J Mol Evol (2000) 51:382–390
DOI: 10.1007/s002390010101

JOURNAL OF MOLECULAR EVOLUTION

The Structure and Gene Repertoire of an Ancient Red Algal Plastid Genome

Gernot Glöckner,[1] André Rosenthal,[1] Klaus Valentin[2]

[1] IMB Jena, Dept. of Genome Analysis, Beutenbergstr. 11, 07745 Jena Germany
[2] Institut für Pflanzenphysiologie, Justus-Liebig-Universität Gieβen, Heinrich Buff Ring 34-36, Gieβen, Germany

Received: 3 June 2000 / Accepted: 18 July 2000

Abstract. Photosynthetic eukaryotes can, according to features of their chloroplasts, be divided into two major groups: the red and the green lineage of plastid evolution. To extend the knowledge about the evolution of the red lineage we have sequenced and analyzed the chloroplast genome (cp-genome) of *Cyanidium caldarium* RK1, a unicellular red alga (AF022186). The analysis revealed that this genome shows several unusual structural features, such as a hypothetical hairpin structure in a gene-free region and absence of large repeat units. We provide evidence that this structural organization of the cp-genome of *C. caldarium* may be that of the most ancient cp-genome so far described. We also compared the cp-genome of *C. caldarium* to the other known cp-genomes of the red lineage. The cp-genome of *C. caldarium* cannot be readily aligned with that of *Porphyra purpurea,* a multicellular red alga, or *Guillardia theta* due to a displacement of a region of the cp-genome. The phylogenetic tree reveals that the secondary endosymbiosis, through which *G. theta* evolved, took place after the separation of the ancestors of *C. caldarium* and *P. purpurea.*

We found several genes unique to the cp-genome of *C. caldarium.* Five of them seem to be involved in the building of bacterial cell envelopes and may be responsible for the thermotolerance of the chloroplast of this alga. Two additional genes may play a role in stabilizing the photosynthetic machinery against salt stress and detoxification of the chloroplast. Thus, these genes may be unique to the cp-genome of *C. caldarium* and may be required for the endurance of the extreme living conditions of this alga.

Key words: Chloroplast — *Cyanidium caldarium* — Structure — Evolution — Sequence — Gene repertoire

Introduction

It is assumed that plastids, the photosynthetic organelles of plants, are the remnants of an endosymbiosis event between an amoeboid eukaryote and a cyanobacterium (Lopez-Garc and Moreira 1999). Some different algal lineages were formed by a secondary endosymbiosis event, in which an unicellular algal species was engulfed by another amoeboid eukaryote (Winhauer et al. 1991). Based on the morphology and the chlorophyll type of the organelle, the plant kingdom can be divided into two evolutionary lineages: the red lineage, comprising several algal groups, and the green lineage, comprising green algae and higher plants (Whatley 1981).

During the process of accommodation to the host the genome of the endosymbiont was reduced to a small set of genes. Some genes of the original endosymbiont needed for the function and maintenance of the chloroplast were transferred to the nucleus, and other genes remained in the chloroplast. However, most genes of the endosymbiont were lost. Most of the remaining genes of the chloroplast genome (cp-genome) are involved in photosynthesis and encode many of the protein subunits needed for the building of ribosomes. The chloroplast genome also provides a full set of tRNA and rRNA genes. Furthermore, the cp-genomes also contain some

Correspondence to: G. Glöckner; *e-mail:* gernot@imb-jena.de

specific genes varying even in closely related species (Martin et al. 1998).

Functional plastid genomes in the green lineage do not differ very much in their gene content. They comprise about 100 ± 20 genes (Sugiura 1995). In this lineage the plastid genome of *Nephroselmis olivacea,* which has recently been sequenced, has the largest gene repertoire with 127 protein coding genes. From the different gene losses during evolution of the green plastids, it was estimated that the ancestral green cp-genome contained about 137 protein coding genes (Turmel et al. 1999). The cp-genome of *Epifagus virginiana,* a plant with nonphotosynthetic chloroplasts encodes only 25 proteins, since the photosynthetic machinery and the corresponding genes are not longer needed (Wolfe et al. 1992). Thus, loss of function is accompanied by the loss of genes in the cp-genome. Here we will show that special environmental conditions may also have influenced the gene content of cp-genomes.

In the red lineage some cp-genomes were previously analyzed, namely, that of *Odontella sinensis*, *Porphyra purpurea*, and *Guillardia theta. O. sinensis*, a diatom, is derived from a secondary endosymbiosis event (Kowallik et al. 1995; Wang et al. 1997), which may have caused major changes in gene content and organization compared to other members of the red lineage. A comparison of these three genomes revealed that there exist large regions of conserved order (Douglas 1998). To allow a more detailed analysis of cp-genomes within the red lineage we decided to sequence and analyze the cp-genome of the red alga *Cyanidium caldarium* RK1. In some aspects, this unicellular red alga may be regarded as a living fossil. It is a unique photosynthetic eukaryote enduring the most extreme living conditions. It thrives all over the world in a restricted environment, in hot and acid springs with temperatures well above 45°C, pH 1, and high salinity (Doemel and Brock 1970). These living conditions are a strong evolutionary pressure lasting without changes since this alga evolved. This may have led to a conservation of its chloroplast genome. Here we report on the detailed analysis of the cp-genome (Accession no. AF022186) of *C. caldarium* and compare it with the other members of the red lineage.

Materials and Methods

DNA Isolation. Total cellular DNA was isolated using standard procedures and separated into nuclear and plastid fractions by CsCl density gradient centrifugation (1.1 g CsCl/ml, 120,000 *g*, 70.1 Ti Rotor, 65 h) in the presence of Hoechst dye (33342, 0.1 mg/ml).

Library Construction. Ten micrograms of the chloroplast DNA obtained were sonicated two times 5 s each using a Sonicator (Heat Systems) to break it into pieces. Protruding ends of the sheared DNA were filled with T4 Polymerase (NEB). To obtain fragments in the range of 1 to 3 kb the DNA was then separated on an 0.8% agarose gel (6 V/cm for 4 h). The region of fragments of the desired length was cut out and the agarose removed using the Jetsorb kit (Genomed). The fragments were then ligated into the SmaI site of M13mp18 (Craxton 1993). A second library with pUC18 as vector was constructed using random fragments of plastidal DNA completely digested with EcoRI.

Sequencing and Assembly. The M13mp18 template DNA for sequencing was recovered using standard methods (Mardis 1994). The templates then were cycle sequenced using Dye terminators (Amersham). Sequencing data of 2400 clones were collected using ABI377 sequencers. Assembly of the reads was performed using the gap4 program (Staden 1996). One hundred random EcoRI fragments cloned in pUC18 were sequenced from both ends to check the correctness of the assembly. Primer walks on sequencing templates and sequences obtained from PCR products filled the remaining three gaps. Two of the gaps were smaller than 10 bp. The last gap was located in the region containing a potential secondary DNA structure. To obtain a reliable sequence in this critical region 23 sequencing reactions on all templates available (one EcoRI fragment, five M13mp18 clones, two PCR products) were performed. The final mean coverage on the total cp-genome was 4.4. For each single base at least one forward and one reverse or two reads with the same direction of good quality were present.

Sequence Analysis. Open reading frames (ORFs) were detected with the Wisconsin Sequence Analysis Package verison 9 (Genetics Computer Group). The translated ORFs were blasted (Altschul et al. 1990) against the trembl and Swissprot Databases. Gene names were assigned only when a significant similarity was reached over the whole length of the protein. In some cases high similarities in conserved cores were sufficient to assign gene names (see Table 1 and 2). ORFs longer than 25 amino acids with no homology to other genes were only defined if they did not overlap other known genes or structures. tRNA genes were detected using tRNAScan (Lowe and Eddy 1997).

Phylogenetic analysis was performed using a concatenated gene set (*psbA, psaC, psbD, atpH, petB, psaA, psaB, atpB, rps12, psbC, psbE, psbF, psbB, psbL, atpA, psbI, rpl14, rpl36, psbT, psbH, psbJ, rpl16, rps19, psbK, rpl2, psaJ, psbN, petG, rps18, rps11, rps7, rps4, rps14, rps2, rps8, rpoB, rps3, rpoC1, rpl20, rpl22, atpE, ycf4, ycf9, rpoC2, atpF*) aligned with Clustal W. Neighbor joining methods as well as most parsimonious analysis was performed with PHYLIP after bootstrapping the data set. Maximum likelihood analysis was performed using puzzle (Strimmer and von Haeseler 1997).

Supplementary material can be viewed at *http://genome.imb-jena.de/~gernot/cyanidium.html.*

Results

Structural Organization of the cp-Genome of C. caldarium

The 164,921-bp circular cp-genome of *C. caldarium* is made up of 67.3% AT nucleotides. Two pairs of small direct repeat units (A/A′; B/B′) exist, which separate the cp-genome into two halves of 85.135 bp and 78.643 bp, respectively. The units A and B together make 386 bp and do not code for any gene. Between the units A and B a potential hairpin loop with a stem of 67 bases and a loop of 6 bases is inserted (Fig. 1). The stem contains a G homopolymer run of eight guanines. The free energy of the whole stem is −81.0. Additionally, the region adjacent to the unit A (bp 57–98) is considerable G rich (66%) and contains one additional short potential hairpin loop. Altogether, the region around the potential stem-

Table 1. Unique genes (compared to *P. purpurea*) of the *C. caldarium* cp-genome with similarities to genes with known functions

		Similar to			
Gene	Classification	Organism	Accession no.	ORF identity	Identity (similarity)
hisH	amino acid biosynthesis	*Synechocystis, (C. paradoxa)*	D64004	slr0084	42 (58)
lpxC	cell envelope	*Synechocystis*	D90902	sll1508	37 (54)
lpxA	cell envelope	*Synechocystis*	D64002	sll0379	50 (70)
thdF	detoxification	*Synechocystis*	D90910	sll1615	36 (62)
menD	biosynthesis of cofactors	*Synechocystis*	D64002	sll0603	27 (47)
menF	biosynthesis of cofactors	*Synechocystis*	D90911	slr0817	26 (49)
menE	biosynthesis of cofactors	*Synechocystis*	D64001	slr0492	29 (44) core
menB	biosynthesis of cofactors	*Synechocystis*	D90906	sll1127	58 (78)
menA	biosynthesis of cofactors	*Synechocystis*	D90911	slr1518	24 (43) core
menC	biosynthesis of cofactors	*Haemophilus influenzae*	P44961	—	31 (51) core
cobA	biosynthesis of cofactors	*Synechocystis*	D64002	sll0378	42 (62)
lipB	biosynthesis of cofactors	*Synechocystis*	D90915	slr0994	31 (54)
glmS	central intermediary metabolism	*Synechocystis*	D90900	sll0220	42 (62)
desA	fatty acid metabolism	*Synechocystis*	D90906	sll1468	54 (68)

Table 2. Unique genes (compared to *P. purpurea*) of the *C. caldarium* cp-genome with unknown functions

			Similar to			
Gene	Classification	Length (aa)	Organism	Accession no.	ORF identity	Identity (similarity) %
ccrf1		712	*Plasmodium chabaudi*	AF019972		20 in parts
ccrf2		31				
ycf82	cell envelope ?	384	*Synechocystis*	D90911	slr1508	46 (65) whole length
ycf83		136	*Synechocystis*	D64002	slr0204	28 (49) whole length
ccrf3		36				
ccrf4		26				
ccrf5		55				
ycf84		397	*Synechocystis*	D90907	slr0882	25 (45) whole length
ccrf6		28				
ccrf7		71				
ccrf8		43				
ccrf9		86				
ccrf10	regulatory functions ?	27	*B. subtilis*	U55043_8	_8	50 in parts
ccrf11		45				

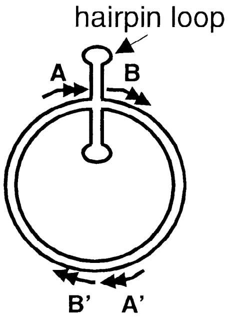

Fig. 1. The overall organization of the *C. caldarium* cp-genome. The regions shown are not drawn to scale.

loop including the repeat units A and B comprises 1200 bp and is free of long ORFs. Thus, this region is the largest gene-free region on the cp-genome. The boundaries of this segment are built by the genes for *ycf27* and *psbD.*

Interestingly, in *P. purpurea* these two genes are also located adjacent to each other, but separated only by 206 bp. Even though the two genes are transcribed in different directions in *C. caldarium* and in the same direction in *P. purpurea,* this can be called a conserved synteny according to the definitions by Clark (1999). The counterpart region of the hairpin loop segment contains the two direct repeat units (A′;B′) directly attached to each other and comprises 463 bp between the *rpl19* and *clpC* genes.

Comparison of the Gene Repertoire in C. caldarium *and* P. purpurea

On the cp-genome (Fig. 2) of *C. caldarium* we found 232 genes and ORFs. As a consequence thereof the intergenic spacer regions are very small with an average length of 71 bp. The cp-genome encodes for 199 proteins and a sufficient set of rRNAs (3) and tRNAs (30). The translation of two of the genes (*rps20, ycf52*) starts with GUG

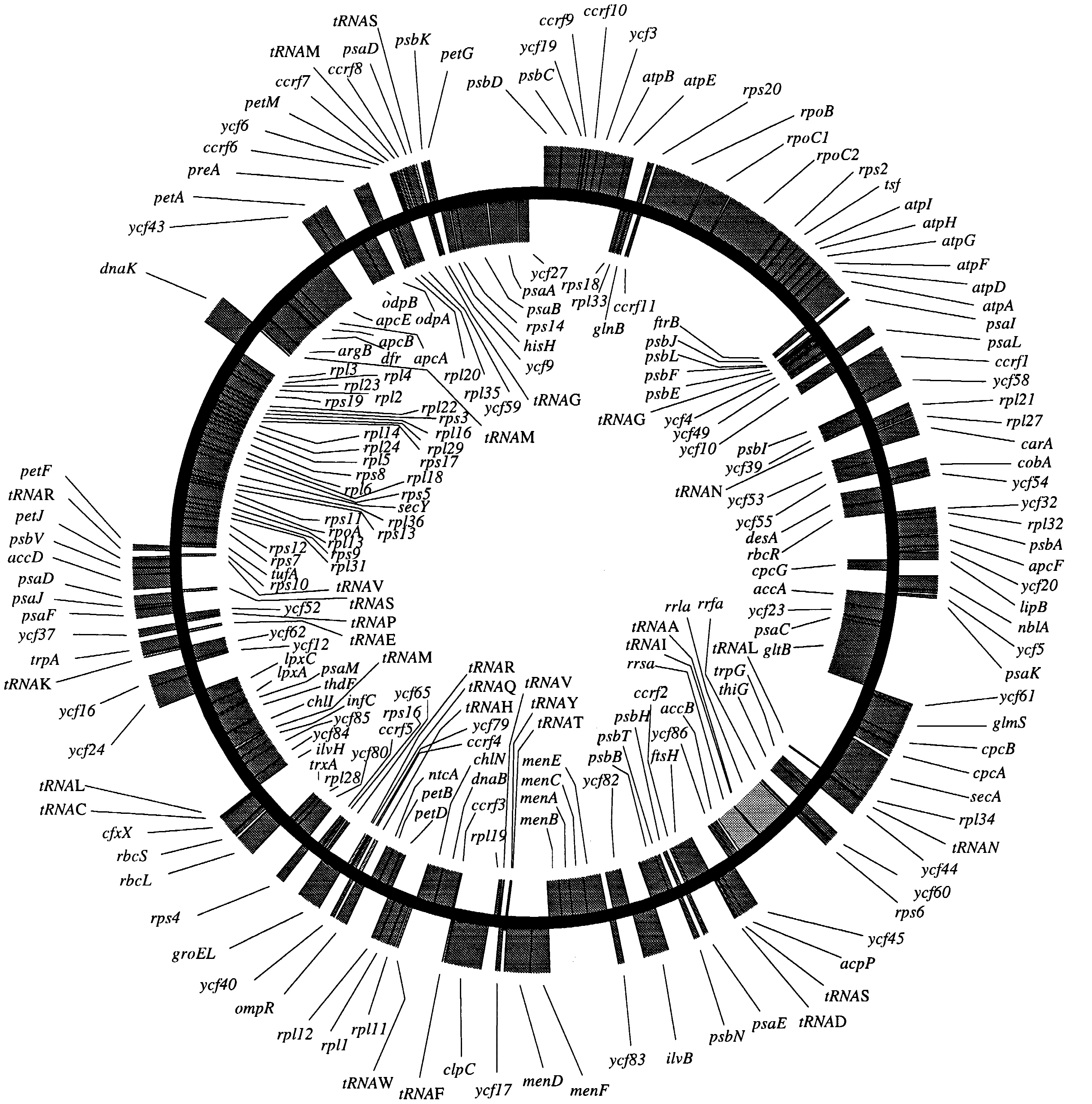

Fig. 2. The circular gene map of the cp-genome of *C. caldarium.* Genes transcribed in + direction are drawn outside the circle, genes in-direction inside the circle as dark gray boxes. The rRNA and tRNA genes are drawn as light gray boxes.

instead of AUG. One gene, *ycf20,* uses UUG. The overall coding capacity of *C. caldarium* is comparable to that of *P. purpurea* (Accession no. U38804), a multicellular red alga, with 251 genes and ORFs. While their general coding capacity is very similar, there are significant variations in the chloroplast gene sets of both red algae.

C. caldarium lacks 39 genes that are present in *P. purpurea* (*basI, chlB, chlL, cpeA, cpeB, dsbD, fabH, infB, moeB, pbsA, pgmA, psbX, rne, rpl9, rps1, syfB, syh, upp, ORF58, ORF62, ORF111, ORF121, ORF148, ORF287, ORF491, ORF621, ycf7, ycf21, ycf22, ycf29, ycf33, ycf34, ycf35, ycf36, ycf38, ycf56, ycf57, ycf63, ycf64*). Many of these 39 genes are also not present in most of the other chloroplast genomes so far described (Martin et al. 1998; Stoebe et al. 1998). On the other hand *C. caldarium* contains 28 genes and not yet defined ORFs that have not been detected in any other plastid genome. Two additional genes (*hisH* and *ycf49*) are only present in a glaucocystophyte alga containing cyanelles, which are suggested to be “primitive” plastids, *Cyanophora paradoxa* (Helmchen et al. 1995). In summary, 30 protein coding genes and ORFs are present in the cp-genome of *C. caldarium* that are not found in the cp-genome of *P. purpurea.* The gene products of 14 of these genes have a known function as shown in Table 1, while the function of the gene products of the remaining 16

Table 3. Conserved regions between *C. caldarium* and *P. porphyra* comprising more than two coding genes in both organisms

Region	Number of genes	Genes in *C. caldarium*	Region (bp)	Polarity	Unique *C. caldarium* genes	Additional genes in *P. purpurea*
1	4	*ycf27, psbD, psbC, ycf19*	4784	−		*upp, ycf17*
2	3	*ycf3, atpB, atpE*	3014	+		
3	16	*rps18, rpl33, glnB, rps20, ccrf11, rpoB, rpoC1, rpoC2, rps2, tsf, atpI, atpH, atpG, atpF, atpD, atpA*	16,181	−	*ccrf11*	
4	6	*ftrB, psaI, psbJ, psbL, psbF, psbE*	1320	+		
5	3	*ycf4, ycf49, psaL*	1204	+	*ycf49*	
6	5	*ycf10, ccrf1, ycf58, psbI, ycf39*	4820	+	*ccrf1*	*orf111, orf621, orf62*
7	9	*carA, ycf53, ycf55, cobA, ycf54, desA, ycf30, ycf32, rpl32*	6416	+	*desA, cobA*	
8	13	*cpcG, ycf18, ycf5, psaK, accA, ycf23, psaC, gltB, ycf61, glmS, cpcB, cpcA, secA*	12,759	+	*glmS*	*ycf21, pgmA, ycf22, cpeB, cpeA*
9	5	*ycf44, trpG, thiG, ycf60, rps6*	3780	−		
10	4	*accB, ycf86, ycf45, acpP*	2120	−	*ycf86*	
11	7	*ftsH, ccrf2, psaE, psbH, psbN, psbT, psbB*	4475	−	*ccrf2*	
12	4	*rpl19, clpC, ccrf3, dnaB*	5048	+	*ccrf3*	
13	13	*rpl11, rpl1, rpl12, petD, petB, ntcA, ompR, ycf40, ccrf4, psbW, groEl, ycf65, rps16*	7727	+	*ccrf4*	*orf148, syh, rps1, syfB*
14	9	*rpl28, trxA, rbcL, rbcS, cfxX, ilvH, ycf84, ycf85, infC*	7138	+	*ycf85, ycf84*	*ycf34*
15	3	*ycf12, ycf62, trpA*	2139	−		*ycf64*
16	8	*ycf37, ycf52, psaF, psaJ, apcD, accD, psbV, petJ*	4799	+		*psbX, fabH*
17	30	*rps10, tufA, rps7, rps12, rpl31, rps9, rpl13, rpoA, rps11, rps13, rpl36, secY, rps5, rpl18, rpl6, rps8, rpl5, rpl24, rpl14, rps17,l rpl29, rpl16, rps3, rpl22, rps19, rpl2, rpl23, rpl4, rpl3, dnaK*	16,998	+		
18	10	*apcB, apcA, apcE, ycf43, petA, odpB, odpA, preA, rpl20, rpl35*	9505	+		
19	4	*ycf59, ccrf6, ycf6, petM*	1566	−	*ccrf6*	
20	5	*ycf9, psbK, petG, hisH, rps14*	1663	+	*hisH*	

ORFs is unknown (Table 2). Four of the five genes, which are similar to *Synechocystis* genes (Kaneko et al. 1996), were given new ycf numbers (*ycf82–ycf85*; W. Löffelhardt, personal communication). The other 11 ORFs show no or only slight similarities to ORFs from other prokaryotic organisms. Therefore, some of these ORFs may not represent true genes. One gene with unknown function is shared only between *P. purpurea* and *C. caldarium* and has been given the name *ycf86* (Löffelhardt, personal communication).

Conserved Order

An alignment of the cp-genomes of *C. caldarium* and *P. purpurea* reveals that large parts of the two genomes are similarly arranged over their total length (not shown). In Table 3 the 20 segments of conserved order making up more than two genes are listed. The respective polarity of these segments is given according to their Genbank entries (U38804 for *P. purpurea* and AF022186 for *C. caldarium*). Only one gene (*ycf17*) located in a conserved order segment in *C. caldarium* appears to have been moved from that postiion in *P. purpurea.*

Many of the unique genes of both species are embedded in segments of conserved order. Fourteen of the 30 novel genes in *C. caldarium* are enclosed in such large regions. On the other hand, 17 of the 39 unique genes of *P. purpurea* show also up in such syntenic regions (Table 3). When the locations of the single genes outside the conserved order segments of both cp-genomes were compared we observed five genes that probably were transferred from their original position. These genes have different positions and neighbors in both organisms (Table 4). Additionally, one entire *C. caldarium* region of 25 kb does not have the same position in the cp-genome as the corresponding region in *P. purpurea.* This region contains the genes of the regions 14 to 16 in Table 3 and the genes *ycf16, ycf24, lpxC, lpxA, psaM, thdF, chlI, ycf80,* and *rps4.* Yet in the cp-genomes of *G. theta* and *P. purpurea,* this region is in a colinear context (Douglas 1998).

Phylogenetic Position of the C. caldarium *Chloroplast*

We performed a bootstrapped phylogenetic analysis with the neighbor joining method using the same gene set as it was used by Martin et al. (1998). In the resulting tree *C. caldarium* robustly clusters with *P. purpurea* and *G. theta*. According to this tree, *G. theta* builds an assembly with *P. purpurea* and does not branch before the two red algae. *O. sinensis,* supposed to be derived from a secondary endosymbiosis, branches before the red algae and

Table 4. Translocated genes

		C. caldarium		*P. purpurea*	
#	Genes	Position	Neighbors	Position	Neighbors
1	*ycf26*	143258–145153	*apcB, argB*	181393–183363	*psbE, orf263*
2	*chlI*	109777–110838	*thdF, infC*	184540–185610	*orf263, psaM*
3	*chlN*	90499–91686	*rpl11, dnaB*	1095–2402	*chlL, orf491*
4	*ycf17*	85058–85189	*rpl19, menD*	143858–144004	*psbC, orf198*
5	*rpl34*	57400–57537	*orf437, secA*	4047–4187	*chlN, ycf37*

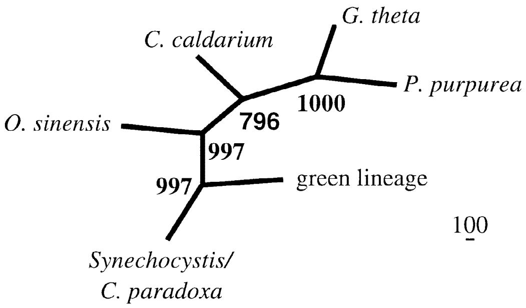

Fig. 3. Phylogenetic position of *C. caldarium.* A neighbor joining tree derived from a concatenated gene set as described in methods is shown. Bootstrap values are indicated at the branches. *Synechocystis* was used as root, the branching of *C. paradoxa* could not be resolved using this data set.

the cryptophyta. To confirm this result we performed additional analyses using maximum likelihood methods. The trees obtained with these methods gave the same topology as that shown in Fig. 3.

Unique C. caldarium *Genes with Known Functions*

The genes *hisH*, *cobA*, *desA*, and *glmS* are inserted in segments on the cp-genome, which show a conserved order between *P. purpurea* and *C. caldarium.* The other 10 of the 14 known genes unique to *C. caldarium* are not embedded in such regions.

A cluster of six genes is responsible for the biosynthesis of menaquinone (vitamin K). Menaquinone is supposed to be the secondary electron acceptor in the Photosystem I (Bittl et al. 1997; Hauska 1988). The biosynthesis of menaquinone requires at least seven reactions (Sharma et al. 1996). The enzyme required for the seventh step has only been described for *Escherichia coli.* The gene for this protein has no counterpart in *Synechocystis*, a cyanobacterium. Thus, this enzyme has to be encoded by a different gene in *Synechocystis* and may be represented by one of the four ORFs of *C. caldarium* with high homology to *Synechocystis* genes with unknown function. The plastid of *C. caldarium* thus seems to posses a nucleus-independent menaquinone biosynthesis. Additionally, the *C. caldarium* cp-genome encodes a *thdF* gene. In *E. coli* the gene product is involved in detoxification and metabolization of thiophene derivatives (Abdulrashid and Clark 1987). These sulfur-containing heterocycles are likely to occur in sulfurous, acidic, and hot springs, the natural habitat of *C. caldarium.* The gene for desaturation of fatty acids (*desA*) is also unique to *C. caldarium.* Recently it was shown that *desA* is important for the tolerance of the photosynthetic machinery to salt stress (Allakhverdiev et al. 1999).

Three genes, *lpxA*, *lpxC*, and *lipB* are involved in the synthesis and secretion of lipid A, the lipid moiety of lipopolysaccharids (Akatsuka et al. 1997; Frenken et al. 1993; Goldman et al. 1992; Stingele et al. 1996). An additional unique ORF with a weak homology to a galactosetransferase (*ycf82*) may be needed for the synthesis of the saccharide moiety of this molecule. Lipopolysaccharids are a main component of bacterial cell envelopes. No other photosynthetic eukaryote described so far contains these genes for the production of lipopolysaccharids in the cp-genome. The gene product of *glmS* is one of the main components of the cell envelope in *Thermus thermophilus* (Fernandez-Herrero et al. 1995). This gene may thus also play a role in plastid envelope synthesis in *C. caldarium.*

Discussion

Phylogeny

The phylogenetic tree shown in Fig. 3 implicates that the red lineage is subdivided into two branches. The formation of organisms such as *G. theta* took place after the separation of *C. caldarium* and *P. purpurea* from a common ancestor, whereas the endosymbiotic ancestor of *O. sinensis* was formed before that event. This tree is supported by the finding that the cp-genomes of *P. purpurea* and *G. theta* can be aligned throughout the whole genome, whereas one region in *C. caldarium* is displaced. Furthermore, *C. caldarium* lacks the intein, which is present in the *dnaB* gene in *P. pupurea* as well as *G. theta* (Douglas 1998).

Structure

A common feature of most green chloroplast genomes is a large, inverted, repeated region containing at least the rRNA genes (Maier et al. 1995; Sugiura 1995; Tsudzuki et al. 1992). Commonly, these repeat units divide the chloroplast genome in a large and a small single copy region. Even *Nephroselmis olivacea*, a green alga, shows the quatripartite structure characteristic for higher green plants (Turmel et al. 1999). Interestingly, a unicellular member of the green lineage, *Chlorella vulgaris*, contains no larger repeat regions (Wakasugi et al. 1997), whereas *Euglena gracilis* strain Z, which most likely evolved from a secondary endosymbiosis, has three tandemly arranged direct repeat units containing the rRNA genes (Hallik et al. 1993).

C. caldarium and *P. purpurea*, the two known red algae, contain direct repeats. On the other hand, members of the red lineage evolved from a secondary endosymbiosis (*O. sinensis, G. theta*) contain inverted repeats. Thus, in both the red and the green lineages repeats of various sizes and in both orientations exist.

Turmel et al. (1999) found striking similarities in gene partitioning among *C. paradoxa, G. theta, N. olivacea*, and even land plants. Due to that widespread existence of uniform structural patterns in chloroplasts it was postulated that the inverted repeats in cp-genomes have a single origin (Turmel et al. 1999). Yet, the occurrence of direct repeats in red algae may be a primary event. The evolution of an inverted repeat after a secondary endosymbiosis as it was found in *G. theta* can be traced back to a recombination of the rRNA cistrons (Douglas 1998). Thus, inverted repeats would have evolved several times. Otherwise, the direct repeats of the two red algae have to be acquired independently after their separation from a common ancestor. This can be inferred from the phylogenetic analysis, which shows that *G. theta*, containing inverted repeats, evolved after the separation of *P. purpurea* and *C. caldarium* from a common ancestor.

The lack of large repeats in *C. caldarium* may represent an archaic form of the cp-genome organization, since hairpin structures together with small repeat elements can be the seed for the development of larger direct repeats (Cohen et al. 1994; Ohshima et al. 1992).

All the genes, where a translocation event is very likely (Table 4), are located in the vicinity of the second copy of the rRNA cluster in *P. purpurea.* Since these genes are distributed over the whole cp-genome in *C. caldarium* they have to be transferred from their original position by other mechanisms than serial inversions. Obviously, the gain or loss of the second copy of the rRNA cluster may have led to these translocations. We cannot exclude that further genes not included in conserved order segments were also translocated during the evolution of the two red algal branches, but there is only weak evidence for such an event.

The region of 1200 bp including the hypothetical hairpin structure is gene-free in an otherwise gene-rich genome. This may be a hint for a role of this region in regulatory functions and/or replication. This region was the only one for which no shotgun clone could be found in the shotgun library. By using several primers located near and in the potential stem we were able to obtain the sequence of this region. The difficulty in cloning and sequencing this part of the cp-genome may be due to the presence of a stable secondary structure in this region. For the replication in mitochondria a hairpin structure is necessary, although no conserved sequence seems to be required (Clayton 1992). Suppose the structure observed is such an origin of replication: there is only one present in this cp-genome as it is in mitochondria. The very long stem of the hypothetical hairpin structure could be necessary for the replication at elevated temperatures, because temperature can influence the replication efficiency, if hairpin structures are involved in this process (Berkhout et al. 1997).

Conserved Order in Large Segments

The inversions observed between the cp-genomes of *P. purpurea* and *C. caldarium* lead to a fragmentation of the alignment of the cp-genomes. The fragments include parts of the genome ranging from 1.2 to 17 kb (Table 3). Yet, most parts are at a comparable position in both cp-genomes. A single region of 25 kb is not located at the same place in both algae. Thus, this region must have been translocated in one branch after the separation of the two branches in the red lineage.

Several unique genes of both organisms are located in regions of conserved order. They can be placed in joined segments (Table 3). A joint gene map would comprise the common genes and 31 of the 68 unique genes of both species. The combination of all genes of both species in and around such syntenic regions may partially resemble the ancestral red cp-genome. Information on additional species, which are as far related as *P. purpurea* and *C. caldarium*, may help totally reconstruct this ancestral red cp-genome.

Environmental Conditions May Have Influenced the Loss or Maintenance of Plastidal Genes During Evolution

Six genes comprise the *men* gene cluster. It is not obvious why at least six of the seven genes needed for menaquinone synthesis were retained in the cp-genome as a cluster. Possibly the clustering of the genes could be the first step for the translocation of all genes together to the nucleus of the eukaryote. The process of translocation then may have been stopped due to problems occurring in connection with the transcriptional regulation of the genes in the nucleus or the transport of the proteins to the chloroplast. There is no evidence that the *men* gene cluster represents alien genes translocated to the chloroplast. The codon usage can be a marker for such events (Mrazek and Karlin 1999). Yet in this cluster it does not deviate significantly from the codon usage of all genes of the cp-genome (data not shown).

A second group of genes comprises genes for lipopolysaccharid synthesis (*lipB*, *lpxA*, *lpxC*, and *ycf82*), which in turn is needed for building bacterial outer membranes. The target for the lipopolysaccharides produced in the chloroplast could therefore either be the cell envelope of the eukaryote or the plastid membrane. A fifth gene detected, *glmS*, is also involved in building cell envelopes.

The presence of these five genes in the cp-genome of *C. caldarium* suggests that this alga contains lipopolysaccharide-like membrane constituents. Thus, this chloroplast would have retained ancient features.

It was shown in *E. coli* that mutations in *lpxA* or *lpxC* cause a temperature-sensitive phenotype (Vuorio and Vaara 1995). Thus, thermotolerance of *C. caldarium* may be mediated through this group of genes.

Two genes of the cp-genome (*thdF* and *desA*) seem to be involved in stress endurance. These genes may have been also retained in the chloroplast due to their importance for survival under the extreme environmental living conditions of this species.

Additionally, the presence of so many genes (19) with similarities in photosynthetic organisms only to *Synechocystis* and *C. paradoxa* genes leads us to the following conclusion: These genes are the remnants of an ancient cp-genome, and they are mainly maintained due to the extreme living conditions of *C. caldarium*. Thus, only cp-genomes of close relatives, which live under the same conditions, but no other cp-genome may contain these genes.

References

Abdulrashid N, Clark DP (1987) Isolation and genetic analysis of mutations allowing the degradation of furans and thiophenes by *Escherichia coli*. J Bacteriol 169:1267–1271

Akatsuka H, Binet R, Kawai E, Wandersman C, Omori K (1997) Lipase secretion by bacterial hybrid ATP-binding cassette exporters: molecular recognition of the LipBCD, PrtDEF, and HasDEF exporters. J Bacteriol 179:4754–4760

Allakhverdiev SI, Nishiyama Y, Suzuki I, Tasaka Y, Murata N (1999) Genetic engineering of the unsaturation of fatty acids in membrane lipids alters the tolerance of *Synechocystis* to salt stress. Proc Natl Acad Sci USA 96:5862–5867

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Berkhout B, Klaver B, Das AT (1997) Forced evolution of a regulatory RNA helix in the HIV-1 genome. Nucleic Acids Res 25:940–947

Bittl R, Zech SG, Fromme P, Witt HT, Lubitz W (1997) Pulsed EPR structure analysis of photosystem I single crystals: localization of the phylloquinone acceptor. Biochemistry 36:12001–12004

Clark MS (1999) Comparative genomics: the key to understanding the Human Genome Project. Bioessays 21:121–130

Clayton DA (1992) Transcription and replication of animal mitochondrial DNAs. Int Rev Cytol 141:217–232

Cohen S, Hassin D, Karby S, Lavi S (1994) Hairpin structures are the primary amplification products: a novel mechanism for generation of inverted repeats during gene amplification. Mol Cell Biol 14: 7782–7791

Craxton M (1993) Cosmid sequencing. Methods Mol Biol 23:149–167

Doemel WN, Brock TD (1970) The upper temperature limit of *Cyanidium caldarium*. Arch Mikrobiol 72:326–332

Douglas SE (1998) Plastid evolution: origins, diversity, trends. Curr Opin Genet Devel 8:655–661

Fernandez-Herrero LA, Badet-Denisot MA, Badet B, Berenguer J (1995) glmS of *Thermus thermophilus* HB8: an essential gene for cell-wall synthesis identified immediately upstream of the S-layer gene. Mol Microbiol 17:1–12

Frenken LG, de Groot A, Tommassen J, Verrips CT (1993) Role of the *lipB* gene product in the folding of the secreted lipase of *Pseudomonas glumae*. Mol Microbiol 9:591–599

Goldman RC, Capobianco JO, Doran CC, Matthysse AG (1992) Inhibition of lipopolysaccharide synthesis in *Agrobacterium tumefaciens* and *Aeromonas salmonicida*. J Gen Microbiol 138:1527–1533

Hallik RB, Hong L, Drager RG, Favreau MR, Monfort A, Orsat B, Spielmann A, Stutz E (1993) Complete sequence of *Euglena gracilis* chloroplast DNA. Nucleic Acids Res 21:3537–3544

Hauska G (1988) Phylloquinone in photosystem I: are quinones the secondary electron acceptors in all types of photosynthetic reaction centers? Trend Biochem Sci 13:415–416

Helmchen TA, Bhattacharya D, Melkonian M (1995) Analyses of ribosomal RNA sequences from glaucocystophyte cyanelles provide new insights into the evolutionary relationships of plastids. J Mol Evol 41:203–210

Kaneko T, Sato S, Kotani H, Tanaka A, Asamizu E, Nakamura Y, Miyajima N, Hirosawa M, Sugiura M, Sasamoto S, Kimura T, Hosouchi T, Matsuno A, Muraki A, Nakazaki N, Naruo K, Okumura S, Shimpo S, Takeuchi C, Wada T, Watanabe A, Yamada M, Yasuda M, Tabata S (1996) Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. DNA Res 3:109–136

Kowallik KV, Stoebe B, Schaffran I, Kroth-Pancic P, Freier U (1995) The chloroplast genome of a chlorophyll a+c- containing alga, *Odontella sinensis*. Plant Mol Biol Rep 13:336–342

Lopez-Garc P, Moreira D (1999) Metabolic symbiosis at the origin of eukaryotes. Trends Biochem Sci 24:88–93

Lowe TM, Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25:955–964

Maier RM, Neckermann K, Igloi GL, Kossel H (1995) Complete sequence of the maize chloroplast genome: gene content, hotspots of divergence and fine tuning of genetic information by transcript editing. J Mol Biol 251:614–628

Mardis ER (1994) High-throughput detergent extraction of M13 subclones for fluorescent DNA sequencing. Nucleic Acids Res 22: 2173–2175

Martin W, Stoebe B, Goremykin V, Hansmann S, Hasegawa M, Kowallik KV (1998) Gene transfer to the nucleus and the evolution of chloroplasts. Nature 393:162–165

Mrazek J, Karlin S (1999) Detecting alien genes in bacterial genomes. Ann NY Acad Sci 870:314–329

Ohshima A, Inouye S, Inouye M (1992) In vivo duplication of genetic elements by the formation of stem-loop DNA without an RNA intermediate. Proc Natl Acad Sci USA 89:1016–1020

Sharma V, Hudspeth ME, Meganathan R (1996) Menaquinone (vitamin K2) biosynthesis: localization and characterization of the *menE* gene from *Escherichia coli*. Gene 168:43–48

Staden R (1996) The Staden sequence analysis package. Mol Biotechnol 5:233–241

Stingele F, Neeser JR, Mollet B (1996) Identification and characterization of the *eps* (Exopolysaccharide) gene cluster from *Streptococcus thermophilus* Sfi6. J Bacteriol 178:1680–1690

Stoebe B, Martin M, Kowallik KV (1998) Distribution and nomenclature of protein-coding genes in 12 sequenced chloroplast genomes. Plant Mol Biol Rep 16:243–255

Strimmer K, von Haeseler A (1997) Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment. Proc Natl Acad Sci USA 94:6815–6819

Sugiura M (1995) The chloroplast genome. Essays Biochem 30:49–57

Tsudzuki J, Nakashima K, Tsudzuki T, Hiratsuka J, Shibata M, Wakasugi T, Sugiura M (1992) Chloroplast DNA of black pine retains a residual inverted repeat lacking rRNA genes: nucleotide sequences of *trnQ, trnK, psbA, trnI* and *trnH* and the absence of *rps16.* Mol Gen Genet 232:206–214

Turmel M, Otis C, Lemieux C (1999) The complete chloroplast DNA sequence of the green alga *Nephroselmis olivacea:* insights into the architecture of ancestral chloroplast genomes. Proc Natl Acad Sci USA 96:10248–10253

Vuorio R, Vaara M (1995) Comparison of the phenotypes of the *lpxA* and *lpxD* mutants of *Escherichia coli.* FEMS Microbiol Lett 134: 227–232

Wakasugi T, Nagai T, Kapoor M, Sugita M, Ito M, Ito S, Tsudzuki J, Nakashima K, Tsudzuki T, Suzuki Y, Hamada A, Ohta T, Inamura A, Yoshinaga K, Sugiura M (1997) Complete nucleotide sequence of the chloroplast genome from the green alga *Chlorella vulgaris:* the existence of genes possibly involved in chloroplast division. Proc Natl Acad Sci USA 94:5967–5972

Wang SL, Liu XQ, Douglas SE (1997) The large ribosomal protein gene cluster of a cryptomonad plastid: gene organization, sequences and evolutionary implications. Biochem Mol Biol Int 41:1035–1044

Whatley JM (1981) Chloroplast evolution—ancient and modern. Ann NY Acad Sci 361:154–165

Winhauer T, Jager S, Valentin K, Zetsche K (1991) Structural similarities between psbA genes from red and brown algae. Curr Genet 20:177–180

Wolfe KH, Morden CW, Palmer JD (1992) Function and evolution of a minimal plastid genome from a nonphotosynthetic parasitic plant. Proc Natl Acad Sci USA 89:10648–10652