Mol Gen Genet (2000) 263: 152–158 © Springer-Verlag 2000

ORIGINAL PAPER

A. C. Barbrook · C. J. Howe

Minicircular plastid DNA in the dinoflagellate *Amphidinium operculatum*

Received: 8 September 1999 / Accepted: 10 November 1999

Abstract Plastid DNA was purified from the dinoflagellate *Amphidinium operculatum*. The genes *atpB*, *petD*, *psaA*, *psbA* and *psbB* have been shown to reside on single-gene minicircles of a uniform size of 2.3–2.4 kb. The *psaA* and *psbB* genes lack conventional initiation codons in the expected positions, and may use GTA for translation initiation. There are marked biases in codon preference. The predicted PsbA protein lacks the C-terminal extension which is present in all other photosynthetic organisms except *Euglena gracilis*, and there are other anomalies elsewhere in the predicted amino acid sequences. The non-coding regions of the minicircles contain a "core" region which includes a number of stretches that are highly conserved across all minicircles and modular regions that are conserved within subsets of the minicircles.

Key words Minicircle · Dinoflagellate · Plastid DNA · Codon bias · Endosymbiosis

Introduction

Plastids of plant and algal cells contain a genome that encodes a small but significant fraction of the proteins present in the plastid, including many components of the photosynthetic electron transfer chain and also the protein synthetic machinery. This genome is a remnant of the ancestral photosynthetic prokaryote(s) which gave rise to plastids by endosymbiosis, the rest of the ancestral genome having been lost or transferred to the nucleus (Martin et al. 1998; Race et al. 1999). Plastid DNA has been characterized from a broad range of plants and eukaryotic algae, including red and brown algae. The genome has been fully sequenced from several green plants and algae, as well as from the red alga *Porphyra purpurea*, the diatom *Odontella sinensis*, the glaucophyte *Cyanophora paradoxa*, and the cryptophyte *Guillardia theta* (Douglas and Penny 1999). In all these organisms, the genome occurs as a large circular molecule, typically of the order of 120–150 kb long. The last major group of algae whose plastid genome has remained essentially uncharacterized until recently are the dinoflagellates, a diverse group of organisms including both photosynthetic and non-photosynthetic members. Photosynthetic dinoflagellates have peridinin- or chlorophyll *a*/*c*-based light-harvesting systems. They are responsible for ecologically important phenomena such as red tides, and many strains produce toxins. They may also have been responsible for the development of important oil deposits (Lee 1989). The organization of the dinoflagellate nuclear genome is anomalous, with large amounts of DNA organized into permanently condensed chromosomes lacking conventional histones (Lee 1989). DNA was detected in dinoflagellate plastids by Kowallik and Haberkorn (1971), and Boczar et al. (1991) fractionated DNA from the dinoflagellates *Glenodinium* sp., *Protogonyaulax tamarensis* and *P. catenella* by density-gradient centrifugation. They identified one fraction as plastid DNA on the basis of its hybridization to genes found in the plastid in other groups, and estimated the size of the plastid genome at 114–125 kb. By contrast, Zhang et al. (1999) recently reported a study of the plastid genome from the dinoflagellate *Heterocapsa triquetra* and found that it was composed of small circular molecules of 2–3 kb. Limited analysis of a small number of other species (including *H. pygmaea*, the same as the *Glenodinium* sp. above) suggested the same was true in those, so there remains some uncertainty as to how widespread the minicircle form may be. We report here the purification and analysis of plastid DNA from the dinoflagellate *Amphidinium operculatum*. We found that it also comprises minicircular molecules, that show both

Communicated by R. G. Herrmann

A. C. Barbrook (✉) · C. J. Howe
Department of Biochemistry and Cambridge Centre for Molecular Recognition, University of Cambridge,
Tennis Court Road, Cambridge CB2 1QW, UK
E-mail: acb18@mole.bio.cam.ac.uk; Fax: +44-1223-333345

similarities to, and differences from, the *H. triquetra* genome.

Materials and methods

Culture conditions

Axenic *A. operculatum* was obtained from the Culture Collection of Algae and Protozoa (Ambleside, Cumbria, UK). This dinoflagellate was cultured in a 1:1 mixture of Antia's Medium and E26 + biotin (Tompkins et al. 1995) at 18 °C on a 16-h light/8-h dark cycle. During the light cycle cultures were illuminated with fluorescent light of moderate intensity.

DNA isolation

A. operculatum cells were harvested by centrifugation (2500 × *g*, 10 min) and frozen and thawed prior to resuspension in 5 ml of buffer containing 100 mM TRIS-borate pH 8.0, 50 mM sodium EDTA. SDS was then added to a final concentration of 7% (w/v) and the suspension heated to 60 °C for 10 min, after which an equal volume of deionised water was added. To remove pigment/protein complexes the suspension was first extracted with two volumes of phenol buffered with TRIS-HCl (pH 8.0), then twice with two volumes of buffered phenol/chloroform (1:1, v/v) and finally with two volumes of chloroform. Nucleic acids were then precipitated by addition of 1/10 volume of 3 M sodium acetate (pH 5.5) and two volumes of ethanol. After 2 h at −20 °C the nucleic acids were pelleted by centrifugation (10,000 ×*g*, 15 min) and vacuum dried. The pellet was gently resuspended in TE (10 mM TRIS-HCl, 1 mM sodium EDTA, pH 8.0).

For centrifugation in ethidium bromide (EtBr)-CsCl density gradients, resuspended DNA was made up to 10 ml with TE and CsCl was added (1 g/ml). EtBr was added to a final concentration of 100 μg/ml. The CsCl lysate was centrifuged at 175,000 ×*g* for 16 h at 25 °C using a vertical angle rotor. Fluorescent DNA bands were removed from the gradient using a needle (1.1 mm diameter) attached to a syringe.

Preparative Hoechst 33258-CsCl density centrifugation was done essentially as described by Aldrich and Cattolico (1981). Thus, for H33258-CsCl density gradients, CsCl was added to the DNA suspensions to give a density of 1.69 g/cm^3. Sarkosyl was added to give a 3% w/v solution and H33258 was added at a final concentration of 250 μg/ml. H33258-CsCl lysates were centrifuged at 70,000 ×*g* for 40 h at 25 °C using a fixed-angle rotor. Fluorescent DNA bands were removed from the density gradient as before.

EtBr or Hoechst dye was removed from DNA fractions by repeated extraction with TE-saturated butan-1-ol. Fractions were then diluted twofold in TE. DNA was precipitated by the addition of two volumes of ethanol and incubation at −20 °C overnight. After centrifugation (10,000 ×*g*, 15 min), DNA pellets were washed twice with 70% ethanol. Pellets were then dried and resuspended in a minimal volume of TE.

DNA analysis

DNA isolated from *A. operculatum* was digested in aliquots of 2–5 μg with restriction enzymes, according to the enzyme supplier's instructions. Agarose gel electrophoresis and Southern analysis were carried out using methods described by Sambrook et al. (1989). Fluorescein-labelled probes were prepared from isolated DNA fragments or PCR products using a random prime labelling system (Amersham). Probes for *psbAII* and *psbDII* were gifts from Prof. S. Golden (Texas A and M University). Membranes were probed and hybridization visualized using either enhanced chemiluminescence (ECL, Amersham) or the Gene Images CDP-Star detection module (Amersham), depending on the sensitivity required. PCR was carried out using Qiagen *Taq* polymerase according to the manufacturer's instructions. Annealing temperature and $MgCl_2$ concentration were varied to optimize amplification of the correct PCR product.

Gene cloning, sequencing and analysis

A. operculatum DNA fragments were cloned into the pUC19 vector, and clones containing putative plastid genes were identified by colony hybridisation using methods described by Sambrook et al. (1989). Sequencing of plasmids or PCR products was carried out using an Applied Biosystems automatic DNA sequencer. Sequencing of PCR products was used to verify restriction sites used for cloning. DNA sequences were assembled using the programs in the GCG Wisconsin package (Version 10.0, Genetics Computer Group, Madison, Wis.). The sequences were then analysed using a combination of the GCG Wisconsin package and the STADEN set of programs. ClustalW was used to generate multiple alignments of sequences (Thompson et al. 1994). Phylogenetic trees were generated from aligned protein sequences using either the Splitstree program (Huson 1998) or PAUP* 4.0 (Swofford 1998). Regions of alignments with gaps were removed prior to analysis.

Results

DNA extraction, fractionation and probing

High-molecular-weight DNA was extracted from *A. operculatum* using the protocol described, and fractionated by caesium chloride density-gradient centrifugation in the presence of ethidium bromide. Two fractions were obtained. These were separated and fractionated by caesium chloride density-gradient centrifugation in the presence of Hoechst 33258. The upper fraction after the first round of centrifugation was designated Fraction 1. The lower fraction (the denser fraction, containing more DNA) was separated into two separate fractions by centrifugation in the presence of the Hoechst dye. These were designated fractions 2 (upper, less dense) and 3 (lower, more dense).

The DNA fractions were incubated with a range of restriction enzymes and the products analysed by agarose gel electrophoresis. Fraction 1 gave discrete banding patterns with *Bam*HI, *Bgl*II, *Hin*dIII and *Sac*I, with a preponderance of fragments of approximately 2–4 kb. Fraction 2 gave a smear with all the enzymes used (*Bgl*II, *Eco*RI, *Hin*dIII, *Pst*I and *Pvu*II). Fraction 3 gave discrete banding patterns with *Bam*HI, *Bgl*II, *Hin*dIII, *Nde*I, *Pst*I, *Pvu*II and *Sac*I, but could not be digested efficiently with *Eco*RI. DNA from Fraction 1 hybridized strongly with probes made from the genes for the Photosystem II components PsbAII and PsbDII from *Synechococcus* sp. PCC7942 (Bustos and Golden 1992), but DNA from the other fractions did not (data not shown). Southern blots of DNA from Fraction 1 digested with a range of restriction enzymes were screened with the *psbAII* probe (Fig. 1). Unexpectedly, the probe hybridized to two bands in uncleaved DNA, with mobilities corresponding to linear sizes of 2.7 kb and 1.7 kb. The same bands were detected in DNA treated with *Bcl*I, *Kpn*I, *Pst*I or *Bgl*II. The probe hybridized to a

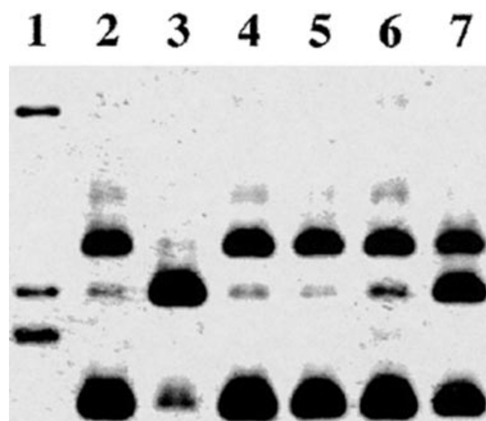

Fig. 1 Hybridization of DNA from Fraction 1 with a probe for *psbA*. Lane 1 contains markers (sizes 4.4, 2.3, 2.0 kb), lane 2 contains uncleaved DNA and the other lanes contain DNA incubated with *Bam*HI (lane 3), *Bcl*I (lane 4), *Kpn*I (lane 5), *Pst*I (lane 6), and *Sac*I (lane 7)

band with a mobility corresponding to a linear size of 2.3 kb in samples digested with either of *Bam*HI or *Sac*I. A 2.3-kb band was also detected faintly in the uncleaved DNA and in DNA treated with the first four enzymes, and bands of 2.7 kb and 1.7 kb were detected faintly in the *Bam*HI and *Sac*I digests. A probe for a *psaA* gene cloned from Fraction 1 (see below) was hybridized to digests of Fraction 1 DNA. It hybridized to bands with mobilities corresponding to 2.7 kb and 2.3 kb in *Bam*HI and *Sac*I digests, and 2.4 kb in a *Bgl*II digest (data not shown).

Gene cloning and analysis

DNA from Fraction 1 was digested separately with *Bam*HI, *Sac*I or *Hin*dIII. The restriction fragments obtained were cloned into pUC19, and colonies containing recombinant plasmids were screened with heterologous probes. Sequence analysis of hybridizing colonies showed that they contained *psbA*, *psaA* and *atpB* sequences. The sequences showed a conserved 49-bp element common to all three clones (see below). This region was then used to probe the pUC19 library, and two further clones were selected. Sequence analysis showed that they contained plasmids carrying the genes *psbB* and *petD*.

Discussion

Organization into minicircles

Probes for genes which are located in chloroplasts in other organisms hybridized to Fraction 1 of the *A. operculatum* DNA. The protein products predicted from these genes did not show any evidence of chloroplast transit peptides (von Heijne et al. 1989), so it seems likely that Fraction 1 corresponds to plastid DNA. Fraction 2, which does not yield discrete bands on restriction digestion, may represent nuclear DNA, and Fraction 3 may comprise mitochondrial or some other extrachromosomal species of DNA. The results obtained in the Southern hybridization experiments with *psbAII* are not consistent with a linear organization of the *psbA* locus in *A. operculatum* plastid DNA, even allowing for the possibility of tandem repeats of the gene. The only feasible interpretation is that the gene is present on a small circle of 2.3 kb (the size obtained on digestion with *Bam*HI or *Sac*I), with the undigested circle existing in forms with mobilities similar to linear molecules of 2.7 kb and 1.7 kb. The weak hybridization to a band of 2.3 kb in uncleaved DNA and in some other lanes is presumably due to DNA which has been linearized by shearing, non-specific nuclease digestion or partial restriction digestion. The data for *psaA* can likewise be interpreted to indicate that the gene is present on a 2.4-kb minicircle. Although Southern hybridization data were not obtained for the other genes, sequences determined from cloned restriction fragments gave circularly permuted forms of the expected coding sequence, also indicating a minicircular form for the DNA. All the minicircles were of a similar size, 2.3–2.4 kb. The organization into minicircles parallels that reported by Zhang et al. (1999) for *H. triquetra*. However, there seems to be less size variation in *A. operculatum*, where the *psaA* and *psbA* genes are on minicircles of 2,443 bp and 2,311 bp, respectively, compared to 3,005 bp and 2,151 bp in *H. triquetra*. As with *H. triquetra*, no minicircle was found to contain more than one gene. So, although the *atpB* gene normally overlaps the *atpE* gene in chloroplast genomes (Howe et al. 1985), there was no trace of *atpE* sequence on the *atpB* minicircle. Our finding of the *atpB* and *petD* genes adds to the list of genes identified on minicircles, but no genes have yet been found for proteins other than components of the light reactions of photosynthesis. Ribulose *bis*-phosphate carboxylase has previously been shown to be encoded in the nucleus of the dinoflagellate *Gonyaulax* (Morse et al. 1995), so it seems likely that the coding capacity of dinoflagellate plastids is much reduced compared to that of plastids in other species.

Features of the coding sequences

The coding regions show several notable features. Although *atpB*, *petD* and *psbA* have ATG codons close to the position where translation initiation would be expected by comparison with homologues, this is not the case for *psaA* or *psbB*. The N-terminus of PsbB is generally well conserved, and this conservation would be maintained in *A. operculatum* if initiation were to take place at a GTA codon (Fig. 2). The *psaA* gene also has a GTA codon close to the expected site of translation initiation. The codon preferences are very unusual and also differ between *A. operculatum* and *H. triquetra*, as

Fig. 2a,b Sequence alignments for predicted sequences of (**a**) PsbB showing proposed initiation at GTA and (**b**) PsbA showing C-terminal truncation in the dinoflagellates and *E. gracilis*. Organisms and Accession Nos. (for *psbB* and *psbA*, respectively, where two are given) are as follows: *A. operculatum* (AJ250263, AJ250262), *H. triquetra* (AF130034, AF130033), *Guillardia theta* (O78511,O78446), *Odontella sinensis* (P49471, P49460), *Cyanidium caldarum* (O19928, P24725), *Anabaena* sp. PCC7120 (P20093, two *psbA* genes – P46242 and P31694), *Chlamydomonas reinhardtii* (P37255, P07753), *Marchantia polymorpha* (P06412, P06402), *Euglena gracilis* (P14813, P06631), *Heterosigma carterae* (Q32389), *Porphyra purpurea* (P51212), *Antithamnion* sp. (P24625), *Cyanophora paradoxa* (P12719), *Nicotiana tabacum* (P04848), *Oryza sativa* (P12094), *Synechococcus elongatus* (two *psbA* genes – P35876, P35877), *Prochlorococcus marinus* CCMP1375 (P46895), *Synechocystis* sp. PCC6714 (P14660), *Synechocystis* sp. PCC6803 (two *psbA* genes – P07826, P16033) *Fremyella diplosiphon* (P07063), *Prochlorothrix hollandica* (P15191)

a)

```
Amphidinium operculatum     VRLPWFRVHIVVLNDPGRLISVHLMHTGLISGWAGLMALYELIVTDP
Heterocapsa triquetra       MRLPWFRVHIVILNDPGRLISVHIMHTALVAGWAAVMTLYELIILDP
Guillardia theta            MGLPWYRVHTVVLNDPGRLIAVHLMHTALVAGWAGSMALYELAVFDP
Odontella sinenis           MALPWYRVHTVVLNDPGRLIAVHLMHTALVAGWAGSMALYELAVFDP
Cyanidium caldarium         MALPWYRVHTVVLNDPGRLISVHLMHTALVSGWAGSMALYELAVFDP
Anabaena sp.                MGLPWYRVHTVVLNDPGRLISVHLMHTALVAGWAGSMALYELAIYDP
Chlamydomonas reinhardtii   MGLPWYRVHTVVINDPGRLISVHLMHTALVSGWAGSMALFEISVFDP
Marchantia polymorpha       MGLPWYRVHTVVLNDPGRLIAVHLMHTALVSGWAGSMALYELAVFDP
Euglena gracilis            MGLPWYRVHTVVLNDPGRFISVHLMHTALVSGWAGSMALYELAIFDP
                            : ***:*** *::*****:*:**:***.*::***. *:*:*: : **
```

b)

```
Amphidinium operculatum         DILNRADLGIEVMHERNAHNFPLDLA
Heterocapsa triquetra           DIVNRADLGMEVMHERNAHNFPLDLA
Euglena gracilis                DIINRANLGMEVMHERNAHNFPLDLA
Odontella sinenis               DIINRADLGMEVMHERNAHNFPLDLASGDVLPVALNAPAVNG
Heterosigma carterae            DIINRADLGMEVMHERNAHNFPLDLASNEVLPVAVNAPAVNG
Porphyra purpurea               DIINRANLGMEVMHERNAHNFPLDLASGESLPVALTAPAVNG
Guillardia theta                DILNRANLGIEVMHERNAHNFPLDLASGESLPVALTAPAVIG
Cyanophora paradoxa             DIINRANLGMEVMHERNAHNFPLDLASGEVMPVALTAPSINA
Nicotiana tabacum               DIINRANLGMEVMHERNAHNFPLDLAAIE-------APSTNG
Oryza sativa                    DIINRANLGMEVMHERNAHNFPLDLAALE-------VPSLNG
Marchantia polymorpha           DIINRANLGMEVMHERNAHNFPLDLAAVE-------APAVNG
Chlamydomonas reinhardtii       DIINRANLGMEVMHERNAHNFPLDLASTN-------SSSNN
Cyanidium caldarium             DILNRANLGIEVMHERNAHNFPLDLASEVSLPVALNKVEING
Synechococcus elongatus (1)     DIINRANLGMEVMHERNAHNFPLDLASAESAPVAMIAPSING
Synechococcus elongatus (2)     DIINRANIGIEVMHERNAHNFPLDLASGELAPVAMIAPSIEA
Prochlorococcus marinus         DVLNRANLGMEVMHERNAHNFPLDLAAAESTSVALVAPSIG
Synechocystis PCC6714           DVLNRANIGFEVMHERNAHNFPLDLASGEQAPVALTAPAING
Synechocystis PCC6803 (1)       DVINRANIGFEVMHERNVHNFPLDLASGDAQMVALNAPAIEG
Synechocystis PCC6803 (2)       DVLNRANIGFEVMHERNAHNFPLDLASGEQAPVALTAPAVNG
Fremyella diplosiphon           DVINRANLGMEVMHERNAHNFPLDLAAGEVAPVALTAPAING
Anabaena sp. (1)                DIINRANLGMEVMHERNAHNFPLDLAAGEVAPVALTAPAING
Anabaena sp. (2)                DIINRANLGMEVMHERNAHNFPLDLAAGEVAPVAISAPAING
Prochlorothrix hollandica       DILNRANLGFEVMHERNAHNFPLDLAAVK-------APSIIG
                                *::***::*:*******.********
```

exemplified by the figures shown for *psaA* and *psbA* in Fig. 3. For example, within the NCN family, 96.5% of the codons are NCA/T in *H. triquetra*, although only 59% are NCA/T in *A. operculatum*. Both genes in both organisms show an overwhelming preference for GGT as a glycine codon, yet by contrast TTC is overwhelmingly preferred over TTT for phenylalanine. In general the biases are more marked for *psbA* than for *psaA* genes, which may be a consequence of differing levels of expression for the two genes. These patterns are more complex than is generally the case for plastids in other groups, where there is a more general bias towards A or T in the third position (Lockhart et al. 1992). The *petD* gene of *A. operculatum* uses a TAA termination codon, whereas the others all have TAG. As with *H. triquetra*, TGA is not observed, but the preference between TAA and TAG is reversed.

Comparison of the *atpB* sequence with others shows that there is a particularly large insertion, of 41 codons, which maintains the ORF and corresponds to amino acid residues 143–183. The rest of the predicted AtpB sequence is very similar to other AtpB sequences, with the exception of the N-terminal region, which usually forms a six-stranded β-barrel (Abrahams et al. 1994), where there appears to be considerable variation. The *psaA* gene shows a number of deletions of a few codons when compared with other *psaA* sequences. These generally occur in regions encoding hydrophilic sections of the protein rather than transmembrane-spanning regions. The consequent reduction in size of the coding region allows it to fit within the space available for protein coding (after allowing for the "core" sequence discussed below) in the 2.4-kb circle. This situation is much more marked than that in *H. triquetra*, where the *psaA* gene shows fewer deletions, and the gene is located on a rather larger minicircle. It will be interesting to see if other dinoflagellate plastid genes for large proteins have been similarly trimmed down. A particularly interesting deletion occurs in the *psbA* gene. In other organisms, with the exception of *Euglena gracilis*, the protein has a C-terminal extension of several amino acids which is post-translationally cleaved by a specific

F	TTT	7	7	**S**	TCT	26	27	**Y**	TAT	17	12	**C**	TGT	5	3
		1	0			10	18			8	9			0	0
F	TTC	42	37	**S**	TCC	0	17	**Y**	TAC	6	11	**C**	TGC	0	1
		35	32			0	15			5	4			0	1
L	TTA	41	0	**S**	TCA	30	3	*	TAA	1	0	*	TGA	0	0
		5	0			19	0			1	0			0	0
L	TTG	2	13	**S**	TCG	0	18	*	TAG	0	1	**W**	TGG	24	19
		1	0			0	0			0	1			9	10
L	CTT	23	36	**P**	CCT	15	9	**H**	CAT	25	25	**R**	CGT	8	9
		19	23			3	6			8	3			8	9
L	CTC	12	19	**P**	CCC	0	1	**H**	CAC	9	9	**R**	CGC	1	0
		9	11			0	0			3	8			0	1
L	CTA	0	12	**P**	CCA	8	9	**Q**	CAA	22	4	**R**	CGA	0	0
		0	0			10	6			4	3			0	0
L	CTG	0	8	**P**	CCG	0	3	**Q**	CAG	9	18	**R**	CGG	0	0
		0	0			0	0			2	5			0	0
I	ATT	42	18	**T**	ACT	22	8	**N**	AAT	5	14	**S**	AGT	13	9
		14	10			9	7			2	0			1	0
I	ATC	18	17	**T**	ACC	0	14	**N**	AAC	15	10	**S**	AGC	8	2
		15	17			0	2			20	19			2	1
I	ATA	1	1	**T**	ACA	8	18	**K**	AAA	2	1	**R**	AGA	11	0
		0	0			2	2			0	0			1	0
M	ATG	16	11	**T**	ACG	0	3	**K**	AAG	19	10	**R**	AGG	1	8
		10	9			0	0			1	1			0	0
V	GTT	31	9	**A**	GCT	36	19	**D**	GAT	21	13	**G**	GGT	66	39
		16	5			26	15			8	8			35	34
V	GTC	13	16	**A**	GCC	2	5	**D**	GAC	5	7	**G**	GGC	4	5
		6	6			0	0			2	1			0	0
V	GTA	2	19	**A**	GCA	22	19	**E**	GAA	8	3	**G**	GGA	1	4
		0	7			9	16			12	5			0	0
V	GTG	1	13	**A**	GCG	3	19	**E**	GAG	3	7	**G**	GGG	1	2
		0	2			0	1			0	8			0	0

Fig. 3 Codon usage table for dinoflagellate *psaA* and *psbA* genes. For each codon the number of occurrences in the *psaA* genes of *H. triquetra* (*top left*) and *A. operculatum* (*top right*) and the *psbA* genes of *H. triquetra* (*bottom left*) and *A. operculatum* (*bottom right*) are given

C-terminal protease (Bowyer et al. 1992). The functional significance of this extension is unclear, but a failure to remove it when present interferes with photosynthetic function (Bowyer et al. 1992; Nixon et al. 1992). The region encoding the extension is absent from *psbA* of both *A. operculatum* and *H. triquetra*, as well as from *Euglena* (Fig. 2).

Phylogenetic analysis of *psbA* and *psaA* genes using split decomposition (Fig. 4) grouped the two dinoflagellate plastids together. Similar results were obtained with other methods of tree recovery, such as Neighbor-Joining (not shown). Although this grouping might be expected, the branches leading to the two dinoflagellates are exceptionally long, and the consequent long-branch attraction may lead to an unrealistically close grouping of the dinoflagellates (Felsenstein 1978). It is not clear why there should be such an apparently high substitution rate in this group, but it may be a consequence of the strong codon biases. A more detailed phylogenetic analysis will be published elsewhere.

Core element

The non-coding regions of the minicircles characterized from *H. triquetra* showed a very highly conserved core of 188 bp around a run of 9 A residues. This was separated by two less well conserved regions from two highly conserved flanking sections centred on runs of 9 G residues. The organisation of the non-coding regions in *A. operculatum* (Fig. 5) is significantly different. There is a block of 49 nucleotides which is identical across all five minicircles compared. Although this contains the sequence 5′-AGAGAAAAA-3′ at the centre, which is reminiscent of the A9 core in *H. triquetra*, the sequences immediately around it show no similarity in the two species. The 49-nucleotide block is preceded by a 19-nucleotide block which is identical across all the *A. operculatum* minicircles. Outside these are regions showing a "modular" organization, where there are blocks of sequence identity across some, but not all, of the minicircles. Comparison of the core region for three independent clones of *atpB*, two of *psaA* and four of *psbA*, did not reveal any differences between the minicircles for a given gene. The whole core covers approximately 400 nucleotides, which is similar to, but slightly smaller than, the "9G-9A-9G" region described for *H. triquetra* by Zhang et al. (1999). It is possible that the sequences conserved across all the *A. operculatum* minicircles are functional, perhaps in initiation of recombination or replication, and that the sequence modules are generated by gene conversion between different minicircles. Gene conversion could

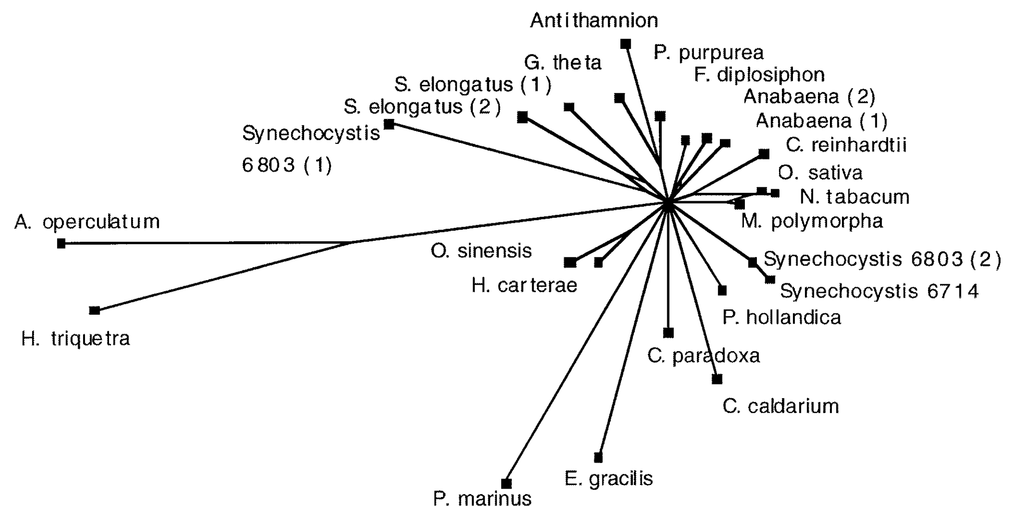

Fig. 4 Splitstree analysis of PsbA sequences. Organisms and accession numbers are as for Fig. 2

Fig. 5 Organization of the core regions on five minicircles. Regions that are conserved between two or more minicircles are shaded similarly. The 5′-AGAGAAAAA-3′ sequence is marked by *asterisks*

```
psaa   CAGA---AACGT--ACGAAGGATAACGAAGGCTGTTGAAGTCACAATCGACAGGA-GTGA  54
atpb   AAGA---AACGT--ACGAAGGATAACGAAGGCTGTTGAAGTCACAATCGACAGGA-GTGA  54
psba   CAAA---GACGA--TCGAAGGATACGGAAGCGTAGAGAATAGACAATCGGGAGACTATCA  55
petd   TTAATTTAATGTTGACGAAGGATAACGAAGGCTGTTGAAGTCACAATCGACAGGA-GTGA  59
psbb   CTAA----ACGT--ACGAAGGATAACGAAGGCTGTTGAAGTCACAATCGACAGGA-GTGA  53

psaa   TTAACTGAGGACAAATGTTGTCAGACATGAGGTTGACCAGGTCAT-TCATAGGCTCTCCG  114
atpb   TTAACTGAGGACAAATGTTGTCAGACAACCCCTCGTCCTGGTCATGTGATAGGCTTCTCG  114
psba   TTAGTCCTGGACAAATGTTGTCAGACATGAGGTTGACCAGGTCAT-TCATAGGCTCTCCG  115
petd   TTAACTGAGGACAAATGTTGTCAGACAACCCCTCGTCCTGGTCATGTGATAGGCTTCTCG  119
psbb   TTAACTGAGGACAAATGTTGTCAGACAACCCCTCGTCCTGGTCATGTGATAGGCTTCTCG  113

psaa   GTCATTTTGTTCCATCTCTACCCCAGTAGAGAAAAATCCAGGTCATATCATAGGAGATGG  174
atpb   ATGAAACTGTTCCATCTCTACCCCAGTAGAGAAAAATCCAGGTCATATCATAGGAGTGAG  174
psba   GTCATTTTGTTCCATCTCTACCCCAGTAGAGAAAAATCCAGGTCATATCATAGGAGATGG  175
petd   ATGAAACTGTTCCATCTCTACCCCAGTAGAGAAAAATCCAGGTCATATCATAGGAGATAG  179
psbb   ATGAAACTGTTCCATCTCTACCCCAGTAGAGAAAAATCCAGGTCATATCATAGGAGATAG  173
                                 *********

psaa   AACTGAGAGATCGAGA-GAACGAAGACAGAAAGGAGGTGAAAGGTAGTTGGCATGTTGAT  234
atpb   AATGTAACAATCGAGAAGAATGAATGTAGAA-GGTGGTAAATCATAGGGAGC-TGGAGAT  232
psba   AACTGAGAGATCGAGA-GAACGAAGACAGAAAGGAGGTGAAAGGTAGTTGGCATGTTGAT  235
petd   AAATGAATGA-CGAGA---ACGAAGACAGAATGACGTTGA-----AGAGAAC-----GAT  225
psbb   AAATGAATGA-CGAGA---ACGAAGACAGAATGACGTTGA-----AGAGAAC-----GAT  219

psaa   CTGGCGAGT---CAGAGGCATCAAACAGGAACGAAGGTAATAGAATGAATGATAATTGAT  291
atpb   CTCGCAATACTGCAGAAGGCTCAGACAGGGACGA---------------TGATAATTGAT  277
psba   CTGGCGAGT---CAGAGGCATCAAACAGGAACGAAGGTAATAGAATGAATGATAATTGAT  292
petd   TAGGTAATA-TGTAGAAACGACAATGAAAGTCCTC-----------------TAAAGAAG  267
psbb   TAGGTAATA-TGTAGAAACGACAATGAAAGATAAA---------------GATAACTTGG  263

psaa   GAAACCACAATTGA-------TAAAGA---------TAAATAAGATGAAACACTCCCT--  333
atpb   GAAACGACAATTGA-------TAAAGA---------TAAATAAGATGAAACACTCCCT--  319
psba   GAAACGACAATGAAAGTCCTCTAAAGAAGAGGCAACTAATAAAGATGAAACATATCTCGG  352
petd   AGGCAACTAATAAA-------GATGAAACATTCCGACTACATG---TCGAAGGCCTCTTG  317
psbb   AAGGTGCCCCTCAA-------CACGGTTCATCCCTTCATGGGAGGTTCGAAGTCACTGAT  316

psaa   -------CCTGATATGAAGGAAGGGATTAATTAACAAATGCCTCCTGGTTAGGCCAGAAG  385
atpb   -------CCTGATATGAAGGAAGGGATTAATTAACAAATGCCTCCTGGTTAGGCCAGAAG  372
psba   GGGTGCGCCTAAGTCAGGCTAAGAAAGCATCAAACAGATACAACCGCTGGGCAGGTCAAC  412
petd   GTTAGGCTCTCAGACAGGTGACCGTGCCACCGAGCATGTTATCACCTTGAAGAGATACCA  377
psbb   GGAACTTAGAGATACAATTGGTCTCGATCATAACAATCAATGCCACCTGGAAGGGCCAGC  376

psaa   TATACCAGGAAA--GGTTCGACATCGTCCAGCCACATTGT  423
atpb   TATACCAGGAAATGAGTTGATAAAAGCGGTCTCGGGGGTG  412
psba   AAGTGATTAATCCTGATCTAAGGATGATTACGACCTGCGG  452
petd   GCGCTCTATCTGGTTCGGTAGCTGTTGACTCCTCTTCTCC  417
psbb   AGTACTCCAGGAAGGTAGCCACAGTGCTCCAAACCGGTAT  416
```

probably result from the formation of recombination intermediates initiated within the absolutely conserved regions.

The analysis presented here shows that the remarkable organization into single-gene minicircles reported for *H. triquetra* is found in broadly similar form in another dinoflagellate species, although the *A. operculatum* minicircles seem to be of a more uniform size than those from *H. triquetra*. This study adds to the list of genes that have been identified on minicircles, but the list is still notably restricted to genes for proteins directly involved with the light reactions of photosynthesis. It will be interesting to see if this is an artefact of the limited number of minicircles characterized so far. The organization of the core region in *A. operculatum* shows some features in common with that in *H. triquetra*, but there are also significant differences, and comparison with other dinoflagellate plastid genomes may help in understanding the structure and function of this unusual section of the genome. Many fundamental questions remain completely open, such as whether GTA is indeed used as an initiation codon, whether editing occurs (which might account for the apparent use of GTA as an initiation codon), why the minicircles are of more or less uniform size, what the consequences of trimming of protein genes to fit this size may be, whether the minicircles replicate independently, and how this unusual genome organization evolved.

Acknowledgements We thank the Broodbank Trust and BBSRC for financial support, and Dr. Peter Lockhart for helpful discussions. This work was carried out in accordance with the Genetically Modified Organisms (Contained Use) Regulations 1992, as amended in 1996. The sequences described here have been deposited in the EMBL database under the Accession Nos. AJ250262-6 (*psbA*, *psbB*, *psaA*, *petD*, *atpB*, respectively).

References

Abrahams JP, Leslie AGW, Lutter R, Walker JE (1994) Structure at 2.8 Å resolution of F_1-ATPase from bovine heart mitochondria. Nature 370: 621–628

Aldrich J, Cattolico RA (1981) Isolation and characterization of chloroplast DNA from the marine chromophyte, *Olisthodiscus luteus*: electron microscopic visualisation of isomeric molecular forms. Plant Physiol 68: 641–647

Boczar BA, Liston J, Cattolico RA (1991) Characterization of satellite DNA from three marine dinoflagellates (Dinophyceae): *Glenodinium* sp. and two members of the toxic genus *Protogonyaulax*. Plant Physiol 97: 613–618

Bowyer JR, Packer JCL, McCormack BA, Whitelegge JP, Robinson C, Taylor MA (1992) Carboxyl-terminal processing of the D1-protein and photoactivation of water-splitting in photosystem II. J Biol Chem 267: 5424–5433

Bustos SA, Golden SS (1992) Light-regulated expression of the *psbD* gene family in *Synechococcus* sp. strain PCC7942: evidence for the role of duplicated *psbD* genes in cyanobacteria. Mol Gen Genet 232: 221–230

Douglas SE, Penny SL (1999) The plastid genome of the cryptophyte alga, *Guillardia theta*: complete sequence and conserved synteny groups confirm its common ancestry with red algae. J Mol Evol 48: 236–244

Felsenstein J (1978) Cases in which parsimony or compatibility methods will be positively misleading. Syst Zool 27: 401–410

Howe CJ, Fearnley IM, Walker JE, Dyer TA, Gray JC (1985) Nucleotide sequences of the genes for the alpha, beta and epsilon subunits of wheat chloroplast ATP synthase. Plant Mol Biol 4: 333–345

Huson DH (1998) Splitstree: a program for analyzing and visualizing evolutionary data. Bioinformatics 14: 68–73

Kowallik KV, Haberkorn G (1971) The DNA-structures of the chloroplast of *Prorocentrum micans* (Dinophyceae). Arch Microbiol 80: 252–261

Lee RE (1989) Phycology (2nd edn). Cambridge University Press, Cambridge, UK

Lockhart PJ, Howe CJ, Bryant DA, Beanland TJ, Larkum AWD (1992) Substitutional bias confounds inference of cyanelle origins from sequence data. J Mol Evol 34: 153–162

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

Morse D, Salois P, Markovic P, Hastings JW (1995) A nuclear-encoded form II RuBisCO in dinoflagellates. Science 268: 1622–1624

Nixon PJ, Trost JT, Diner BA (1992) Role of the carboxy terminus of polypeptide-D1 in the assembly of a functional water-oxidizing manganese cluster in photosystem-II of the cyanobacterium *Synechocystis* sp. PCC6803 – assembly requires a free carboxyl group at C-terminal position 344. Biochemistry 31: 10859–10871

Race HL, Herrmann RG, Martin W (1999) Why have organelles retained genomes? Trends Genet 15: 364–370

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual (2nd edn). Cold Spring Habor Laboratory Press, Cold Spring Harbor, N.Y.

Swofford DL (1998) PAUP*. Sinauer Associates, Sunderland, Mass.

Thompson JD, Higgins DG, Thompson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22: 4673–4680

Tompkins J, DeVille MM, Day JG, Turner MF (1995) Culture collection of algae and protozoa: catalogue of strains, Appendix 1. National Environmental Research Council, London, pp 144–173

Von Heijne G, Steppuhn J, Herrmann RG (1989) Domain structure of mitochondrial and chloroplast targeting peptides. Eur J Biochem. 180: 535–545

Zhang Z, Green BR, Cavalier-Smith T (1999) Single gene circles in dinoflagellate chloroplast genomes. Nature 400: 155–159