The Plant Cell, Vol. 8, 539–553, March 1996 © 1996 American Society of Plant Physiologists

# Rubisco in Marine Symbiotic Dinoflagellates: Form II Enzymes in Eukaryotic Oxygenic Phototrophs Encoded by a Nuclear Multigene Family

**Rob Rowan,[a,b,1,2] Spencer M. Whitney,[c,2] Amanda Fowler,[c] and David Yellowlees [c,3]**

[a] Australian Institute of Marine Science, PMB No. 3, Townsville MC, Queensland 4810, Australia
[b] Smithsonian Tropical Research Institute, Apartado 2072, Balboa, Republic of Panama
[c] Department of Molecular Sciences, James Cook University of North Queensland, Townsville, Queensland 4811, Australia

**Genes encoding ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) were cloned from dinoflagellate symbionts (*Symbiodinium* spp) of the giant clam *Tridacna gigas* and characterized. Strikingly, *Symbiodinium* Rubisco is completely different from other eukaryotic (form I) Rubiscos: it is a form II enzyme that is ~65% identical to Rubisco from *Rhodospirillum rubrum* (Rubisco forms I and II are ~25 to 30% identical); it is nuclear encoded by a multigene family; and the predominantly expressed Rubisco is encoded as a precursor polyprotein. One clone appears to contain a predominantly expressed Rubisco locus (*rbcA*), as determined by RNA gel blot analysis of *Symbiodinium* RNA and sequencing of purified Rubisco protein. Another contains an enigmatic locus (*rbcG*) that exhibits an unprecedented pattern of amino acid replacement but does not appear to be a pseudogene. The expression of *rbcG* has not been analyzed; it was detected only in the minor of two taxa of *Symbiodinium* that occur together in *T. gigas*. This study confirms and describes a previously unrecognized branch of Rubisco's evolution: a eukaryotic form II enzyme that participates in oxygenic photosynthesis and is encoded by a diverse, nuclear multigene family.**

## INTRODUCTION

*Symbiodinium* is a genus of phototrophic, peridin-containing dinoflagellates that live as endosymbionts (zooxanthellae) in a wide variety of tropical marine invertebrates (reviewed in Trench, 1987, 1992). For example, all reef-building corals harbor—and are dependent on—*Symbiodinium*. As in any phototrophic system, carbon assimilation is central to the energetics of zooxanthella–invertebrate symbioses. Photosynthetic organisms fix $CO_2$ into organic compounds via the photosynthetic carbon reduction (Calvin–Benson) cycle. The initial and rate-limiting step, in which ribulose-1,5-bisphosphate (RuBP) and $CO_2$ are converted to two molecules of 3-phosphoglycerate, is catalyzed by ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco; EC 4.1.1.39). Labeling experiments confirmed that Rubisco is the primary $CO_2$ fixation enzyme in *Symbiodinium* freshly isolated from a coral (Streamer et al., 1993).

Despite their ecological significance, relatively little is known about carbon assimilation in dinoflagellates. "Typical" phototrophic dinoflagellates uniquely contain peridinin as the major light-harvesting carotenoid; the few exceptional species may or do contain chloroplasts acquired from other algal classes (Gibbs, 1990; Whatley, 1993). Active Rubisco has not been purified from peridinin-containing dinoflagellates, owing to the enzyme's instability after cell lysis, but inactive protein was purified after labeling it with the tight-binding inhibitor $^{14}$C-carboxyarabinitol bisphosphate (Whitney and Yellowlees, 1995). Surprisingly, an initial characterization of proteins from *Symbiodinium* and from the free-living species *Amphidinium carterae* indicated similarity with the bacterial form II Rubisco (Whitney et al., 1995). Recently, a form II Rubisco was also found in the free-living dinoflagellate *Gonyaulax polyedra* (Morse et al., 1995). All other known eukaryotic Rubiscos are form I enzymes.

Structurally (reviewed in Tabita, 1988; Hartman and Harpel, 1994) and catalytically (Jordan and Ogren, 1981), the differences between form I and form II Rubiscos are considerable. Of particular significance is the much lower specificity of known form II enzymes for $CO_2$ versus $O_2$ (Jordan and Ogren, 1981), restricting their role in $CO_2$ fixation to anaerobic environments (Tabita, 1988). Suppression of Rubisco's oxygenase activity is a major theme in the evolution of oxygenic photosynthesis (e.g., Hatch, 1971; Jordan and Ogren, 1981; Beardall, 1989; Read and Tabita, 1994), and the involvement of a form II enzyme poses a unique situation. As such, dinoflagellates might provide unique insights into the evolution of photosynthesis.

Here, we report the isolation and characterization of two loci from *Symbiodinium*—one encoding Rubisco and the other

[1] Permanent address: University of Guam Marine Laboratory, U.O.G. Station, Mangilao, Guam 96923, USA.
[2] These authors contributed equally to this work.
[3] To whom correspondence should be addressed.

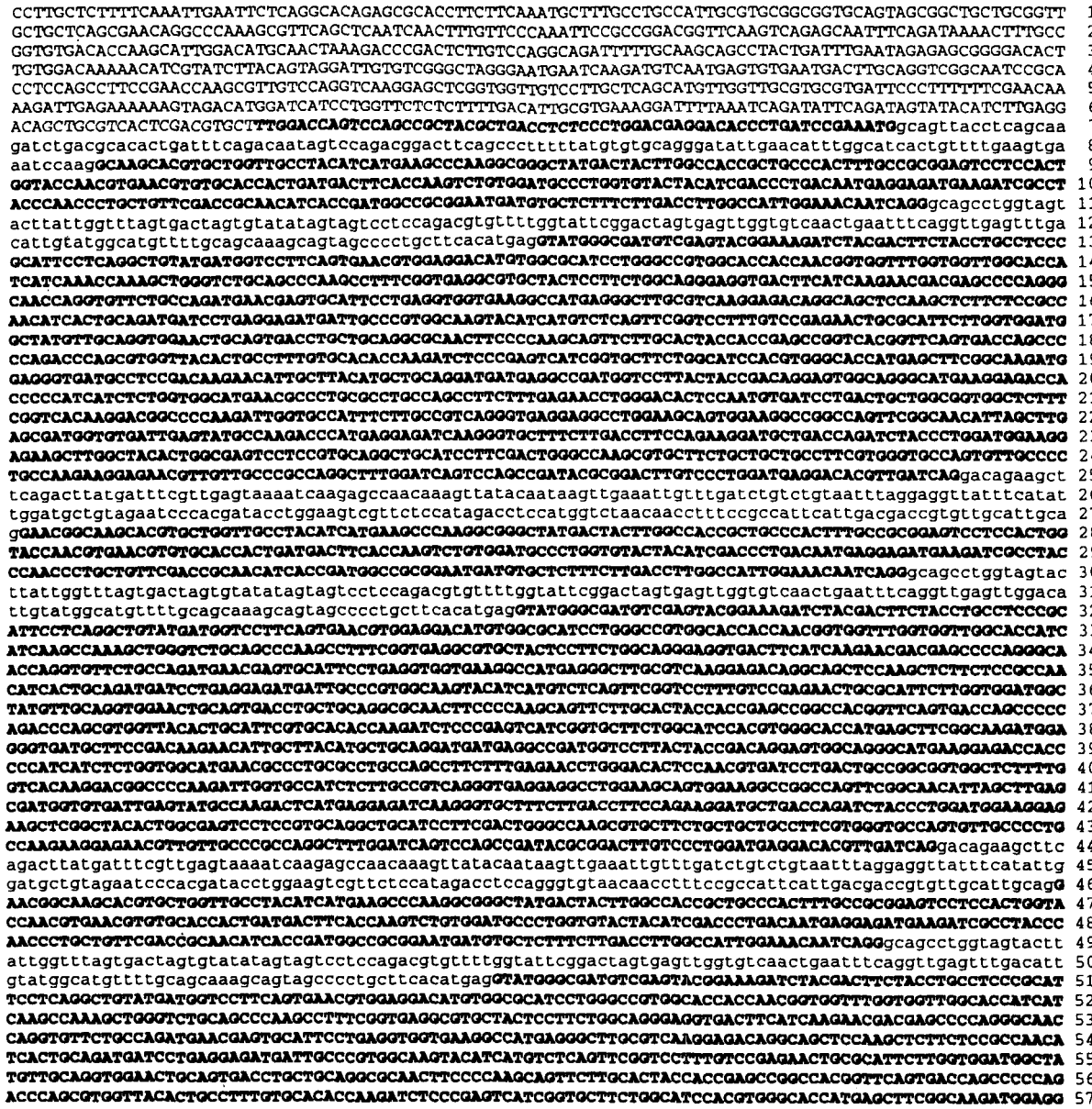

**Figure 1.** Sequence of the *rbcA* Locus.

The identified ORF is presented in boldface uppercase letters. Other letters indicate 5′ and 3′ flanking sequences (uppercase) and introns (lowercase).

encoding a "Rubisco-like" protein. In every respect, these differ from Rubisco genes in other eukaryotes: *Symbiodinium* Rubisco is clearly a form II enzyme that is ~65% identical with that from *Rhodospirillum rubrum* (form I and II Rubiscos are only ~25 to 30% identical [Narang et al., 1984]); it is nuclear encoded by a multigene family; and a predominantly expressed enzyme is encoded as a precursor polyprotein. The inferred amino acid sequence of one *Symbiodinium* gene violates the "rules" established for all other Rubiscos, but it does not appear to be a pseudogene. This Rubisco-like gene demonstrates the evolutionary capacity of a multigene family.

Thus, both known forms of Rubisco participated in the evolution of eukaryotic photosynthesis, and they have followed different genetic pathways. Whereas the origin(s) of eukaryotic form I Rubiscos is discussed with the origin(s) of the plastids that encode them (Martin et al., 1992; Morden et al., 1992; Palmer, 1993), the history of nuclear-encoded form II enzymes is open to wider speculation. The genetic complexity of *Symbiodinium* Rubisco may reflect the greater evolutionary potential of a nuclear-encoded as opposed to a plastid-encoded gene, it may be relevant to the physiology of endosymbiosis, and it should provide valuable material for research into this important enzyme.

## RESULTS

Amino acid sequences from each of eight Rubisco peptides could be aligned with the form II Rubisco of *R. rubrum* (alignment not shown). From these data, a polymerase chain reaction (PCR) amplification strategy was designed so that if *Symbiodinium* and *R. rubrum* genes were exactly colinear, it would yield 819 bp of a corresponding *Symbiodinium* gene. Instead, the PCR product was ~1000 bp (data not shown). Restriction enzyme–digested *Symbiodinium* DNA contained many fragments that hybridized to this PCR product (data not shown, but see below); BamHI fragments of ~20 and 14 kb were obtained in the vector EMBL3. The two clones, designated rbcA and rbcG, respectively, have different restriction maps (data not shown).

### Analysis of rbcA

Rubisco-encoding sequences were localized by DNA gel blotting to an ~10-kb EcoRI fragment of rbcA (data not shown), which was sequenced in its entirety. Part of this sequence is presented in Figure 1. The Rubisco-encoding locus (*rbcA*), as interpreted here, is summarized schematically in Figure 2. The predicted protein sequence is presented in Figure 3.

Protein-encoding regions were inferred by similarity (after conceptual translation) to *R. rubrum* Rubisco. They occur in three tandem copies (designated genes *rbcA1*, *rbcA2*, and *rbcA3*, in the 5′ to 3′ direction) that, with the exception of one nonsynonymous substitution in *rbcA3* (Asn replaced by Lys; codons at positions 2069, 3967, and 5866 in Figure 1), encode identical proteins. There is good agreement between the predicted amino acid sequence and the peptide sequence data (122 matches of 125 amino acids that could be determined with confidence; of seven additional positions at which two amino acids seemed equally likely, *rbcA* encodes one of those residues in six instances; 128/132 matches = 97%; peptide sequence data presented within parentheses in Figure 3). Each gene is interrupted by two introns. Intron 2 accounts for the increased size of the PCR product, compared with the prediction from a bacterial gene (see above).

The N and C termini of the protein encoded by *rbcA* were assigned from protein sequence data (Figure 3; Whitney et al., 1995) and from an in-frame termination codon (TAG; the universal genetic code has been used in all analyses) in *rbcA3*, respectively. The N-terminal amino acid was not determined unequivocally (Gly or Arg seemed likely); the *rbcA* sequence specifies Leu (TTG) at this position. Although this codon might initiate translation, it seems probable that N termini are instead created by proteolytic processing of a larger precursor: *rbcA1*, *rbcA2*, and *rbcA3* are separated by two identical 69-bp "spacers" that link all three genes into one open reading frame (ORF). Thus, *rbcA* appears to encode a polyprotein of at least (the actual translation start site remains unspecified) 1501 amino acids (Figure 3), which is processed to yield three Rubisco polypeptides of 485 amino acids each; their calculated molecular mass (53.2 kD) agrees well with the estimate of 56 kD for purified Rubisco subunits (Whitney and Yellowlees, 1995).

Introns (designated by lowercase letters in Figure 1) were confirmed by sequencing a cDNA clone that encodes one complete Rubisco polypeptide. The clone was obtained by PCR amplification of oligo(dT)-primed *Symbiodinium* cDNA, using the 5′ and 3′ terminal sequences of *rbcA3* for the primers and sequenced completely (S.M. Whitney, unpublished data). The 5′ intron (intron 1) is identical in position and similar in sequence (two substitutions) in *rbcA2* and *rbcA3*. Intron 1 in *rbcA1* is different both in sequence and (apparently) in position, with a shift

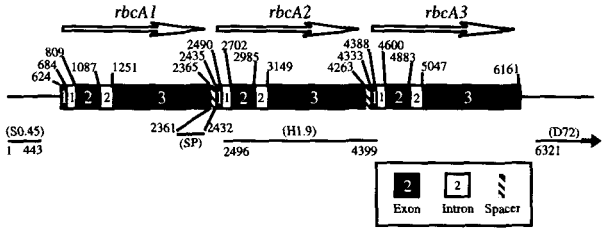


**Figure 2.** Schematic Diagram of the *rbcA* Locus.

Bars identify exons (solid), introns (open), and spacer peptides (hatched), as indicated. Hybridization probes (S0.45, SP, H1.9, and D72) are diagrammed as lines. Numbers give the DNA sequence positions (as given in Figure 1) of the first (5′) and last (3′) nucleotides in the exons (and of the nucleotides that flank spacer peptide-encoding sequences) and hybridization probes; probe D72 extends beyond the presented sequence by 3268 bp.

```
                                                                    45
LDQSSRYADLSLDEDTLIRNGKHVLVAYIMKPKAGYDYLATAAHF
(GDQSSRYADLSLDEDTLIRNG)        (MKPKAGYDYLATAAHF)
R                                          G T
                                                                    90
AAESSTGTNVNVCTTDDFTKSVDALVYYIDPDNEEMKIAYPTLLF
                    (KSVDALVYYIDPDNEEMKIAYPTLLF
                     RG
                                                                   135
DRNITDGRGMMCSFLTLAIGNNQGMGDVEYGKIYDFYLPPAFLRL
DRNITDG)
                                                                   180
YDGPSVNVEDMWRILGRGTTNGGLVVGTIIKPKLGLQPKPFGEAC
          (MGRILGRGTTNGGLV)
                    D L
                                                                   225
YSFWQGGDFIKNDEPQGNQVFCQMNECIPEVVKAMRACVKETGSS

                                                                   270
KLFSANITADDPEEMIARGKYIMSQFGPLSENCAFLVDGYVAGGT
              (MIARGKYXTDQFG)
                                                                   315
AVTCCRRNFPKQFLHYHRAGHGSVTSPQTQRGYTAFVHTKISRVI
                                                                   360
GASGIHVGTMSFGKMEGDASDKNIAYMLQDDEADGPYYRQEWQGM
              (MEGDASDKNIAYMLQDDEA)
                                                                   405
KETTPIISGGMNALRLPAFFENLGHSNVILTAGGGSFGHKDGPKI
          (MNALRLPAFFENLGXXNV)
                                                                   450
GAISCRQGEEAWKQWKAGQFGNISLSDGVIEYAKTHEEIKGAFLT
                                                        485
FQKDADQIYPGWKEKLGYTGESSVQAASFDWAKRASAAAFVGASV
APAKKENVVARQALDQSSRYADLSLDEDTLIRNGKHVLVAYIMKP
KAGYDYLATAAHFAAESSTGTNVNVCTTDDFTKSVDALVYYIDPD
NEEMKIAYPTLLFDRNITDGRGMMCSFLTLAIGNNQGMGDVEYGK
IYDFYLPPAFLRLYDGPSVNVEDMWRILGRGTTNGGLVVGTIIKP
KLGLQPKPFGEACYSFWQGGDFIKNDEPQGNQVFCQMNECIPEVV
KAMRACVKETGSSKLFSANITADDPEEMIARGKYIMSQFGPLSEN
CAFLVDGYVAGGTAVTCCRRNFPKQFLHYHRAGHGSVTSPQTQRG
YTAFVHTKISRVIGASGIHVGTMSFGKMEGDASDKNIAYMLQDDE
ADGPYYRQEWQGMKETTPIISGGMNALRLPAFFENLGHSNVILTA
GGGSFGHKDGPKIGAISCRQGEEAWKQWKAGQFGNISLSDGVIEY
AKTHEEIKGAFLTFQKDADQIYPGWKEKLGYTGESSVQAASFDWA
KRASAAAFVGASVAPAKKENVVARQALDQSSRYADLSLDEDTLIR
NGKHVLVAYIMKPKAGYDYLATAAHFAAESSTGTNVNVCTTDDFT
KSVDALVYYIDPDNEEMKIAYPTLLFDRNITDGRGMMCSFLTLAI
GNNQGMGDVEYGKIYDFYLPPAFLRLYDGPSVNVEDMWRILGRGT
TNGGLVVGTIIKPKLGLQPKPFGEACYSFWQGGDFIKNDEPQGNQ
VFCQMNECIPEVVKAMRACVKETGSSKLFSANITADDPEEMIARG
KYIMSQFGPLSENCAFLVDGYVAGGTAVTCCRRNFPKQFLHYHRA
GHGSVTSPQTQRGYTAFVHTKISRVIGASGIHVGTMSFGKMEGDA
SDKNIAYMLQDDEADGPYYRQEWQGMKETTPIISGGMNALRLPAF
FENLGHSKVILTAGGGSFGHKDGPKIGAISCRQGEEAWKQWKAGQ
FGNISLSDGVIEYAKTHEEIKGAFLTFQKDADQIYPGWKEKLGYT
GESSVQAASFDWAKRA
```

**Figure 3.** Predicted Protein Product of the *rbcA* Locus.

The ORF identified in Figure 1 encodes three Rubisco subunits of 485 amino acids each, separated by spacer peptides (underlined). Sequences obtained directly from Rubisco peptides are presented in parentheses below the first (N-terminal) subunit (where two residues seemed equally likely, both are given; X, residue not identified). Peptide sequences in italics represent the PCR primers used to obtain a partial genomic clone. Amino acids given in boldface letters indicate the positions of introns in *rbcA*, as given in Figures 1 and 2.

by 5 bp in the 3′ direction (relative to its position in *rbcA2/3*), generating agreement with the cDNA and the N-terminal protein sequences. Intron 2 is identical in *rbcA1*, *rbcA2*, and *rbcA3*.

RNA gel blot analyses presented in Figure 4 support the above interpretation of *rbcA*. Probe H1.9, which includes all three exons, both introns, and the intergenic spacer (probes are diagrammed in Figure 2), detected an abundant (implied by the strength of the signal; see Figure 4 legend for details) RNA of ~5200 nucleotides. The same (size of) RNA was also detected by probe S0.45, which includes 5′ flanking sequences, by probe SP, which includes most of the 69-bp intergenic spacer and 4 bp of exon 3, and by probe D72, which includes 3′ flanking sequences. (Probe D72, which extends 3268 bp beyond the sequence presented in Figure 1, also hybridizes with an RNA of ~1600 nucleotides; this RNA has not been investigated.) As given in Figure 1, the ORF encoded by *rbcA* is 4503 nucleotides, which is ~700 nucleotides smaller than this RNA. This difference can be accounted for by transcribed sequences on the 5′ and 3′ side (detected by probes S0.45 and D72, respectively). In addition, the efficient selection (~95%) of this RNA by chromatography on oligo(dT) cellulose (data not shown) implies a poly(A) tail.

Sequence differences among *rbcA1*, *rbcA2*, and *rbcA3* are conspicuous only at the 5′ end of *rbcA1*. These include a different intron 1 and nine synonymous nucleotide substitutions (in 20 codons) in exon 1. In comparison, exon 1 is identical in *rbcA2* and *rbcA3*, all three exon 2 sequences (94 codons) are identical, and exon 3 sequences exhibit 11 synonymous substitutions (*rbcA1* versus *rbcA2*) or 14 synonymous and one replacement substitutions (*rbcA1* versus *rbcA3* and *rbcA2* versus *rbcA3*) in a total of 372 codons. The nonrandom distribution of nucleo-

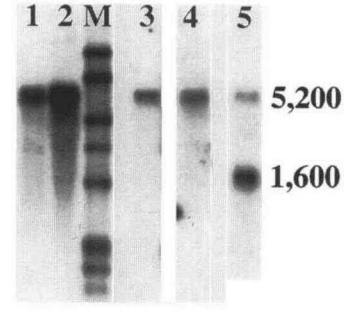


**Figure 4.** RNA Gel Blot Analysis of *Symbiodinium* RNA.

Lane M contains RNA size standards (9488, 6225, 3911, 2800, 1898, 872, and 562 nucleotides from top to bottom). Other lanes contain 1.6 μg (lane 1) or 3.2 μg (lanes 2 to 5) of total *Symbiodinium* RNA, which was probed with H1.9 (lanes 1 and 2), S0.45 (lane 3), SP (lane 4), or D72 (lane 5). The hybridized blots were washed in 0.25 × SSPE (lanes 1 and 2), 0.5 × SSPE (lanes 3 and 5), or 1 × SSPE (lane 4) and exposed to film either at room temperature for 20 min (lanes 1 and 2) or at −80°C with an intensifying screen for 48 hr (lane 3), 87 hr (lane 4), or 72 hr (lane 5). Estimated sizes of the hybridizing RNAs are 5200 nucleotides (all lanes) and 1600 nucleotides (lane 5 only).

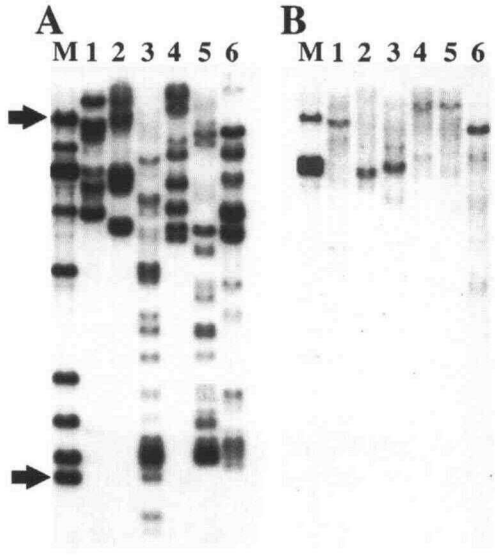


**Figure 5.** DNA Gel Blot Analysis of *Symbiodinium* DNA.

Lanes M contain DNA size standards obtained by digesting genomic λ clones rbcA and rbcG with various enzymes. Approximate lengths in **(A)** are, top to bottom, 20 (arrow), 14, 10, 7, 5, 3, 2.3, 1.9, and 1.7 (arrow) kb; in **(B)**, 20, 11.6, and 10 kb. Other lanes contain ~4 μg total zooxanthella DNA digested with BamHI (lanes 1), EcoRI (lanes 2), HindIII (lanes 3), SacI (lanes 4), KpnI (lanes 5), or PvuII (lanes 6).
**(A)** A blot probed with H1.9 was washed in 0.25 × SSC and exposed to film with an intensifying screen at −80°C for 72 hr.
**(B)** The same blot in **(A)** was probed with D72, washed in 0.25 × SSC, and exposed to film with an intensifying screen at −80°C for 98 hr.

tide variation among the three genes, clustered near the 5′ end of the locus, focuses attention on exon 1 of *rbcA1*. Probably, it is incomplete, as diagrammed in Figures 1 and 2, and is preceded by another exon with which it encodes a transit peptide (see Discussion).

*rbcA* hybridized to multiple restriction fragments of *Symbiodinium* DNA on gel blots. An example, using the genomic DNA from which rbcA and rbcG were isolated, is presented in Figure 5. The complex hybridization patterns of probe H1.9 (Figure 5A) suggest that *rbcA* is part of a multigene family. Alternatively, these patterns could represent partial digestion of a single Rubisco locus. The modified bases in *Symbiodinium* DNA (Blank et al., 1988), which in theory could vary from cell to cell, might possibly render this DNA only "partially digestible." Increased concentrations of restriction enzymes and longer digestion times (with fresh enzyme added at intervals) had no effect on the hybridization patterns (data not shown). Another possibility is that individual clams contain genetically heterogeneous populations of zooxanthellae (see Rowan and Knowlton, 1995), with allelic variation at one Rubisco locus among these "strains" yielding the multiple restriction fragments.

The hybridization of genomic digests to probe D72, which is located 3′ to the *rbcA* genes (see Figure 2), did not support the hypothesis of partial digestion. The predominance of single bands (Figure 5B) of the size predicted from the rbcA restriction map (data not shown) implies that the corresponding DNA was well digested. Minor bands of smaller sizes could represent other loci. Whether these represent different loci within one organism or allelic variation at *rbcA* among several taxa cannot be determined from these data.

The second hypothesis was tested by analyzing a clonal culture of a zooxanthella obtained from *Tridacna squamosa* (initial efforts to obtain clonal cultures from *T. gigas* were not successful). As shown in Figure 6A, genomic DNA from this clone exhibits about as much complexity as do freshly isolated zooxanthellae, demonstrating that multiple *rbcA*-hybridizing restriction fragments are a feature of one *Symbiodinium* genome. Moreover, this one alga contains loci similar to both *rbcA* and *rbcG* (see below). Probe D72 did not hybridize to genomic DNA from the clonal isolate, probably because it is a different species of *Symbiodinium* (see below).

As described above, *rbcA* exhibits characteristics of a nuclear gene. DNA gel blot hybridizations confirmed this suspicion. Quantitative data presented in Figure 7A demonstrate that *rbcA*-hybridizing sequences are equally abundant in the DNA obtained from highly purified nuclei and from whole cells. If these sequences were nonnuclear, DNA prepared from purified nuclei should have a lower specific signal. The qualitative data presented in Figure 7B demonstrate that the restriction fragment complexity is equivalent in both cases, indicating that all (as opposed to most) *rbcA*-hybridizing DNA fragments are

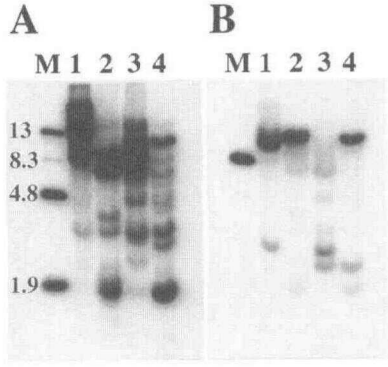


**Figure 6.** DNA Gel Blot Analysis of Total DNA from a Clonal *Symbiodinium* Culture.

Lanes contain ~3 μg of genomic DNA digested with XhoI (lanes 1), ApaI (lanes 2), SalI (lanes 3), or BamHI (lanes 4). Lanes M contain the *rbcA* probe sequence (~12 pg) on cloned fragments of ~13, 4.8, and 1.9 kb and the *rbcG* probe sequence (~12 pg) on a cloned fragment of ~8.3 kb. Duplicate gels (one-half of a digest of ~6 μg of genomic DNA was run on each) were blotted and hybridized separately. The extent of cross-hybridization between *rbcA* and *rbcG* is shown in lanes M.
**(A)** One blot was hybridized with the *rbcA* probe H1.9, washed in 0.25 × SSC, and exposed to film with an intensifying screen at −80°C for 32 hr.
**(B)** The other blot was hybridized with the *rbcG* probe S3.2 and processed as given in **(A)**.

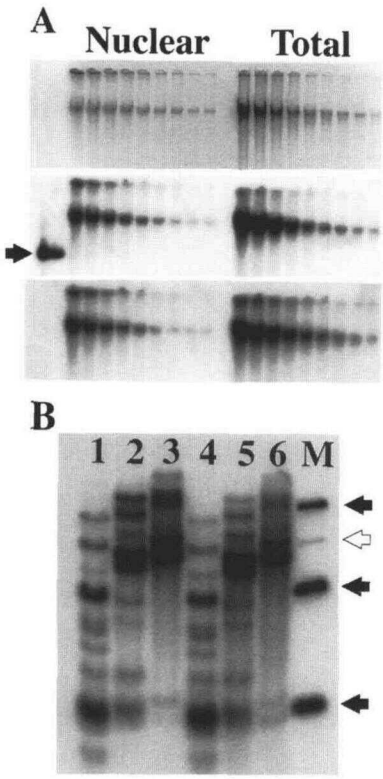


**Figure 7.** Localization of *rbcA* to the Nuclear Genome.

**(A)** Quantitative DNA gel blot analysis. Serial dilution series (by a factor of 0.7) of undigested nuclear (Nuclear) and undigested total cell (Total) DNA (both pretreated with ribonuclease) were electrophoresed on the same gel. At top is the negative image of the ethidium bromide–stained gel before denaturation and blotting. In the middle is the blot probed with H1.9 and washed with 0.25 × SSC; signals in the first two lanes (arrow), which do not appear in the other panels, reveal ~25 and 12 pg, respectively, of the probe sequence (in an ~13-kb linearized plasmid). At the bottom is the same blot reprobed with 178-6 (a cloned, *Symbiodinium* nuclear ribosomal RNA gene), confirming equally efficient transfer of both nuclear and total DNAs to the hybridization membrane.

**(B)** Qualitative DNA gel blot analysis of nuclear (lanes 1 to 3) and total cell (lanes 4 to 6) DNA. Lanes contain ~2.5 μg of DNA digested with HindIII (lanes 1 and 4), PvuII (lanes 2 and 5), and BamHI (lanes 3 and 6). Lane M contains the *rbcA* probe sequence (~12 pg) on cloned fragments of ~13, 4.8, and 1.9 kb (solid arrows) and the *rbcG* probe sequence (~12 pg) on a cloned fragment of ~8.3 kb (open arrow). The blot was probed with H1.9, washed in 0.25 × SSC, and exposed with an intensifying screen at −80°C for 36 hr. The lesser relative intensities of larger fragments in lanes 4 to 6, as compared with lanes 1 to 3, are explained by the better integrity of the nuclear DNA, as seen in **(A)**.

nuclear. Complementary and control hybridizations for these analyses, using purified chloroplast DNA and/or cloned sequences known to be chloroplast encoded, were precluded by our failure to purify chloroplast DNA convincingly from *Symbiodinium* and by the complete lack of molecular data on the chloroplast genomes of peridinin-containing dinoflagellates.

Taken together, the DNA gel blot analyses of Rubisco gene organization, including those presented below, indicate that *rbcA* is part of a nuclear, multigene family in *Symbiodinium*.

### Analysis of rbcG

Rubisco-encoding sequences were localized by DNA gel blotting to a 5.1-kb XbaI-BamHI fragment of rbcG (data not shown). This fragment was cloned, and deletion subclones were screened (by sequencing) for homology with *rbcA*. One putative gene (*rbcG*) was located and sequenced; 3.1 kb of these data are presented in Figure 8. Using the same interpretive reasoning as above, *rbcG* has an intron/exon organization like that of *rbcA*, summarized in Figure 9.

The N terminus of the putative *rbcG*-encoded protein was assigned by apparent homology with the *rbcA*-encoded protein. The first in-frame termination codon in the *rbcG* ORF extends well beyond (64 codons) that of *rbcA*. Compared with *rbcA*, intron 1 is larger and intron 2 is smaller in *rbcG*; there is no obvious similarity in intron sequences between the two loci. Intron 1 has been assigned to a more 5′ position in *rbcG*, owing to a termination signal (TGA) at what would otherwise be the seventeenth codon in exon 1. Intron 2 occurs at the same position in both loci.

On RNA gel blots, the *rbcG*-derived probe S3.2 (Figure 9) apparently hybridizes with the same RNA that was detected by *rbcA*, albeit with decreased efficiency as would be expected from the sequence divergence between the two loci (comparison not shown). The smaller size of the putative *rbcG* ORF (one gene versus three genes) suggests that a corresponding mRNA should also be smaller, but no such RNA was detected. One explanation was provided by analyzing the genomic DNA obtained from a variety of zooxanthella isolates.

Because *rbcA* and *rbcG* cross-hybridize weakly (shown by reciprocally probing the cloned loci; Figure 6), the two loci are distinguishable on DNA gel blots. Although *rbcA* (or *rbcA*-like) loci were always detected (e.g., Figure 5A), *rbcG*-related loci were not detected convincingly (weak signals were observed, but these can be attributed to other, e.g., *rbcA*-like, loci) in *Symbiodinium* isolated from *T. gigas*. However, both loci were detected, with comparable efficiency (as determined from cloned *rbcA* and *rbcG* DNAs that were included on the blot as hybridization controls), in the clonal *Symbiodinium* isolate (Figure 6). This demonstrates that both *rbcA*-like and *rbcG*-like loci are present in a single *Symbiodinium* genome. The simple hybridization patterns of *rbcG* (Figure 6B), compared with the more complex patterns of *rbcA* hybridization (Figure 6A), imply that the latter represent multiple loci rather than products of partial digestion.

This discrepancy between zooxanthellae freshly isolated from *T. gigas* (in which only *rbcA* could be demonstrated) and the zooxanthella cultured from *T. squamosa* (in which both loci were detected) correlates with a taxonomic difference. By an analysis of small ribosomal subunit RNA (srRNA) genes (data not presented), zooxanthellae obtained from *T. gigas* were iden-

```
AACATGCCAGGGAAGAAGCTTTTTCTACAGGTTGCTAGCCTCCAGTTAAACCGAGGTGTCCAGCTTTGAGTCCTGAGTGAGCCGAGAGATGTCAAAAGTC  100
AAACCGCAGAGCACTTTTCACTTCTGCGCTGACACCCGTGGCACGCCTTGCGCACGAAAAGATGCCAGAGCCGGAAGAGGAGCCTGCAGAGGCCGCAGTG  200
GCACAGGACCCTTCTTTGCCGGAACAGCAGGTTTCGGGTGATGCAAAGTCCTCCTTGTCAGGTTCTGCAGATGAAAGTTTGGGCTGAACTCCGACCCATT  300
GAAAGCATCTAAGCCTTTGCAGGAAGATGCCGTGGCCTGTCGTACCTTGGAACAGTCCAGCCGCTATGCCGACCTCACTCTGGATGAGGCTGGGTgaggg  400
                                                  L  E  Q  S  S  R  Y  A  D  L  T  L  D  E  A  G
ctggacgggttgcccagcatcctccattcagtcacattcgtatgtttgtgctgttcgttccttcaggcggggtgttctctctgtctctctctcccgccttt  500
tctctttgctttccctggtctctaatgttacaaagccagcaatgggttgcctgggaagacacaggaatcctttttggttgggcagagtgttgctcgtgaagt  600
tgtgccgcagattcgcatctcgtgtcgtcgtgcattgggtcttgtctagttgggctaggctgcgtccagctacccgggggtttaccgatttgacccacagt  700
tgaattccgtcatgctacgatgaacccgcagctcggaaccttgtgcaggcgaccttggtaaagcaactcgctgcaaggcagtacagctttagttccgcaa  800
cttgtgtcactttaccgtgggtgctgagtctgccagttccactcatgctggCCTTAAGGAACGGGGAGCATGTGCTGGTCGCATATATCATGAAGCCGAA  900
                                                  S  L  R  N  G  E  H  V  L  V  A  Y  I  M  K  P  K
AATGGGTTACGATTACCTGGCCACAGCCGCCCATTTCGCTGCTGAGTCTTGTACTGGCGCCAGAAGCACAGACGACGCCAAGGAGTCAGCAAACCGCAGTA 1000
  M  G  Y  D  Y  L  A  T  A  A  H  F  A  A  E  S  C  T  G  A  R  S  T  D  D  A  K  E  S  A  N  A  V
GTGTACTACATCGACGCGGACAGTCAGGAGATGAGAATTGCCTACCCGACCGTGCTCTTCGACAGCAACCTCACCGATGGCCGCGACATGGTGAGGTCCT 1100
 V  Y  Y  I  D  A  D  S  Q  E  M  R  I  A  Y  P  T  V  L  F  D  S  N  L  T  D  G  R  D  M  V  R  S
TCCTGACCCTCGCCATTGGCAACACTCAGGgtcccgggcttggtgtcagaaatcggagcctgaaatggttggagtaactccttgctccatggccgactga 1200
F  L  T  L  A  I  G  N  T  Q
ctgagGTATGGAAGATGTCGAGTTCGGCAAGATCTACGACTTCTACTTGCCCCCGTCGTTCTTGCGACTACAGGAAAGCCCATCCGTGAACACGATGTGG 1300
     G  M  E  D  V  E  F  G  K  I  Y  D  F  Y  L  P  P  S  F  L  R  L  Q  E  S  P  S  V  N  T  M  W
CGCATTCTTGACAAGGGCGCCAGCAATGGAGGCCTTGCGGCAGGTGCCATGAAGCCCAAGCTTGGCATGCAGCCAAAACCTTTCCGTGCGGCCTCCTATG 1400
 R  I  L  D  K  G  A  S  N  G  G  L  A  A  G  A  M  K  P  K  L  G  M  Q  P  K  P  F  R  A  A  S  Y
CGCTCTGGCAAGGTGGAGCATACACCACAGCCGACGTGCAGGCAAACCAGGCCCCCACCCAAGCAAGTGAGGGCATTCCGGAGGTCGTGAAGGCCATGAG 1500
A  L  W  Q  G  G  A  Y  T  T  A  D  V  Q  A  N  Q  A  P  T  Q  A  S  E  G  I  P  E  V  V  K  A  M  R
GTCGTGCGTCCAAGGGACTGATGTGGCTAAGCGGCATTCGGGTGGCAACTACACTGATGAACCGAAGGAGATGGATGCTCGTGGAAAGTACGTCCTGTCG 1600
  S  C  V  Q  G  T  D  V  A  K  R  H  S  G  G  N  Y  T  D  E  P  K  E  M  D  A  R  G  K  Y  V  L  S
CAGTTTGGACCTTTCTCCGGAAATTGCGTTTCCTTGGTGGCAGGATATGCAGCGGGAGGCAGTGCCCGGCACACTTTCCCCAGGCAGTTCGTTCATTACC 1700
 Q  F  G  P  F  S  G  N  C  V  S  L  V  A  G  Y  A  A  G  G  S  A  R  H  T  F  P  R  Q  F  V  H  Y
ATCGTGCGGGACAAGCATCCACAAGCAGCCCCCAGACCCAACGCGGCTACTCTGCTTTCGTGCACACCAAGATCTCCCAAGTCATCAGCACCACCAGCAT 1800
H  R  A  G  Q  A  S  T  S  S  P  Q  T  Q  R  G  Y  S  A  F  V  H  T  K  I  S  Q  V  I  S  T  T  S  I
CCACGCCGACACCATGAGCTTCGGCAAGACGCAGGGTGATTCCCCCGACAAGAGTATCGCTGTCATGCTGCAGGACGACGCGGCCGATGCCCGTCTCTGC 1900
  H  A  D  T  M  S  F  G  K  T  Q  G  D  S  P  D  K  S  I  A  V  M  L  Q  D  D  A  A  D  A  R  L  C
AATCAGGAGTGGGAAGGCATGAGGCAAAATGCTCCCATTATTTCGGGTAGCATGAATGCTCTCCGCCTGCCGGCTTTCTTTGAGACCCTTGGCAACTCCA 2000
 N  Q  E  W  E  G  M  R  Q  N  A  P  I  I  S  G  S  M  N  A  L  R  L  P  A  F  F  E  T  L  G  N  S
ACGTGATCCTGACTGCCGACAGCGGTTCTTTCGGCCACAAGGACGGCCCTGCGTCTGGTGCAATCGCTTGCCGCCAAGTTGAGGAAGCATGGAAGGCCTG 2100
N  V  I  L  T  A  D  S  G  S  F  G  H  K  D  G  P  A  S  G  A  I  A  C  R  Q  V  E  E  A  W  K  A  W
GAGGTCTGGACAGTATGGCAACGTCAGCTTGAGCGATGGTGTGGTTGAGTTCGCCAAGACCCACGAGGAGATCAAGGGTGCCTTCCTCACTTTCCCGAAG 2200
  R  S  G  Q  Y  G  N  V  S  L  S  D  G  V  V  E  F  A  K  T  H  E  E  I  K  G  A  F  L  T  F  P  K
GATGCAGACGAGATATACCCGGGCTGGAAGGAGAAGCTCGGCTGCACGGCCGAGGCGTCTGTACCCCCGGCAAGCTTTGTGAATGCAAAGATCTCTACAG 2300
 D  A  D  E  I  Y  P  G  W  K  E  K  L  G  C  T  A  E  A  S  V  P  P  A  S  F  V  N  A  K  I  S  T
CCTCCTCAGCCGCAGTAGCCACGGCGCGCACCACTATGAATGCCGCGAAGATGGCCTCGCAAAGCACAGCAGGATCAGCGGTCAACCCGTACACCGGTGG 2400
A  S  S  A  A  V  A  T  A  R  T  T  M  N  A  A  K  M  A  S  Q  S  T  A  G  S  A  V  N  P  Y  T  G  G
ACTCAAAAGTATCCACCCTGCCAGCAGCAGCAACACGGCACGAACCGTGCTCCTCTCGCGGCAGGGGCGCGAGGCTGCGCTTCAGGACTGAACAAGAATA 2500
  L  K  S  I  H  P  A  S  S  S  N  T  A  R  T  V  L  L  S  R  Q  G  R  E  A  A  L  Q  D  *
AATAATAACCATGCCACGAACCCTCCGTAAAGCCCCGACCCTCAAAAAAGTCAGAGCAGGATTCGCCCAAGCAGAACCAAGACGTCCGCATGGCGTCCCA 2600
GACCGCTCCCCAGGCGGGATTGACACTTGAGCTCATCGACCCTGGTCTTTCGTCACCCATCCTTCCATTGTAGGGCCAGTTGTTCTGTGAGACTGAGGTG 2700
AATTCCAGCGACAAGAAGATGAGCGTGAGAATCGAGCGTGCCGGCACAGACTCACAGCTGTCAGCCACACGGTTGGCCAGAAGCAGCCACGATGCCAGTG 2800
GCGGGTCGCGCACCCTGATGGGAACAGGAAGCACTTCGAGTGATGTCCCTCTACAGCGGCGTGGGTCCATGTCTTTGCTGCAGGACTGCGGCAGTCATCG 2900
CAGTCTAAGATGAGCCACGCCACACGTTGTCAAACACAGTCAGCGCAGGATCAGCGGTCAACCCGTACACCGGTGGACTCAAAAGTATCCACCCTGCCAG 3000
CAGCAGCAACACAGGCACGAACCGTGCTCCTCTCGCGGCAGGGGCGCGAGGCGTCGCTTCAGGACTGACAGAATAATTAATAACCATGCCACGAACCCTC 3100
```

**Figure 8.** Sequence of the *rbcG* Locus and Its Hypothetical Protein Product.

The DNA sequence is presented in uppercase (protein-encoding and flanking DNA) and lowercase (introns) letters. The conceptual translation product is presented below the ORF (the asterisk indicates termination codon).

tified as *Symbiodinium* type C or occasionally as mixtures of *Symbiodinium* type C and a lesser (estimated at <10% of total) amount of *Symbiodinium* type A. (See Rowan and Powers [1991a, 1991b] and Rowan and Knowlton [1995] for details of taxonomic identification.) Post facto analysis revealed that *rbcA* and *rbcG* had been cloned from a mixed population of types A and C. The clonal isolate from *T. squamosa* is *Symbiodinium* type A. We conclude that *Symbiodinium* type A—not *Symbiodinium* type C—contains *rbcG*; both types of *Symbiodinium* contain *rbcA*-like loci. It is impossible to detect unambiguously (by blot hybridization) *rbcG*-like genes or RNA transcripts in zooxanthellae from *T. gigas* because these algae are overwhelmingly *Symbiodinium* type C. Of course, *rbcG* can be detected in such zooxanthellae by molecular cloning, as demonstrated here. We have not yet had the opportunity to analyze RNA from *Symbiodinium* type A.

Conceptual translations of the *rbcA* and *rbcG* ORFs, a *G. polyedra* Rubisco cDNA sequence (Morse et al., 1995), and the homologous sequence from *R. rubrum* are compared in Figure 10 (see below).

## DISCUSSION

This study confirms and extends previous analyses of Rubisco from *Symbiodinium* (Whitney and Yellowlees, 1995; Whitney et al., 1995). Again, it is important to address the possibility that our data originate from contaminants.

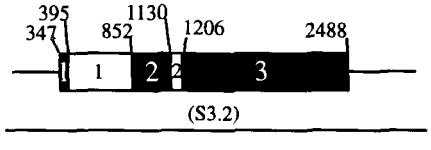


**Figure 9.** Schematic Diagram of the *rbcG* Locus.

The locus is discussed in the text, and the diagram is in the same format as given in Figure 2. The hybridization probe S3.2, a SacII fragment, corresponds to the sequence in Figure 8 plus ~100 bp of additional genomic DNA at either end.

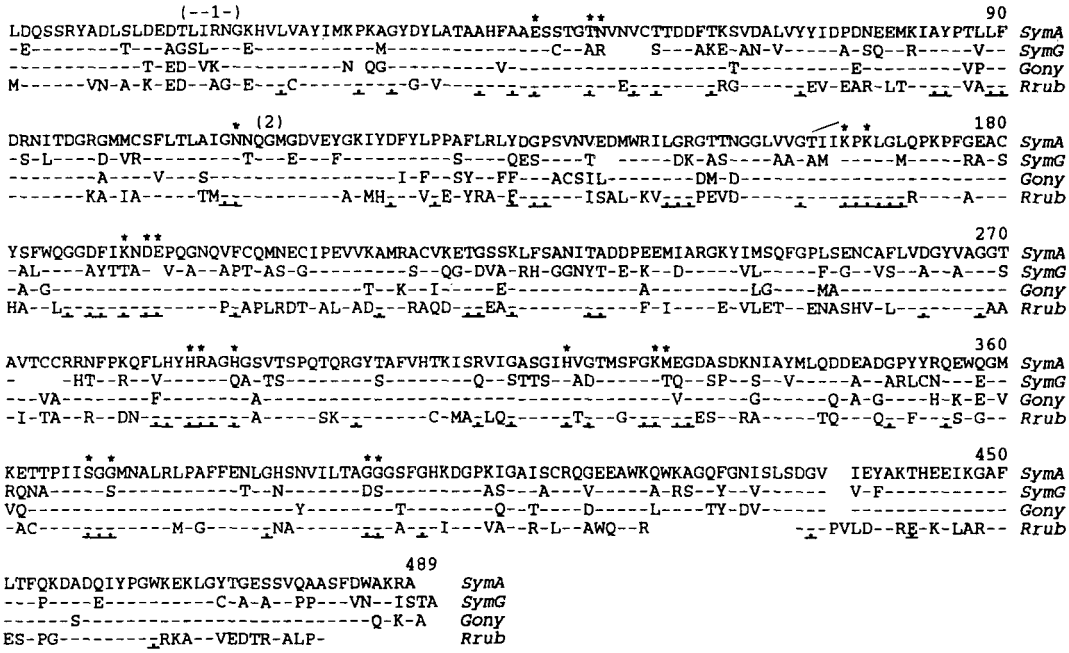


**Figure 10.** Comparison of *Symbiodinium*, *G. polyedra*, and *R. rubrum* Rubisco Protein Sequences.

The full conceptual translation (universal genetic code) of *rbcA1* (from Figure 1) is presented (*SymA*); the translation of *rbcA2* is identical, and that of *rbcA3* differs by the replacement of Lys at position 387. Conceptual translations of *rbcG* (*SymG*; from Figure 8), a *G. polyedra* cDNA (*Gony*), and *R. rubrum* Rubisco (*Rrub*) are presented in comparison (dashes indicate identity; replacements are identified; spaces indicate apparent deletions). (−1-) and (2) identify the locations of introns 1 and 2, respectively, in the *Symbiodinium* genes. Asterisks identify active site residues, as discussed in the text. Underlined dots below the alignment identify residues that are conserved between form I and form II Rubiscos; these are the residues identified by Wagner et al. (1988), less six that are not conserved in form II Rubiscos recently characterized from *Hydrogenovibrio marinus* (Yaguchi et al., 1994) or *Thiobacillus denitrificans* (GenBank accession number L37437).

Zooxanthellae freshly isolated from giant clams appeared unialgal by light microscopy and free of bacterial and animal cell contamination by epifluorescence microscopy after 4′,6-diamidino-2-phenylindole (DAPI) staining. Purified zooxanthella nuclei were similarly free of detectable bacteria, and the localization of all *rbcA*-hybridizing sequences to these nuclei (Figure 7) excludes cytoplasmic bacterial endosymbionts as the source of *rbcA*. Nuclear bacterial endosymbionts (e.g., Silva, 1978) remain a formal possibility, but none has been reported among several ultrastructural studies of *Symbiodinium* nuclei (e.g., Blank and Trench, 1985; Blank, 1987). Also, the high abundance and apparent polyadenylation of *rbcA*-hybridizing RNA and the introns in *rbcA* and *rbcG* are characteristics of eukaryotic, not prokaryotic, organisms. Analyses of srRNA genes confirmed that the zooxanthellae we studied were *Symbiodinium* spp, as expected (Taylor, 1969).

Giant clams conveniently provide large amounts of healthy and exceptionally clean *Symbiodinium* but not without a drawback. Zooxanthellae are not necessarily clonal within an individual host or taxonomically equivalent between different individuals of the same host species. This has been demonstrated for corals (Rowan and Knowlton, 1995) and giant clams, including *T. gigas* (R. Rowan, unpublished data). Analyses of srRNA genes assigned zooxanthellae from *Tridacna* spp to two types within the genus *Symbiodinium*, but because this classification is undoubtedly superficial (Rowan, 1991), different preparations of genomic DNA or RNA could never be assumed to represent the same (or even one) species of symbiont. This caveat restricted empirical analyses of zooxanthella Rubisco genes to general features.

The general features of *Symbiodinium* Rubisco, however, are most interesting. Its similarity to prokaryotic form II enzymes (~54 to 65%; see below), which are only ~25 to 30% similar to form I Rubisco large (L) subunits (Narang et al., 1984), reveals an evolutionary origin among the former, like that of the free-living dinoflagellate *G. polyedra* (Morse et al., 1995). All other known eukaryotic Rubiscos are homologous to the prokaryotic form I enzyme (Martin et al., 1992; Morden et al., 1992). Immunological data indicate that *Symbiodinium* expresses only form II Rubisco (Whitney and Yellowlees, 1995), and genes encoding form I Rubisco could not be detected (by PCR) in *Symbiodinium* DNA, despite considerable effort (R. Rowan, unpublished data). The distinction between Rubisco forms I and II is fundamental (reviewed in Tabita, 1988; Hartman and Harpel, 1994). Form I enzymes are hexadecamers of equal numbers of large (L) and small (S) subunits ($L_8S_8$); form II

enzymes are multimers of one polypeptide ($L_N$) that is homologous to the form I L subunit. Forms I and II Rubisco probably diverged from a common ancestor long ago (Andersen and Caton, 1987), before the emergence of (extant) photosynthetic eukaryotes.

*Symbiodinium* Rubisco is also remarkable in being encoded by a nuclear, multigene family. Formally, *rbcA* (Figures 1 to 3) demonstrates three genes at one locus, and the analysis of a clonal culture (Figure 6) demonstrated at least two loci per *Symbiodinium* genome. DNA gel blot hybridizations (e.g., Figures 5 to 7) suggest that the actual number is larger. Other eukaryotic Rubisco L subunits are encoded by a single locus (which may be duplicated) on chloroplast genomes (Palmer, 1985; Newman and Cattolico, 1990; Reith, 1995). Preliminary studies on *A. carterae* (using *rbcA* as a probe) also imply a nuclear, multigene organization of a form II Rubisco (data not presented; see Whitney and Yellowlees [1995] and Whitney et al. [1995] for related data). Morse et al. (1995) presented indirect evidence that form II Rubisco in *G. polyedra* is nuclear encoded but did not address gene structure or organization. Thus, a nuclear-encoded form II Rubisco may be characteristic of typical (peridinin-containing) photosynthetic dinoflagellates (Morse et al., 1995).

Amino acid sequence data (Figure 3) and RNA gel blot analyses (Figure 4) imply that *rbcA*, or similar loci, encodes the predominant Rubisco in *Symbiodinium* symbiotic with *T. gigas*. The few outright discrepancies between the peptide sequences and the translated *rbcA* ORF (four of 132 amino acids) and ambiguities in the peptide sequences (seven of 132 amino acids) may represent taxonomic variation (genomic clones and purified protein were obtained from different populations of zooxanthellae), the expression of other Rubisco-encoding loci, or both. At the RNA level, *rbcA* (and/or similar loci) are abundantly expressed (Figure 4). An estimate of Rubisco in *Symbiodinium* of 1 to 2% of soluble protein (Whitney and Yellowlees, 1995), based on bound $^{14}$C-carboxyarabinitol bisphosphate, may be regarded as a minimum because only activated enzyme binds this inhibitor (Hartman and Harpel, 1994).

Apparently, *rbcA* encodes a precursor polyprotein. An alternative interpretation, that it encodes a stable polycistronic mRNA, is considered unlikely because such mRNAs are unknown in eukaryotes, whereas polyproteins are known. Notably, these include nuclear-encoded chloroplast proteins in the green alga *Euglena* (reviewed in Houlné and Schantz, 1993), namely, proteins of the light-harvesting complexes I (Houlné and Schantz, 1988) and II (Muchhal and Schwartzbach, 1992), and the Rubisco (form I) S subunits (Chan et al., 1990; Tessier et al., 1995). These *Euglena* loci encode mature proteins separated by decapeptide spacers that exhibit a bipartite structure: the N-terminal half is relatively hydrophobic and rich in Ala; the C-terminal half is relatively polar (Houlné and Schantz, 1993). The *rbcA* spacer peptide (underlined in Figure 3) is larger (23 residues) but structurally similar. This motif, which is reminiscent of bacterial signal sequences, may direct the proteolytic processing of the precursor (Schiff et al., 1991; Houlné and Schantz, 1993).

A form II Rubisco in *G. polyedra* (Morse et al., 1995) also seems to be encoded as a precursor polyprotein. A partial cDNA sequence (GenBank accession number L41063) encodes a Rubisco ORF preceded by the 41 C-terminal amino acids of that ORF and a 20-residue peptide similar to (~50% identity) the *Symbiodinium rbcA* "spacer peptide." This sequence could represent slightly more than one-third of a *G. polyedra* homolog of *rbcA* (i.e., the C terminus of *rbcA2*, spacer, and all of *rbcA3*); Morse et al. (1995) did not mention the size of Rubisco mRNA in *G. polyedra*.

Assuming that *Symbiodinium* Rubisco is a chloroplast (stroma) protein, the *rbcA*-encoded precursor should contain a transit peptide (Keegstra et al., 1989) preceding *rbcA1*. In-frame termination codons and the absence of potential initiation (ATG) codons in the immediate vicinity suggest that part of the hypothetical transit peptide is encoded by a separate, upstream exon. RNA gel blot analyses (Figure 4, lane 3) demonstrated upstream sequences in *rbcA* mRNA. The predicted peptide sequence preceding *rbcA1* (VTRRA ↓—the C terminus of a hypothetical transit peptide) resembles that which precedes *rbcA2/3* (VARQA ↓—the C terminus of the spacer peptide); this is consistent with a common mechanism of proteolysis (↓) at all three mature N termini. The Rubisco S subunit precursor polyprotein in *Euglena*, for which the transit peptide sequence has been defined, exhibits an analogous pattern of conserved residues (Chan et al., 1990). An *rbcG*-encoded protein should also contain a transit peptide, and the preceding peptide/amino termini of the two precursors are somewhat similar (*rbcA*, EDSCVTRRA↓L; *rbcG*, EDAVACRT↓L).

Transit peptides of chloroplast-targeted proteins in *Euglena* are longer than those of higher plants, perhaps because *Euglena* chloroplasts are surrounded by three, as opposed to two, membranes (Chan et al., 1990; Shashidhara et al., 1992). Because this arrangement also exists in most dinoflagellates (Gibbs, 1990), including *Symbiodinium* (Blank, 1987), a long transit peptide is not unanticipated. One could also speculate that the triple-membrane condition favors the organization of abundant, nuclear-encoded chloroplast proteins into precursor polyproteins (Chan et al., 1990).

A second Rubisco locus, *rbcG*, contains only one Rubisco-encoding sequence. This locus was detected only in *Symbiodinium* A, which occurs as a minor symbiont in *T. gigas* (R. Rowan, unpublished data). SrRNA sequence data from *Symbiodinium* types A and C imply that they are not closely related (Rowan and Powers, 1991b, 1992; Rowan and Knowlton, 1995), so a qualitative genetic difference is not surprising. Lacking a good supply of (pure) *Symbiodinium* A, we were unable to verify the expression of *rbcG* by RNA gel blot analysis. Likewise, nuclear DNA from *Symbiodinium* A has not been tested, but *rbcG* is probably a nuclear gene, given an exon/intron organization like that of *rbcA*.

Introns in *rbcA* (and in *rbcG*) apparently represent the first intron sequences reported from dinoflagellates; *rbcA* intron

sequences have been confirmed from cDNA clones (S.M. Whitney, unpublished data). Like eukaryotic spliceosomal introns in general, these *Symbiodinium* introns begin with G and end in AG, but the second nucleotide (typically T) is C or A. *rbcG* introns are hypotheses based on the homology of *rbcG* and *rbcA* exons. Both ends of *rbcG* intron 2 are typical of eukaryotic spliceosomal introns; intron 1 ends are similar (GA; GG). None of the *Symbiodinium* exon–intron junctions exhibits the unusual base pairing inferred for *Euglena* polyprotein pre-mRNAs (Tessier et al., 1995).

Conceptual translations of the *rbcA* and *rbcG* ORFs, a *G. polyedra* Rubisco cDNA (Morse et al., 1995), and *R. rubrum* Rubisco (Narang et al., 1984) are compared in Figure 10. Here, 62 of the 64 additional C-terminal residues encoded by *rbcG* (Figure 8) are removed for simplicity.

The *rbcA*-encoded protein aligns to *R. rubrum* Rubisco with one insertion (13 residues), one deletion (two residues), an eight-residue C-terminal extension, and an overall identity of 64.4% (scoring each deleted/inserted residue as one difference and ignoring the C-terminal extension). The two proteins differ most near the (inferred) C terminus, which is a well-recognized feature of Rubiscos (e.g., Andersen and Caton, 1987; Wagner et al., 1988). The *rbcA*-encoded and *G. polyedra* proteins are of equal length (485 residues), align with a single one-residue gap (position 32), and are 85.6% identical.

Amino acid residues that should be important to Rubisco function have been identified from sequence comparisons, three-dimensional structural analysis, and in vitro mutagenesis (reviewed in Brändén et al., 1991; Hartman and Harpel, 1994). Seventy-four residues are conserved among both form II and form I (L subunit) molecules (see Figure 7 legend). Only two of these (positions 136 and 429 in the *R. rubrum* sequence) are different in the *rbcA*-encoded protein. All 19 residues of the active site (residues within 5 Å of an enzyme-bound, transition-state analog in crystals [Brändén et al., 1991; Hartman and Harpel, 1994]) are conserved. Thus, the protein encoded by *rbcA* resembles a typical and predictably functional Rubisco.

Rubisco catalyzes both the carboxylation and oxygenation of RuBP (Andrews et al., 1973), at relative rates governed by an intrinsic property of the enzyme (the specificity factor, $\tau$, measuring the relative efficiency of carboxylation), temperature, and by the local concentrations of $CO_2$ and $O_2$ (reviewed in Woodrow and Berry, 1988). Bacterial form II Rubiscos have the lowest known values of $\tau$ (Jordan and Ogren, 1981), but because they are expressed in anoxic environments (Tabita, 1988), their high affinity for $O_2$ does not impede net carbon fixation. In contrast, *Symbiodinium* is an oxygenic phototroph, living in a sometimes hyperoxic environment (e.g., Crossland and Barnes, 1977; Dykens and Shick, 1982). A typical form II Rubisco, acting at equilibrium with these conditions, would not yield net carbon fixation (cf. Pierce et al., 1989).

Because *Symbiodinium* does fix $CO_2$ (Muscatine, 1980; Streamer et al., 1993), either its Rubisco's $\tau$ is dramatically greater than that of the prokaryotic form II enzymes, or oxygenase activity is suppressed in vivo by a $CO_2$-concentrating mechanism (Hatch, 1971; Beardall, 1989), or both. Unfortunately, the instability of *Symbiodinium* Rubisco in vitro (Whitney and Yellowlees, 1995) has obviated the determination of its kinetic properties. Natural variation in $\tau$ is a well-established fact (Jordan and Ogren, 1981; Read and Tabita, 1994), and the ~35% nonidentity of the *rbcA*-encoded and *R. rubrum* Rubiscos implies that their $\tau$ values could differ. The participation of form II Rubisco in oxygenic photosynthesis is no less novel than its occurrence in a eukaryote, and the mechanism(s) of oxygenase suppression may also be novel.

The protein encoded by *rbcG* exhibits 66.2% identity with that encoded by *rbcA* (15 of the differences are missing residues, distributed among five deletions; Figure 10) but only 47.6% identity with *R. rubrum* Rubisco. Only 45 of the 74 highly conserved residues (above) remain, a percentage (61%) that is similar to the overall identity between *rbcA*- and *rbcG*-encoded proteins. This suggests that *rbcG* has diverged without typical constraints.

Moreover, one-half (9 of 19) of the predicted active site residues are replaced in the *rbcG*-encoded protein. Foremost is Lys-191 (replaced by Thr), by which Rubisco is activated upon carbamylation of its side chain (Lorimer et al., 1976). Carbamylated Lys-191, Glu-194 (deleted in *rbcG* or replaced by Val), and Asp-193 (which is present) coordinate the metal ion cofactor (usually $Mg^{2+}$) at the active site (Brändén et al., 1991; Hartman and Harpel, 1994). Unactivated (noncarbamylated) Rubisco binds RuBP in the inverse orientation (Brändén et al., 1991; Zhang et al., 1994) and catalyzes neither its carboxylation nor oxygenation (Hartman and Harpel, 1994). All five of the residues that seem to orient bound RuBP (and reaction intermediates) via hydrogen bonds to the $P_1$ phosphate oxygens (Thr-53, Asn-54, Gly-370, Gly-393, and Gly-394) are also replaced in the *rbcG*-encoded protein. In contrast, the $P_2$ binding site (Arg-288 and His-321) remains intact. The other active-site replacements (Thr-330 and Glu-291) involve residues of as yet unspecified function.

Although it is difficult to imagine the *rbcG*-encoded protein as a "normal" Rubisco, its complete ORF (after removing two introns, as was done for *rbcA*)—despite numerous nucleotide substitutions and at least five internal deletions—with codon usages similar to *rbcA* (comparison not presented), argues that *rbcG* is maintained as a polypeptide-encoding gene. It is tempting to rationalize this paradox within the nuclear location of *Symbiodinium* Rubisco genes. There, gene duplication and divergence (Ohta, 1980), and perhaps frequent recombination, should allow *Symbiodinium* Rubisco to evolve at rates (and in ways) that are not available to chloroplast-encoded or prokaryotic Rubiscos. For example, a homomultimeric structure (Tabita, 1988; Hartman and Harpel, 1994) might diversify into heteromultimeric forms. With an active site created from the N-terminal domain of one subunit and the C-terminal domain of another (Larimer et al., 1987; Brändén et al., 1991) and inactivating substitutions in *rbcG* (the absence of Lys-191 and Glu-194; a negative charge in the $P_1$ binding site from Asp-393) restricted to the C-terminal domain, the polypeptide might still contribute to a functional, heteromultimeric enzyme (cf. Larimer et al., 1987).

Whatever its role (if any), *rbcG*'s bizarre sequence emphasizes the possibilities for diversification within a nuclear multigene family. This may be relevant to physiological issues faced by *Symbiodinium.* Multiple loci, expressed independently and encoding different catalytic properties, might provide one mechanism for acclimating to a complex environment (e.g., high versus low $O_2$ within the host [Crossland and Barnes, 1977; Dykens and Shick, 1982] and symbiotic versus free living [Steele, 1977; Taylor, 1983]). Indeed, bacteria that contain two nonidentical Rubisco genes do (Tabita, 1988) or may (Viale et al., 1990; Stoner and Shively, 1993; Yaguchi et al., 1994) preferentially use one or the other under different conditions. Also, an enhanced capacity for Rubisco evolution might facilitate *Symbiodinium's* radiation into diverse hosts and habitats (cf. Chang et al., 1983; Rowan and Powers, 1991b; Rowan and Knowlton, 1995). For studies that utilize natural variation to investigate structure–function relationships (e.g., Read and Tabita, 1994), *Symbiodinium* might provide a wealth of material.

Eukaryotes acquired photosynthesis from prokaryotic endosymbionts that became organelles—the chloroplasts (Palmer, 1993; Whatley, 1993). Rubisco genes excepted, plastome sequence data support the hypothesis that all (studied) chloroplasts derive from one primary eukaryote–cyanobacterium endosymbiosis (see Cavalier-Smith, 1992; Kowallik, 1992; Martin et al., 1992; Morden et al., 1992; Reith and Munholland, 1993; Reith, 1995). Eukaryotic Rubiscos apparently derive from two sources: a cyanobacterium (for green algae, land plants, and *Cyanophora paradoxa*) and an α- or β-purple bacterium (for nongreen algae). This conflict suggests a single ancestral cyanobacterium with two Rubisco proteins (Martin et al., 1992) or that a purple bacterial Rubisco gene invaded during plastid evolution (Morden et al., 1992). *Symbiodinium* and *G. polyedra* (Morse et al., 1995) reveal a third participant and consequently imply an ancestor with three distinct Rubisco genes, an additional gene transfer event, or an independent origin of chloroplasts among the peridinin-containing dinoflagellates.

There are no sequence data from chloroplasts of peridinin-containing dinoflagellates, and any discussion of their origin is speculative (Gibbs, 1990; Cavalier-Smith, 1992; Whatley, 1993). Because *Symbiodinium* Rubisco is not plastome encoded, the current data do not address this issue directly. Of the few form II Rubiscos that have been characterized, those from the α-purple bacteria (*R. rubrum* and *Rhodobacter* [*Rhodopseudomonas*] *sphaeroides*) are most similar to those from *Symbiodinium* (Table 1). Note that *R. sphaeroides* also contains a form I Rubisco that groups with those encoded by nongreen algal chloroplasts (Martin et al., 1992; Morden et al., 1992). If parsimony is used for reconstructing the evolution of eukaryotic photosynthesis (Cavalier-Smith, 1992), this observation draws attention to the α-purple (rather than the β-purple) bacteria as the source of Rubisco in all nongreen algae, including dinoflagellates.

*Symbiodinium* Rubisco could represent one of many genes that translocated from an evolving plastid genome to the nuclear genome (Palmer, 1993) and then diversified (e.g, Dean et al., 1989). If the chloroplasts of peridinin-containing dinoflagellates represent a reduced eukaryotic endosymbiont (Gibbs, 1990; Whatley, 1993), this might have occurred elsewhere, followed by the further translocation from endosymbiont nucleus to host (dinoflagellate) nucleus. Alternatively, *Symbiodinium* Rubisco may never have been plastome- (or plastid progenitor–) encoded; it may represent an unrelated gene transfer event. Potential sources would include endosymbiotic α-purple bacteria (Fenchel and Bernard, 1993) and the α-purple bacterial ancestor of mitochondria (Cavalier-Smith, 1992). Only one fact is obvious: in acquiring photosynthesis, eukaryotes have also acquired the entire range of diversity in prokaryotic Rubisco.

**Table 1.** Percentage Identities among Prokaryotic Form II Rubiscos and the *Symbiodinium rbcA*-Encoded Protein[a]

| | *Rrub* (%) | *Rsph* (%) | *Thio* (%) | *Hyvi* (%) |
|---|---|---|---|---|
| *SymA* | 64.4 | 64.7 | 57.9 | 53.7 |
| *Rrub* | | 77.3 | 69.6 | 63.7 |
| *Rsph* | | | 66.1 | 61.6 |
| *Thio* | | | | 58.7 |

[a] Amino acid sequences from *Symbiodinium rbcA (SymA*; Figure 10), the α-purple bacteria *Rhodospirillum rubrum (Rrub*; Narang et al., 1984), and *Rhodobacter sphaeroides (Rsph*; Wagner et al., 1988), the β-purple bacterium *Thiobacillus denitrificans (Thio*; GenBank accession number L37437), and the hydrogen bacterium *Hydrogenovibrio marinus (Hyvi*; Yaguchi et al., 1994) were aligned pairwise, as in Figure 10. The percentage of alignment positions (between the N terminus and the C terminus of the shorter sequence) that were identical in both sequences is given; apparent deletions were scored as nonidentities for the number of positions missing.

## METHODS

Unless indicated otherwise, enzymes were from New England Biolabs (Beverly, MA) and chemicals were from Sigma. Protocols and recipes for DNA and RNA analysis were as given in Ausubel et al. (1989).

### Isolation, Culture, and Identification of Zooxanthellae

Cultured clams (*Tridacna gigas*) from Pioneer Bay, Orpheus Island Research Station (Queensland, Australia) were held in outdoor recirculating tanks at James Cook University. Zooxanthellae were isolated from mantle tissue by repeated washing with filtered (0.45 or 0.22 μm) sea water, as described by Whitney and Yellowlees (1995). The absence of contamination by animal tissue or bacteria was confirmed by examining 4′,6-diamidino-2-phenylindole (DAPI)–stained subsamples (Coleman, 1980) by epifluorescence and phase contrast microscopy. Except for those presented in Figure 6, all data were obtained from freshly isolated zooxanthellae.

A clonal culture of *Symbiodinium* was established from *T. squamosa* collected at Davies Reef, Great Barrier Reef, Australia. Algae were obtained by grinding a small biopsy of mantle tissue in filtered (0.22

μm) sea water, filtering the homogenate through a nylon mesh (50 μm), and washing five times in sterile sea water. Single cells were isolated by hand picking. The clone used in this study (designated SQ3A) was grown in f/2 medium (Guillard, 1975) under fluorescent lights (80 to 100 $\mu E\ m^{-2}\ sec^{-1}$) on a 12-hr-light/12-hr-dark photoperiod at room temperature (23 to 28°C), with transfers every 3 to 5 weeks. Each subculture was treated with antibiotics (penicillin G, 100 μg/mL; streptomycin sulfate, 50 μg/mL; chloramphenicol, 10 μg/mL) for the first 3 days (two treatments of 36 hr each), as described by McLaughlin and Zahl (1959). For DNA, 1-L cultures were harvested by centrifugation after 4 to 6 weeks of growth.

The identity of zooxanthellae was established by analyzing small ribosomal subunit RNA (srRNA) genes, as described by Rowan and Knowlton (1995) and Rowan and Powers (1991a, 1991b), using universal polymerase chain reaction (PCR) primers. Diagnoses were based on the restriction fragment patterns obtained with DpnII, HaeIII, and TaqI by using previously cloned and sequenced *Symbiodinium* srRNA genes (Rowan and Powers, 1992; Rowan and Knowlton, 1995) as reference.

### Protein Purification and Sequencing

Ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) was purified from zooxanthellae as described by Whitney and Yellowlees (1995). Peptides were generated with cyanogen bromide (Promega Corporation, Madison, WI) and with endoproteinase Lys-C (sequencing grade; Promega) and isolated on membranes, using the Probe-Design peptide separation system (Promega), according to the manufacturer's protocols. Peptide sequences were determined using an Applied Biosystems protein sequencer (model no. 477A).

### Isolation of DNA from Zooxanthellae

Zooxanthellae were suspended in approximately two volumes of STE (0.15 M NaCl, 10 mM Tris-HCl, and 1 mM EDTA, pH 8.0) and frozen and ground to a fine powder under liquid nitrogen. This was mixed gently into ~15 volumes of preheated (60°C) 5 M guanidine thiocyanate, 50 mM DTT, 50 mM Tris-HCl, 20 mM EDTA, pH 7.5, and 1% (w/v) sarkosyl. After 30 min at 60°C, 0.25 g of CsCl was added per milliliter and allowed to dissolve, and the lysate was centrifuged at 5900*g* for 5 min at 20°C to pellet debris. Clarified lysate was layered onto 4-mL cushions of 5.7 M CsCl prepared in STE in SW41 centrifuge tubes (Beckman Instruments, Fullerton, CA), and the filled tubes were centrifuged in a Beckman SW41 rotor at 32,000 rpm at 15°C for 24 hr. DNA, banding above the cushion, was recovered and diluted with approximately eight volumes of a 1:1 (w/w) mixture of STE and CsCl. This was centrifuged to density equilibrium in a Beckman Ti 65 rotor at 15°C, and the purified DNA was recovered and dialyzed exhaustively against TE (10 mM Tris-HCl and 0.5 mM EDTA, pH 8.0) at 4°C. Yields averaged ~120 μg of DNA from $10^9$ cells.

For comparison with nuclear DNA (below), whole-cell DNA was prepared as follows. Frozen and powdered zooxanthellae (as given above) were mixed into ~20 volumes of prewarmed (60°C) 0.15 M NaCl, 50 mM EDTA, 10 mM Tris-HCl, pH 7.8, and 1% (w/v) SDS. After 30 min at 60°C, the lysate was placed at 45°C, proteinase K (Promega) was added to a final concentration of 0.25 mg/mL, and the lysate was incubated for 8 hr. Next, NaCl was added to a final concentration of 0.8 M, hexadecyltrimethylammonium bromide was added to 1% (w/v), and the lysate was incubated at 65°C for 1 hr with occasional mixing. After sequential extractions with equal volumes of chloroform, phenol-chloroform (1:1 [v/v]), and chloroform again (all at room temperature), nucleic acids were precipitated with ethanol. The precipitate was washed in 70% ethanol, air dried, dissolved in TE, and stored at 4°C. Yields averaged ~80 μg of DNA from $10^9$ cells. Contaminating RNA was removed by digestion with ribonuclease A (Sigma; 50 μg/mL final concentration).

### Isolation of Nuclear DNA from Zooxanthellae

Nuclei were obtained from freshly isolated zooxanthellae, using a method adapted from Rizzo and Nooden (1973). All steps were at 0 to 4°C except as indicated.

Approximately $3 \times 10^8$ washed cells were suspended in 5 mL of nuclear isolation buffer (NIB; 0.25 M sucrose, 10 mM Tris, pH 7.2, 5 mM $CaCl_2$, 5% [w/v] Dextran T-40 [Pharmacia, Uppsala, Sweden], and 2.5% [w/v] Ficoll [Type 400]). The slurry was added dropwise to liquid nitrogen in a mortar, ground to a fine powder, and then added to 20 mL ice-cold NIB plus Triton X-100 (0.12% [v/v]). The frozen paste was stirred until thawed and then sonicated at low power two or three times for 10 sec each. Cells were monitored by staining with DAPI and epifluorescence microscopy: liberated nuclei stained intensely and exhibited the typical dinokaryotic morphology (Spector, 1984). Good preparations exhibited one to two free nuclei per remaining intact cell, and further sonication decreased the yield of nuclei.

The sonicate was mixed with 6 mL of 2.2 M sucrose prepared in TCT (10 mM Tris, pH 7.2, 5 mM $CaCl_2$, and 0.1% [v/v] Triton X-100) and centrifuged at 5900*g* for 15 min. The pellet (nuclei, cells, and cell debris) was completely resuspended in 18 mL of NIB plus 0.1% [v/v] Triton X-100. One-fourth of this sample was layered onto 9 mL of 2.2 M sucrose in TCT in each of four SW41 centrifuge tubes (Beckman Instruments). The interface between the two sucrose layers was stirred, and the samples were centrifuged in an SW41 rotor at 8000 rpm for 45 min. Nuclear pellets were recovered by removing the overlaying gradient in aliquots, washing the sides of the tubes (several times) with ice-cold TCTS (TCT plus 0.15 M NaCl) after each removal. The nuclei were resuspended in 4 mL of NIB and appeared to be pure by epifluorescence (DAPI-stained) and phase contrast microscopy.

Purified nuclei were collected by centrifugation at 5900*g* for 6 min and resuspended in 5 mL of 0.1 M NaCl and 50 mM EDTA, pH 8.0. At room temperature, SDS was added to 1% (w/v), proteinase K was added to 0.25 mg/mL, and the lysate was incubated at 50°C for 3 to 4 hr. After extracting three times with an equal volume of phenol-chloroform (1:1 [v/v]), nucleic acids were recovered by ethanol precipitation, washed in 70% ethanol, resuspended in 2 to 3 mL of TE, and stored at 4°C.

### RNA Isolation

Several methods were used to isolate zooxanthella RNA, with varying degrees of success; the following protocol consistently yielded relatively undegraded RNA.

Working quickly at room temperature, zooxanthellae were isolated from a single clam with one wash in filtered sea water. The cells were frozen and ground (as above) and blended into ~30 volumes of a 2:1 (v/v) emulsified mixture of LTE buffer (0.4 M LiCl, 50 mM Tris-HCl, 25 mM EDTA, pH 7.5) and phenol (preheated to 50°C), using a tissue homogenizer (Tekmar, Cincinnati, OH) until nonviscous. At room temperature, one-third volume of chloroform was added, and homogenization was repeated. After centrifugation at 5900*g* for 15 min at room temperature, the aqueous phase was recovered, and the organic

phase/interface was extracted with one-half volume of LTE, which, after recentrifugation, was combined with the first aqueous phase. This was extracted twice more with a 1:1 (v/v) mixture of phenol-chloroform, and nucleic acids were precipitated by adding 2.5 volumes of ethanol.

The precipitate was dissolved in ~25 times the original cell volume of HES buffer (20 mM Hepes, 5 mM EDTA, pH 7.2, and 1% [w/v] sarkosyl; pretreated with diethyl pyrocarbonate). This was layered onto 5-mL cushions of 5.7 M CsCl in HE buffer (HES minus sarkosyl) in SW41 centrifuge tubes, and the filled tubes were centrifuged in a Beckman SW41 rotor at 32,000 rpm at 15°C for 24 hr. RNA pellets were collected and resuspended in HES, and this solution was extracted twice with an equal volume of phenol-chloroform (1:1 [v/v]). After that, the RNA was stored as an ethanol precipitate at −20°C. Yields averaged ~0.4 mg of RNA from $10^9$ cells.

### DNA Gel Blot and RNA Gel Blot Hybridizations

DNAs were electrophoresed through 0.7% agarose (Seakem LE; FMC BioProducts, Rockland, ME) gels in TAE buffer, and RNAs were electrophoresed through 1% agarose formaldehyde gels, as described by Ausubel et al. (1989). Nucleic acids were transferred to nylon membranes (Hybond N; Amersham, Arlington Heights, IL) by capillary action. DNA blots were prehybridized and hybridized in aqueous solutions at 65°C; RNA blots were prehybridized and hybridized in 50% formamide solutions at 40°C. Blots were washed in dilutions (see legends to Figures 4 to 7) of SSC (1 × SSC is 0.15 M NaCl, 0.015 mM sodium citrate, pH 7.5) or SSPE (1 × SSPE is 0.15 M NaCl, 10 mM sodium phosphate, 1 mM EDTA, pH 7.5) at the temperature of hybridization. RNA length markers (Promega) were visualized by reprobing the blots with $^{32}$P-labeled cDNA that was obtained by random-primed reverse transcription of marker RNA.

Probes S0.45, SP, H1.9, and D72 (Figure 2) were isolated by molecular cloning (confirmed by sequencing) and then obtained as restriction fragments or as PCR-amplified (using universal forward and reverse sequencing primers) inserts. Probe SP was isolated as a PvuII fragment that included 396 bp of vector (pGEM-3Zf [+]; Promega); the corresponding PvuII fragment from nonrecombinant vector DNA did not hybridize to zooxanthella RNA in parallel controls (data not shown). Probe S3.2 (Figure 9) was a SacII restriction fragment from a larger genomic clone. For quantitative DNA gel blot hybridizations (Figure 7), a nuclear-encoded srRNA gene from *Symbiodinium* (probe 178-6) was used to measure *Symbiodinium* DNA independently on the membrane; it was obtained by PCR amplification from an M13 clone (Rowan and Knowlton, 1995). Probe DNAs were purified by agarose gel electrophoresis and GeneClean II (Bio 101, Vista, CA), labeled with $^{32}$P-dATP (Amersham) to > 5 × $10^8$ cpm/μg, and used at a concentration of $10^6$ cpm per mL of hybridization solution.

### Isolation of a Partial Rubisco Gene by Using the PCR

Sequences of Rubisco peptides (given in Figure 3) were used to design oligonucleotide primers for PCR-amplifying part of the corresponding gene. Targeted amino acid sequences were PDNEEMKI and YMLQDDEA; the corresponding primers were 5′-CC(N)GA(CT)AA(CT)-GA(AG)GA(AG)ATGAA(AG)AT-3′ and 5′-GC(CT)TC(AG)TC(AG)TC(CT)-TG(N)A(AG)CAT(AG)TA-3′, respectively, where N represents an equimolar mixture of G, A, T, and C. PCR reactions are as follows: 2 μM in each primer; 200 μM each in dGTP, dATP, dTTP, and dCTP (Boehringer Mannheim); 50 mM KCl; 1.5 mM $MgCl_2$; 10 mM Tris-HCl, pH 8.3; 0.01% (w/v) gelatin; 0.05% [v/v] Tween 20; and contained ~15 ng of zooxanthella DNA and 2.5 units of AmpliTaq DNA polymerase (Roche Molecular Systems, Branchburg, NJ) in a volume of 40 or 80 μL. Thirty cycles of 94°C for 1 min, 50°C for 1 min, 72°C for 1 min (8 min on the last cycle) yielded a product that was purified using GeneClean II and cloned into the vector pGEM-3zf(+) (Promega).

### Isolation of Genomic Clones

Approximately 4 μg of zooxanthella DNA was digested with various restriction enzymes and analyzed by DNA gel blotting to identify fragments for cloning into λ vectors. Then, ~250 μg of DNA were digested with BamHI and size fractionated in sucrose gradients. Appropriate fractions were identified by agarose gel electrophoresis and blot hybridization, ligated to BamHI-digested EMBL3 vector DNA (Promega), and packaged in vitro (Amersham); the resulting subgenomic library was plated on bacterial strain KW251 (Promega). Plaques (2 × $10^5$) were screened by filter hybridization to the cloned PCR product.

### Clone Characterization and DNA Sequencing

DNA was isolated from recombinant bacteriophage grown on lawns of KW251 and mapped with several restriction enzymes, and fragments that hybridized to the PCR clone were identified by DNA gel blotting. These were cloned into either pGEM3zf(+) (Promega) or pBluescriptII SK− or KS+ (Stratagene) as required and then deleted progressively from one end, using exonuclease III and mung bean nuclease. DNAs were sequenced on both strands primarily from these sets of deletion subclones, using Sequenase kits (United States Biochemical); gaps were filled using additional primers and additional subclones. Sequence data have been submitted to GenBank (accession numbers are U43532 for rbcA and U43533 for rbcG).

## ACKNOWLEDGMENTS

We thank Dennis Shaw (John Curtin School of Medical Research, Australian National University) for all peptide sequencing; T.J. Andrews, Catherine Lovelock, and Dennis Shaw for their interest and comments; Chao-lun Chen, David Miller, and Dimitri Odorico for sharing time, enthusiasm, and reagents on many occasions; Paul Levine and G. Jason Smith for hospitality in Pacific Grove, CA; and Andrew Baker for hospitality in Miami, FL. This work was supported by fellowships from the Australian Institute of Marine Science and from the Smithsonian Institution (to R.R.), by an Australian Postgraduate Research Award (to S.M.W.), by a Nevitt Bequest Fellowship (to A.F.), and by grants from James Cook University and the Australian Research Council (to D.Y.).

Received July 18, 1995; accepted January 4, 1996.

## REFERENCES

**Andersen, K., and Caton, J.** (1987). Sequence analysis of the *Alcaligenes eutrophus* chromosomally encoded ribulose bisphosphate carboxylase large and small subunit genes and their gene products. J. Bacteriol. **169,** 4547–4558.

**Andrews, T.J., Lorimer, G.H., and Tolbert, N.E.** (1973). Ribulose diphosphate oxygenase. I. Synthesis of phosphoglycolate by fraction-1 protein of leaves. Biochemistry **12,** 11–18.

**Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., and Struhl, K.,** eds (1989). Current Protocols in Molecular Biology. (New York: John Wiley and Sons).

**Beardall, J.** (1989). Photosynthesis and photorespiration in marine phytoplankton. Aquat. Bot. **34,** 105–130.

**Blank, R.J.** (1987). Cell architecture of the dinoflagellate *Symbiodinium* sp. inhabiting the Hawaiian stony coral *Montipora verrucosa*. Mar. Biol. **94,** 143–155.

**Blank, R.J., and Trench, R.K.** (1985). Speciation and symbiotic dinoflagellates. Science **229,** 656–658.

**Blank, R.J., Huss, V.A.R., and Kersten, W.** (1988). Base composition of DNA from symbiotic dinoflagellates: A tool for phylogenetic classification. Arch. Microbiol. **149,** 515–520.

**Brändén, C.-I., Lindqvist, Y., and Schneider, G.** (1991). Protein engineering of rubisco. Acta Crystallogr. B **47,** 824–835.

**Cavalier-Smith, T.** (1992). The number of symbiotic origins of organelles. BioSystems **28,** 91–106.

**Chan, R.L., Keller, M., Canaday, J., Weil, J.-H., and Imbault, P.** (1990). Eight small subunits of *Euglena* ribulose 1-5 bisphosphate carboxylase/oxygenase are translated from a large mRNA as a polyprotein. EMBO J. **9,** 333–338.

**Chang, S.S., Prezlin, B.B., and Trench, R.K.** (1983). Mechanisms of photoadaption in three strains of symbiotic dinoflagellate *Symbiodinium microadriaticum*. Mar. Biol. **76,** 219–229.

**Coleman, A.W.** (1980). Enhanced detection of bacteria in natural environments by fluorochrome staining of DNA. Limnol. Oceanogr. **25,** 948–951.

**Crossland, C.J., and Barnes, D.J.** (1977). Gas-exchange studies with staghorn coral *Acropora acuminata* and its zooxanthellae. Mar. Biol. **40,** 185–194.

**Dean, C., Pichersky, E., and Dunsmuir, P.** (1989). Structure, evolution, and regulation of *RbcS* genes in higher plants. Annu. Rev. Plant Physiol. Plant Mol. Biol. **40,** 415–439.

**Dykens, J.A., and Shick, J.M.** (1982). Oxygen production by endosymbiotic algae controls superoxide dismutase activity in their animal host. Nature **297,** 579–580.

**Fenchel, T., and Bernard, C.** (1993). A purple protist. Nature **362,** 300.

**Gibbs, S.P.** (1990). The evolution of algal chloroplasts. In Experimental Phycology, Vol. 1, Cell Walls and Surfaces, Reproduction, Photosynthesis, W. Wiessner, D.G. Robinson, and R.C. Starr, eds (Berlin: Springer-Verlag), pp. 145–157.

**Guillard, R.R.L.** (1975). Culture of phytoplankton for feeding marine invertebrates. In Culture of Marine Invertebrate Animals, W.L. Smith and M.H. Chanley, eds (New York: Plenum), pp. 29–59.

**Hartman, F.C., and Harpel, M.R.** (1994). Structure, function, regulation, and assembly of D-ribulose-1,5-bisphosphate carboxylase/oxygenase. Annu. Rev. Biochem. **63,** 197–234.

**Hatch, M.D.** (1971). Mechanism and function of the $C_4$-pathway of photosynthesis. In Photosynthesis and Photorespiration, M.D. Hatch, C.B. Osmond, and R.O. Slatyer, eds (New York: John Wiley and Sons), pp. 139–152.

**Houlné, G., and Schantz, R.** (1988). Characterization of cDNA sequences for LHCI apoproteins in *Euglena gracilis*: The mRNA encodes a large precursor containing several consecutive divergent polypeptides. Mol. Gen. Genet. **213,** 479–486.

**Houlné, G., and Schantz, R.** (1993). Expression of polyproteins in *Euglena*. Crit. Rev. Plant Sci. **12,** 1–17.

**Jordan, D.B., and Ogren, W.L.** (1981). Species variation in the specificity of ribulose biphosphate carboxylase/oxygenase. Nature **291,** 513–515.

**Keegstra, K., Olsen, L.J., and Theg, S.M.** (1989). Chloroplastic precursors and their transport across the envelope membranes. Annu. Rev. Plant Physiol. Plant Mol. Biol. **40,** 471–501.

**Kowallik, K.V.** (1992). Origin and evolution of plastids from chlorophyll-*a*+*c*-containing algae: Suggested ancestral relationships to red and green algal plastids. In Origins of Plastids, R.A. Lewin, ed (New York: Chapman and Hall), pp. 223–263.

**Larimer, F.W., Lee, E.H., Mural, R.H., Soper, T.S., and Hartman, F.C.** (1987). Inter-subunit location of the active site of ribulosebisphosphate carboxylase/oxygenase as determined by in vivo hybridization of site-directed mutants. J. Biol. Chem. **262,** 15327–15329.

**Lorimer, G.H., Badger, M.R, and Andrews, T.J.** (1976). The activation of ribulose-1,5-bisphosphate carboxylase by carbon dioxide and magnesium ions. Equilibria, kinetics, a suggested mechanism, and physiological implications. Biochemistry **15,** 529–536.

**Martin, W., Somerville, C.C., and Loiseaux-de Goer, L.** (1992). Molecular phylogenies of plastid origins and algal evolution. J. Mol. Evol. **35,** 385–404.

**McLaughlin, J.J.A., and Zahl, P.A.** (1959). Axenic culture of zooxanthellae from various invertebrate hosts. Ann. NY Acad. Sci. **77,** 55–72.

**Morden, C.W., Delwiche, C.F., Kushel, M., and Palmer, J.D.** (1992). Gene phylogenies and the endosymbiotic origin of plastids. BioSystems **28,** 75–90.

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

**Muchhal, U.S., and Schwartzbach, S.D.** (1992). Characterization of a *Euglena* gene encoding a polyprotein precursor to the light-harvesting chlorophyll a/b protein of photosystem II. Plant Mol. Biol. **18,** 287–299.

**Muscatine, L.** (1980). Productivity of zooxanthellae. In Primary Productivity in the Sea, P.G. Falkowski, ed (New York: Plenum Press), pp. 381–402.

**Narang, F., McIntosh, L., and Somerville, C.** (1984). Nucleotide sequence of the ribulosebisphosphate carboxylase gene from *Rhodospirillum rubrum*. Mol. Gen. Genet. **193,** 220–224.

**Newman, S.M., and Cattolico, R.A.** (1990). Ribulose bisphosphate carboxylase in algae: Synthesis, enzymology and evolution. Photosynth. Res. **26,** 69–85.

**Ohta, T.** (1980). Evolution and Variation of Multigene Families. (Berlin: Springer-Verlag).

**Palmer, J.D.** (1985). Comparative organizations of chloroplast genomes. Annu. Rev. Genet. **19,** 325–354.

**Palmer, J.D.** (1993). A genetic rainbow of plastids. Nature **364,** 762–763.

**Pierce, J., Carlson, T.J., and Williams, J.G.K.** (1989). A cyanobacterial mutant requiring the expression of ribulose bisphosphate carboxylase from a photosynthetic anaerobe. Proc. Natl. Acad. Sci. USA **86,** 5753–5757.

**Read, B.A., and Tabita, F.R.** (1994). High substrate specificity factor ribulose bisphosphate carboxylase/oxygenase from eukaryotic marine algae and properties of recombinant cyanobacterial rubisco containing "algal" residue modifications. Arch. Biochem. Biophys. **312,** 210–218.

**Reith, M.** (1995). Molecular biology of rhodophyte and chromophyte plastids. Annu. Rev. Plant Physiol. Plant Mol. Biol. **46,** 549–575.

**Reith, M., and Munholland, J.** (1993). A high-resolution gene map of the chloroplast genome of the red alga *Porphyra purpurea*. Plant Cell **5,** 465–475.

**Rizzo, P.J., and Nooden, L.D.** (1973). Isolation and chemical composition of dinoflagellate nuclei. J. Protozool. **20,** 666–672.

**Rowan, R.** (1991). Molecular systematics of symbiotic algae. J. Phycol. **27,** 661–666.

**Rowan, R., and Knowlton, N.** (1995). Intraspecific diversity and ecological zonation in coral–algal symbiosis. Proc. Natl. Acad. Sci. USA **92,** 2850–2853.

**Rowan, R., and Powers, D.A.** (1991a). The molecular genetic identification of symbiotic dinoflagellates (zooxanthellae). Mar. Ecol. Prog. Ser. **71,** 65–73.

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

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

**Schiff, J.A., Schwartzbach, S.D., Osafune, T., and Hase, E.** (1991). Photocontrol and processing of LHCP II apoprotein in *Euglena*: Possible role of Golgi and other cytoplasmic sites. J. Photochem. Photobiol. B **11,** 219–236.

**Shashidhara, L.S., Lim, S.H., Shackleton, J.B., Robinson, C., and Smith, A.G.** (1992). Protein targeting across the three membranes of the *Euglena* chloroplast envelope. J. Biol. Chem. **267,** 12885–12891.

**Silva, E. da S.** (1978). Endonuclear bacteria in two species of dinoflagellates. Protistologica **14,** 113–119.

**Spector, D.L.** (1984). Dinoflagellate nuclei. In Dinoflagellates, D.L. Spector, ed (New York: Academic Press), pp. 107–147.

**Steele, R.D.** (1977). The significance of zooxanthella-containing pellets extruded by sea anemones. Bull. Mar. Sci. **27,** 591–594.

**Stoner, M.T., and Shively, J.M.** (1993). Cloning and expression of the D-ribulose-1,5-bisphosphate carboxylase/oxygenase form II gene from *Thiobacillus intermedius* in *Escherichia coli*. FEMS Microbiol. Lett. **107,** 287–292.

**Streamer, M., McNeil, Y.R., and Yellowlees, D.** (1993). Photosynthetic carbon dioxide fixation in zooxanthellae. Mar. Biol. **115,** 195–198.

**Tabita, F.R.** (1988). Molecular and cellular regulation of autotrophic carbon dioxide fixation in microorganisms. Microbiol. Rev. **52,** 155–189.

**Taylor, D.L.** (1969). Identity of zooxanthellae isolated from Pacific Tridacnidae. J. Phycol. **5,** 336–340.

**Taylor, F.J.R.** (1983). Possible free-living *Symbiodinium microadriaticum* (*Dinophyceae*) *in tide pools in southern Thailand*. In Endocytobiology, Vol. 2, H.E.A. Schenk and W. Schwemmler, eds (Berlin: de Gruyter), pp. 1009–1014.

**Tessier, L.H., Paulus, F., Keller, M., Vial, C., and Imbault, P.** (1995). Structure and expression of *Euglena gracilis* nuclear *rbcS* genes encoding the small subunits of the ribulose 1,5-bisphosphate carboxylase/oxygenase: A novel splicing process for unusual intervening sequences? J. Mol. Biol. **245,** 22–33.

**Trench, R.K.** (1987). Dinoflagellates in non-parasitic symbioses. In Botanical Monographs, Vol. 21, The Biology of Dinoflagellates, F.J.R. Taylor, ed (Oxford: Blackwell), pp. 530–570.

**Trench, R.K.** (1992). Microalgal–invertebrate symbiosis, current trends. In Encyclopedia of Microbiology, Vol. 3, J. Lederberg, ed (New York: Academic Press), pp. 129–142.

**Viale, A.M., Kobayashi, H., and Akazawa, T.** (1990). Distinct properties of *Escherichia coli* products of plant-type ribulose-1,5-bisphosphate carboxylase/oxygenase directed by two sets of genes from the photosynthetic bacterium *Chromatium vinosum*. J. Biol. Chem. **265,** 18386–18392.

**Wagner, S.J., Stevens, S.E., Jr., Nixon, B.T., Lambert, D.H., Quivey, R.G., Jr., and Tabita, F.R.** (1988). Nucleotide and deduced amino acid sequence of the *Rhodobacter sphaeroides* gene encoding form II ribulose-1,5-bisphosphate carboxylase/oxygenase and comparison with other deduced form I and form II sequences. FEMS Microbiol. Lett. **55,** 217–222.

**Whatley, J.M.** (1993). The endosymbiotic origin of chloroplasts. Int. Rev. Cytol. **44,** 259–299.

**Whitney, S.M., and Yellowlees, D.** (1995). Preliminary investigations *into the structure and activity of ribulose bisphosphate carboxylase* from two photosynthetic dinoflagellates. J. Phycol. **31,** 138–146.

**Whitney, S.M., Shaw, D.C., and Yellowlees, D.** (1995). Evidence that some dinoflagellates contain a ribulose-1,5-bisphosphate carboxylase/oxygenase related to that of the α-proteobacteria. Proc. R. Soc. Lond. B **259,** 271–275.

**Woodrow, I.E., and Berry, J.A.** (1988). Enzymatic regulation of photosynthetic $CO_2$ fixation in $C_3$ plants. Annu. Rev. Plant Physiol. Plant Mol. Biol. **39,** 533–594.

**Yaguchi, T., Seong, Y.-C., Igarashi, Y., and Kodama, T.** (1994). Cloning and sequence of the $L_2$ form of rubisco from a marine obligately autotrophic hydrogen-oxidizing bacterium *Hydrogenovibrio marinus* strain MH-110. Biosci. Biotech. Biochem. **58,** 1733–1737.

**Zhang, K.Y.J., Cascio, D., and Eisenberg, D.** (1994). Crystal structure of the unactivated ribulose 1,5-bisphosphate carboxylase/oxygenase complexed with a transition state analog, 2-carboxy-D-arabinitol 1,5-bisphosphate. Protein Sci. **3,** 64–69.