ISEP XIV

Highly organized structure in the non-coding region of the *psbA* minicircle from clade C *Symbiodinium*

Robert B. Moore,[1] Katherine M. Ferguson,[1] William K. W. Loh,[2] Ove Hoegh-Guldberg[2] and Dee A. Carter[1]

[1]School of Molecular and Microbial Biosciences, University of Sydney, NSW 2006, Australia

[2]Centre for Marine Sciences, University of Queensland, Qld 4072, Australia

Correspondence
Dee A. Carter
d.carter@mmb.usyd.edu.au

The chloroplast genes of dinoflagellates are distributed among small, circular dsDNA molecules termed minicircles. In this paper, we describe the structure of the non-coding region of the *psbA* minicircle from *Symbiodinium*. DNA sequence was obtained from five *Symbiodinium* strains obtained from four different coral host species (*Goniopora tenuidens*, *Heliofungia actiniformis*, *Leptastrea purpurea* and *Pocillopora damicornis*), which had previously been determined to be closely related using LSU rDNA region D1/D2 sequence analysis. Eight distinct sequence blocks, consisting of four conserved cores interspersed with two metastable regions and flanked by two variable regions, occurred at similar positions in all strains. Inverted repeats (IRs) occurred in tandem or 'twin' formation within two of the four cores. The metastable regions also consisted of twin IRs and had modular behaviour, being either fully present or completely absent in the different strains. These twin IRs are similar in sequence to double-hairpin elements (DHEs) found in the mitochondrial genomes of some fungi, and may be mobile elements or may serve a functional role in recombination or replication. Within the central unit (consisting of the cores plus the metastable regions), all IRs contained perfect sequence inverses, implying they are highly evolved. IRs were also present outside the central unit but these were imperfect and possessed by individual strains only. A central adenine-rich sequence most closely resembled one in the centre of the non-coding part of *Amphidinium operculatum* minicircles, and is a potential origin of replication. Sequence polymorphism was extremely high in the variable regions, suggesting that these regions may be useful for distinguishing strains that cannot be differentiated using molecular markers currently available for *Symbiodinium*.

INTRODUCTION

Unigenic DNA minicircles of 2–3 kbp that encode plastid gene functions have been found in a number of different peridinin-containing dinoflagellates, including species of *Heterocapsa* (Zhang *et al.*, 1999, 2002), *Amphidinium operculatum* (Barbrook & Howe, 2000; Barbrook *et al.*, 2001), *Amphidinium carterae* (Hiller, 2001) and *Protoceratium reticulatum* (Zhang *et al.*, 2002). The non-coding regions of the minicircles are highly divergent across dinoflagellate genera, except that short stretches of a single nucleotide are frequent. A common format across dinoflagellate genera and species is that the non-coding regions contain conserved core regions, usually of between two and four in number, separated by variable regions (Zhang *et al.*, 2002; Howe *et al.*, 2003). In a given species, two or more cores are often identical to each other, seemingly duplicated or triplicated within the same minicircle, such as the two 9G regions of *Heterocapsa triquetra* (Zhang *et al.*, 1999, 2002), and the three 5G regions of *Heterocapsa pygmeae* (Zhang *et al.*, 2002).

It has been shown that conserved core regions are shared between all the minicircles present in one culture of

This paper was presented at the XIVth meeting of the International Society for Evolutionary Protistology in Vancouver, Canada, 19–24 June 2002.

Abbreviations: C, M and V regions, conserved, metastable and variable regions; DHE, double-hairpin element; OTI, One Tree Island; poly-A, poly-adenine; *Sym-Gt*, symbiont of *Goniopora tenuidens*; *Sym-Ha*, symbiont of *Heliofungia actiniformis*; *Sym-Pd*, symbiont of *Pocillopora damicornis*; *Tm*, *Tridacna maxima*; twin IR, twin inverted repeat.

The GenBank accession numbers for the full non-coding DNA sequences of *psbA* minicircles from *Sym-Gt1*, *-Gt2*, *-Ha*, *-Lp* and *-Pd* are AY160085–AY160089, respectively. The GenBank accession number for the complete *psbA* ORF sequence from culture *Tm*8.2 is AY160084.

Phylogenetic information on relatedness of selected OTI zooxanthellae is available as supplementary material in IJSEM Online.

Heterocapsa triquetra, implying that recombination maintains the sequences of the core regions (Zhang *et al.*, 1999, 2002). Likewise, core regions are identifiable across all the minicircles from *Amphidinium operculatum*, again implying intra-circle recombination (Barbrook & Howe, 2000; Howe *et al.*, 2003), but the sequence of these cores is not alignable with those from *Heterocapsa triquetra*. A hypothesis of species-specific conserved cores has been tested on two minicircles, *psbA* and 23S rDNA, from three further species of *Heterocapsa* and from the distantly related *A. carterae*, and was supported in each case (Zhang *et al.*, 2002). While common cores were apparent within the two sequenced minicircles of each culture, there was a lack of core homology between species of *Heterocapsa* (Zhang *et al.*, 2002). Only cores that contain inverted repeats or have a high A+T content have been assigned putative roles, the former as replication origins, recombination sites, integron-like sites, or DNA segregation loci, and the latter as replication origins (Zhang *et al.*, 1999, 2002; Barbrooke & Howe, 2000; Howe *et al.*, 2003). Inverted repeats have not before been found in cores that are present in duplicate or triplicate (Zhang *et al.*, 2002).

An area of interest in dinoflagellate ecology is the indispensable role played by symbiotic dinoflagellates (commonly termed 'zooxanthellae') in providing fixed carbon to their metazoan hosts, which include scleractinian corals, tridacnid clams and others (Goodson *et al.*, 2001; Trench, 1993; Taylor, 1983). It is desirable to assess biodiversity within zooxanthellae, in the context of dwindling coral reef habitat worldwide because of a possible role of zooxanthellae in the coral-bleaching process (Brown, 1997; Hoegh Guldberg, 1999; Walther *et al.*, 2002; Downs *et al.*, 2002; Baker, 2003; Hughes *et al.*, 2003).

Symbiodinium spp. (*gen.* Freudenthal 1962) are the dominant zooxanthellae in tropical and equatorial waters (Kevin *et al.*, 1969; Trench, 1993; Rowan, 1998). Only a handful of species have been named in the genus *Symbiodinium* [Freudenthal, 1962; Trench & Blank, 1987; Trench & Thinh, 1995 (emend. LaJeunesse, 2001); Trench, 1997], even though the genus includes numerous strains separated by large genetic distances (Rowan, 1998; LaJeunesse, 2001; Rodriguez-Lanetty, 2003; Baker, 2003). The genus is currently divided into seven highly divergent clades, designated A–G based on nuclear and plastid rDNA sequences (Rowan & Powers, 1992; Carlos *et al.*, 1999; Baillie *et al.*, 2000; LaJeunesse, 2001; Pochon *et al.*, 2001; Santos *et al.*, 2002, 2003; Rodriguez-Lanetty, 2003; Baker, 2003). Clade C is dominant in corals of the Great Barrier Reef, and studies based on rDNA sequence analysis have found considerable diversity occurs within this clade (Carter, 2000; Loh *et al.*, 2001; van Oppen *et al.*, 2001; LaJeunesse *et al.*, 2003; Rodriguez-Lanetty, 2003). Phenotypic subgroups within clades have been distinguished, based on cell size, and cell surface in culture (reviewed by LaJeunesse, 2001). It is hoped that such studies, together with DNA phylogeny will gradually yield finer classification of *Symbiodinium* spp. to approximately the species level.

We sought to develop a fine-scale DNA marker for use in *Symbiodinium* genotyping and chose the non-coding region of the *psbA* minicircle. The *psbA* gene encodes the D1 protein of photosystem II, a quinone-binding protein and regulator of photosynthetic flux, that is present in all algal phyla (Pakrasi, 1995; Warner, 1999; Singh, 2000). Takishita *et al.* (2003) established by *in situ* hybridization that *Symbiodinium psbA* mRNA is present only in the plastid. We report here on the genetic structure of the non-coding region of the *psbA* minicircle from selected *Symbiodinium* isolates obtained from different coral host species. The non-coding region features a high density of G+C-rich inverted repeats (IRs) within core sequence blocks that are conserved among zooxanthellae from different hosts. All of the G+C-rich IRs occur as tandem IR pairs with no intervening bases separating the two abutting IRs. Between closely related strains, there is a tendency for these 'twin' IRs to be present or absent as a complete module. The twin IRs may be mobile elements and may have the ability to form 'double-hairpin' structures. Each doublet appears to act as a single unit of selection.

METHODS

Sampling, DNA extraction and nomenclature. Four common coral species were chosen for the analysis of the *psbA* sequence: *Goniopora tenuidens*, *Heliofungia actiniformis*, *Leptastrea purpurea* and *Pocillopora damicornis*. Our previous analyses using rDNA LSU D1/D2 sequencing found the first three of these species housed genetically similar zooxanthellae, whereas the *P. damicornis* zooxanthellae were more distantly related (Carter *et al.*, 2000; see supplementary material in IJSEM Online). Two colonies of *Goniopora tenuidens*, one each of *Leptastrea purpurea* and *Pocillopora damicornis* and one individual polyp of *Heliofungia actiniformis* were collected from One Tree Island (OTI) atoll at the southern tip of the Great Barrier Reef, Australia. Whole tissue, including zooxanthellae, was obtained from the *P. damicornis* and *L. purpurea* colonies by blowing material from the coral skeleton, using a triggered airgun connected to a SCUBA tank as the air supply. The resultant slurry was augmented with 1 ml 20 % DMSO, 0·25 M EDTA; pH 8·0, saturated NaCl, for preservation and transport. In the case of *Heliofungia actiniformis* and *G. tenuidens*, tissue was removed by cutting off a number of long tentacles from the host into an Eppendorf tube containing 1 ml of the DMSO/EDTA/NaCl solution. Prior to DNA extraction, the preserved samples were spun in an Eppendorf tube in a bench-top microfuge (~12 000 r.p.m.), and the pellet was resuspended in TE buffer (10 mM Tris/HCl, 1 mM EDTA; pH 8·0) containing 1 % SDS. The mixture was then incubated at 65 °C for 1 h. Proteinase-K (Bioline) was subsequently added to a final concentration of 0·5 mg ml^{-1}, followed by an overnight incubation at 37 °C. Two phenol/chloroform : isoamyl-alcohol (25 : 24 : 1, by vol.) extractions and one chloroform : isoamyl-alcohol (24 : 1) extraction were then performed. The DNA was precipitated by the addition of 0·6 vols 2-propanol and centrifugation at 12 000 r.p.m. for 10 min. The pellet was washed with cold 70 % ethanol before resuspension in 30 µl sterile distilled water.

Symbiodinium nomenclature, adopted for the purposes of this paper, is that the zooxanthella genus is shortened to *Sym* and this is followed by the initials of the host coral, e.g. the symbiont of *Heliofungia actiniformis* is *Sym-Ha*, that of *P. damicornis* is *Sym-Pd*, etc.

Table 1. Oligonucleotide primers for the *psbA* minicircles of *Symbiodinium* clades A and C

Primer	Sequence	Minicircle region*	Base position
PCR primers for *Sym-Tm*8.2†			
psbAF6-Forw	5′-GARCACAACATHYTNATGCAYCC	*psbA*	565–587‡
psbAL1-Rev	5′-CRTGCATWACTTCCATWCC	"	978–997‡
4.6-Forw	5′-GGTTTAAACTTCAACCAATCCATCC	"	895–919‡
4.3-Rev	5′-CTCCTGCACTTTCTGCAWGAAG	"	667–688
PCR primers for *Sym-Ha, -Gt, -Lp, -Pd* non-coding region			
7.4-Forw	5′-GCATGAAAGAAATGCACACAACTTCCC	*psbA*	993–1019‡
7.8-Rev	5′-GGTTCTCTTATTCCATCAATATCTACTG	"	170–197‡
Internal forward sequencing primer for *Sym-Gt*			
10.11-Forw	5′-ACGAAGTGTRACCGAGAAAG	C1	286–319§
Internal forward sequencing primers for *Sym-Ha, -Lp*			
10.7-Forw	5′-ACGAAGTGTRACCG	C1	286–299§
18.1-Forw	5′-TTGNAGCCGGGAACKAGC	C3	543–560§
Internal reverse sequencing primer for *Sym-Ha, -Gt, -Lp*			
10.20-Rev	5′-TAGGCCAAAATTAGCGGCA	C1	322–340§

*Nomenclature of minicircle regions is as specified in Fig. 1.

†Nomenclature of *Symbiodinium* strains is as described in Methods.

‡Base numbering for these primers refers to the clade A *psbA* gene sequence, GenBank accession no. AY160084.

§Base numbering for these primers refers to Fig. 1.

PCR conditions and development of primers. Degenerate primers psbAF6 and psbAL1 (Table 1) were designed to anneal to conserved regions in an alignment of dinoflagellate (*Heterocapsa triquetra* AF130033), stramenopile, Cryptophyta, Rhodophyta, Glaucocystophyceae, Euglenozoa, Viridiplantae and cyanobacterial *psbA* gene sequences. These primers were used to amplify 433 bp of *psbA* DNA from cultured clade A *Symbiodinium* strain *Tm*8.2, (source: *Tridacna maxima*, OTI). Unialgal culture *Tm*8.2, produced in our laboratory, was previously analysed by rDNA sequencing and comparison to published *Symbiodinium* sequences, confirming its identity (data not shown). The resulting partial *psbA* sequence was then used to design primers 4.6 and 4.3 (Table 1), which face out from the coding region. The remaining coding DNA of the *psbA* gene was thereby amplified from the clade A culture. The clade A amplicon obtained with primers 4.6 and 4.3 was ~2 kb in length indicating that the complete *psbA* minicircle of that culture is ~2·2 kb in length. Approximately 570 bp at each end of this amplicon were sequenced, reaching the termini of the gene. The full *psbA* gene sequence was used to design primers 7.4 and 7.8 (Table 1). These anneal to regions found to be conserved between culture *Tm*8.2 and the organisms listed above and are located near the N- and C-termini of the *psbA* gene, facing out toward the non-coding region. The 7.4/7.8 primer pair proved capable of amplifying the non-coding region of the *psbA* minicircle from clade C *Symbiodinium* spp. as well as clade A *Symbiodinium* spp. The PCR conditions for all amplifications were: 94 °C 2 min; then 35–40 cycles of 94 °C for 10 s, 55 °C for 30 s and 72 °C for 2 min; followed by a single extension at 72 °C for 10 min, using a Perkin Elmer Gene Amp 2400 thermocycler. Reactants in each 100 µl reaction were: 20 pmol each primer, combined with ~10 ng genomic template DNA and a mix consisting of 1 µl AmpliTaq (Perkin Elmer), 40 µg BSA, 0·1 mM (final concn) of each of dATP, dCTP, TTP and dGTP, 10 µl of 10× PCR buffer [1 M Tris/HCl; pH8·3, 5 M KCl, 150 mM $MgCl_2$, 1 % (w/v) gelatin; Sambrook *et al.*, 1989] and autoclaved MilliQ water.

Sequencing. Amplicons were purified using the GFX PCR DNA and Gel Band Purification kit (Amersham Pharmacia Biotech). These were directly sequenced with each of the amplification primers. In the case of amplicons containing the non-coding side of the *psbA* minicircle from *Sym-Ha*, *-Lp* and *-Gt*, further sequences were obtained using two or more internal primers (Table 1). All sequencing was done at the Australian Genomic Research Facility, University of Queensland, Australia.

Sequence analysis. Sequence chromatograms were edited (and reverse complemented if necessary) using Chromas 1.45 (Conor McCarthy, available from http://www.technelysium.com.au/chromas145-95.zip). Initial alignments were done using the program CLUSTALX (Thompson *et al.*, 1997). The gap-opening and gap-extension penalties were lowered, because the conserved minicircle cores and some of the non-conserved regions of the minicircles consist of repeated sequences that yielded many equally tentative alignments unless gap penalties were relaxed. Extensive realignments by hand were performed using the sequence-analysis suite Bioedit (Hall, 1999; available from http://www.mbio.ncsu.edu/BioEdit/bioedit.html). Inverted repeats (IRs) were identified by visual inspection. Double-hairpin structure of IRs was tested using the MFOLD web server (Zuker, 2003; available at http://www.bioinfo.rpi.edu/applications/mfold).

RESULTS

PCR amplification and sequence alignments

The full-length *psbA* gene of clade A *Symbiodinium* from a tridacnid clam was amplified and yielded a 1035 bp gene sequence. Subsequently, primers 7.4 and 7.8 successfully amplified the entire non-coding side of the *psbA* minicircle from *Sym-Ha* (1202 bp), *Sym-Lp* (1133 bp), *Sym-Gt* (965–1018 bp) and *Sym-Pd* (882 bp), which included 43 bp of

the C-terminus and 191 bp of the N-terminus of the *psbA* gene in each case (including primers). After initial alignment by CLUSTALX, extensive alignment by hand was necessary due to the presence of multiple repeated sequences and large indels. All minicircle fragments that were amplified with primer pair 7.4/7.8 yielded unambiguous chromatograms when sequenced with the primers of Table 1. We interpreted this as indicating that only a single minicircle variant encodes the full-length *psbA* gene in each of the clade C zooxanthella strains analysed.

Structure of the minicircle non-coding region

An alignment of sequences from the non-coding side of the *psbA* minicircle is presented in Fig. 1. Assignment of regions as variable (V), conserved (C) and metastable (M),

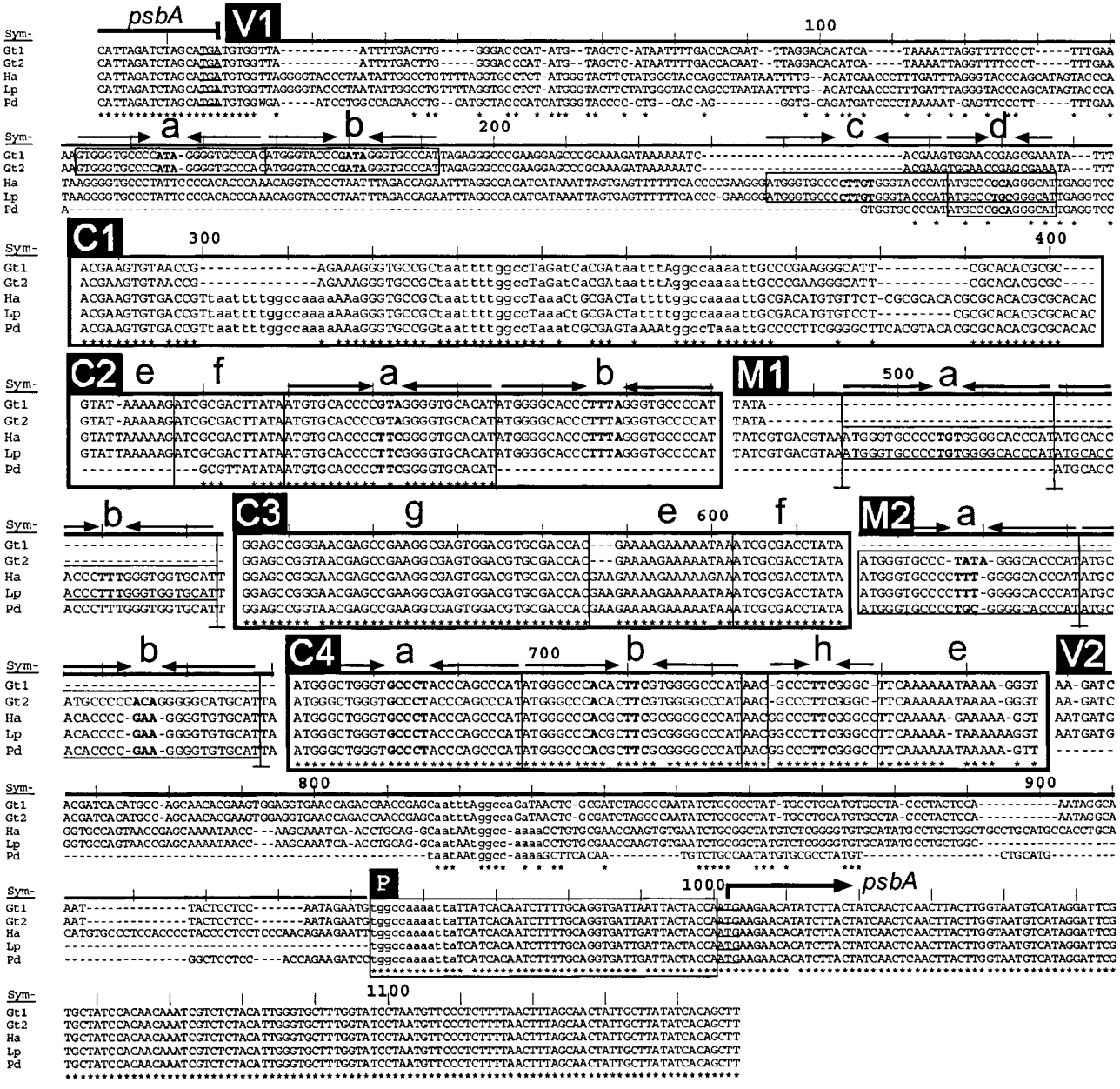

Fig. 1. Complete alignment of the non-coding region of *psbA* minicircles from five isolates of *Symbiodinium* clade C. Start and stop codons of the *psbA* gene are underlined. Bases conserved in all five isolates are asterisked. Each arrow represents one side of a single inverted repeat. Bold bases are intra-IR spacers not constituting part of a palindrome. Motifs within core (C), metastable (M) and variable (V) regions are named alphabetically in the order in which they occur except that analogous regions within C2, C3 and C4 are given a common suffix.

is based principally on the comparison of *Sym-Ha*, *-Gt* and *-Lp* sequences, as the sequence of *Sym-Pd* is very divergent (having many large deletions compared to the orthologous sequences from the other three hosts). The outlier status of *Sym-Pd* as detected in this study is consistent with differences already noted between *Sym-Pd* and these other OTI zooxanthellae based on rDNA sequencing (Carter *et al.*, 2000; and supplementary material in IJSEM Online).

There are four main conserved or core regions, C1–4, excluding the presumed promoter region (P) of the *psbA* gene, which is 48 bp in length, continuous with the gene and conserved across all the clade C symbionts studied. Two highly conserved regions, C2 and C4, are similar to each other, but not identical. Region C3 is also highly conserved but the overall sequence is not related to that of C2 and C4, except that a poly-A sequence is found in each, and a 13 bp motif (ATCGCGACYTATA), designated f, is common between C2 and C3. Region C1 is only semi-conserved, with a sequence completely unrelated to those of the other C regions. Two variable regions, V1 and V2, lie between the set of cores and the *psbA* gene. The sequences within any one V region cannot be unambiguously aligned across the four zooxanthellae types. The two remaining regions are those that separate C2 and C4 from C3. The sequences of these two regions are related to those of the C2 and C4 regions, but they are each either present in full or completely absent, depending on the zooxanthella strain analysed. As such they are termed metastable or modular (M) regions.

Inverted repeats (IRs) lie in regions V1, C2, C4, M1 and M2. A term that describes the majority of the IRs in the C and M regions is ATG–CAT IRs, being those with ATG and CAT invariant as the first three and last three bases respectively. Five of the ATG–CAT IRs are conserved across *Sym-Ha*, *-Lp* and *-Gt*, and these lie in the C2 and C4 regions (Fig. 1). In region C2 is a pair of IRs designated C2a and C2b. A similar pair in region C4 are designated C4a and C4b. IR C2a directly abuts IR C2b; similarly IR C4a directly abuts IR C4b. The term 'twin IR' describes this abutment. Within all the C1, C4, M1 and M2 regions studied, the twin IR form creates a palindromic sequence at the position of abutment: CATATG. One additional small IR, designated C4h, lies downstream of the twin IR C4a–b. The C2 and C4 regions can be considered to qualitatively mirror each other, given that a run of adenines (C2e) flanks C2 at the upstream side, and another run of adenines (C4e) flanks C4 at the downstream side. At the centre of the observed partial symmetry of the minicircle is C3e, consisting of the conserved adenine-rich sequence GAAAAGAAAAA (positions 589 to 599 in Fig. 1).

The arrangement of the C2 and C4 regions, relative to the central adenine-rich region C3e, can be interpreted in the sense that the sum C2 + C3 + C4 approximates a single large symmetrical unit, but may have intervening metastable regions present as well. The metastable regions, M1 and M2, contain twin IRs (M1a–b and M2a–b respectively), but are each without a flanking poly-A motif, and without an accompanying small IR such as C4h. In this sense the variation present in M1 and M2 seems to be constrained. Presence/absence switching (modularity) of M1 and M2 suggests that the twin IRs in these regions are highly evolved in that a twin pair may be gained or lost only as a unit. Likewise indicating a refined process, IRs in the M1 and M2 regions do not contain mismatches in the inverse sequence relationship, even though the IR sequence may change between zooxanthella strains. Inverse sequence relationships in C2 and C4 IRs are also perfect. Even though sequence variation occurs between zooxanthellae with regard to these core IRs, they are always perfect inversions. Perfect inversions constituting twin IR sequences are not observed outside the partially symmetrical unit C2 + M1 + C3 + M2 + C4. There are twin IRs in the V1 region of the minicircle from *Sym-Gt*, *-Ha* and *-Lp*, but these IRs (V1a–V1d) are all either short, or composed of slightly imperfect inverse sequences.

In analysing the lack of inverse sequence mismatch in the core IRs, a special case must be made for IR C4b; it is an unusual core IR because there is a conserved insertion of a single adenine (position 706, Fig. 1) in the left-hand side of the IR, making it asymmetrical. After the inserted base A, the IR structure resumes for another 3 bases (positions 707 to 709 in Fig. 1). It is assumed that these 3 offset bases of IR C4b are functional, because the inverse relationship at these positions is retained in *Sym*-Gt relative to that of *Sym-Ha* and *-Pd*, even though point mutations have occurred [i.e. CAC/GTG (*Sym-Gt*) versus CGC/GCG (*Sym-Ha*, *-Lp*, *-Pd*)].

Some core IRs are interrelated in the sequence of the 3–5 bp intra-IR spacers that lie centrally within any given IR (Fig. 1). There is conservation of the spacer TTC within many of the core IRs of the *Symbiodinium* isolates studied. All isolates have the spacer TTC within C4b and C4h. *Sym-Ha*, *-Lp* and *-Pd* isolates have spacer TTC within C2a. Conservation of the spacer TTC in two unrelated IR types (h and b) may indicate a function for this particular spacer motif.

Following the observation that each twin IR constitutes a unit and that this unit occurs many times in a sequence, an alignment of all the twin IRs was generated (Fig. 2). The alignment indicates that all units but C4ab are related. When C4ab is reverse-complemented, the match does not improve significantly (data not shown). Since sequence likeness occurs between all units except C4, these may be considered to have had a common mode of origination, or to have been duplicates of each other that have since diverged.

Precise abutment of C3f with M2a imitates the junction of C2f and C2a (Figs 1 and 3). This recurring juxtaposition might imply a relationship between the poly-A stretches in the e regions and twin IRs, because f regions intervene and

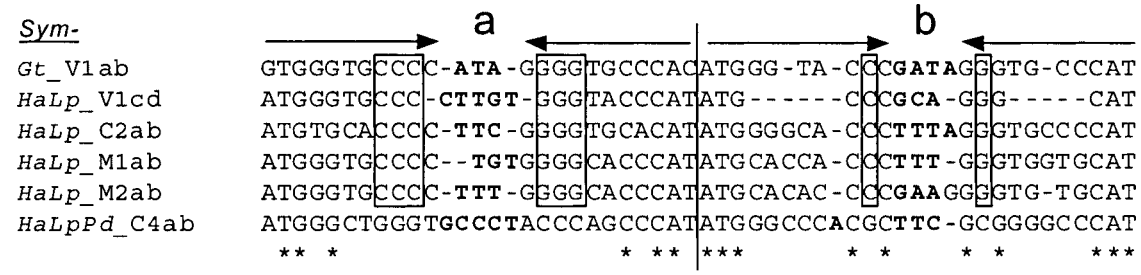

Fig. 2. Alignment of representative twin IRs that occur in the *psbA* minicircles of five *Symbiodinium* clade C isolates. Boxed regions are common to the majority, but different in C4a–b. Bases with asterisks are common to all sequences shown. Bold bases are intra-IR spacers. Each arrow represents one side of a single IR.

are an exact length. Considering the recurrence of the order e–f–a–b, a high level of organization is noted in the *Symbiodinium* minicircles studied here (Fig. 3). First-order organization is the presence of IRs. Second order is the precise abutment of two IRs to make a twin IR. Third order is the abutment of a twin unit against the f region (ATCGCGACYTATA) and the accompanying poly-A stretch (specifically C2e–f abuts C2a–b, and C3e–f abuts M2a–b). Fourth order is the duplication of the third order complex (e–f–a–b) about an intervening unique region, C3g. A tendency for the third order complex to include a second a–b repeat is noted (Fig. 3), but is apparently non-essential since not all isolates contain these extra repeats.

Conserved sequence elements also exist in the C1 region, including apparent relics of two to three copies of an 18 bp palindrome TAATTTTGGCCAAAATTA (relics conforming to this consensus are indicated by lower-case letters on Fig. 1). Two elements such as the 18 bp palindrome also occur in other parts of the minicircles: a possible relic in V2 (lower case, positions 816–831), and an invariantly conserved 12 bp relative TGGCCAAAATTA located at the extreme 5′ end of the putative *psbA* promoter (lower case, positions 953–964).

DISCUSSION

The *psbA* minicircles from clade C *Symbiodinium* spp. share features common to other dinoflagellate minicircles but also possess unique features, making them of interest from biological, functional and evolutionary viewpoints. One notable feature of *Symbiodinium* clade C minicircles that is shared with the minicircles of other dinoflagellates is that, like those of *Amphidinium operculatum* and *Heterocapsa triquetra*, they possess tracts of poly-A. An alignment of the poly-A section from *A. operculatum* minicircles and two regions from *Symbiodinium*: C2e–f and C3e–f, is shown in Fig. 4. The e–f region of C3 in the *Symbiodinium* sequence has particularly high similarity to the *A. operculatum* sequence. The poly-A tract from *Heterocapsa triquetra* minicircles has been proposed by Zhang *et al.* (1999, 2002) to be a putative origin of replication, as A + T-rich sequences are more easily opened to create a replication fork than are G + C-rich sequences. It is plausible that one or more of the *Symbiodinium* poly-A tracts could serve as origin(s) of replication for the minicircle.

The feature that particularly distinguishes *Symbiodinium* clade C minicircles from those of other dinoflagellates is the

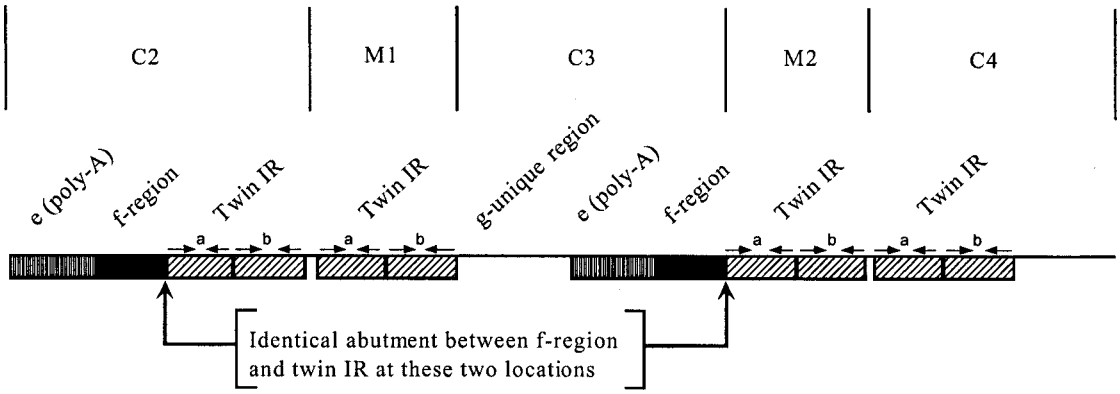

Fig. 3. Schematic interrelationship of f regions and inverted repeats in the *psbA* minicircle of clade C *Symbiodinium* (region labels are in accordance with Fig. 1). The scheme is generalized. The exceptions to this arrangement are that: twin IR M1a-b is present only in *Sym-Ha* and *-Lp* isolates; twin IR M2a–b is conserved in all isolates except *Sym-Gt1*.

```
                 *** *R***R***  *R *Y***
Sym C2e-f      CGTATAAAAAGATCGCGACTTATA
Amph           AGTAGAGAAAAATCCAGG-TCATA
Sym C3e-f      AGAAAAGAAAAATCGCGACTTATA
               ** *R*********  *R *Y***
```

Fig. 4. Comparison of a poly-A sequence motif present in *Amphidinium operculatum* minicircles with that found in the C2e–f and C3e–f regions of the *Symbiodinium* clade C *psbA* minicircle. Asterisks indicate identity of top or bottom sequence compared to middle sequence. Conservative changes are noted as R (purine) or Y (pyrimidine).

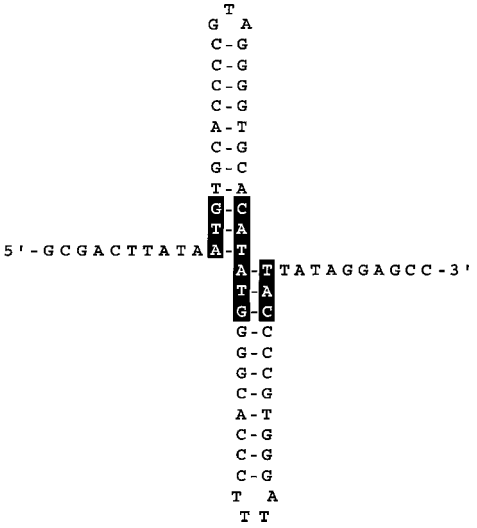

Fig. 5. Schematic of the proposed base pairings within one DNA strand of a twin IR, to potentially form a double-hairpin structure. This would be repeated on the complementary strand and result in four hairpins, which cannot easily be represented in two dimensions. The sequence depicted is the a–b region present in the C2 block from *Sym-Gt1*. Boxed ATG and CAT bases are invariant in all the twin IRs present in the C and M regions of all the minicircles in this study. Each of the twin IRs in Fig. 1 can be represented by a structure very similar to this, generally with 11 bp per putative stem, and always with no unpaired bases at the junction of the two stems.

presence of multiple IRs, rather than a single IR, and also their arrangement, such that each IR is abutted to another very similar IR. IRs have been found in other dinoflagellate minicircles (Zhang *et al.*, 1999, 2002), but not at this density nor in twin formation. A notable feature of the IRs in the *Symbiodinium* minicircles is that they can be metastable even between isolates having otherwise identical minicircle sequences, and obtained from a single host species. The minicircles from *Sym-Gt1* and *-Gt2* had identical non-coding sequences except for an indel consisting of a twin IR, M2a–b. A possible explanation for this is that minicircle IRs can form loop regions during minicircle replication or recombination, and thus may be lost or gained as a unit. Similar modular behaviour of a twin IR is apparent in the M1 region, where the *Sym-Ha* and *Sym-Lp* minicircles possess a twin IR in region M1 while *Sym-Gt1* and *–Gt2* do not.

DNA sequence motifs similar to the twin IRs reported here have been found in the mitochondrial genomes of a range of taxonomically diverse fungal species (Paquin & Lang, 1996; Paquin *et al.*, 1997, 2000). These motifs, termed double-hairpin elements (DHEs), range in size from 26–79 bases and assume a secondary structure consisting of a 3–5 base loop in each of two adjacent, helical stems, with extensive G–C pairing in at least one stem. Secondary structure prediction (Zuker, 2003) on our twin IRs indicates they have the potential to form double-hairpin structures (Fig. 5). Like the fungal DHEs, the twin IRs are G+C-rich in the proposed stem regions, with loops of 3–5 bases. Given the possibility of extruded hairpins, it is noted that GNA (and its complementary strand sequence TNC) is present in many of the most highly conserved IRs in this study (C2a, C4b and C4h; see Fig. 5), and GNA has been shown to be a thermodynamically favoured loop sequence in DNA hairpins (Hirao *et al.*, 1992; Dai *et al.*, 1997).

Paquin *et al.* (2000) argued that DHEs are mobile elements, as multiple copies are present in all of the mitochondrial genomes where they have been detected, they are scattered in distribution and they vary in number between closely related species. The iteration of twin IRs in the *Symbiodinium psbA* minicircle and their sporadic occurrence in closely related strains is also consistent with a mobility hypothesis. Each twin IR form appears to act as a single unit of selection, since lone IRs of the ATG–CAT type have not been found in this set of sequences.

Twin IRs may be parasitic mobile elements that have accumulated in the non-coding region of this plastid minicircle. Alternatively, they may have a functional role, which is suggested by their very high level of conservation compared to other regions of the *Symbiodinium* minicircle. One possible function of the twin IRs is that they could facilitate recombination between minicircles of the genomic set. In this scenario, recombination would act to maintain the nucleotide sequence and thereby maintain replicative and other essential functions of the minicircles. IRs are known to be recombinogenic in several systems, including a plastid (Kawata *et al.*, 1997), plant and fungal mitochondria (Gross *et al.*, 1989; Lupold *et al.*, 1999), fungal and bacterial genomes (Farah *et al.*, 2002; Holmes *et al.*, 2003), plasmids (Francia *et al.*, 1997, 1999) and phage (Mertens *et al.*, 1988; Smith-Mungo *et al.*, 1994). In the current study, support for frequent recombination comes from the observation that mismatches within IR arms do not occur in the C2, C4, M1 and M2 regions, even though these IR sequences vary among strains. In contrast, IRs in the V regions contain mismatches. Strict maintenance of inverse matching in the C and M regions indicates that although these loci mutate rapidly, there is some stabilizing influence,

such as gene conversion, which occurs within them but not within the V regions. Further, recombination could be invoked to explain the loss of non-abutting IRs C2b and M1a in *Sym-Pd*, on the basis that these two IRs may have been loops at the time of recombination with another minicircle.

If the model of recombination among cores is correct (Zhang *et al.*, 1999, 2002; Barbrook & Howe, 2000) then the hypervariable sequences of the *Symbiodinium* V regions would indicate that these do not homogenize or recombine nearly as often as the cores. It is possible that the V regions act as buffer zones lowering the probability that recombination will extend into the protein-coding regions of minicircles. Such buffers might help prevent the accumulation of truncated proteins, and of hybrid, potentially non-functional proteins at valuable sites within the plastid apparatus.

Another hypothetical function for the twin IRs is that they may be part of the essential machinery of replication origins. Zhang *et al.* (1999) hypothesized that A + T-rich regions of *Heterocapsa triquetra* minicircles may be involved in DNA unwinding, and stated that a nearby IR element (of 40 bp) might be part of the putative replication origin. In the *Symbiodinium* minicircles, poly-A sections and twin IRs do occur together in some cases (Fig. 3), giving some support to a functional linkage between the two sequence types.

The results of this analysis of the *Symbiodinium psbA* minicircle open up a range of questions to be explored: First, do clade C minicircles possess their high level of organization as a derived feature or an ancestral one? *Symbiodinium* spp. belong in the family Symbiodiniaceae, which has been recently included in the order Suessiales (Fensome *et al.*, 1993) based on ultrastructural data (Loeblich & Sherley, 1979), and was subsequently supported by rDNA data (Saunders *et al.*, 1997; Montresor *et al.*, 1999). The divergence of Symbiodiniaceae from *Polarella glacialis* Montresor, the only other cultured Suessiale, has been estimated at ~200 Mya (Montresor *et al.*, 1999). It would be interesting to sequence the minicircles of *Polarella glacialis*, to establish whether any of the non-coding motifs observed in *Symbiodinium* minicircles are of ancient origin.

Second, will the level of organization seen here also be found in other minicircles from *Symbiodinium*? It is our intention to sequence additional minicircles encoding other genes, from each *Symbiodinium* strain characterized here. Comparison of 23S minicircles, for instance, against the sequence of the corresponding *psbA* minicircles will allow confirmation of core locations, and will enable us to test the hypothesis that recombination acts to prevent inverse sequence mismatches in the core regions. In the set of minicircles from the zooxanthella of a single host species, e.g. *Sym-Gt*, the expectation is to find homogeneous perfect inverses in core regions. By comparison, subtly different, but equally perfect matches would be expected for cores across the set of minicircles from a closely related zooxanthella strain, e.g. *Sym-Ha*. Likewise, we will be able to establish whether the buffer zone hypothesis for the V1 and V2 regions is correct. The expectation is that in a single zooxanthella cell, identical V1 and V2 regions would not be shared between any two minicircles containing different genes.

Third, are our results typical of all *Symbiodinium* species, or only those belonging to clade C? Analysis of the plastid minicircles across the remaining six clades should reveal the extent to which this unusual genetic organization is found.

Finally, will the *psbA* minicircle be useful as a molecular marker for *Symbiodinium*? The data presented here indicate that the *psbA* non-coding region can distinguish between closely related clade C *Symbiodinium* isolates, such as *Sym-Ha* and *Sym-Lp*, that could not be differentiated using sequencing analysis of the D1/D2 region of the LSU rDNA gene (see supplementary material). Similar to the LSU rDNA gene, the *psbA* minicircle may allow different levels of differentiation to be obtained depending on the region chosen: the V regions may be used to distinguish closely related strains, with the C and M regions used to assess diversity in more distant strains and species. Further studies using a wide range of *Symbiodinium* isolates are planned to test the utility of this novel sequence.

ACKNOWLEDGEMENTS

This work was supported by ARC grant A10009205 to D. C. and O. H.-G. We thank To-Ha Loi for rDNA LSU D1/D2 sequencing data and Tien Bui for technical assistance.

REFERENCES

Baker, A. C. (2003). Flexibility and specificity in coral – algal symbiosis: diversity, ecology, and biogeography of *Symbiodinium*. *Annu Rev Ecol Evol Syst* Review in Advance, doi: 10.1146/annurev.ecolsys.34.011802.132417.

Barbrook, A. C. & Howe, C. J. (2000). Minicircular plastid DNA in the dinoflagellate *Amphidinium operculatum*. *Mol Gen Genet* **263**, 152–158.

Barbrook, A. C., Symington, H., Nisbet, R. E. R., Larkum, A. & Howe, C. J. (2001). Organisation and expression of the plastid genome of the dinoflagellate *Amphidinium operculatum*. *Mol Genet Genomics* **266**, 632–638.

Baillie, B. K., Belda-Baillie, C. A. & Maruyama, T. (2000). Conspecificity and Indo-Pacific distribution of *Symbiodinium* genotypes (Dinophyceae) from giant clams. *J Phycol* **36**, 1153–1161.

Brown, B. E. (1997). Coral bleaching: causes and consequences. *Coral Reefs* **16** (suppl.), S129–S138.

Carlos, A. A., Baillie, B. K., Kawachi, M. & Maruyama, T. (1999). Phylogenetic position of *Symbiodinium* (Dinophyceae) isolates from Tridacnids (Bivalvia), Cardiids (Bivalvia), a sponge (Porifera), a soft coral (Anthozoa), and a free living strain. *J Phycol* **35**, 1054–1062.

Carter, D. A., Gava, N., Loi, T. H., Loh, W. & Hoegh Guldberg, O. (2000). Genetic diversity of symbiotic dinoflagellates (''zooxanthellae'') inhabiting different scleractinian coral species. *Australian Society for Microbiology: Annual Scientific Meeting*, Cairns Queensland Australia.

Dai, X., Greizerstein, M. B., Nadas-Chinni, K. & Rothman-Denes, L. B. (1997). Supercoil-induced extrusion of a regulatory DNA hairpin. *Proc Natl Acad Sci U S A* **94**, 2174–2179.

Downs, C., Fauth, J., Dustan, P., Bemiss, J. & Woodley, C. (2002). Oxidative stress and seasonal coral bleaching. *Free Radic Biol Med* **33**, 533.

Farah, J. A., Hartsuiker, E., Mizuno, K., Ohta, K. & Smith, G. R. (2002). A 160-bp palindrome is a Rad50. Rad32-dependent mitotic recombination hotspot in *Schizosaccharomyces pombe*. *Genetics* **161**, 461–468.

Fensome, R. A., Taylor, F. J. R., Norris, G., Sarjeant, W. A. S., Wharton, D. I. & Williams, G. L. (1993). A classification of living and fossil dinoflagellates. In *Micropaleontology Special Publication no. 7*. Hanover, PA: Sheridan Press.

Francia, M. V., Avila, P., de la Cruz, F. & García Lobo, J. M. (1997). A hot spot in plasmid F for site-specific recombination mediated by Tn21 integron integrase. *J Bacteriol* **179**, 4419–4425.

Francia, M. V., Zabala, J. C., de la Cruz, F. & Garcia Lobo, J. M. (1999). The IntI1 integron integrase preferentially binds single-stranded DNA of the *attC* site. *J Bacteriol* **181**, 6844–6849.

Freudenthal, H. D. (1962). *Symbiodinium* gen. nov. and *Symbiodinium microadriaticum* sp. nov., a zooxanthella. Taxonomy, life cycle and morphology. *J Protozool* **9**, 45–52.

Goodson, M. S., Whitehead, L. F. & Douglas, A. E. (2001). Symbiotic dinoflagellates in marine Cnidaria: diversity and function. *Hydrobiologia* **461**, 79–82.

Gross, S. R., Levine, P. H., Metzger, S. & Glaser, G. (1989). Recombination and replication of plasmid-like derivatives of a short section of the mitochondrial chromosome of *Neurospora crassa*. *Genetics* **121**, 693–701.

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* **41**, 95–98.

Hiller, R. G. (2001). 'Empty' minicircles and *petB*/*atpA* and *psbD*/*psbE* ($cytb_{559}$ α) genes in tandem in *Amphidinium carterae* plastid DNA. *FEBS Lett* **505**, 449–452.

Hirao, I., Nishimura, Y., Tagawa, Y., Watanabe, K. & Miura, K. (1992). Extraordinarily stable mini-hairpins: electrophoretical and thermal properties of the various sequence variants of d(GCGAAAGC) and their effect on DNA sequencing. *Nucleic Acids Res* **20**, 3891–3896.

Hoegh-Guldberg, O. (1999). Coral bleaching, climate change and the future of the world's coral reefs. *Mar Freshw Res* **50**, 839–866.

Holmes, A. J., Holley, M. P., Mahon, A., Nield, B., Gillings, M. & Stokes, H. W. (2003). Recombination activity of a distinctive integron-gene cassette system associated with *Pseudomonas stutzeri* populations in soil. *J Bacteriol* **185**, 918–928.

Howe, C. J., Barbrook, A. C., Koumandou, V. L., Nisbet, R. E. R., Symington, H. A. & Wightman, T. F. (2003). Evolution of the chloroplast genome. *Philos Trans R Soc Lond B Biol Sci* **358**, 99–106 (discussion pp. 106–107).

Hughes, T. P., Baird, A. H., Bellwood, D. R. & 14 other authors (2003). Climate change, human impacts, and the resilience of coral reefs. *Science* **301**, 929–933.

Kawata, M., Harada, T., Shimamoto, Y., Oono, K. & Takaiwa, F. (1997). Short inverted repeats function as hotspots of intermolecular recombination giving rise to oligomers of deleted plastid DNAs (ptDNAs). *Curr Genet* **31**, 179–184.

Kevin, M. J., Hall, W. T., McLaughlin, J. A. & Zahl, P. A. (1969). *Symbiodinium microadriaticum* Freudenthal, a revised taxonomic description, ultrastructure. *J Phycol* **5**, 341–350.

LaJeunesse, T. C. (2001). Investigating the biodiversity, ecology, and phylogeny of endosymbiotic dinoflagellates in the genus *Symbiodinium* using the ITS region: in search of a "species" level marker. *J Phycol* **37**, 866–880.

LaJeunesse, T., Loh, W. K. W., van Woesik, R., Hoegh-Guldberg, O., Schmidt, G. W. & Fitt, W. K. (2003). Low symbiont diversity in southern Great Barrier Reef corals, relative to those of the Caribbean. *Limnol Oceanogr* **48**, 2046–2054.

Loeblich, A. R., III & Sherley, J. L. (1979). Observations on the theca of the motile phase of free-living and symbiotic isolates of *Zooxanthella microadriatica* (Freudenthal) comb. nov. *J Mar Biol Assoc UK* **59**, 195–205.

Loh, W. K. W., Loi, T. H., Carter, D. A. & Hoegh-Guldberg, O. (2001). Genetic variability of the symbiotic dinoflagellates from the wide ranging coral species *Seriatopora hystrix* and *Acropora longicyathus* in the Indo-West Pacific. *Mar Ecol Prog Ser* **222**, 97–107.

Lupold, D. S., Caoile, A. G. & Stern, D. B. (1999). Genomic context influences the activity of maize mitochondrial *cox2* promoters. *Proc Natl Acad Sci U S A* **96**, 11670–11675.

Mertens, G., Klippel, A., Fuss, H., Blocker, H., Frank, R. & Kahmann, R. (1988). Site-specific recombination in bacteriophage Mu: characterization of binding sites for the DNA invertase Gin. *EMBO J* **7**, 1219–1227.

Montresor, M., Procaccini, G. & Stoecker, D. K. (1999). *Polarella glacialis* gen. nov., sp. nov. (Dinophyceae): Suessiaceae are still alive! *J Phycol* **35**, 186–197.

Pakrasi, H. B. (1995). Genetic analysis of the form and function of photosystem I and photosystem II. *Annu Rev Genet* **29**, 755–776.

Paquin, B. & Lang, B. F. (1996). The mitochondrial DNA of *Allomyces macrogynus*: the complete genomic sequence from an ancestral fungus. *J Mol Biol* **255**, 688–701.

Paquin, B., Laforest, M. J., Forget, L., Roewer, I., Wang, Z., Longcore, J. & Lang, B. F. (1997). The fungal mitochondrial genome project: evolution of fungal mitochondrial genomes and their gene expression. *Curr Genet* **31**, 380–395.

Paquin, B., Laforest, M. J. & Lang, B. F. (2000). Double-hairpin elements in the mitochondrial DNA of Allomyces: evidence for mobility. *Mol Biol Evol* **17**, 1760–1768.

Pochon, X., Pawlowski, J., Zaninetti, L. & Rowan, R. (2001). High genetic diversity and relative specificity among *Symbiodinium*-like endosymbiotic dinoflagellates in soritid foraminiferans. *Mar Biol* **139**, 1069–1078.

Rodriguez-Lanetty, M. (2003). Evolving lineages of *Symbiodinium*-like dinoflagellates based on ITS1 rDNA. *Mol Phylogenet Evol* **28**, 152–168.

Rowan, R. (1998). Diversity and ecology of zooxanthellae on coral reefs. *J Phycol* **34**, 407–417.

Rowan, R. & Powers, D. A. (1992). Ribosomal RNA sequences and the diversity of symbiotic dinoflagellates (zooxanthellae). *Proc Natl Acad Sci U S A* **89**, 3639–3643.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Santos, S. R., Taylor, D. J., Kinzie, R. A., III, Hidaka, M., Sakai, K. & Coffroth, M. A. (2002). Molecular phylogeny of symbiotic dinoflagellates inferred from partial chloroplast large subunit (23S)-rDNA sequences. *Mol Phylogenet Evol* **23**, 97–111.

Santos, S. R., Gutierrez-Rodriguez, C. & Coffroth, M. A. (2003). Phylogenetic identification of symbiotic dinoflagellates via length heteroplasmy in domain V of chloroplast large subunit (cp23S)-ribosomal DNA sequences. *Mar Biotechnol (NY)* **5**, 130–140.

Saunders, G. W., Hill, D. R. A., Sexton, J. P. & Andersen, R. A. (1997). Small subunit ribosomal RNA sequences from selected dinoflagellates: testing classical evolutionary hypotheses with molecular systematic methods. In *Origins of Algae and their Plastids*, pp. 237–259. Edited by D. Bhattacharya. Vienna: Springer-Verlag.

Singh, M. (2000). Turnover of D1 protein encoded by *psbA* gene in higher plants and cyanobacteria sustains photosynthetic efficiency to maintain plant productivity under photoinhibitory irradiance. *Photosynthetica (Prague)* **38**, 161–169.

Smith-Mungo, L., Chan, I. T. & Landy, A. (1994). Structure of the P22 *att* site. Conservation and divergence in the lambda motif of recombinogenic complexes. *J Biol Chem* **269**, 20798–20805.

Stokes, H. W., O'Gorman, D. B., Recchia, G. D., Parsekhian, M. & Hall, R. M. (1997). Structure and function of 59-base element recombination sites associated with mobile gene cassettes. *Mol Microbiol* **26**, 731–745.

Takishita, K., Ishikura, M., Kioke, K. & Maruyama, T. (2003). Comparison of phylogenies based on nuclear-encoded SSU rDNA and plastid-encoded *psbA* in the symbiotic dinoflagellate genus *Symbiodinium*. *Phycologia* **42**, 285.

Taylor, D. L. (1983). The coral algal symbiosis. In *Algal Symbiosis, a Continuum of Interaction Strategies*. Edited by L. J. Goff. New York: Cambridge University Press.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997). The Clausal windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **24**, 4876–4882.

Trench, R. K. (1993). Micro algal-invertebrate symbioses: a review. *Endocytob Cell Res* **9**, 135–175.

Trench, R. K. (1997). Diversity of symbiotic dinoflagellates and the evolution of microalgal-invertebrate symbioses. *Proceedings of the 8th International Coral Reef Symposium, 24–29 June 1996, Panama*, vol. **2**, pp. 1275–1286.

Trench, R. K. & Blank, R. J. (1987). *Symbiodinium-microadriaticum* Freudenthal *Symbiodinium-goreauii* new-species *Symbiodinium-kawagutii* new-species and *Symbiodinium-pilosum* new-species Gymnodinioid dinoflagellate symbionts of marine invertebrates. *J Phycol* **23**, 469–481.

Trench, R. K. & Thinh, L. V. (1995). *Gymnodinium linucheae* sp. nov. the dinoflagellate symbiont of the jellyfish *Linuche unguiculata*. *Eur J Phycol* **30**, 149–154.

van Oppen, M. J. H., Palstra, F. P., Piquet, A. M. T. & Miller, D. J. (2001). Patterns of coral-dinoflagellate associations in *Acropora*: significance of local availability and physiology of *Symbiodinium* strains and host-symbiont selectivity. *Proc Royal Soc Biol Sci Ser B* **268**, 1759–1767.

Walther, G. R., Post, E., Convey, P., Menzel, A., Parmesan, C., Beebee, T. J., Fromentin, J. M., Hoegh-Guldberg, O. & Bairlein, F. (2002). Ecological responses to recent climate change. *Nature* **416**, 389–395.

Warner, M. E., Fitt, W. K. & Schmidt, G. W. (1999). Damage to photosystem II in symbiotic dinoflagellates: a determinant of coral bleaching. *Proc Natl Acad Sci U S A* **96**, 8007–8012.

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

Zhang, Z., Cavalier-Smith, T. & Green, B. R. (2002). Evolution of dinoflagellate unigenic minicircles and the partially concerted divergence of their putative replicon origins. *Mol Biol Evol* **19**, 489–500.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**, 3406–3415.