on land plant chloroplast genomes, one would expect to find among these 75 genes at least six more photosynthetic genes, eight more ribosomal protein genes, 15 tRNA genes, and a few miscellaneous genes (e.g., *clpP* and several ORFs). This suggests that there should be at least 40 more genes encoding proteins with novel functions still to be identified on the *P. purpurea* chloroplast genome.

Noticeably absent from Table 1 are genes encoding subunits of the NADPH dehydrogenase complex. Eleven genes for proteins from this complex are located on the chloroplast genomes of land plants (Arizmendi et al., 1992). One would have expected to encounter at least one of these genes during the relatively random sequencing so far carried out on the *P. purpurea* chloroplast genome. However, with the exception of the diatom *Cyclotella meneghiniana*, no NADPH dehydrogenase genes have been detected on any rhodophyte or chromophyte chloroplast genome, including that of *Cyanophora paradoxa* from which DNA sequence is available for more than 80% of the genome (D. A. Bryant and H. J. Bohnert, unpublished results). In the case of *Cyclotella meneghiniana*, probes for *ndhB* and *ndhD* gave hybridization signals, although several other *ndh* genes did not (Bourne et al., 1992). However, because all three positively hybridizing probes (one from *ndhB* and two from *ndhD*) also contained other genes, these results must be interpreted with caution. Until confirming DNA sequence data are available, it appears prudent to assume that NADPH dehydrogenase genes are absent from the rhodophyte and chromophyte chloroplast lineages.

Organization of Red Algal Chloroplast Genomes

A comparison of the order of known genes in the two previously mapped rhodophyte chloroplast genomes, those of *Griffithsia pacifica* and *P. yezoensis* (Shivji et al., 1992), to that of the *P. purpurea* chloroplast genome is shown in Figure 2. Interestingly, the *P. purpurea* gene order is more similar to that of the more complex Floridiophyte, *G. pacifica,* than to that of the congeneric species, *P. yezoensis.* The only differences between the *P. purpurea* and *G. pacifica* gene orders are the placement of the *rpoA* gene (which gives two, possibly erroneous, hybridization signals in *G. pacifica*) and an inversion between the *psbA* and *cpeBA* genes. One rRNA operon appears to have been lost from this inverted region in the *G. pacifica* chloroplast genome. Conversely, only a few genes align when the two *Porphyra* species are compared. These include *psbC/D*, the *psaA/psaB/tufA* cluster, *atpA*, the right rRNA repeat (although it appears to be shifted by several kilobases in *P. yezoensis*), and *cpcBA*. The left rRNA repeat of *P. yezoensis* is inverted relative to that of *P. purpurea*, and there appears to be a large inversion between *apcAB* and *psbA*. Several genes (*psbB*, *rpoA*, and *cpeBA*) are in completely different positions in the two *Porphyra* species. Although these differences may just reflect rearrangement of the *P. yezoensis* chloroplast genome that occurred after its divergence from a common ancestor with *P. purpurea*, further investigation of the *P. yezoensis* chloroplast genome is needed to confirm this interpretation. Additional characterization of other rhodophyte chloroplast genomes will provide a better understanding of the overall pattern of chloroplast genome evolution within this group.

Gene Clustering

The chloroplast genome of *P. purpurea* contains many genes that appear to be organized in operons. Many of these operons,

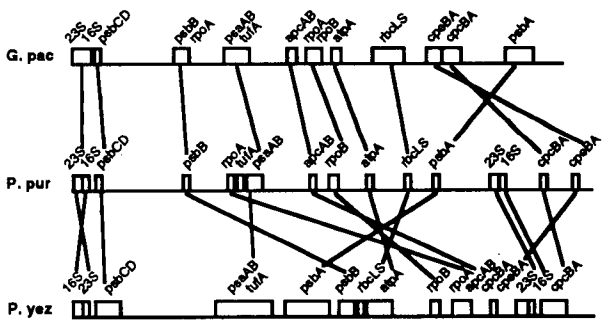

Figure 2. Comparison of Gene Order among the Chloroplast Genomes of *G. pacifica*, *P. purpurea*, and *P. yezoensis.*

The *G. pacifica* (top) and *P. yezoensis* (bottom) maps are redrawn from Shivji et al. (1992). The *P. purpurea* map is shown at center.

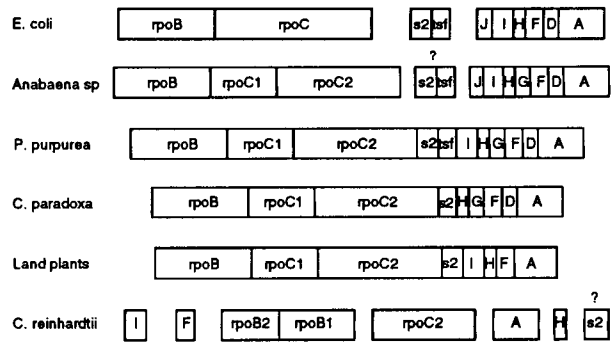

Figure 3. Organization of the *atp/rps2/rpo* Operons of Chloroplasts and Bacteria.

RNA polymerase genes and *tsf* are indicated by the full gene designation, while *rps2* genes are represented by s2 and *atp* genes are denoted by the appropriate capital letter. Genes drawn without any space between them are adjacent in the indicated genome; genes with spaces between them are physically separated by other genes. Genes are drawn proportional to their coding length (excluding introns), while spaces between genes are not proportional. The question marks indicate genes that have not yet been mapped but are assumed to be unlinked to known genes. Data for these operons are from the following: *E. coli*, Falk and Walker (1988); *Anabaena* sp strain PCC 7120, McCarn et al. (1989) and Bergsland and Haselkorn (1991); *P. purpurea*, this work; *C. paradoxa*, M. Annarella, V. L. Stirewalt, and D. A. Bryant, unpublished results; land plants, Shinozaki et al. (1986), Ohyama et al. (1986), and Hiratsuka et al. (1989); *C. reinhardtii*, Woessner et al. (1987) and Fong and Surzycki (1992).

such as *psbD/C*, *psaA/B*, *psbE/F/L/J*, *atpB/E*, and *rpoB/C1/C2*, are conserved in both cyanobacteria and land plant chloroplast genomes. Other *P. purpurea* operons, such as *rbcL/S*, *atpI/H/G/F/D/A*, *cpcB/A*, *cpeB/A*, and *apcE/A/B*, are identical to those found in cyanobacteria, but have been completely lost or reduced due to gene transfer to the nucleus in land plants. More significant with regard to chloroplast evolution may be those gene arrangements in which the genes are widely separated in cyanobacteria but are grouped together in several chloroplast genomes. One example of this is the conserved arrangement of *psbB*, ORF 31, *psbN*, and *psbH* that is present in the chloroplast genomes of land plants, *P. purpurea*, and *C. paradoxa* (V. L. Stirewalt and D. A. Bryant, unpublished results). In cyanobacteria, *psbB* is separated from *psbH* and *psbN* (Vermaas et al., 1987; Lang and Haselkorn, 1989; Mayes and Barber, 1991), suggesting that this operon was formed during the evolution of the ancestral chloroplast.

A second example of a gene arrangement that appears to have been assembled after endosymbiosis is the *rpo/atp* cluster shown in Figure 3. In land plant chloroplasts, the *rpoB/C1/C2* genes are followed by *rps2* and then the reduced *atp* operon containing *atpI/H/F/A*. In cyanobacteria, the *rpo* genes form one operon and the *atp* genes another, whereas *rps2*, which has not yet been characterized in cyanobacteria, is presumably

found at yet a third location. In *E. coli*, the *rpo* and *atp* genes are arranged as separate operons, whereas *rps2* is found in an operon with *tsf*, the gene encoding elongation factor Ts (An et al., 1981). In the chloroplast genome of *P. purpurea*, all three complete operons, *rpoB/C1/C2*, *rps2/tsf*, and *atpI/H/G/F/D/A*, are located together in an arrangement that may be ancestral for chloroplast genomes. Hybridization studies indicate that the *atp* and *rpo* clusters are also closely linked in *G. pacifica* (Shivji et al., 1992), and DNA sequence data indicate that *tsf* is upstream of *atpI* in another rhodophyte, *Antithamnion* sp (Kostrzewa and Zetsche, 1992). To generate the organization of these genes found in land plants from that of *P. purpurea* would require the transfer of *tsf*, *atpG*, and *atpD* to the nucleus. Similarly, the present arrangement in *C. paradoxa* requires the transfer of the *tsf* and *atpI* genes to the nucleus (M. Annarella, V. L. Stirewalt, and D. A. Bryant, unpublished results). In the *Chlamydomonas reinhardtii* chloroplast genome, *rpoC1* is apparently absent, whereas there are two *rpoB*-like genes (Fong and Surzycki, 1992). In addition, the *rpo* and *atp* clusters have been shuffled to scatter these genes around the genome. Such an arrangement makes it difficult to establish what genes from this cluster are still present on the *C. reinhardtii* chloroplast genome, and to understand the events that resulted in the present organization.

Due to a lack of data, the arrangement of the *rpo*, *rps2/tsf*, and *atp* genes in chromophytes is still somewhat unclear. Data from *Cryptomonas* Φ (Douglas, 1992) and *Olisthodiscus luteus* (Shivji et al., 1992) suggest that these genes may be linked in these organisms. However, in the diatom *Odontella sinensis*, where the complete *atpI/H/G/F/D/A* operon has been detected, DNA sequence upstream from *atpI* does not reveal *tsf*, *rps2*, or *rpoC2* (Pancic et al., 1992). Based on hybridization data, the *atp* and *rpo* operons map at least 10 kb apart in both *Vaucheria bursata* (Linne von Berg and Kowallik, 1992) and *Cyclotella meneghiniana* (Bourne et al., 1992). These data indicate that the clustering of the *rpo*, *rps2/tsf*, and *atp* genes may occur in some chromophytes, but not others.

Primitive Characteristics

During the evolution of chloroplasts, the cyanobacterial endosymbiont appears to have reduced its genome through the transfer of genes to the host and the loss of nonessential genes. Presumably, the ancestral chloroplast(s) that resulted from this process maintained many of the characteristics of the prokaryotic endosymbiont(s) and had not yet established the characteristics of highly evolved chloroplasts (e.g., enlarged, identical repeats, and a reduced number of tRNA genes [assuming the endosymbiont started with a complete set]). We have demonstrated elsewhere that the chloroplast genome of *P. purpurea* appears to be more cyanobacterium-like than those of land plants in having short, nonidentical rRNA repeats (M. Reith and J. Munholland, manuscript submitted). Further evidence from the gene mapping studies presented here confirms the primitive nature of the *P. purpurea* chloroplast genome. First, the gene-coding capacity of the *P. purpurea* genome is greater than that of other chloroplast genomes. A similar number of genes as is found in land plant chloroplast genomes has already been identified in the *P. purpurea* genome even though ~40% of the genome remains to be investigated. Among the *P. purpurea* genes that are not found on land plant chloroplast genomes are three tRNA genes. This observation suggests the possible presence of a complete, or nearly complete, set of tRNA genes, a characteristic one might expect to find in the ancestral chloroplast(s).

Another primitive characteristic of the *P. purpurea* chloroplast genome is the absence of any introns in the 80 genes that have been completely sequenced. This situation is similar to that of eubacteria where introns are rare and have only been recently detected in *trnL*(UAA) from several cyanobacteria (Kuhsel et al., 1990) and in two different tRNA genes from two purple bacteria (Reinhold-Hurek and Shub, 1992). Interestingly, introns are absent from *trnL*(UAA) in *P. purpurea*, as well as in several other chloroplast genomes (Kuhsel et al., 1990), suggesting that there may have been multiple losses (or gains?) of these introns in the course of chloroplast evolution. A further primitive characteristic of the *P. purpurea* chloroplast genome is the presence of genes encoding a transcriptional regulatory system. This observation suggests that the *P. purpurea* chloroplast has maintained control over the expression of some of its genes, unlike the situation in land plants where all chloroplast regulatory proteins appear to be encoded in the nucleus. In addition, the *P. purpurea* chloroplast genome has maintained more cyanobacterial operons than any other known chloroplast genome. These characteristics make a compelling argument for the primitive nature of the *P. purpurea* chloroplast genome and its similarity to cyanobacterial genomes.

A Monophyletic Origin of Plastids

Although the increased coding capacity of rhodophyte and chromophyte chloroplast genomes has been interpreted as evidence for a polyphyletic origin of chloroplasts (Kostrzewa and Zetsche, 1992), this and other primitive characteristics do not distinguish whether chloroplasts evolved in a monophyletic (Cavalier-Smith, 1982) or polyphyletic (Whatley and Whatley, 1981) fashion (for review, see Gray, 1991). To support a polyphyletic origin of chloroplasts, it is necessary to identify characteristics common to a chloroplast and its presumed prokaryotic ancestor, but different from other such pairs. On the other hand, a monophyletic origin of chloroplasts requires the identification of characteristics that are shared among all three types of chloroplasts, but that differ in cyanobacteria. To date, only a single molecular characteristic supporting a polyphyletic origin of chloroplasts has been identified: the presence, in chlorophytes and *Prochlorothrix hollandica*, of a seven–amino acid deletion at the C terminus of the *psbA* gene product (Morden and Golden, 1989). However, phylogenetic analyses of *psbA* amino acid sequences do not support a close

relationship between *P. hollandica* and chlorophyte chloroplasts and thus question the validity of this character (Gray, 1989). On the other hand, the clustering of the *rpo/rps2/atp* genes and the *psbB/N/H* genes in chloroplasts, but not cyanobacteria, are two examples that support a monophyletic origin of chloroplasts. Although the arrangement of both groups of genes in chromophyte chloroplasts requires further investigation, these examples provide a strong link between rhodophyte and chlorophyte chloroplasts. Other data establish a close relationship between rhodophyte and chromophyte chloroplasts. These include phylogenetic analyses of genes located on chloroplast genomes (e.g., Valentin and Zetsche, 1990; Douglas and Turner, 1991) and the close relationship between red algae and the nucleomorph of *Cryptomonas* Φ as detected by the analysis of 18S rRNA sequences (Douglas et al., 1991). Taken together, these observations begin to make a case for a monophyletic origin of chloroplasts.

Several recent investigations provide additional support for a monophyletic origin of plastids. Molecular phylogenetic analyses (Witt and Stackebrandt, 1988; Palenik and Haselkorn, 1992; Urbach et al., 1992) investigating the relationship of prochlorophytes and *Heliobacterium chlorum*, the proposed ancestors of chlorophyte and chromophyte chloroplasts, respectively (Whatley and Whatley, 1981; Margulis and Obar, 1985), to cyanobacteria and chloroplasts have failed to demonstrate the relationships expected for a polyphyletic origin of plastids. In addition, these studies indicate that the prochlorophytes themselves do not form a lineage distinct from the cyanobacteria. This observation has led Bryant (1992) to suggest that the prokaryotic ancestor of chloroplasts utilized, under different environmental conditions, both phycobilisomes and chlorophyll *a/b* light–harvesting antennae, and that one or the other system was subsequently lost during chloroplast evolution.

For the existing data to be consistent with a monophyletic origin of chloroplast evolution, at least two difficulties must first be addressed. The apparent absence of NADPH dehydrogenase genes in the rhodophyte/chromophyte lineage can be resolved if one assumes that these genes were present in the ancestral chloroplast but were transferred to the nucleus or lost early in the evolution of the rhodophyte/chromophyte lineage. More difficult to explain is the higher degree of similarity of rhodophyte/chromophyte Rubisco subunits to those of chemolithotrophic β-purple bacteria such as *A. eutrophus* than to those of cyanobacteria (for review, see Martin et al., 1992). This situation seems to require a lateral transfer of genes into the chloroplast genome of the rhodophyte/chromophyte lineage not long after its separation from the chlorophyte lineage, but after the separation of the branch leading to the *C. paradoxa* chloroplast (because the latter has cyanobacterium-like Rubisco subunits) (Martin et al., 1992; Douglas, 1992). Alternatively, both types of Rubiscos might have been present in the ancestral chloroplast genome with the differential loss of one or the other Rubisco type occurring after the establishment of each lineage (Martin et al., 1992). Reconstructing the establishment of the different Rubisco types in the different chloroplast lineages is not likely to be easily determined.

Assuming a monophyletic origin of chloroplasts, the primitive nature of the *P. purpurea* chloroplast genome suggests that further analysis of rhodophyte chloroplast genomes should provide key information on both the characteristics of the ancestral chloroplast and the process of chloroplast evolution. Particularly interesting aspects of rhodophyte chloroplast genomes that are still to be determined include the extent to which photosynthetic and gene expression-related genes have been maintained on these genomes and the complete spectrum of metabolic functions encoded. In addition, further analyses of chlorophyte and chromophyte algal chloroplast genomes are required to understand the evolutionary relationships between all three groups and to substantiate the monophyletic origin of chloroplasts.

METHODS

Methods for DNA purification, cloning, DNA gel blot hybridization, polymerase chain reaction (PCR), and DNA sequencing were as described previously (Reith and Munholland, 1991, 1993). Direct sequencing of PCR products was performed according to the method of Bachmann et al. (1990). Data bank searches and similarity analysis were done with the FASTA software package (Pearson and Lipman, 1988). Only genes, including open reading frames (ORFs), that showed high FASTA (>100) and RDF2 (>10) scores in comparisons to known genes are identified in Figure 1 and Table 1.

ACKNOWLEDGMENTS

We thank Dr. Don Bryant for pointing out the separation of the *psbB* and *psbN/H* genes in cyanobacteria and for communication of results prior to publication. We would also like to thank Adrienne Seel, Christine Crossman, Hui Xu, Dr. Rama K. Singh, and Dr. Paul X.-Q. Liu for contributions to the DNA sequencing and Drs. Don Bryant, Susan Douglas, Michael Gray, Ron MacKay, and Mark Ragan for comments on the manuscript. This is National Research Council of Canada publication number 34840.

Received December 7, 1992; accepted February 22, 1993.

REFERENCES

An, G., Bendiak, D.S., Mamelak, L.A., and Friesen, J.D. (1981). Organization and nucleotide sequence of a new ribosomal operon in *Escherichia coli* containing the genes for ribosomal protein S2 and elongation factor T_s. Nucl. Acids Res. **9,** 4163–4172.

Arizmendi, J.M., Runswick, M.J., Skehel, J.M., and Walker, J.E. (1992). NADH:ubiquinone oxidoreductase from bovine heart

mitochondria. A fourth nuclear encoded subunit with a homologue encoded in chloroplast genomes. FEBS Lett. **301,** 237–242.

Bachmann, B., Lüke, W., and Hunsmann, G. (1990). Improvement of PCR amplified DNA sequencing with the aid of detergents. Nucl. Acids Res. **18,** 1309.

Baldauf, S.L., Manhart, J.R., and Palmer, J.D. (1990). Different fates of the chloroplast *tufA* gene following its transfer to the nucleus in green algae. Proc. Natl. Acad. Sci. USA **87,** 5317–5321.

Bergsland, K.J., and Haselkorn, R. (1991). Evolutionary relationships among eubacteria, cyanobacteria and chloroplasts: Evidence from the *rpoC1* gene of *Anabaena* sp. strain PCC 7120. J. Bacteriol. **173,** 3446–3455.

Bourne, C.M., Palmer, J.D., and Stoermer, E.F. (1992). Organization of the chloroplast genome of the freshwater centric diatom *Cyclotella meneghiniana*. J. Phycol. **28,** 347–355.

Bryant, D.A. (1992). Puzzles of chloroplast ancestry. Curr. Biol. **2,** 240–242.

Bryant, D.A., Schluchter, W.M., and Stirewalt, V.L. (1991). Ferredoxin and ribosomal protein S10 are encoded on the cyanelle genome of *Cyanophora paradoxa*. Gene **98,** 169–175.

Cavalier-Smith, T. (1982). The origins of plastids. Biol. J. Linn. Soc. **17,** 289–306.

Christopher, D.A., and Hallick, R.B. (1989). *Euglena gracilis* chloroplast ribosomal protein operon: A new chloroplast gene for ribosomal protein L5 and description of a novel organelle intron category designated group III. Nucl. Acids Res. **17,** 7591–7608.

Copertino, D.W., and Hallick, R.B. (1991). Group II twintron: An intron within an intron in a chloroplast cytochrome b-559 gene. EMBO J. **10,** 433–442.

Douglas, S.E. (1991). Unusual organization of a ribosomal protein operon in the plastid genome of *Cryptomonas* Φ: Evolutionary considerations. Curr. Genet. **19,** 289–294.

Douglas, S.E. (1992). Eukaryote-eukaryote endosymbioses: Insights from studies of a cryptomonad alga. BioSystems, **28,** 57–68.

Douglas, S., and Turner, S. (1991). Molecular evidence for the origin of plastids from a cyanobacterium-like ancestor. J. Mol. Evol. **33,** 267–273.

Douglas, S.E., Murphy, C.A., Spencer, D.A., and Gray, M.W. (1991). Cryptomonad algae are evolutionary chimaeras of two phylogenetically distinct unicellular eukaryotes. Nature **350,** 148–151.

Erdmann, R., Wiebel, F.F., Flessau, A., Rytka, J., Beyer, A., Fröhlich, K.-U., and Kunau, W.-H. (1991). *PAS1*, a yeast gene required for peroxisome biogenesis, encodes a member of a novel family of putative ATPases. Cell **64,** 499–510.

Falk, G., and Walker, J.E. (1988). DNA sequence of a gene cluster coding for subunits of the F_0 membrane sector of ATP synthase in *Rhodospirillum rubrum*. Biochem. J. **254,** 109–122.

Fong, S.E., and Surzycki, S.J. (1992). Chloroplast RNA polymerase genes of *Chlamydomonas reinhardtii* exhibit an unusual structure and arrangement. Curr. Genet. **21,** 485–497.

Gottesman, S., Squires, C., Pichersky, E., Carrington, M., Hobbs, M., Mattick, J.S., Dalrymple, B., Kuramitsu, H., Shiroza, T., Foster, T., Clark, W.P., Ross, B., Squires, C.L., and Maurizi, M.R. (1990). Conservation of the regulatory subunit for the Clp ATP-dependent protease in prokaryotes and eukaryotes. Proc. Natl. Acad. Sci. USA **87,** 3513–3517.

Gray, M.W. (1989). The evolutionary origins of organelles. Trends Genet. **5,** 294–299.

Gray, M.W. (1991). Origin and evolution of plastid genomes and genes. In Cell Culture and Somatic Cell Genetics of Plants, Vol. 7A, The Molecular Biology of Plastids, L. Bogorad and I.K. Vasil, eds (San Diego: Academic Press), pp. 303–330.

Gray, M.W., and Doolittle, W.F. (1982). Has the endosymbiont hypothesis been proven? Microbiol. Rev. **46,** 1–42.

Hiratsuka, J., Shimada, H., Whittier, R., Ishibashi, T., Sakamoto, M., Mori, M., Kondo, C., Honji, Y., Sun, C., Meng, B., Li, Y., Kanno, A., Nishizawa, Y., Harai, A., Shinozaki, K., and Sugiura, M. (1989). The complete sequence of the rice (*Oryza sativa*) chloroplast genome: Intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. Mol. Gen. Genet. **217,** 185–194.

Kessler, U., Maid, U., and Zetsche, K. (1992). An equivalent to bacterial *ompR* genes is encoded on the plastid genome of red algae. Plant Mol. Biol. **18,** 777–780.

Kostrzewa, M., and Zetsche, K. (1992). Large ATP synthase operon of the red alga *Antithamnion* sp. resembles the corresponding operon in cyanobacteria. J. Mol. Biol. **227,** 961–970.

Kuhsel, M.G., Strickland, R., and Palmer, J.D. (1990). An ancient group I intron is shared by eubacteria and chloroplasts. Science **250,** 1570–1573.

Kusian, B., Yoo, J.-G., Bednarski, R., and Bowien, B. (1992). The Calvin cycle enzyme pentose-5-phosphate 3-epimerase is encoded within the *cfx* operons of the chemoautotroph *Alcaligenes eutrophus*. J. Bacteriol. **174,** 7337–7344.

Lang, J.D., and Haselkorn, R. (1989). Isolation, sequence and transcription of the gene encoding the photosystem II chlorophyll binding protein, CP-47, in the cyanobacterium *Anabaena* 7120. Plant Mol. Biol. **13,** 441–457.

Li, S.-J., Rock, C.O., and Cronan, J.E., Jr. (1992). The *dedB* (*usg*) open reading frame of *Escherichia coli* encodes a subunit of acetyl-coenzyme A carboxylase. J. Bacteriol. **174,** 5755–5757.

Linne von Berg, K.-H., and Kowallik, K.V. (1992). Structural organization of the chloroplast genome of the chromophytic alga *Vaucheria bursata*. Plant Mol. Biol. **18,** 83–95.

Maid, U., and Zetsche, K. (1992). A 16-kb small single-copy region separates the plastid DNA inverted repeat of the unicellular red alga *Cyanidium caldarium*: Physical mapping of the IR-flanking regions and nucleotide sequences of the *psbD-psbC*, *rps16*, 5S rRNA and *rpl21* genes. Plant Mol. Biol. **19,** 1001–1010.

Margulis, L., and Obar, R. (1985) *Heliobacterium* and the origin of chrysoplasts. BioSystems **17,** 317–325.

Martin, W., Somerville, C.C., and Loiseaux-de Goër, S. (1992). Molecular phylogenies of plastid origins and algal evolution. J. Mol. Evol. **35,** 385–404.

Mayes, S.R., and Barber, J. (1991). Primary structure of the *psbN-psbH-petC-petA* gene cluster of the cyanobacterium *Synechocystis* PCC 6803. Plant Mol. Biol. **17,** 289–293.

McCarn, D.F., Whitaker, R.A., Alam, J., Vrba, J.M., and Curtis, S.E. (1989). Genes encoding the alpha, gamma, delta and four F_0 subunits of ATP synthase constitute an operon in the cyanobacterium *Anabaena* sp. strain PCC 7120. J. Bacteriol. **170,** 3448–3458.

Meijer, W.G., Arnberg, A.C., Enequist, H.G., Terpstra, P., Lidstrom, M.E., and Dijkhuizen, L. (1991). Identification and organization of

carbon dioxide fixation genes in *Xanthobacter flavus* H4-14. Mol. Gen. Genet. **225,** 320–330.

Michalowski, C.B., Löffelhardt, W., and Bohnert, H.J. (1991a). An ORF323 with homology to *crtE,* specifying prephytoene pyrophosphate dehydrogenase, is encoded by cyanelle DNA in the eukaryotic alga *Cyanophora paradoxa.* J. Biol. Chem. **266,** 11866–11870.

Michalowski, C.B., Flachmann, R., Löffelhardt, W., and Bohnert, H.J. (1991b). Gene *nadA,* encoding quinolinate synthetase, is located on the cyanelle DNA from *Cyanophora paradoxa.* Plant Physiol. **95,** 329–330.

Morden, C.W., and Golden, S.S. (1989). *psbA* genes indicate common ancestry of prochlorophytes and chloroplasts. Nature **337,** 382–385.

Neumann-Spallart, C., Brandtner, M., Kraus, M., Jakowitsch, J., Bayer, M.G., Maier, T.L., Schenk, H.E.A., and Löffelhardt, W. (1990). The *petFI* gene encoding ferredoxin I is located close to the *str* operon on the cyanelle genome of *Cyanophora paradoxa.* FEBS Lett. **268,** 55–58.

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

Orsat, B., Monfort, A., Chatellard, P., and Stutz, E. (1992). Mapping and sequencing of an actively transcribed *Euglena gracilis* chloroplast gene (*ccsA*) homologous to the *Arabidopsis thaliana* nuclear gene *cs(ch-42).* FEBS Lett. **303,** 181–184.

O'Toole, P.W., Logan, S.M., Kostrzynska, M., Wadström, T., and Trust, T.J. (1991). Isolation and biochemical and molecular analyses of a species-specific protein antigen from the gastric pathogen *Helicobacter pylori.* J. Bacteriol. **173,** 505–513.

Palenik, B.P., and Haselkorn, R. (1992). Multiple evolutionary origins of prochlorophytes, the chlorophyll *b*–containing prokaryotes. Nature **355,** 265–267.

Palmer, J.D. (1991). Plastid chromosomes: Structure and evolution. In Cell Culture and Somatic Cell Genetics of Plants, Vol. 7A, The Molecular Biology of Plastids, L. Bogorad and I.K. Vasil, eds (San Diego: Academic Press), pp. 5–53.

Pancic, P.G., Strotmann, H., and Kowallik, K.V. (1992). Chloroplast ATPase genes in the diatom *Odontella sinensis* reflect cyanobacterial characters in structure and arrangement. J. Mol. Biol. **224,** 529–536.

Pearson, W.R., and Lipman, D.J. (1988). Improved tools for biological sequence analysis. Proc. Natl. Acad. Sci. USA **85,** 2444–2448.

Reinhold-Hurek, B., and Shub, D.A. (1992). Self-splicing introns in tRNA genes of widely divergent bacteria. Nature **357,** 173–176.

Reith, M. (1992). *psaE* and *trnS*(CGA) are encoded on the plastid genome of the red alga *Porphyra umbilicalis.* Plant Mol. Biol. **18,** 773–775.

Reith, M. (1993). A β-ketoacyl-acyl carrier protein synthase III gene (*fabH*) is encoded on the chloroplast genome of the red alga *Porphyra umbilicalis.* Plant Mol. Biol. **21,** 185–189.

Reith, M., and Munholland, J. (1991). An Hsp70 homolog is encoded on the plastid genome of the red alga *Porphyra umbilicalis.* FEBS Lett. **294,** 116–120.

Reith, M., and Munholland, J. (1993). Two amino-acid biosynthetic genes are encoded on the plastid genome of the red alga *Porphyra umbilicalis.* Curr. Genet. **23,** 59–65.

Richard, M., and Bellemare, G. (1990). Nucleotide sequence of *Chlamydomonas moewusii* chloroplastic tRNA-Thr. Nucl. Acids Res. **18,** 3061.

Sasaki, Y., Nagan, Y., Morioka, S., Ishikawa, H., and Matsuno, R. (1989). A chloroplast gene encoding a protein with one zinc finger. Nucl. Acids Res. **17,** 6217–6227.

Scaramuzzi, C.D., Hiller, R.G., and Stokes, H.W. (1992). Identification of a chloroplast-encoded *secA* gene homologue in a chromophytic alga: Possible role in chloroplast protein translocation. Curr. Genet. **22,** 421–427.

Shen, J.-R., Ikeuchi, M., and Inoue, Y. (1992). Stoichiometric association of extrinsic cytochrome c_{550} and 12 kDa protein with a highly purified oxygen-evolving photosystem II core complex from *Synechococcus vulcanus.* FEBS Lett. **301,** 145–149.

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, B.Y., Sugita, M., Deno, H., Kamogashira, T., Yamada, K., Kusuda, J., Takaiwa, F., Kato, A., Tohdoh, N., Shimada, H., and Suguira, M. (1986). The complete nucleotide sequence of the tobacco chloroplast genome: Its gene organization and expression. EMBO J. **5,** 2043–2049.

Shivji, M.S., Li, N., and Cattolico, R.A. (1992). Structure and organization of rhodophyte and chromophyte plastid genomes: Implications for the ancestry of plastids. Mol. Gen. Genet. **232,** 65–73.

Smith, A.G., Wilson, R.M., Kaethner, T.M., Willey, D.L., and Gray, J.C. (1991). Pea chloroplast genes encoding a 4-kDa polypeptide of photosystem I and a putative enzyme of C_1 metabolism. Curr. Genet. **19,** 403–410.

Stock, J.B., Ninfa, A.J., and Stock, A.M. (1989). Protein phosphorylation and regulation of adaptive responses in bacteria. Microbiol. Rev. **53,** 450–490.

Suzuki, J.Y., and Bauer, C.E. (1992). Light-independent chlorophyll biosynthesis: Involvement of the chloroplast gene *chlL* (*frxC*). Plant Cell **4,** 929–940.

Tartaglia, L.A., Storz, G., Brodsky, M.H., Lai, A., and Ames, B.N. (1990). Alkyl hydroperoxide reductase from *Salmonella typhimurium.* Sequence and homology to thioredoxin reductase and other flavoprotein disulfide oxidoreductases. J. Biol. Chem. **265,** 10535–10540.

Torian, B.E., Flores, B.M., Stroeher, V.L., Hagen, F.S., and Stamm, W.E. (1990). cDNA sequence analysis of a 29-kDa cysteine-rich surface antigen of pathogenic *Entamoeba histolytica.* Proc. Natl. Acad. Sci. USA **87,** 6358–6362.

Urbach, E., Robertson, D., and Chisholm, S.W. (1992). Multiple evolutionary origins of prochlorophytes within the cyanobacterial radiation. Nature **355,** 267–269.

Valentin, K., and Zetsche, K. (1990). Rubisco genes indicate a close phylogenetic relation between the plastids of Chromophyta and Rhodophyta. Plant Mol. Biol. **15,** 575–584.

Vermaas, W.F., Williams, J.G., and Arntzen, C.J. (1987). Sequencing and modification of *psbB,* the gene encoding the CP-47 protein of photosystem II, in the cyanobacterium *Synechocystis* 6803. Plant Mol. Biol. **8,** 317–326.

Wang, S., and Liu, X.-Q. (1991). The plastid genome of *Cryptomonas* Φ encodes an Hsp70-like protein, a histone-like protein, and an acyl carrier protein. Proc. Natl. Acad. Sci. USA **88,** 10783–10787.

Whatley, J.M., and Whatley, F.R. (1981). Chloroplast evolution. New Phytol. **87,** 233–247.

Willey, D.L., and Gray, J.C. (1990). An open reading frame encoding a putative haem-binding polypeptide is cotranscribed with the pea chloroplast gene for apocytochrome *f*. Plant Mol. Biol. **15,** 347–356.

Witt, D., and Stackebrandt, E. (1988). Disproving the hypothesis of a common ancestry for the *Ochromonas danica* chrysoplast and *Heliobacterium chlorum*. Arch. Microbiol. **150,** 244–248.

Woessner, J.P., Gillham, N.W., and Boynton, J.E. (1987). Chloroplast genes encoding subunits of the H^+-ATPase complex of *Chlamydomonas reinhardtii* are rearranged compared to higher plants. Sequence of the *atpE* gene and location of *atpF* and *atpI* genes. Plant Mol. Biol. **8,** 151–158.

Yamamoto, T., Matsui, Y., Natori, S., and Obinata, M. (1989). Cloning of a housekeeping-type gene (*MER5*) preferentially expressed in murine erythroleukemia cells. Gene **80,** 337–343.