'Empty' minicircles and petB/atpA and psbD/psbE (cytb_{559} α) genes in tandem in *Amphidinium carterae* plastid DNA[1]

Roger G. Hiller*

Department of Biology, Macquarie University, Macquarie, NSW 2109, Australia

Received 8 June 2001; revised 20 August 2001; accepted 27 August 2001

First published online 7 September 2001

Edited by Marc Van Montagu

Abstract ***Amphidinium carterae*** **minicircle chloroplast DNA was separated from total DNA by centrifugation through a sucrose/NaCl gradient. Sequences of minicircles with psbA and 23S rRNA contained a common region of 67 bp. Primers designed from this generated numerous polymerase chain reaction products of 1.5–2.6 kb. These contained psaA and psaB as one gene/circle, and petB/atpA and psbD/psbE as two genes/circle. 'Empty' minicircles of 1.7–2.5 kb containing no identifiable genes or parts of genes were more abundant than gene-containing circles. From 15 minicircles a minimum common region of 48 bp was identified, with little identity to that from other dinoflagellate minicircles.** © **2001 Published by Elsevier Science B.V. on behalf of the Federation of European Biochemical Societies.**

Key words: Minicircle; Plastid DNA; Tandem gene organisation; Dinophycea; *Amphidinium carterae*

1. Introduction

It is now generally accepted that chloroplasts are derived from cyanobacterial endosymbionts of eukaryotic protists. The residual genome of the endosymbiont forms the chloroplast DNA, which in most algal and plant groups is circular and in the range of 100–200 kb. This circular DNA encodes many of the genes for components of chloroplast replication, protein synthesis and the photosynthetic apparatus. Although there have been many minor gene losses and rearrangements in the course of evolution [1], all groups show remarkable consistency in the organisation of their chloroplast genome. There is an earlier report [2] that this conservative DNA organisation is characteristic of dinoflagellates, but three recent papers have provided compelling evidence this is not so in at least some species [3–5]. In these species chloroplast DNA is reduced to a number of minicircles of 2–3 kb, each containing a single gene, or part of a gene. Genes comprising the core of both photosystems (e.g. psaA, psbA etc.) are present, as are some components of the ATPase, the cytb_6/*f* complex, 23S rRNA and 16S rRNA. The total number of genes so far identified is only 11 (compared with 50–200 in other chloroplast genomes) and there appears to be a complete absence of tRNA and ribosomal protein genes, as well as genes encoding small proteins of photosystem 1 (PSI), PSII and the ATPase. This investigation of the plastid DNA-encoded genes of *Amphidinium carterae* confirms their occurrence on minicircles but adds two important properties. These are the tandem arrangements of genes, which are not normally adjacent, atpA with petB, and psbD with psbE, and an abundance of related but 'empty' minicircles.

2. Materials and methods

A. carterae CS21 was cultured as previously described [6]. Cells were harvested by centrifugation at 3000×*g* using a Beckman J-10 rotor. Total DNA was isolated using sodium dodecyl sulphate and proteinase-K [7] and after spooling was further purified using the Wizard DNA cleanup protocol (Promega). To separate the circles from high molecular mass nuclear DNA, spooled crude DNA was applied to a gradient made by freezing and thawing a 20% sucrose solution in 1 M NaCl buffered by 0.1 M Tris–HCl pH 7.5 containing 1 mM EDTA. The gradients were centrifuged for 16 h at 32 000 rpm using a Beckman SW41 rotor and fractions collected by piercing the bottom of the tube. Fractions containing chloroplast-encoded genes were identified by Southern blotting using a probe made to the coding region of psbA of *Heterocapsa* [3] as well as to genes isolated in this study. Fractions towards the top of the gradient positive to chloroplast genes were combined, the DNA precipitated by addition of two volumes of ice cold ethanol and washed once with 70% ethanol. DNA was digested with *Sau*3A and cloned into *Bam*HI cut pUC18. Twelve randomly selected clones containing inserts of 0.4–1.0 kb were sequenced. From the sequence of a clone, containing part of psbA, adjacent primers were made, which would only give product if psbA is on a minicircle. A similar procedure was used to obtain a 23S rRNA gene using sequence information provided for *A. carterae* by Dr B.R. Green. From the sequences of psbA and 23S rRNA, a region of identity including 5′-TTCTGGTCAAACTAGTCAATTTGGGT-GCGAGTTTGGAATCTCAGCTCGATTCTCATAGG-3′ was identified. Universal forward 5′-TTTGGAATCTCAGCTCGATTC-3′ and reverse 5′-CTCGCACCCAAATTGACT-3′ back to back primers were used to generate by polymerase chain reaction (PCR) the gene-containing and empty minicircles comprising this paper. Standard PCR conditions were an initial cycle of 94°C for 1 min followed by 35 cycles of 94°C for 1 min, 52°C for 1 min, 72°C for 4 min. PCR products were cloned into pGEM-T (Promega) and clones containing inserts of >1.5 kb were sequenced using the automatic dye terminator system (ABI 377). The circular nature of the sequences was confirmed by PCR reactions across the region covered by the original universal circle primers. For Southern blotting DNA was transferred to Hybond-N+ nylon membrane (Amersham) and detected with fluorescein-labelled probes and antifluorescein alkaline phosphatase together with CDP-star reagent (NEN Life Science) according to

*Fax: (61)-2-98508245.
E-mail address: rhiller@rna.bio.mq.edu.au (R.G. Hiller).

[1] The sequences reported in this paper have been deposited in the EMBL data base under numbers AJ311628–AJ311633, AJ307009–AJ3079016 and AJ318067.

Abbreviations: PSI, photosystem 1; PCR, polymerase chain reaction

0014-5793/01/$20.00 © 2001 Published by Elsevier Science B.V. on behalf of the Federation of European Biochemical Societies.
PII: S0014-5793(01)02871-X

the manufacturer's protocols. Open reading frames (ORFs) and nucleotide matches were determined using standard NCBI programmes (http://www.ncbi.nlm.nih.gov/gorf/gorf.html and http://www.ncbi.nlm.nih.gov/blast/bl2seq/bl2.html).

3. Results and discussion

Southern blots of total *A. carterae* DNA using psbA or 23S rRNA as probe showed, as previously reported [3,4], bands at ~2.5 kb and 1.5 kb rather than labelling of high molecular mass DNA expected for other groups of algae (results not shown). It was not possible to separate *A. carterae* satellite DNA (which includes minicircle chloroplast DNA) by the Hoechst 33258 dye method [6] or that of Scaramuzzi et al. [7]. Instead a sucrose/NaCl gradient was utilised for separating a minicircle fraction from genomic DNA essentially as described [8] for separating DNA into different size classes. Southern blotting showed that minicircle DNA is found towards the top of the gradient and high molecular mass DNA towards the bottom. A representative result is shown in Fig. 1 where fractions from the top of the gradient are probed with petB/atpA. Fractions 12, 13 and 14 are strongly positive and some fractionation of the two bands is apparent. The position of lower band in the gradient is consistent with it being a supercoiled plasmid. DNA from fractions 4 and 13 was concentrated 20× by precipitation and analysed by a dot blot. Fraction 13 was strongly positive to the petB/atpA probe whereas fraction 4 was negative as expected if these chloroplast genes are confined to minicircles.

Complete psbA and 23S rRNA minicircle sequences were used to design universal circle primers. These generated numerous PCR products of 1.5–2.5 kb with either total DNA or gradient fractions 12 plus 13 as target. From these products, additional chloroplast-encoded genes were cloned: psaA, psaB, petB, atpA, psbD and psbE (cytb_{559} α) (Table 1). PetB/atpA were on one circle and psbD/psbE on another. This was confirmed by PCR using a reverse primer to the coding region of the first gene and a forward primer to the coding region of the second gene (Fig. 2). The resulting single PCR product was of the predicted size and identical in sequence to that obtained originally. When separate PCR reactions were performed using back to back primers made to each gene, petb/atpA and psbD/psbE as in circles 9 and 26 respectively, products indistinguishable as to mobility were obtained. In addition total DNA was restricted with *Kpn*I, which is predicted to cut circle 9 containing petB and atpA at one site. A Southern blot of cut and uncut DNA was probed with each gene separately and the cut DNA gave a single band of similar mobility to the upper part of the uncut DNA (results not shown). The last results are entirely consistent with circles 9 and 26 each containing a pair of genes but do not prove that they do. Proof that this pairing of genes is not some extraordinary PCR artefact could be obtained by direct cloning of circle DNA digested with *Kpn*I to obtain circle 9 containing petB and atpA or *Sca*I to get circle 26 containing psbD and psbE.

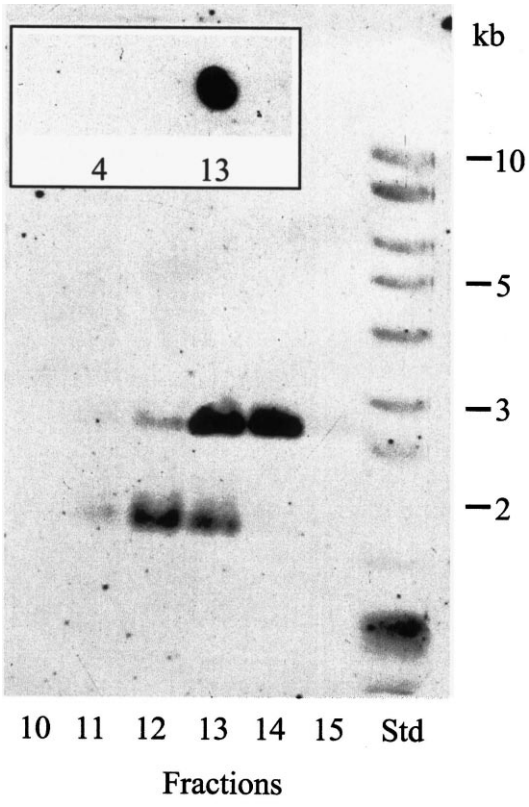

Fig. 1. Fractions 10–15 of *A. carterae* DNA separated on a sucrose/salt gradient. The blot was probed with part of the petB and atpA genes from circle 9. Inset: a dot blot of concentrated fractions 4 and 13 from the same gradient. The top of gradient was fraction 16.

Circles, which encoded no identifiable protein even as a fragment, were found and have been designated 'empty'. Nine 'empty' circles have been completely sequenced and share differing levels of identity with non-coding regions of the gene-containing circles (Table 2). Although some contain ORFs of >200 bases, no pattern of putative products is apparent. On the basis of the PCR data, empty circles are more abundant than circles containing genes as random sequencing of the PCR-derived clones gave only four gene-containing circles but at least 12 different 'empty' circles.

These results may be slightly biased due to the properties of Taq polymerase, as 'empty' circles average ~2 kb whereas gene-containing circles average 2.5 kb. Taq polymerase is relatively ineffective for long range PCR and it may be significant that neither psbA nor 23S rRNA genes was recovered using the universal circle primers. Each gene-containing circle was obtained once and only one of the 'empty' circles was

Table 1
Properties of DNA minicircles containing chloroplast genes

Circle	1	7	9	26	85	23S
Gene	psbA	psaB	petB atpA	psbD psbE	psaA	23S rRNA
Base pairs	2520	2306	2606	2369	2556	2713
Start codon	ATG	GTN? TTG?	GTG GTN?	ATG ATG	ATG GTA?	–
Stop codon	TAG	TAG	TAA TAG	TAA TAA	TAG	–
Outside coding region (bp)	1498	446	480	684	520	–

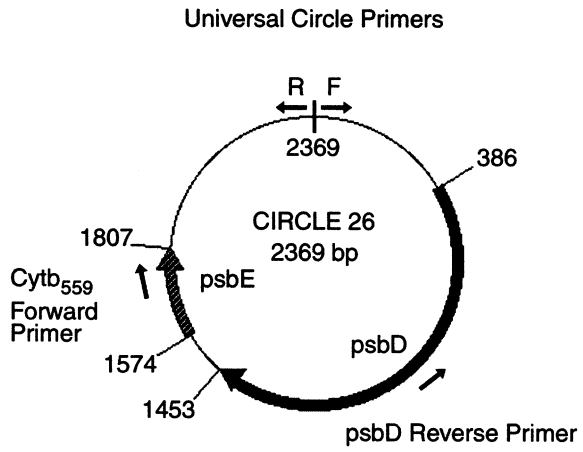

Fig. 2. Representation of circle 26 with psbD and psbE genes in tandem. The position of the forward and reverse universal primers is shown together with the positions of gene-specific primers used to confirm the tandem circular arrangement of psbD and psbE (cytb_{559} α).

found in two clones. These results are consistent with equal copies of each minicircle as is also the case in *A. operculatum* (A. Barbrook, personal communication). However, Zhang et al. [3] noted that psbA turned up many times in their randomly selected clones. They also found circles containing truncated psbC and 16S rRNA. Circle 33 contains a small part of a 16S rRNA gene but apart from this I found no fragments of any gene. It seems that the empty circles in *A. carterae* are discrete and do not reflect rampant homologous recombination or PCR artefacts.

The derived products of the core PSI genes, psaA and psaB, are quite extraordinary. For both sequences, alignment with other psaA and psaB proteins is only possible if large gaps are introduced (results not shown). Most of these gaps correspond to loop regions between the putative transmembrane region: the N-terminal region is also aberrant. The derived protein sequence from psaA of *A. operculatum* (but not the non-coding region) is almost identical [4] to that reported here so it is unlikely that the sequences are those of pseudogenes. It was speculated [4] that truncation of the psaA gene might result from adapting encoding a large protein to the confines of a minicircle. Given the 23S rRNA minicircle of 2.7 kb and the 3.1 kb psaA minicircle in *Heterocapsa triquetra* [3], this seems somewhat implausible.

A. operculatum apparently utilises GTA as a start codon in psbB [4] and ATA and TTG in addition to ATG have been suggested as start codons in *Heterocapsa* [3]. GTG is clearly a candidate in petB in the present work (Fig. 3). For atpA and psaB it is impossible to assign a start codon. Assuming the intergenic space between petB and atpA is similar to that between psbD and psbF (70–120 bp), and there is similarity with atpA from *Heterocapsa*, we arrive at a sequence of GCN

Amphidinium	**VGFIYDWCEE**
Heterocapsa	**MGFIYDWSEE**
Porphyra	**MSKIYDWFEE**
Odontella	**MGKVYDWFEE**
Chlorella	**MGKVYDWFEE**
	··:·****

Fig. 3. Alignment of the N-terminus of cytb_6/*f* subunits encoded by petB. GTG encodes valine at the N-terminus in *A. carterae*. * identity; : similarity. The sequences used are *Heterocapsa* [3], *Porphyra* [20], *Odontella* [21], *Chlorella* [22].

(Fig. 4). There is no in frame GTN, ATA or TTG upstream of the alignment with the *Heterocapsa* sequence. Downstream there is no ATG in the first 150 bases but there are seven in frame GTN and one TTG. A similar problem arises with the start codon for psaB in circle 7. Aligning the derived sequence with that of *Heterocapsa* gives no appropriately located ATG or ATA but there is a wide choice of in frame GTN and TTG (results not shown). ATA does not seem to be used at all in *A. carterae*. In petB, atpA, psaB, psbD and psbF combined, isoleucine is encoded by 48 ATT and 45 ATC.

My results are in general agreement with previous reports [3,4] of the plastid DNA of dinoflagellates comprising a limited number of genes, one to each minicircle. In addition I found two circles, each of which contained two genes separated by ~80–120 bp. The genes atpA/petB and psbD/psbE are not normally adjacent, so their proximity in *A. carterae* may be fortuitous; a recent addition to the circles' repertoire. It is uncertain if these associations are normal in minicircle DNA, as neither psbD nor psbE was found previously [3,4] although in *Heterocapsa* atpA is not associated with petB.

Since the empty circles contain no identifiable coding features analysis is difficult. Circles 2 and 10 have long stretches of near identity to the non-coding region of psbA (Table 2). This region includes that common to psbA and 23S rRNA circles. Circle 25 has only 48 bp identity in the common region which, perhaps fortuitously, is almost the same as that in *A. operculatum*. Zhang et al. [3] found 79 clones derived from *Heterocapsa* satellite DNA contained nothing identifiable but do not indicate how many have identity with non-coding regions of gene-containing circles or if they were parts of minicircles at all.

This report and [3–5] have noted highly conserved core regions in gene-containing circles and this also applies to the 'empty' circles. In *H. triquetra* the conserved region of 188 bp contained runs of nine A and nine G. No such motifs are present in *A. operculatum* but a sequence 5′-AGA-GAAAAA-3′ was noted in the conserved 49 bp block. There is little similarity between the core region of *A. carterae* and

Table 2
Properties of nine 'empty' plastid DNA minicircles

Circle	2	10	11	14	15	17	25	33[a]	82
Base pairs	1708	1742	2097	1855	2252	1995	2253	2533	1991
ORF >150 bp	7	6	3	2	7	3	2	3	4
Largest ORF (bp)	348	345	207	177	297	270	273	243	255
Common region/difference vs psbA (bp/number)	471/4	315/1	84/0	155/0	147/0	116/8	48/0	78/0	102/7

[a]Circle 33 contains a short piece of a 16S rRNA gene.

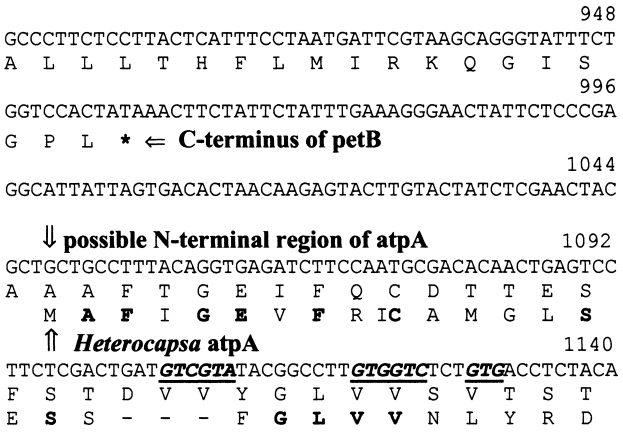

Fig. 4. Part of circle 9 with petB and atpA genes. Bold face italics, possible start codons. The proposed N-terminus of *Heterocapsa* atpA protein [3] is shown for comparison.

that of either *A. operculatum* or *H. triquetra* except for the seven base identity 5′-TCATAGG-3′ in the first two. BlastN searches of the sequences reported here gave matches with short non-coding regions of individual *A. operculatum* minicircles but no pattern was apparent. From a comparison [5] of 23S rRNA and psbA sequences from several species of dinoflagellates it was concluded that the minicircles are undergoing rapid evolution. This is consistent with the coding regions of psbA and psaA for *A. carterae* and *A. operculatum* being almost identical but the non-coding regions are highly varied. For 23S rRNA the sequence reported for *A. carterae* by Zhang et al. [5] is ~600 bp shorter than that found here. The difference lies in a number of blocks, inserted in the sequence from the CS21 strain compared to that used in [5]. There must be some doubt as to the status of the 23S rRNA gene reported here, as only a few hundred bases are colinear with that of *Heterocapsa* or any other organism besides other strains of *Amphidinium*. *Heterocapsa* 23S rRNA also diverges considerably from other 23S rRNAs and it is possible that all the reported dinoflagellate sequences represent truncated forms of this RNA.

Plasmid-like DNA molecules have been reported from several algal and plant groups in addition to the dinoflagellates [9–15] but few have been completely characterised. In *Cylindrotheca* [11] there are two circular plasmids of ~4 kb which hybridise to chloroplast DNA but they appear to contain no known chloroplast genes or parts. pCf1 has six putative ORFs of >150 bp with three of them being >300 bp. One of them, ORF218, had significant identity with bacterial resolvases and the largest (1449 bp) with hypothetical proteins derived from *Guillardia* and *Marchantia* chloroplast genomes.

In *Chenopodium album* a mitochondrial plasmid of 1309 bp, which replicates by a rolling circle mechanism, has been described [15]. Although this plasmid encodes no known mitochondrial genes it contains three ORFs of 208, 243 and 465 bp. More comparable with dinoflagellate minicircles are those from the mesozoan animal *Dicema* [16]. These are ~1.7 kb and contained COI, COII and COIII as one gene/circle with no evidence for a maxicircle of DNA containing CO genes. Reduction in mitochondrial/plastid genome sizes with formation of minicircles has possibly arisen independently on several occasions.

Barbrook and Howe [4] comment that there remain many questions about dinoflagellate plastid DNA: 'does RNA editing occur?'; 'why are the circles of only 2–3 kb?'; 'can the circles replicate independently?'; and 'how did they evolve?'. To these I would add, 'what is the functional role of the 'empty' circles?' Are they extreme examples of 'selfish' DNA [17,18] where an evolutionary advantage [19] of contracting the chloroplast genome has been subverted to propagate circular nonsense DNA on an unprecedented scale or do they play an important but unknown role in dinoflagellate plastid DNA?

Acknowledgements: I thank Beverley Green, Zhadou Zang, Adrian Barbrook and Chris Howe for helpful discussions, gene probes and communication of results prior to publication. Financial support was provided by the Australian Research Council and Macquarie University. This research was in accordance with the regulations of the Commonwealth of Australia, the State of New South Wales and Macquarie University governing the use of genetically modified organisms and recombinant DNA.

References

[1] Martin, W., Stoebe, B., Goremykin, V., Hansmann, S., Hasagawa, M. and Kowallik, K. (1998) Nature 393, 162–165.
[2] Boczar, B.A., Liston, J. and Cattolico, R.A. (1991) Plant Physiol. 97, 613–618.
[3] Zhang, Z., Green, B.R. and Cavalier-Smith, T. (1999) Nature 400, 155–159.
[4] Barbrook, A.C. and Howe, C.J. (2000) Mol. Gen. Genet. 263, 152–158.
[5] Zhang, Z., Green, B.R. and Cavalier-Smith, T. (2000) J. Mol. Evol. 51, 26–40.
[6] Aldrich, J. and Cattolico, R.A. (1981) Plant Physiol. 68, 641–647.
[7] Scaramuzzi, C.D., Stokes, H.W. and Hiller, R.G. (1992) Plant Mol. Biol. 18, 467–476.
[8] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning: a Laboratory Manual, 2nd edn., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
[9] Ebert, C., Tymms, M.J. and Schweiger, H.-G. (1985) Mol. Gen. Genet. 200, 187–192.
[10] La Claire, J.W., Loudenslager, C.M. and Zuccarello, G.C. (1998) Curr. Genet. 34, 204–211.
[11] Hildebrand, M., Hasagawa, P. and Ord, R.W. et al. (1992) Plant Mol. Biol. 19, 759–770.
[12] Heizmann, P., Ravel-Chaupis, P. and Nigon, V. (1982) Curr. Genet. 6, 119–122.
[13] Leon, P., Walbot, V. and Bedinger, P. (1989) Nucleic Acids Res. 17, 4089–4099.
[14] Backert, S., Nielsen, B.L. and Börner, T. (1997) Trends Plant Sci. 2, 477–483.
[15] Backert, S., Kunnimalaiyaan, M., Börner, T. and Nielsen, B.L. (1998) J. Mol. Biol. 284, 1005–1015.
[16] Watanabe, K.I., Bessho, Y., Kawasaki, H. and Hori, H. (1999) J. Mol. Biol. 2, 645–660.
[17] Orgel, L.E. and Crick, F.H.C. (1980) Nature 284, 604–607.
[18] Doolittle, W.F. and Sapienza, C. (1980) Nature 284, 601–603.
[19] Race, H.L., Herrmann, R.G. and Martin, W. (1999) Trends Genet. 15, 364–370.
[20] Reith, M.E. and Mulholland, J. (1995) Plant Mol. Biol. Rep. 13, 333–335.
[21] Kowallik, K.V., Stoebe, B., Schaffran, I., Kroth-Pancic, P. and Freier, U. (1995) Plant Mol. Biol. Rep. 13, 336–342.
[22] 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. and Sugiura, M. (1997) Proc. Natl. Acad. Sci. USA 94, 5967–5972.