http://www.elsevier.de/protis
Published online date 17 October 2013

Protist

ORIGINAL PAPER

The Chloroplast Genome of a *Symbiodinium* sp. Clade C3 Isolate

Adrian C. Barbrook[a,1], Christian R. Voolstra[b], and Christopher J. Howe[a]

[a]Department of Biochemistry, University of Cambridge, Building O, Downing Site, Tennis Court Road, Cambridge, CB2 1QW, UK
[b]Red Sea Research Center, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia

Submitted June 6, 2013; Accepted September 17, 2013
Monitoring Editor: Michael Melkonian

Dinoflagellate algae of the genus *Symbiodinium* form important symbioses within corals and other benthic marine animals. Dinoflagellates possess an extremely reduced plastid genome relative to those examined in plants and other algae. In dinoflagellates the plastid genes are located on small plasmids, commonly referred to as 'minicircles'. However, the chloroplast genomes of dinoflagellates have only been extensively characterised from a handful of species. There is also evidence of considerable variation in the chloroplast genome organisation across those species that have been examined. We therefore characterised the chloroplast genome from an environmental coral isolate, in this case containing a symbiont belonging to the *Symbiodinium* sp. clade C3. The gene content of the genome is well conserved with respect to previously characterised genomes. However, unlike previously characterised dinoflagellate chloroplast genomes we did not identify any 'empty' minicircles. The sequences of this chloroplast genome show a high rate of evolution relative to other algal species. Particularly notable was a surprisingly high level of sequence divergence within the core polypeptides of photosystem I, the reasons for which are currently unknown. This chloroplast genome also possesses distinctive codon usage and GC content. These features suggest that chloroplast genomes in *Symbiodinium* are highly plastic.

Key words: *Symbiodinium*; minicircle; dinoflagellate; chloroplast genome; diversity.

Introduction

Dinoflagellates are an important group of eukaryotic algae. They have a widespread global distribution and are found in both freshwater and marine environments. Dinoflagellates are significant primary producers, ranking second in importance behind diatoms in oceanic ecosystems (Field et al. 1998). Several species are remarkable in having profound environmental effects such as the ability to produce extensive algal blooms, many of which are capable of producing potent toxins with correspondingly devastating effects (Lee 2008). In contrast members of the genus *Symbiodinium* are renowned for forming symbioses with corals and other benthic marine animals (Rowan and Powers 1991). The symbioses which they with corals are the basis of exceptionally diverse ecosystems. However, there is increasing concern about the stability of these ecosystems and whether they are able to withstand projected changes of climate and the impact of other human interventions (Hughes et al. 2003).

[1]Corresponding author; fax +44 1223 333345
e-mail acb18@cam.ac.uk (A.C. Barbrook).

http://dx.doi.org/10.1016/j.protis.2013.09.006

Dinoflagellates show a surprising litany of unusual or even unique biological properties (Bayer et al. 2012; Lin 2011). Focusing solely on genetic aspects, all three of the genomes found within dinoflagellates can be said to be exceptional. For example, the nuclear genome of dinoflagellates varies enormously in size between species and possesses permanently condensed chromosomes, whilst containing relatively few nucleosomes despite possessing histone genes (Bayer et al. 2012; Lin 2011; Roy and Morse 2012). The mitochondrial genome, despite containing very few genes, achieves remarkable complexity arising from the presence of extensive non-coding inverted repeats and numerous gene isoforms, many of which require considerable post-transcriptional modification (Nash et al. 2008; Waller and Jackson 2009). The chloroplast genome of dinoflagellates is also highly anomalous (Howe et al. 2008). In dinoflagellates with chloroplasts containing the carotenoid peridinin as an accessory pigment, which are believed to be the ancestral form, the chloroplast genome is highly reduced in gene content and fragmented into several plasmid-like DNA minicircles. The genes that remain part of the chloroplast genome and are located on the minicircles mostly encode key components of the major photosynthetic complexes. The minicircles are usually between 2-3 kbp in size and contain a single gene (Howe et al. 2008). However, exceptions to this standard form exist. In *Adenoides eludens* there is evidence for larger single gene minicircles of 5 kbp as well as two gene minicircles of up to 12 kbp (Nelson and Green 2005). A two gene minicircle has also been reported from *Ceratium horridum* (Laatsch et al. 2004). Unusually, in this case the two genes, *ycf16* and *ycf24*, are not directly related to photosynthetic reactions; in fact they encode ABC transporter family proteins. Additionally, a recent study has suggested that these genes from *Ceratium horridum* and as well as the genes *rpl28* and *rpl33*, putative minicircle sequences from *Pyrocystis lunula*, are likely to have been acquired by horizontal gene transfer (HGT) (Moszczynski et al. 2012). These genes are doubly unusual as they have been found on minicircles only in these species and not in any other dinoflagellates, and whether these four genes are genuinely of dinoflagellate origin remains controversial. In *Amphidinium* species minicircles have been characterised that carry as many as four genes (Barbrook et al. 2006a). However, the size of these minicircles nevertheless conforms to the 2-3 kbp norm. In *Alexandrium tamarense* and *Alexandrium catenella* multiple variant forms of the *psbA* and *psbD* genes exist, in addition to a standard copy that is presumed to be functional, with each variant found on its own minicircle (Iida et al. 2009, 2010). The variant *psbA* genes of *A. tamarense* have been shown to be transcribed and the transcripts edited despite none of them being able to encode a functional protein because of the presence of various insertions and deletions (Iida et al. 2009). In *Lingulodinium polyedrum* chloroplast genes may not be on minicircles at all, since at least the *psbA* gene hybridises to DNA of much higher molecular weight (Wang and Morse 2006a). It therefore appears that the organisation of genes within the peridinin-containing dinoflagellates shows considerable diversity. Whilst minicircles have been isolated from a relatively large number of dinoflagellate species (Howe et al. 2008), for most species only one or two minicircles have been characterised. This has left a largely incomplete picture of the overall genome organisation in each case.

The coding content of the dinoflagellate chloroplast genome has largely been estimated following extensive minicircle characterisation from just a handful of species, including *Amphidinium* (two strains) and *Heterocapsa* (Barbrook and Howe 2000; Barbrook et al. 2001, 2006a; Hiller 2001; Nelson et al. 2007; Nisbet et al. 2004; Zhang et al. 1999, 2001). These data have been used in conjunction with EST data from several species, which indicate transfer to the nucleus of many genes that are normally chloroplast-located in other algae and plants (Bachvaroff et al. 2004; Hackett et al. 2004), to confirm which genes have been retained in the dinoflagellate chloroplast.

In spite of the ecological importance of symbioses between dinoflagellates and corals, there has been no systematic characterisation of a *Symbiodinium* chloroplast genome, although partial sequences of individual chloroplast genes or minicircles have been generated from a variety of *Symbiodinium* strains and isolates (Barbrook et al. 2006b; LaJeunesse and Thornhill 2011; McGinley et al. 2012; Moore et al. 2003; Santos et al. 2002). A systematic survey of the *Symbiodinium* chloroplast gene complement would have a number of benefits. Earlier studies have shown that the sequences of the *psbA* and *psbD* genes show significant variations at the amino acid level which are not found in other algal species (Iida et al. 2008). These variations could have consequences for physiological function and ultimately to susceptibility to coral bleaching (Warner et al. 1999). By characterising an essentially complete set of minicircles we hope to find out whether potentially significant variation

exists in other key genes and provide a starting point for further studies.

An extensive characterisation of minicircle sequences would also be of value in the further development of molecular markers for different *Symbiodinium* strains as has been suggested by Sampayo and co-authors (Sampayo et al. 2009). We have previously analysed the *psbA* minicircle in *Symbiodinium* with regard to its utility as a molecular marker (Barbrook et al. 2006b). We found that its coding region gave results consistent with those generated in rDNA studies. However, we found the non-coding region to be highly variable, and subsequent studies showed the non-coding region to be a useful marker for the fine-scale differentiation of *Symbiodinium* strains (e.g. LaJeunesse and Thornhill 2011). Given the value of the *psbA* minicircle as a taxonomic marker it seemed important to find out whether other gene minicircles could be employed in this way as well. Some of the other minicircles may have characteristics that are even more suitable than the *psbA* minicircle. For example, in previously studied dinoflagellates the non-coding region of the *psbA* minicircle is often longer than those of the other minicircles and less suitable for easy PCR amplification and sequencing. For instance, the *psbA* minicircle characterised by Barbrook et al. from a clade B *Symbiodinium* had a non-coding region close to 1.5 kbp. In both *Amphidinium* and *H. triquetra* the non-coding regions of the *psaA* and *psaB* minicircles were considerably shorter than the *psbA* minicircle (Barbrook et al. 2006a; Zhang et al. 1999). Although the shorter length of these non-coding regions may result in their offering less taxonomic resolution, the easier PCR amplification and sequencing may make them more useful.

In this paper we present the first extensive set of minicircle sequences from a *Symbiodinium* sp. isolate belonging to clade C3. We find that the gene content is similar to that found in other peridinin-containing dinoflagellates. Although the minicircles from this *Symbiodinium* isolate share many features in common with other dinoflagellates, numerous novel features are also apparent. The results indicate that minicircles other than the *psbA* minicircle may be useful for assessing *Symbiodinium* diversity.

Results

PCR Amplification of Minicircles

We used three strategies to characterise gene-containing minicircles from a DNA extract of in hospite *Symbiodinium*. Firstly degenerate primers were designed based on conserved regions of PsbA, PsbC, PsbD, PetB and chloroplast 23S ribosomal DNA (cp23S rDNA) as determined from alignments based on algal sequences with particular reference made to dinoflagellate sequences (Fig. 1a, primers labelled 1). PCR using these primers was carried out using the sample from *Agaricia* sp. DNA products of the expected size for the various primer pairs were generated and sequenced. BLAST searches indicated that the products were derived from the expected genes. Using these sequences new 'outward' primers were designed that were expected to amplify the remainder of the minicircle sequences (Fig. 1a, primers labelled 2). In PCRs which used these 'outward' primers products were again successfully obtained. Sequencing of these products showed that the non-coding regions of the *psbA*, *psbC*, *psbD*, *petB* and cp23S rDNA minicircles contained sequences of approximately 500 bp in length that shared >70% identity with the equivalent non-coding regions reported previously for *psbA* minicircles of phylotype C (Barbrook et al. 2006b). Thus, as has been reported for other dinoflagellate species, the non-coding regions of minicircles containing different genes have extensive regions, which have high levels of identity within a species, which we designate as the 'core' region (Howe et al. 2008). Outside of the 'core' region only limited identity is observed between minicircles and is typically only shared between pairs of them.

In our second approach, primers were designed based on stretches of the non-coding region that had high percentage identity across the different gene-encoding minicircles (Fig. 1b, primers labelled 1). In PCRs with these primers, multiple products were generated ranging in size from 1.4-2.8 kbp. These products were gel purified and subsequently cloned. The DNA sequence of individual cloned PCR products was determined. In addition to sequences corresponding to previously characterised genes for proteins involved in photosynthesis (*petB*, *psbA*, *psbC*, *psbD*), from which the primer sequences had been designed, further sequences for such genes were obtained, namely *atpA*, *atpB*, *psaA*, *psbB* and chloroplast 16S ribosomal DNA (cp16S rDNA). To obtain complete minicircle sequences of the newly acquired genes, outward facing primers were designed corresponding to the 5′ and 3′ coding regions of the genes (Fig. 1b, primers labelled 2). These primers were used successfully to amplify the complete non-coding regions of the minicircles of the respective genes, which were subsequently sequenced.

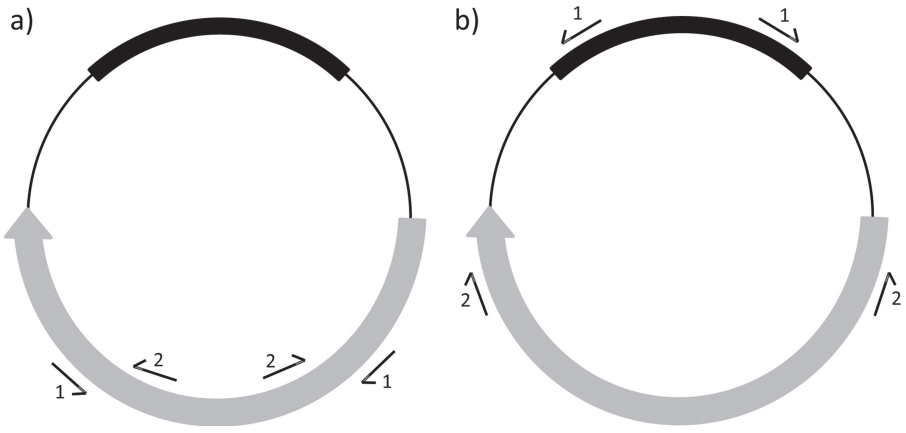

Figure 1. PCR amplification of minicircles. **a**) initial amplification of coding sequence (primers labelled 1) followed by 'outward' amplification (primers labelled 2), **b**) initial amplification using non-coding region primers (primers labelled 1) followed by 'outward' amplification from 5′ and 3′ termini of genes (primers labelled 2). Grey shaded thick stretches with arrow designate gene-coding regions, black thick stretches designate highly conserved non-coding regions ('core' region).

Minicircle sequences from other dinoflagellate species suggested that the *psaB*, *psbE*, *petD*, *psbI*, *rpl28*, *rpl33*, *ycf16* and *ycf24* genes might also be present on minicircles in *Symbiodinium*. However, these genes had not been identified thus far. In the third strategy we made use of transcriptome data, which had been generated from two *Symbiodinium* isolates (Bayer et al. 2012). We searched the EST data and the NCBI TSA database for sequences corresponding to *psaB*, *psbE*, *petD*, *psbI*, *rpl28*, *rpl33*, *ycf16* and *ycf24*. We identified sequences for *psaB*, *psbE*, *petD*, *rpl28*, *rpl33*, *ycf16* and *ycf24* but not *psbI*. The *psbI* sequence has only been identified in one dinoflagellate species, *A. carterae*, and as well as being short is highly divergent (Nisbet et al. 2004). It is therefore unsurprising that we failed to identify sequences corresponding to it. Numerous features of the EST sequences for *rpl28*, *rpl33*, *ycf16* and *ycf24* indicated that in *Symbiodinium* they were located in the nucleus rather than the chloroplast (codon usage, presence of organellar targeting sequences in inferred protein sequence, polyA tail addition, low sequence identity to putative minicircular homologues) and were not consequently further investigated. Using the *Symbiodinium* EST sequences for *psaB*, *psbE* and *petD*, as well as any other dinoflagellate sequences for these genes that were available, we designed primers that would amplify coding regions of these genes. PCRs were carried out that successfully amplified coding regions of all three genes (as Fig. 1a, primers labelled 1). These sequences were then used to design 'outward' primers for the genes (as Fig. 1a, primers labelled 2). PCRs with these primers were then able to amplify the remainder of the three minicircles, which included the conserved non-coding regions.

Minicircle Genome Characteristics

The *Symbiodinium* minicircles had a mean GC content of 41.5% (Table 1). This is mid-way between the percentage GC contents of the plastid minicircles of the two other well characterised dinoflagellate species. However, unlike both *H. triquetra* and the *Amphidinium* species, the percentage GC content of the coding regions (36.83%) is much lower than the non-coding regions (48.24%). This appears to be a consequence of the very high incidence of short inverted repeat sequences in the *Symbiodinium* non-coding region, which are highly GC-rich. The organisation and diversity of the non-coding region will be described in further detail below.

The total length of the identified genome is 27,293 base pairs (Table 1). This is shorter than the total lengths of the characterised genomes of *A. carterae* (both) and *H. triquetra*. However, the larger sizes of the other genomes are primarily due to the existence of numerous 'empty' and chimeric minicircles from these species (Howe et al. 2008; Zhang et al. 2001). We did not identify any of this type of minicircle in *Symbiodinium*. The first and third strategies employed for determination of the minicircle sequences from *Symbiodinium* would not be expected to lead to the discovery of any 'empty' minicircles. However, the second strategy should have amplified 'empty' minicircles assuming their non-coding regions were of sufficient identity at the primer binding sites. Chimeric minicircles should

Table 1. Type and size (bp) of minicircle sequences in the chloroplast genomes studied (* partial cp16S rRNA sequence).

	A. carterae CS21	*A. carterae* CCAP 1102/6	*H. triquetra*	*Symbiodinium* sp. clade C3
total length (bp)	45,815	34,186	42,769	27,293
mean GC content	45.27%	46.46%	37.02%	41.50%
gene minicircles				
psaA	2,558	2,443	3,005	2,788
psaB	2,366	2,363	3,121	2,741
psbA	2,520	2,311	2,151	1,914
psbB	2,327	2,282	2,286	2,232
psbC	2,477	2,341	2,330	2,135
psbD/E/I	2,358	2,369		
psbD			2,644	1,880
psbE			2,214	1,309
petB/atpA	2,606	2,713		
petB			2,204	1,722
atpA			2,444	2,213
petD	2,563	2,416	2,177	1,419
atpB	2,587	2,483	-	2,691
cp23S rRNA	2,713	2,655	3,027	2,839
cp16S rRNA	2,553	2,458	2,563	1,420*
'empty' minicircles				
No.	10	5	1	0
total length of empty minicircles	18,196	7,367	2,012	0
chimeric minicircles	0	0	5	0
total length of chimeric minicircles	0	0	10,625	0
total length excluding empty and chimeric minicircles	27,619	26,819	30,132	27,293

have been amplified by both the first and second strategies, but would have been unlikely in the third strategy as the primer binding sites were located close to both the 5′ and 3′ ends of the genes, one or other of which are usually absent from the chimeric minicircles. Overall our results suggest that empty or chimeric minicircles are less prevalent in *Symbiodinium*, but their presence cannot be excluded. When the empty minicircles are excluded from comparisons of genome length, *Symbiodinium* and the *A. carterae* strains are remarkably similar (Table 1), despite there being significant variation in the sizes of the individual minicircles. The *Amphidinium* minicircles are considerably more uniform in length than those of *Symbiodinium* (Table 1). There is only a 431 bp difference between the largest and smallest gene-containing minicircles in the *Amphidinium* strains, whilst the corresponding figure is 1,530 bp in *Symbiodinium*. This is also a much wider range than found in *H. triquetra*. The wider range is primarily due to the existence of short minicircles in *Symbiodinium* (namely those encoding PsbE, PetD and cp16S rRNA). The shortest minicircle in *Symbiodinium* contains *psbE*. On previously described minicircles *psbE* occurs alongside other genes, either other photosystem II genes in *Amphidinium* or tRNA genes in *Heterocapsa*. However, in *Symbiodinium* we found no evidence for any co-localised genes. If one subtracts the core region sequences this means there is considerably more non-coding than coding DNA on the minicircle (234 coding base pairs versus 596 non-coding base pairs). In previous studies we have suggested that genes such as *psaA* might be shortened by internal deletions to allow them to be accommodated within a 'typical' sized minicircle (Barbrook and Howe 2000; Howe et al. 2008). In *Symbiodinium* where the minicircle sizes are much more variable, similar patterns of deletion are observed, suggesting that gene length alteration is unlikely to be a consequence of minicircle size constraints.

Minicircle Gene Content

All the *Symbiodinium* minicircles appear to contain either a single protein-encoding gene or a ribosomal RNA gene. This is similar to the majority of minicircles that have been characterised in other peridinin-containing dinoflagellates. In *Amphidinium carterae* and *Adenoides eludens* minicircles have been identified that contain more than one protein-coding gene. We found no evidence for such minicircles in *Symbiodinium*. In other species tRNA genes have been found in close association with protein or ribosomal genes. Again we did not find this arrangement on the *Symbiodinium* minicircles. In fact somewhat surprisingly, no tRNA genes were detected on any minicircles. We were able to amplify sequences corresponding to all the genes that we expected to be present on the chloroplast genome including *atpB*, which was not reported in *H. triquetra*. We anticipate that one or more further cp16S rRNA minicircles exist, for reasons we discuss later. The coding content of the *Symbiodinium* minicircles therefore appears to be largely consistent with other studies. We did not detect any ORFs similar to those discovered in the genomes of *A. carterae* (Barbrook et al. 2006a). For most minicircles the obvious gene coding region plus core region covers virtually all of the sequence. However, as implied above, some of the short minicircles, which contain small genes, have stretches of sequence with no apparent function. No sizeable ORFs were determined in these regions. RT-PCR experiments might reveal if some of these regions are transcribed. RT-PCR experiments employing polyA cDNA primers have previously been carried out in *Amphidinium* to verify the existence of such ORFs (Barbrook et al. 2012).

Minicircle Codon Usage

The low GC content of the coding regions is reflected in the *Symbiodinium* codon usage (see Supplementary Table S2). In cases where an amino acid is encoded by multiple possible codons those with a higher AT content are used more frequently. We have previously noted the biased codon preferences of minicircle genes in *A. carterae* and *H. triquetra* (Howe et al. 2003). However, in these species an inconsistent pattern was observed, with some codons which are more GC rich preferred over AT rich ones. The *Symbiodinium* preferences are actually much more similar to those found in more typical plastid genomes. The fact that in the three well characterised dinoflagellate minicircular genomes three differing patterns have been observed seems to indicate minicircle codon preference is highly plastic amongst the peridinin-containing dinoflagellates.

There are also relatively few rare codons in the *Symbiodinium* minicircles studies. Just 8 codons are observed 10 times or fewer (out of 4512 total codons). This is comparable to the *Amphidinium* species in which 10 codons are observed to be correspondingly rare, whereas *Heterocapsa* has 20 instances of rare codons. The sets of rare codons are not highly overlapping among the three species (Fig. 2). Interestingly, *Symbiodinium* uses TGA and TAA exclusively as Stop codons whereas both *H. triquetra* and *A. carterae* use TAA and TAG. In most algal chloroplast genomes a low percentage of Stop codons are TGA, and they are not used at all in *Toxoplasma gondii* (Meurer et al. 2002). The pattern of Stop codons observed on dinoflagellate minicircles is consistent with a requirement for termination factor pRF1 in *H. triquetra* and *A. carterae*, but pRF2 in *Symbiodinium* (Duarte et al. 2012). Again whilst translation termination appears to have been simplified in dinoflagellate chloroplasts an inconsistent or diverse pattern is observed among different taxa.

In two cases, for PsaA and PetD, alternative Start codons [Ile (ATT) and Ile (ATA) respectively] are predicted to be used (see Supplementary Figs S1, S2). Alternative Start codons, including these, have been invoked for numerous minicircle genes from a range of species.

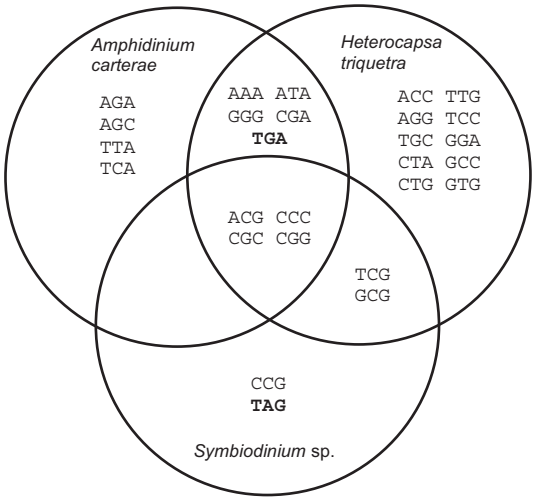

Figure 2. Venn Diagram showing the distribution of rare codons between dinoflagellate species. These codons are found 10 or fewer times in combined minicircle coding regions within a species. STOP codons which are never used within a species are also shown. STOP codons are shown in bold.

Phylogenetics and Sequence Diversity

We generated simple phylogenetic networks using each of the inferred *Symbiodinium* minicircle protein sequences along with all available homologous dinoflagellate sequences and representative sequences from a wide range of algal lineages including cyanobacteria. The networks were assessed for any unexpected phylogenetic associations amongst the minicircle derived sequences. For all protein networks all of the peridinin-containing dinoflagellate sequences grouped as a monophyletic group, albeit with long edge lengths in comparison to other eukaryotic algal sequences. None of the dinoflagellate sequences grouped closely with the other eukaryotic algal or cyanobacterial sequences. A striking feature apparent from these networks is the increased rate of evolution that has occurred within the plastids of the dinoflagellate lineage (Fig. 3). This accelerated evolution has been noted in previous studies (Bachvaroff et al. 2006; Zhang et al. 2000). It is also apparent that the minicircle proteins are not evolving at a uniform rate. Whilst the core proteins of the PSII complex, D1 and D2, show the highest degree of protein conservation relative to other algal sequences, this is still much lower than is typically found in all other eukaryotic algae (Fig. 3). For example, there is more protein sequence diversity amongst known dinoflagellate minicircle PsbA sequences than between a similar sized set of taxa comprising a rhodophyte, a cryptophyte, a haptophyte, a diatom, a euglenophyte, a chlorophyte and a glaucophyte, as shown by the differences in the number of constant amino acids in these groups (see Supplementary Table S3). This high level of diversity is also seen in components of other photosynthetic complexes, even though relatively few examples of dinoflagellate sequences are known (see Supplementary Table S3). The core proteins of PSI are highly divergent, which is surprising given their crucial role in the photosynthetic electron transfer chain. Closer analysis of the protein alignments of PsaA and PsaB shows that this diversity is not equally spread over the length of the proteins. Whilst the C-terminal portion

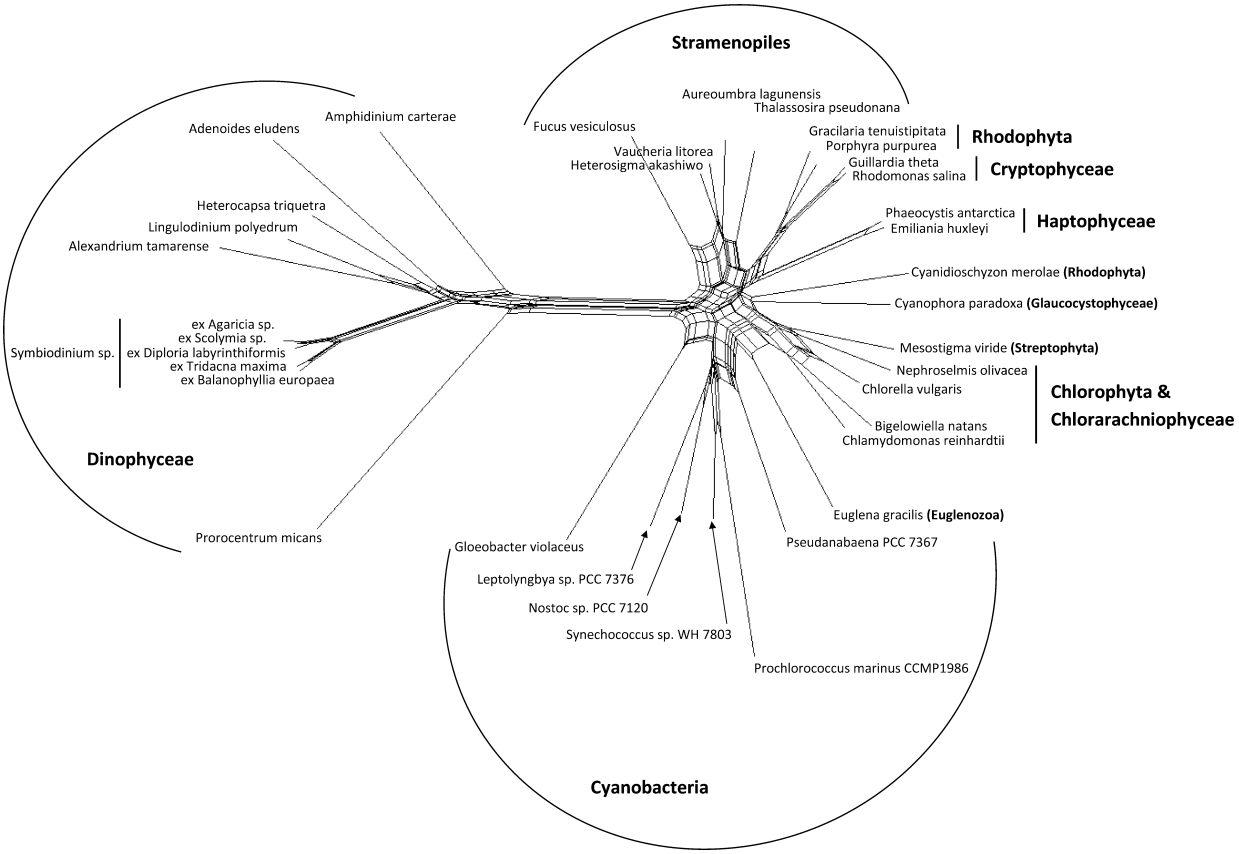

Figure 3. Phylogenetic network constructed from simple protein sequence distances for the Photosystem II D1 protein (PsbA). Taxonomic groupings of dinoflagellates and other photosynthetic protists and cyanobacteria are shown. Taxonomic assignments and nomenclature according to NCBI Taxonomy database.

of both proteins is conserved to a similar extent to D1 and D2 the N-terminus is highly divergent. This pattern of evolution of the core proteins of PSI has previously been noted in certain cyanobacteria, notably members of the genus *Prochlorococcus* (Ting et al. 2009; van der Staay et al. 2000). As in some cyanobacteria, in dinoflagellates the regions which show highest divergence comprise the light-harvesting domains of the PSI. It is possible that this may indicate a change in the light-harvesting properties of this complex. In consequence, differences in light-harvesting properties between different *Symbiodinium* species might comprise part of the physiological differences that these dinoflagellates exert on coral holobiont physiology. Overall the dinoflagellate sequences show a large decrease in the percentage of constant sites (Supplementary Table S3), implying a loss of functional constraints on the protein structure. These are key hallmarks of covariotide evolution of sequences in which some positions in sequences are invariant in certain phylogenetic lineages but not in others (Lockhart et al. 1998).

Inferred Protein Sequences

Looking in detail at the alignments of the dinoflagellate proteins compared with representatives of the major algal groups, some notable features are apparent. The ATP synthase α and β subunits show considerable variation at the N-terminal ends. These sequences correspond to the HAS barrel domain of the subunits, which is a β –barrel domain that forms the 'crown' of the F1 complex. Whilst the domain does not appear to be critical for the catalytic function of the ATPase, it is postulated to be important in the formation of critical interactions that are required for the assembly of the complex (Bakhtiari et al. 1999). We have previously noted the existence of a large insertion (41 amino acids) in the *A. carterae* AtpB sequence within this region (Barbrook and Howe 2000). We find that the *Symbiodinium* sequence has an even larger insertion (82 amino acids) at precisely the same location, which is shortly after the poorly conserved HAS-barrel domain. Other *Symbiodinium* sequences have indels relative to other typical algal sequences. For example, the chloroplast cytochrome *b* (PetB) sequence has a 4 amino acid insertion in a luminal loop region, which would be predicted to lie close to the Q_o site. The sequences of the core proteins of PSI in *Symbiodinium* possess several indels in comparison to typical eukaryotic algal sequences (Table 2). All these indels are located in loop regions between the transmembrane helices and are found in both stromal and luminal loops. The majority of the indels are deletions and the most extensive deletions, >10 amino acids, are found between the same transmembrane helices in both PsaA and PsaB (see Table 2). Sequence length variation in the core components of PSI has been previously reported in a number of cases, notably in cyanobacteria (Vanselow et al. 2009) and most extensively in several species of *Prochlorococcus* (Ting et al. 2009; van der Staay et al. 2000), as well as in dinoflagellates (Barbrook and Howe 2000). In other dinoflagellates PsaA and PsaB sequences have frequent indels which often map to the same positions as the ones found in *Symbiodinium*. However, the lengths of the indels and the sequence similarity surrounding them are highly variable, suggesting that these may be mutational hotspots within the coding region. This suggests that sequences containing these regions might constitute good candidates for molecular markers.

Table 2. Size and position of indels relative to typical algae sequences within PsaA and PsaB. Stromal loops shaded grey. "+" indicates insertion and "-" indicates deletion, numerical value indicates number of amino acids inserted or deleted.

loop	PsaA	PsaB
N-terminal	-9	+17
I-II		
II-III		+20
III-IV	-19	-41
IV-V	-19	-12
V-VI		
VI-VII	-9	
VII-VIII	-17	-18
VIII-IX		
IX-X		-5
X-XI		+2
C-terminal		-4

A striking feature of the coding sequences of the two components of the cytochrome b_6f complex is the presence of an in-frame Stop codon in each. For the PetB polypeptide this is at amino acid position 136 out of 218 and in the PetD polypeptide at position 32 out of 158 (see Supplementary Fig. S2). In both cases this is a TGA codon. TGA is used as a Stop codon at 6 out of 11 predicted translation termination sites for the *Symbiodinium* minicircle genes, so it seems unlikely that the internal Stop codons would be read as sense codons. We assume that the Stop codons are edited to sense codons to allow expression of these proteins. In support of this, the published *Symbiodinium*

transcriptome data lack stop codons at these positions (although the sequences were derived from A and B phylotype strains). Additionally, the *Lingulodinium polyedrum petB* gene also has a stop codon (although in a different position in the sequence) that is modified to a sense codon in the mRNA (Wang and Morse 2006b, Supplementary Data).

Ribosomal RNA Genes

In addition to the protein coding genes we identified two rRNA genes on minicircles. The cp16S rRNA gene is apparently not a full-length sequence. The section we identified is approximately 300 base pairs in length and shares 70% sequence identity with the *H. triquetra* cp16S rRNA sequence. The *H. triquetra* cp16S rRNA gene is highly divergent and has been shown to be fragmented into five pieces (Dang and Green 2009). The *Symbiodinium* sequence corresponds to the first two fragments of the *H. triquetra* cp16S rRNA. Whereas in *H. triquetra* all the pieces that comprise the cp16S rRNA are found adjacent to one another, and in the expected order, it appears in *Symbiodinium* fragments of the cp16S rRNA are localised to separate minicircles. In contrast, the cp23S rRNA appears to be encoded in a continuous sequence on a single minicircle. The *Symbiodinium* and *H. triquetra* cp23S rRNAs show a similar percentage identity to the cp16S rRNAs (≈70%). This high level of identity was observed over a stretch of 1800 base pairs on the minicircle. A somewhat lower percentage identity was observed beyond this region. Given these high levels of sequence identity and other similarities to the rRNAs found in *H. triquetra* we believe the *Symbiodinium* sequences represent functional copies of the chloroplast rRNAs, albeit we anticipate further fragments of the cp16S rRNA to be found.

Minicircle Non-coding Regions

The non-coding regions of the minicircles each contain conserved 'core' sequences of approximately 500 base pairs in this *Symbiodinium* strain. Conserved non-coding regions of *psbA* minicircles from phylotype C have previously been described in some detail (Barbrook et al. 2006b; Moore et al. 2003; Thornhill et al. 2013). The conserved regions in this strain are broadly similar to these. As in other dinoflagellate species they show high levels of identity between the different minicircles, although this varies according to individual minicircles with some pairs of minicircles displaying higher pairwise identity than others. Comparison of the non-coding region between minicircles from this isolate and also with the previously identified *psbA* minicircles further highlights the most conserved sequences. The region previously described as C4 (Barbrook et al. 2006b; Moore et al. 2003) is the most highly conserved (Fig. 4a). The region contains three inverted repeats followed by a $A_{5\text{-}6}$-N-$A_{4\text{-}6}$ motif (with the exception of the *psbE* sequence where the second run of As is interrupted by a C). The low level of nucleotide variation observed invariably maintains the integrity of the inverted repeat sequences. Additionally, it is noticeable that within the C4 sequence identity is maintained within the unpaired regions in the centre of the inverted repeats. Typically in the many inverted repeats (>8) that are found in non-coding regions these unpaired nucleotides are the most variable (Barbrook et al. 2006b). A feature common to the majority of non-coding sequences between the end of the C4 region and the start of the coding regions is a stretch of around 24 base pairs that is highly C-rich (typically 75%) (Fig. 4b). This is found on all minicircles apart from the *petD* and *psbE* ones. The highly conserved P region that immediately preceded the coding region on *psbA* minicircles (Moore et al. 2003) is not conserved among other minicircles and it seems that this may be a gene specific feature, possibly involved in *psbA* transcript regulation. Given its close proximity to the coding region it may be included in the *psbA* transcripts and form part of the 5′ untranslated region.

Since dinoflagellates are believed to harbour multiple copies of a given minicircle (Koumandou and Howe 2007), it is possible that single *Symbiodinium* cells could harbour numerous divergent copies. The study of LaJeunesse and Thornhill (2011) suggests that, at least for the non-coding region of *psbA* minicircles, intragenomic variation does exist albeit at low levels, although the extent of this may vary according to lineage. In our PCR reactions we did not encounter sequence heterogeneity on gene specific minicircles, with one exception, even though the initial DNA preparation was from an environmental sample. For example, multiple bands were not visible in gene specific minicircle amplifications, either of coding or non-coding sequences. In cases where we sequenced multiple independent clones of the same minicircle we found sequences to be identical, excepting occasional single base differences. Such levels of variation have been shown to be readily generated by methodological artefacts associated with PCR and bacterial cloning (LaJeunesse and Thornhill 2011). Our results indicate either a low level or an absence of intra-colony and intra-genomic sequence variation for all but one type of minicircle in this isolate. The exception to

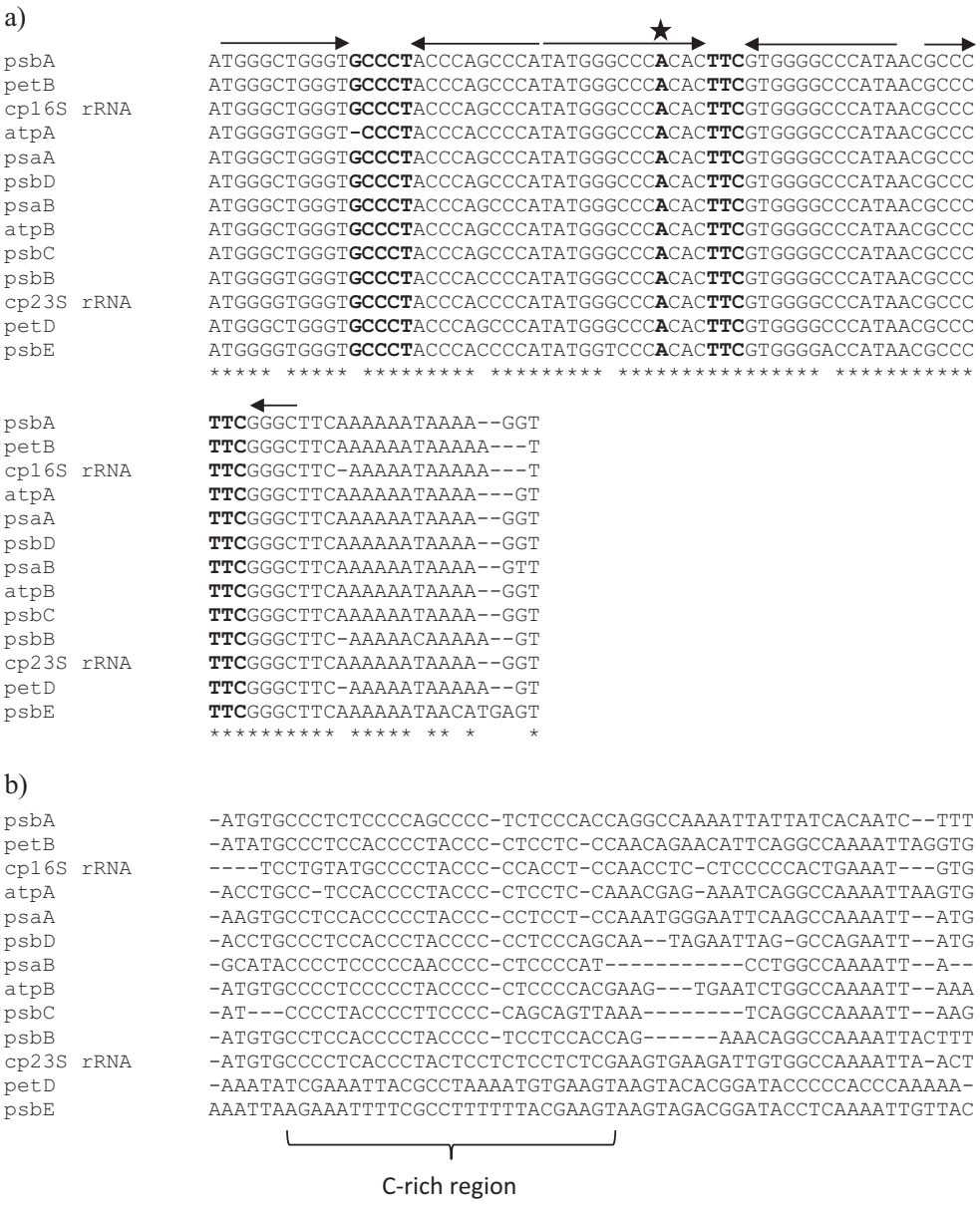

Figure 4. Selected regions of an alignment of *Symbiodinium* ex *Agaricia* sp. non-coding minicircle sequences. **a**) C4 region of 'core' sequences. Arrows indicate inverted repeats. Star shows a single nucleotide within a repeat that does not have a potential base pair. Bold bases indicate unpaired nucleotides between complementary repeat sequences. **b**) C-rich nucleotide sequence. Sequence is found typically <100 bp prior to predicted coding start sites.

this observation was the cp16S rRNA minicircle. We cloned numerous fragments of this minicircle following PCR amplification using primers complementary to the 'core' region. The sequences were identical except for the presence of variable numbers of copies, ranging from six to twelve in different clones, of a repetitive sequence, 5′-TGTGCCCACATGCCTGCCC-3′, which was found outside of the 'core' in the non-coding region of the minicircle. We did not detect this repeat in any other minicircle sequences.

Discussion

We found that the chloroplast genome of our *Symbiodinium* sp. clade C3 isolate is constituted by single-gene minicircles. The size of these

minicircles is largely similar to the majority of minicircles so far identified, although a somewhat wider range of sizes is observed. We did not detect numerous variant or chimaeric forms of minicircles as have been identified in some other species (Howe et al. 2008; Zhang et al. 2001). This *Symbiodinium* chloroplast genome appears to possess a number of distinctive features; for example, the absence of the Stop codon TAG, the inclusion of Stop codons within the gene sequences of the cytochrome b_6f complex components, and the complex architecture of the minicircle core region with its many inverted repeats. There also appears to be a low incidence of 'empty' minicircles. In other respects the genome conforms to characters found in other peridinin-containing dinoflagellates. Characterisation of further *Symbiodinium* isolates from other phylotypes will be necessary to confirm that these are general features found throughout the genus. We believe that sequences we have obtained will be of use in the further study of *Symbiodinium*. For example, some of the unusual protein sequences may be of significance in determining susceptibility to high thermal or light exposure. Perhaps more significantly, the relatively short non-coding sequences of the *psaA* and *psaB* minicircles along with the presence of possible mutational 'hotspots' within their coding regions suggest they may be excellent candidates as molecular markers for analysing *Symbiodinium* diversity and possible targets of positive selection. This may be aided by low levels of intragenomic variation between minicircles coding for the same gene, although further sampling of a range of *Symbiodinium* phylotypes will be needed to verify this.

Methods

Sample origin and DNA extraction: The material for this study was provided as a gift by Professor Angela Douglas. The single sample of *Agaricia* sp. used was collected from Bermuda and was analysed as part of previous research on the rDNA diversity of *Symbiodinium* (Savage et al. 2002). The DNA was extracted as described by Savage et al. (2002). This study showed the zooxanthellae from the sample belonged to Phylotype C. PCR amplification and sequencing of the ITS2 region of the nuclear rDNA verified this, yielding a BLASTN hit with a 1 bp difference to the C1 or C3 subphylotype, (Acc. no. HG515026). This PCR was performed according to Sampayo et al. (2009) with the primers ITSintfor2 5′ GAATTGCAGAACTCCGTG 3′ and ITS2-reverse 5′ GGGATCCATATGCTTAAGTTCAGCGGGT 3′. We subsequently conducted a phylogenetic analysis based on the *psbA* non-coding region, as this marker has been shown previously to provide improved resolution of *Symbiodinium* species diversity (LaJeunesse and Thornhill 2011). Comparison of the *psbA* non-coding portion of our isolate to C1 and C3 sequences taken from (Thornhill et al. 2013) with MAFFT (Katoh and Standley 2013) revealed a much better alignment to C3. A phylogenetic tree constructed from these data (Neighbor Joining, 1,000 bootstraps) showed closest common ancestry with strong support of our isolate to C3.

PCR amplification of minicircles: Primers used in PCR reactions are described in Supplementary Table S1. Standard PCR conditions were an initial cycle of 95 °C for 3 minutes 15 seconds followed by 35 cycles of 95 °C for 45 seconds, an annealing temperature appropriate to primer pair for 45 seconds, 72 °C for 1 to 3 minutes and a final step of 72 °C for 10 minutes. PCR products were purified using either a QIAquick PCR Purification Kit (QIAGEN, Germany) or a MinElute Gel Extraction Kit (QIAGEN, Germany) following agarose gel electrophoresis. PCR products were either sequenced directly or cloned into pGEM-T plasmid vector (Promega, USA) and transformed into *Escherichia coli* prior to sequencing.

DNA sequencing and computational analysis of sequences: Sequencing was carried out using an Applied Biosystems 3730xl DNA Analyser. The identity of PCR products was confirmed by using BLASTN or BLASTX analysis (www.ncbi.nlm.nih.gov/BLAST). Minicircle contigs were assembled and edited using ChromasPro (Technelysium Pty Ltd, Australia). Sequences have been deposited in the EMBL database under the following accession numbers: HG515015-HG515025, HG515027 and HG515028.

Multiple alignments were generated using ClustalX2 (Larkin et al. 2007) and adjusted manually. Artemis and Artemis Comparison Tool (ACT) were used for whole genome analyses of minicircle sequences (Berriman and Rutherford 2003; Carver et al. 2005). For these analyses minicircle sequences were concatenated as linear DNA sequences. The circular sequences were linearised by breaking immediately 5′ of the coding regions. ACT was used to visualise regions of identity between species. Regions of identity were determined by pairwise BLAST. The output of the pairwise BLAST was then used as an input into ACT. Searches for tRNAs were carried out using tRNAscan-SE 1.21 (Schattner et al. 2005). Phylogenetic analysis of sequences was carried out with Splitstree (Huson and Bryant 2006). For phylogenetic analyses of protein sequences regions including gaps were excluded. Distances used were simple protein sequence distances (uncorrected (P) distances) and networks were generated using NeighborNet (Huson and Bryant 2006).

Acknowledgements

The authors wish to thank Professor Angela Douglas for the gift of the DNA extract used in this study. The authors also wish to thank Daniel J. Thornhill and Todd J. LaJeunesse for providing assistance in phylogenetic assignment of the *Symbiodinium* isolate. This work was funded through a grant from the Cambridge - KAUST Academic Excellence Alliance (AEA) program (award number 7000000056) as well as by the Leverhulme Trust.

Appendix A. Supplementary Data

Supplementary material related to this article can be found, in the online version, at http://dx.doi.org/10.1016/j.protis.2013.09.006.

References

Bachvaroff TR, Sanchez-Puerta MV, Delwiche CF (2006) Rate variation as a function of gene origin in plastid-derived genes of peridinin-containing dinoflagellates. J Mol Evol **62**:42–52

Bachvaroff TR, Concepcion GT, Rogers CR, Herman EM, Delwiche CF (2004) Dinoflagellate expressed sequence tag data indicate massive transfer of chloroplast genes to the nuclear genome. Protist **155**:65–78

Barbrook AC, Howe CJ (2000) Minicircular plastid DNA in the dinoflagellate *Amphidinium operculatum*. Mol Gen Genet **263**:152–158

Barbrook AC, Visram S, Douglas AE, Howe CJ (2006b) Molecular diversity of dinoflagellate symbionts of Cnidaria: the *psbA* minicircle of *Symbiodinium*. Protist **157**:159–171

Barbrook AC, Santucci N, Plenderleith LJ, Hiller RG, Howe CJ (2006a) Comparative analysis of dinoflagellate chloroplast genomes reveals rRNA and tRNA genes. BMC Genomics **7**:297

Barbrook AC, Symington H, Nisbet RER, Larkum A, Howe CJ (2001) Organization and expression of the plastid genome of the dinoflagellate *Amphidinium operculatum*. Mol Genet Genomics **266**:632–638

Barbrook AC, Dorrell RG, Burrows J, Plenderleith LJ, Nisbet RE, Howe CJ (2012) Polyuridylylation and processing of transcripts from multiple gene minicircles in chloroplasts of the dinoflagellate *Amphidinium carterae*. Plant Mol Biol **79**:347–357

Bayer T, Aranda M, Sunagawa S, Yum LK, Desalvo MK, Lindquist E, Coffroth MA, Voolstra CR, Medina M (2012) *Symbiodinium* transcriptomes: genome insights into the dinoflagellate symbionts of reef-building corals. PLoS One **7**:e35269

Berriman M, Rutherford K (2003) Viewing and annotating sequence data with Artemis. Brief Bioinform **4**:124–132

Carver TJ, Rutherford KM, Berriman M, Rajandream MA, Barrell BG, Parkhill J (2005) ACT: the Artemis Comparison Tool. Bioinformatics **21**:3422–3423

Dang Y, Green BR (2009) Substitutional editing of *Heterocapsa triquetra* chloroplast transcripts and a folding model for its divergent chloroplast 16S rRNA. Gene **442**:73–80

Duarte I, Nabuurs SB, Magno R, Huynen M (2012) Evolution and diversification of the organellar release factor family. Mol Biol Evol **29**:3497–3512

Field CB, Behrenfeld MJ, Randerson JT, Falkowski P (1998) Primary production of the biosphere: integrating terrestrial and oceanic components. Science **281**:237–240

Hackett JD, Yoon HS, Soares MB, Bonaldo MF, Casavant TL, Scheetz TE, Nosenko T, Bhattacharya D (2004) Migration of the plastid genome to the nucleus in a peridinin dinoflagellate. Curr Biol **14**:213–218

Hiller RG (2001) 'Empty' minicircles and *petB*/*atpA* and *psbD*/*psbE* (*cyt*b_{559} alpha) genes in tandem in *Amphidinium carterae* plastid DNA. FEBS Lett **505**:449–452

Howe CJ, Barbrook AC, Koumandou VL, Nisbet RE, Symington HA, Wightman TF (2003) Evolution of the chloroplast genome. Philos Trans R Soc Lond B Biol Sci **358**:99–106

Howe CJ, Nisbet RE, Barbrook AC (2008) The remarkable chloroplast genome of dinoflagellates. J Exp Bot **59**:1035–1045

Hughes TP, Baird AH, Bellwood DR, Card M, Connolly SR, Folke C, Grosberg R, Hoegh-Guldberg O, Jackson JB, Kleypas J, Lough JM, Marshall P, Nyström M, Palumbi SR, Pandolfi JM, Rosen B, Roughgarden J (2003) Climate change, human impacts, and the resilience of coral reefs. Science **301**:929–933

Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. Mol Biol Evol **23**:254–267

Iida S, Kobiyama A, Ogata T, Murakami A (2008) The D1 and D2 proteins of dinoflagellates: unusually accumulated mutations which influence on PSII photoreaction. Photosynth Res **98**:415–425

Iida S, Kobiyama A, Ogata T, Murakami A (2009) Identification of transcribed and persistent variants of the *psbA* gene carried by plastid minicircles in a dinoflagellate. Curr Genet **55**:583–591

Iida S, Kobiyama A, Ogata T, Murakami A (2010) Differential DNA rearrangements of plastid genes, *psbA* and *psbD*, in two species of the dinoflagellate *Alexandrium*. Plant Cell Physiol **51**:1869–1877

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol **30**:772–780

Koumandou VL, Howe CJ (2007) The copy number of chloroplast gene minicircles changes dramatically with growth phase in the dinoflagellate *Amphidinium operculatum*. Protist **158**:89–103

Laatsch T, Zauner S, Stoebe-Maier B, Kowallik KV, Maier UG (2004) Plastid-derived single gene minicircles of the dinoflagellate *Ceratium horridum* are localized in the nucleus. Mol Biol Evol **21**:1318–1322

LaJeunesse TC, Thornhill DJ (2011) Improved resolution of reef-coral endosymbiont (*Symbiodinium*) species diversity, ecology, and evolution through *psbA* non-coding region genotyping. PLoS One **6**:e29013

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2. 0. Bioinformatics **23**:2947–2948

Lee RE (2008) Phycology. 4th edition Cambridge University Press, UK

Lin S (2011) Genomic understanding of dinoflagellates. Res Microbiol **162**:551–569

Lockhart PJ, Steel MA, Barbrook AC, Huson DH, Charleston MA, Howe CJ (1998) A covariotide model explains apparent phylogenetic structure of oxygenic photosynthetic lineages. Mol Biol Evol **15**:1183–1188

McGinley MP, Aschaffenburg MD, Pettay DT, Smith RT, LaJeunesse TC, Warner ME (2012) Transcriptional response of two core photosystem genes in *Symbiodinium* spp. exposed to thermal stress. PLoS One **7**:e50439

Moore RB, Ferguson KM, Loh WKW, Hoegh-Guldberg O, Carter DA (2003) Highly organised structure in the non-coding

region of the *psbA* minicircle from clade C *Symbiodinium*. Int J Sys Evol Microbiol **53**:1725–1734

Moszczynski K, Mackiewicz P, Bodyl A (2012) Evidence for horizontal gene transfer from bacteroidetes bacteria to dinoflagellate minicircles. Mol Biol Evol **29**:887–892

Nash EA, Nisbet RER, Barbrook AC, Howe CJ (2008) Dinoflagellates: a mitochondrial genome all at sea. Trends Genet **24**:328–335

Nelson MJ, Green BR (2005) Double hairpin elements and tandem repeats in the non-coding region of *Adenoides eludens* chloroplast gene minicircles. Gene **358**:102–110

Nelson MJ, Dang Y, Filek E, Zhang Z, Yu VW, Ishida K, Green BR (2007) Identification and transcription of transfer RNA genes in dinoflagellate plastid minicircles. Gene **392**:291–298

Nisbet RER, Koumandou VL, Barbrook AC, Howe CJ (2004) Novel plastid gene minicircles in the dinoflagellate *Amphidinium operculatum*. Gene **331**:141–147

Rowan R, Powers DA (1991) A molecular genetic classification of zooxanthellae and the evolution of animal-algal symbioses. Science **251**:1348–1351

Roy S, Morse D (2012) A full suite of histone and histone modifying genes are transcribed in the dinoflagellate *Lingulodinium*. Plos One **7**:e34340

Sampayo EM, Dove S, Lajeunesse TC (2009) Cohesive molecular genetic data delineate species diversity in the dinoflagellate genus *Symbiodinium*. Mol Ecol **18**:500–519

Santos SR, Taylor DJ, Kinzie RA III, Hidaka M, Sakai K, Coffroth MA (2002) Molecular phylogeny of symbiotic dinoflagellates inferred from partial chloroplast large subunit (23S)-rDNA sequences. Mol Phylogenet Evol **23**:97–111

Savage AM, Goodson MS, Visram S, Trapido-Rosenthal H, Weidenmann J, Douglas AE (2002) Molecular diversity of symbiotic algae at the latitudinal margins of their distribution: dinoflagellates of the genus *Symbiodinium* in corals and sea anemones. Mar Ecol Prog Ser **244**:17–26

Schattner P, Brooks AN, Lowe TM (2005) The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. Nucleic Acids Res **33**:W686–W689

Ting CS, Ramsey ME, Wang YL, Frost AM, Jun E, Durham T (2009) Minimal genomes, maximal productivity: comparative genomics of the photosystem and light-harvesting complexes in the marine cyanobacterium, *Prochlorococcus*. Photosynth Res **101**:1–19

Thornhill DJ, Lewis AM, Wham DC, LaJeunesse TC (2013) Host-specialist lineages dominate the adaptive radiation of reef coral endosymbionts. Evolution, http://dx.doi.org/10.1111/evo.12270

van der Staay GW, Moon-van der Staay SY, Garczarek L, Partensky F (2000) Rapid evolutionary divergence of Photosystem I core subunits PsaA and PsaB in the marine prokaryote *Prochlorococcus*. Photosynth Res **65**:131–139

Vanselow C, Weber AP, Krause K, Fromme P (2009) Genetic analysis of the Photosystem I subunits from the red alga, *Galdieria sulphuraria*. Biochim Biophys Acta **1787**:46–59

Waller RF, Jackson CJ (2009) Dinoflagellate mitochondrial genomes: stretching the rules of molecular biology. Bioessays **31**:237–245

Wang Y, Morse D (2006a) The plastid-encoded *psbA* gene in the dinoflagellate *Gonyaulax* is not encoded on a minicircle. Gene **371**:206–210

Wang Y, Morse D (2006b) Rampant polyuridylylation of plastid gene transcripts in the dinoflagellate *Lingulodinium*. Nucleic Acids Res **34**:613–619

Warner ME, Fitt WK, Schmidt GW (1999) Damage to photosystem II in symbiotic dinoflagellates: a determinant of coral bleaching. Proc Natl Acad Sci USA **96**:8007–8012

Zhang Z, Cavalier-Smith T, Green BR (2001) A family of selfish minicircular chromosomes with jumbled chloroplast gene fragments from a dinoflagellate. Mol Biol Evol **18**:1558–1565

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

Zhang Z, Green BR, Cavalier-Smith T (2000) Phylogeny of ultra-rapidly evolving dinoflagellate chloroplast genes: a possible common origin for sporozoan and dinoflagellate plastids. J Mol Evol **51**:26–40