The Dissertation Committee for Jay Randall Reichman Certifies that this is the approved version of the following dissertation:

Characterization and Evolution of Peridinin-Chlorophyll *a* Binding Protein Gene Families in Symbiotic Dinoflagellates

Committee:

David M. Hillis, Supervisor

Peter D. Vize, Co-Supervisor

Robert K. Jansen

Bassett Maguire Jr.

Thomas P. Wilcox

Characterization and Evolution of Peridinin-Chlorophyll *a* Binding Protein Gene Families in Symbiotic Dinoflagellates

by

Jay Randall Reichman, B.S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

December, 2002

Dedication

To all my family members present and past who encouraged and supported me emotionally, helped me financially and kept doing so through all seasons long after they ceased to understand what I was working on. And especially to the three women who have had the largest positive impacts on my life, my mother Ann, my wife Melada and my daughter Lauren.

Acknowledgements

Substantial help with this work in the form of molecular training and critical feedback came from Dr. Peter Vize of the University of Calgary and Dr. Thomas Wilcox of the University of Texas at Austin. Thanks to Dr. Robert Trench formerly of UC Santa Barbara for donation of *Symbiodinium* cultures and exchange of ideas. Dr. Eric Lader at Ambion, Inc. provided access to an ABI 7700 and provided valuable assistance with the quantitative real-time PCR experiments. Thanks to Drs. Bassett Maguire and Judy Lang for supervision in the early days when I was trying to figure out what I wanted to work on, and for getting me down on the reefs in the Bahamas where I could really clear my head and think. Paul Thompson and Walter Hokanson from the UT Austin found and/or derived the appropriate statistical equations to use for error propagation. Derrick Zwickl from UT Austin was instrumental in setting up and evaluating phylogenetic analyses. Thanks to lab mates Dr. Thomas Carroll for helping me get that first clone and Dr. Derek Hagman for getting me involved in many underwater science projects at the Flower Garden Banks National Marine Sanctuary and for playing a mean game of rum pong. Undergraduate lab assistants Maria Polycarpo and Tasmin Smith gave considerable help in carrying

the large number of PCR and sequencing reactions were conducted. Thanks to my fellow scuba instructors at the University of Texas Underwater Science and Scuba Diving Program who kept me laughing and took much of the load off of me while I was trying to finish this project. And special thanks Erik and Madonna Reichman for replacing my crashed computer at their own expense and who often baby sat in my hour of need. Funding was made available through the Caribbean Marine Research Center Perry Foundation, the U.S National Oceanic and Atmospheric Administration, Oryx, Inc. and The University of Texas at Austin.

Characterization and Evolution of Peridinin-Chlorophyll *a* Binding Protein Gene Families in Symbiotic Dinoflagellates

Publication No._____

Jay Randall Reichman, Ph.D.

The University of Texas at Austin, 2002

Supervisors: David M. Hillis and Peter D. Vize

Abstract: This dissertation consists of three integrated chapters. Chapter One presents the first genomic characterization of the PCP gene family from a symbiotic dinoflagellate, *Symbiodinium sp.* from *Hippopus hippopus* (*Symbiodinium* 203). *Symbiodinium* 203 has long PCP genes organized like those of *Amphidinium carterae* and *Lingulodinium polyedra*, but with a putative promoter that is different from *L. polyedra*. There are at least 14 distinct coding regions out of 36 ± 12 PCP genes in this family. Diversity of *Symbiodinium* 203's PCP gene family appears to be consequence of low levels of concerted evolution and acts as a primary source of variability in PCP isoforms. Predicted amino acid substitutions in *Symbiodinium* 203's PCP apoproteins result in shifts of isoelectric points, and protein modeling suggests that polymorphic sites may influence light

harvesting of holoproteins. In Chapter Two, the first PCP coding sequences from *S. pilosum*, *Symbiodinium sp*. from *Dichocoenia stokesii*, *S. pulchrorum* and *S. kawagutii* were presented. Diverse PCP gene families occur in all major clades of *Symbiodinium* and in both size classes of the gene. As with *Symbiodinium* 203 in Chapter One, these PCP gene families do not appear to have been homogenized through mechanisms leading to concerted evolution. The predicted PCP apoproteins from *S. pilosum* and *S. kawagutii* have calculated isoelectric focusing points that generally match values previously measured for these species, which supports the hypothesis that genetic polymorphism is the primary source generating differences in PCP isoforms. Protein modeling produced a putative tertiary structure for *S. pilosum* apoproteins and was used to identify polymorphic sites in *S. pilosum* and *S. kawagutii* PCPs that could affect spectral tuning of peridinins. And Chapter Three contains the first phylogenetic analyses of the evolution of dinoflagellate PCP gene families. The objective of this section is to estimate the selective pressure at the codon level within PCP genes. PCP polymorphism is ancient, however, the polymorphism is not maintained by positive selection. Codon sites within PCP genes are evolving under purifying selection and are subjected to net reduced levels of concerted evolution. Isoform diversity is probably selected for within a functional range.

Table of Contents

List of Tables

List of Figures

CHAPTER ONE

Characterization and organization of the peridinin-chlorophyll *a* binding protein (PCP) gene family in *Symbiodinium sp.* from *Hippopus hippopus*: low levels of concerted evolution among PCP genes contribute to PCP isoform diversity

ABSTRACT

Photosynthetic dinoflagellates have evolved unique light harvesting complexes that combine the accessory pigment peridinin and chlorophyll *a* within the hydrophobic center of water-soluble peridinin-chlorophyll *a* binding proteins (PCPs). Most species of dinoflagellates express either 14-17 & 32-35 kDa mature PCP apoproteins and do so in stable combinations of isoforms that differ in isoelectric point (pI). The source (post-translational modification, protein degradation or genetic) and functional significance of PCP isoform variation has remained unclear. PCPs are coded for by multi-gene families. However, previous reports on free-living dinoflagellates conflict over the diversity of PCP genes within gene arrays. I present the first genomic characterization of the PCP gene family from a symbiotic dinoflagellate. *Symbiodinium sp.* from the Pacific bivalve *Hippopus hippopus* (RK Trench culture collection # 203) contains genes for 33 kDa PCP apoproteins that are organized in tandem arrays like those of free-living dinoflagellates *Lingulodinium (Gonyaulax) polyedra* and *Heterocapsa pygmaea.* The *Symbiodinium* 203 PCP cassette consists of 1098 bp coding regions separated

by spacers of variable length that are approximately 900 bp. The spacers contain a conserved upstream sequence similar to the promoter in *L. polyedra.* There is surprising diversity between cloned genomic coding regions. Ka/Ks divergence analyses between paired coding regions generally had values < 1. These differences are not accounted for by recombinations or errors introduced during PCR. Reverse Transcription PCR cDNA clones also show the same pattern of diversity. Quantitative Real-Time PCR and flow cytometry indicate that the *Symbiodinium* 203 genome has 36 ± 12 PCP genes per genome. Thus the PCP gene family of *Symbiodinium* 203 appears to be evolving under little or no of concerted evolution. The predicted proteins have pIs that are within the range of those published for other species of *Symbiodinium*. Post-translational modifications are not necessary to explain the multiple PCP isoforms. Amino acid substitutions were mapped onto the *Amphidinium carterae* PCP crystal structure and identified several polymorphic sites that may influence spectral absorption tuning of chromophores.

INTRODUCTION

Background on Dinoflagellates and PCPs

Dinoflagellates are an ancient and diversified group of organisms that diverged from other eukaryotic lineages at or before the Cambrian (Moldowan and Talyzina, 1998) and underwent extensive radiation to become dominant members of phytoplankton communities by the Jurassic (Loeblich, 1976, 1984). There are roughly 130 extant genera in the division Pyrrophyta with more than

1800 marine and 200 freshwater species of described dinoflagellates (Spector, 1984; Falkowski and Raven, 1997). Among these species, there has been considerable niche diversification. Dinoflagellates have become adapted to several life styles, and there are species that fall into one or more of the following categories: bioluminescent, toxic, free-living, parasitic, symbiotic, non-photosynthetic and photosynthetic. There are at least seven dinoflagellate genera that form endosymbioses with a wide variety of marine invertebrates (Banaszak et al. 1993; Trench 1993). The best known of these mutualism occur between photosynthetic dinoflagellates in the genus *Symbiodinium* and corals, anemones, jellyfish or bivalves. Biogeographic and phylogentic studies have identified at least 3 major clades (A, B and C; Figure 1.1) of *Symbiodinium* based on RFLPs and sequences of ss and ls RNA genes (Rowan and Powers, 1991a & b; 1992, Rowan and Knowlton, 1995; Rowan et al. 1997; Wilcox, 1998). An additional *Symbiodinium* clade E has been justified by comparison of ITS sequences (LaJeunesse and Trench 2000; LaJeunesse, 2001).

Photosynthetic dinoflagellates have evolved unique light-harvesting pigment-proteins that bind the carotenoid peridinin in close proximity to chlorophyll *a,* most often in a 4:1(8:2) molar ratio. These complexes are known as PCPs, and they are unique in terms of their combination of pigments, water solubility, nucleotide sequence and occurrence solely in dinoflagellates. PCPs have been isolated and analyzed for several dinoflagellate species. Examples include PCPs from free-living *A. carterae* (Haxo et al. 1976; Siegelman et al. 1977; Sharples et al. 1996), *Alexandrium cohorticula* (Ogata et al. 1994)

Glenodinium (Heterocapsa) sp. (Roman et al., 1988; Prezelin and Haxo 1976) *L. polyedra* (Prezelin and Haxo 1976), and endosymbionts *S. microadriaticum* (Chang and Trench 1982; 1984; Trench and Blank, 1987; Iglesias-Prieto et al. 1991), *S. goreauii* (Trench and Blank, 1987) *S. pilosum* and *S. kawagutii* (Trench and Blank, 1987; Iglesias-Prieto et al. 1991).

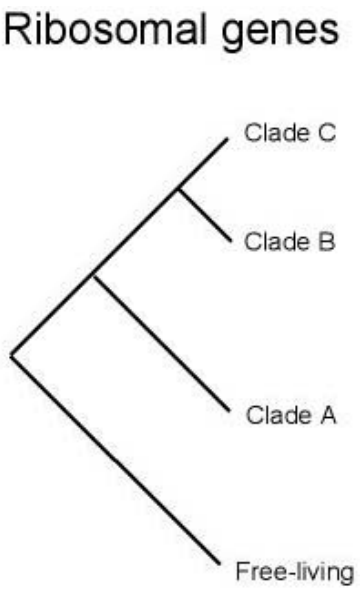


Figure 1.1 General phylogenetic relationship between free-living dinoflagellates and clades of *Symbiodinium* based on large subunit ribosomal sequence (redrawn from Wilcox, 1998).

PCPs also do not show significant structural homology to other chromophore binding proteins (Triplett et al. 1993; Norris and Miller, 1994) including phycobiliproteins or membrane-bound light harvesting complexes. It

should be noted that discovery of dinoflagellate membrane-bound antennae which also utilize peridinin and chlorophyll *a* has prompted the development of more accurate but somewhat inconsistent naming of these complexes. Extrinsic water-soluble and intrinsic membrane-bound PCPs have been respectively referred to as PCP1 and PCP2 (Grossman et al. 1995), sPCP and acpPC (Iglesias-Preito et al, 1991; Hiller et al. 1993; Iglesias-Prieto et al. 1993) and also sPCP and iPCP (Durnford et al. 1999). For simplicity within this work, PCP is used to indicate the water-soluble form.

The properties of PCPs and their chromophores have been extensively characterized. As mentioned above, PCPs are water-soluble light harvesting complexes (Haxo et al. 1976; Prezelin and Haxo, 1976). PCP preproteins contain transit peptides that presumably help translocate the mature polypeptides to the aqueous thylakoid lumen of dinoflagellate chloroplasts. One way that dinoflagellates photo-acclimate is by changing cellular concentration of PCPs (Prezelin, 1976; Roman et al., 1988; Iglesias-Prieto and Trench, 1997), and this occurs through light regulated transcription (ten Lohuis and Miller; 1998). Under low light conditions, PCPs can constitute a major percentage of the protein in a dinoflagellate cell (Jovine et al. 1993). The peridinin within PCPs harvests blue green 435-550 nm light where chlorophyll *a* does not (Seigelman et al. 1977; Larkum, 1996; Moffat, 1996). Peridinin also aids in photo-protection by dissipating excess light energy as heat and by limiting the production of singlet oxygen free-radicals that can disrupt photosynthesis (Larkum, 1996; Lakum and Howe, 1997; Pinto et al. 2000). There is near 100% efficiency of energy transfer

between peridinin and chlorophyll *a* in PCPs (Song, 1976; Damjanovic et al., 2000). Tuning of this transfer is achieved by the specific physical arrangement of the chromophore complex and the surrounding amino acids. A high resolution crystal structure of the PCP trimer from *A. carterae* (Figure 1.2) has indicated that peridinins are held as close as 3.3-3.8 Å of the tetrapyrole rings of the chlorophylls (Figure 1.3). Furthermore, changes in the polarity of the PCP protein environment neighboring the polyene chains and furanic rings of peridinins can modify the spectroscopic properties of these accessory pigments (Figure 1.4).

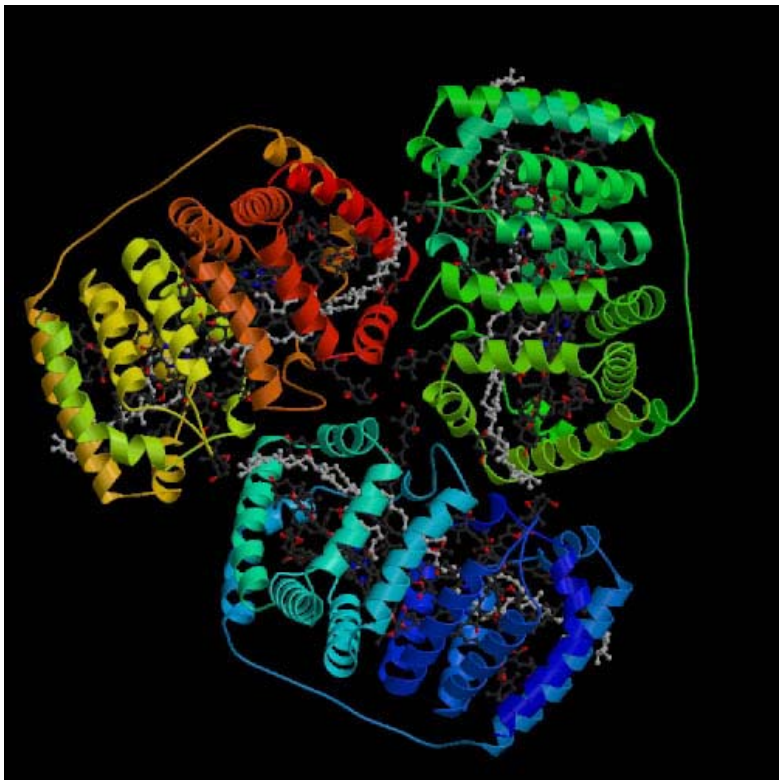


Figure 1.2 *A. carterae* trimer of 35 kDa PCPs. (NCBI PDB ID 1PPR; Hofmann et al. 1996).

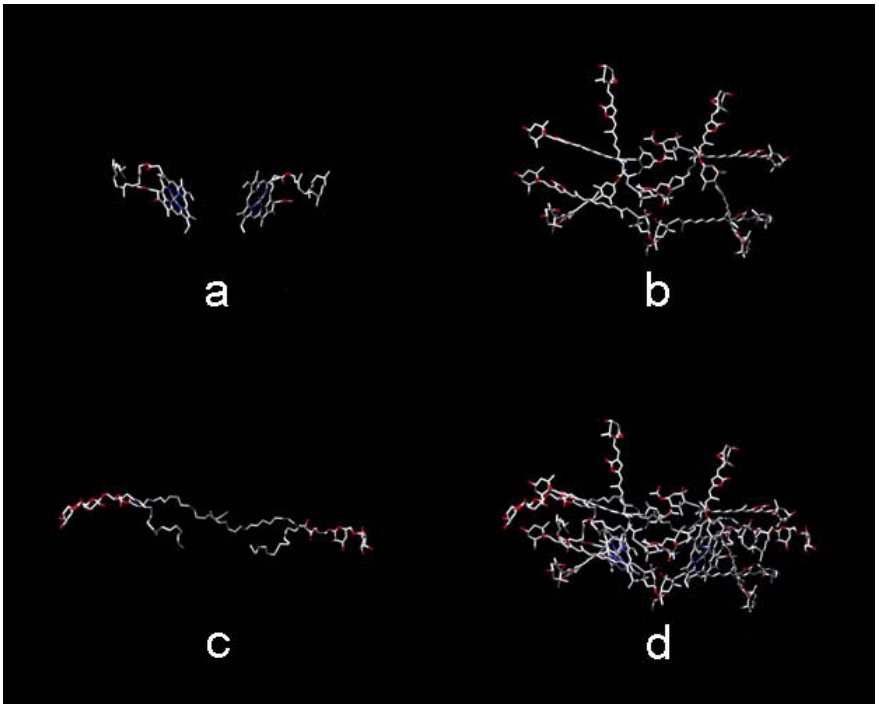


Figure 1.3 Chromophores and lipids within 35 kDa PCPs; (**a)** 2 chlorophyll-*a*. (**b)** 8 peridinin; (**c)** 2 digactosyl diacyl glycerol; (**d)** combined chromophores complex.

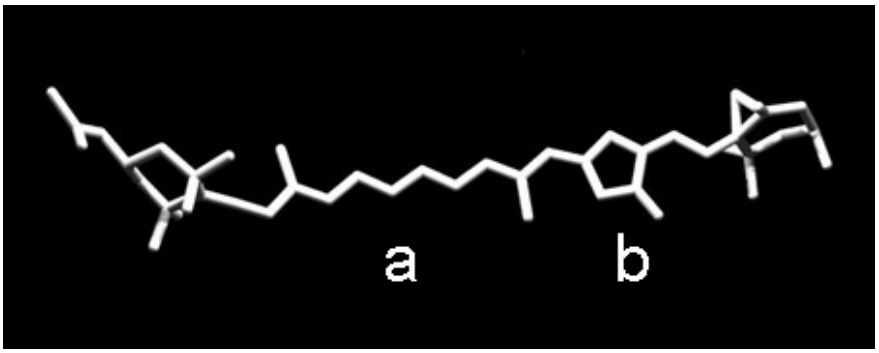


Figure 1.4 Regions of interaction between peridinin and amino acids with polar side chains within PCPs; (**a)** polyene chain and (**b)** furanic ring of peridinin.

Interactions between chromophores and surrounding amino acids are also stabilized by the lipid DGDG which maintains hydrophobicity of the interior of the PCP monomers (Hofmann et al. 1996). Once the light energy is transferred from peridinin and chlorophyll-*a*, it is then passed to membrane-bound light harvesting complexes and then to photosystem II (Damjanovic et al. 2000).

There are two general size classes of PCP apoproteins, 14-17 kDa monomers or homodimers and 31-35 kDa monomers. These are commonly called 15 kDa (short) and 35 kDa (long) PCPs. The axis of symmetry in the amino acid sequence of long PCPs strongly suggests that original gene for long PCPs arose from a duplication and fusion event between genes for short PCPs (Le et al. 1997). This is also supported by the similarity of short PCP gene nucleotide sequence to regions of long PCP genes (Hiller et al. 2001). It is unclear if long PCP genes have evolved more than once.

Many dinoflagellate species express one size class of PCP or the other (Prezelin and Haxo, 1976), however, some express both (Govind et al. 1990). There is evidence that the 35 kDa PCPs of certain dinoflagellates do not contain a single long polypeptide, but rather are constructed with two dimeric short polypeptides (Prezelin and Haxo, 1976). *S. microadriaticum* is an interesting example of a dinoflagellate that has both sizes of PCPs, and its 35 kDa PCP is dimeric (Chang and Trench, 1984). It is possible that the same short polypeptides that form the 15 kDa PCPs in *S. microadriaticum* can also be used, perhaps after post-translational modification, to build the 35 kDa version (personal communication, RK Trench, UC Santa Barbara). It is not yet known if all

dinoflagellates that express both sizes have dimeric 35 kDa PCPs. Whether assembled with short or long polypeptides, PCP holoproteins probably interact with each other while in the thylacoid lumen. The crystal structure of the *A. carterae* PCP showed that its functional holoproteins exists as trimers of long PCPs (Hofmann et al. 1996). Furthermore, it has been suggested that the small PCPs of *H. pygmaea* may also aggregate to form similarly constructed homohexamer (Hiller et al. 2001).

It has often been demonstrated that PCPs occur in multiple isoforms that have distinct isoelectric points (pI) that range from pH 4.5-9.0 (Haxo et al. 1976; Prezelin and Haxo, 1976; Seigelman et al. 1977; Chang and Trench, 1984). Stable combinations of isoforms are expressed in a manner that is inversely proportional to light intensity and these combinations are species specific (Chang and Trench, 1982; 1984; Trench and Blank, 1987; see Figure 1.5). The source of PCP isoform variation, be it post-translational modification (Haxo et al. 1976; Seigelman et al. 1977), protein degradation (Ogata et al 1994) or genetic (Chang and Trench, 1984; Triplett et al. 1993; Sharples et al. 1996; Hiller et al. 2001) is unclear. Individual species may contain PCPs that have unique spectroscopic absorbance maxima (Prezelin and Haxo, 1976; Song et al. 1976). While the spectral properties of different PCP isoforms from *A. carterae* were found to be identical (Haxo et al. 1975), subtle differences were detected in both absorption and fluorescence maxima between individual isoforms from *S. microadriaticum* (Iglesias-Prieto et al. 1991). However, the functional significance is unknown.

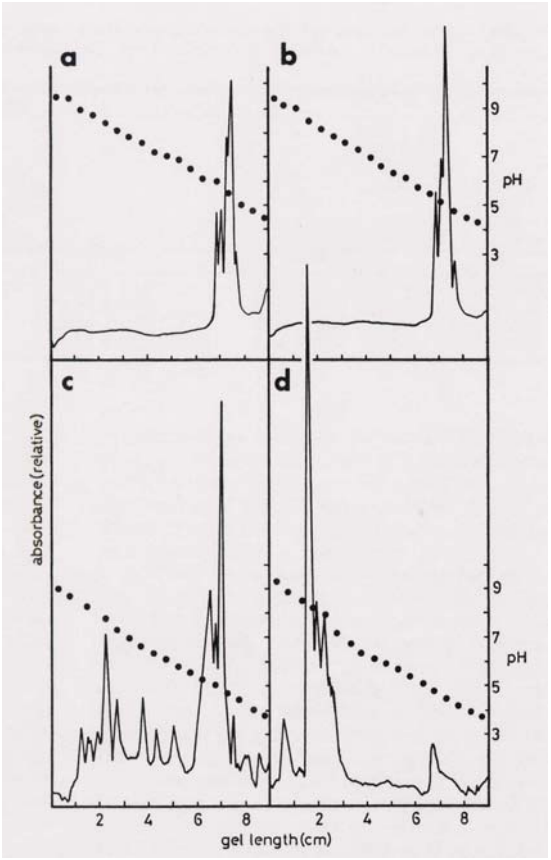


Figure 1.5 Isoelectric focusing patterns from **(a)** *Symbiodinium microadriaticum*; **(b)** *S. goreauii*; **(c)** *S. kawagutii*; and **(d)** *S. pilosum*. Reproduced with permission of the *Journal of Phycology* from Trench and Blank, 1987.

PCPs are coded for by nuclear genes. Just as there are short and long PCPs, two sizes of PCP genes have also been identified. To date, complete short gene coding sequences have come from *H. pygmaea* and *S. muscatinei*, and long gene coding sequences have come from *A. carterae*, *L. polyedra* and *Symbiodinium* from the Pacific coral *Acropora formosa* (Triplett et al. 1993;

Norris and Miller, 1994, Sharples et al. 1996; Hiller et al. 2001, Weis et al. 2002). PCP genes from free-living dinoflagellates have been shown to be intron-less and to exist in tandem arrays (Sharples, et al. 1996; Le et al. 1997; Hiller et al. 2001). Characterization of PCP genes from *L. polyedra* indicates that they occur in approximately 5000 highly conserved copies (Le et al. 1997). By contrast, *H. pygmaea* is estimated to have about 50 PCP gene copies and at least five different versions of these genes (the most diversity described prior to results presented in this paper). Evidence from *A. carterae* or *H. pygmaea* has mounted to suggest that PCP gene families may not be conserved in general and that the expression of different PCP genes is a primary source distinct PCP isoforms (Triplett et al. 1993; Sharples et al. 1996; Hiller et al. 2001). Diversity has also been a common feature reported in other dinoflagellate multigene families including those for luciferin-binding protein (Lee et al. 1993; Machabee' et al. 1994) and Rubisco (Rowan et al. 1996).

The organization and evolution of PCP genes is probably a reflection of the very unusual nature of dinoflagellate genomes. DNA content per dinoflagellate nucleus has been shown to range from 3.2 pg for *A. carterae* to 200 pg in *L. polyedra* (Holm-Hansen, 1969). Depending on species, dinoflagellate nuclei contain as few as 4 or many as 325 chromosomes (Spector, 1984). Changes in ploidy may have contributed to genome variability, but this has not been conclusively demonstrated. Furthermore, the distribution of PCP genes between chromosomes is completely unknown. Dinoflagellate chromosomes stay condensed and associated with nuclear membrane throughout most of the cell

cycle. Their DNA is coiled around non-histone proteins and the DNA contains the modified base 5-hydroxymethyluracil (Spector, 1984 and Rizzo, 1991). The expectation that genes in tandem arrays would tend to become homogenized through gene conversion and then evolve in concert may not be the case for PCP genes.

Detailed information is needed about PCP genes from symbiotic dinoflagellates to lay the groundwork for broader comparative analyses. Several basic questions about PCP genes remain unanswered. Apart from size, which characteristics distinguish PCP genes from different dinoflagellate species? To what degree are PCP gene families composed of divergent coding regions? Do different dinoflagellates species allocated the same relative proportion of their genomes to the expansion of PCP gene families? Does genetic diversity translate to functional differences of PCP isoforms?

This chapter presents the first extensive characterization of the PCP gene family from a symbiotic dinoflagellate, *Symbiodinium* 203 (ls rDNA clade C). Particular emphasis is placed on demonstrating the level PCP gene diversity in *Symbiodinium* 203 and how it acts as a primary source of variability in PCP isoforms. *Symbiodinium* 203 PCP gene structure, organization and copy number are compared to genes from *L. polyedra*. Predicted polypeptide pIs are calculated and compared to published pIs. Amino acid substitutions are mapped onto the *A. carterae* PCP crystal structure to consider functional significance especially with regard to how variation at polymorphic sites may influence the spectral tuning of peridinins.

MATERIALS & METHODS

Algal Cultures

Robert K. Trench donated *Symbiodinium* 203 from his culture collection at UC Santa Barbara. Uni-algal subcultures cultures of were maintained in 10 ml of 0.45 μm filtered Instant Ocean plus either Provasoli's Enriched Seawater (PES) or Guillard's F/2 (Sigma, Inc.). The cultures were grown under full spectrum fluorescent light banks (Sylvania 40W 4100K Designer) at 80 μ mole quanta/m^2/sec. on a 12:12 light to dark cycle at 27°C. Cultures were serially transferred every 3-4 weeks.

L. polyedra (CCMP no.1738) was obtained from the Provasoli - Guillard National Center for Culture of Marine Phytoplankton. One liter subcultures of *L. polyedra* were grown in 2.5 l Fernbach flasks under conditions similar to *Symbiodinium* 203 except that *L. polyedra* was only grown in F/2 and the temperature was maintained at 21ºC.

Nucleic Acid Extractions

Log-phase culture cells were harvested by centrifugation at 10,000g for 1 minute in 1.5 ml microcentrifuge tubes. Total genomic DNA was extracted either by a method previously developed for symbiotic dinoflagellates (Rowan and Powers, 1991b) or by a modification of the DNAeasy tissue extraction kit (Qiagen, Inc.), 1999 protocol. Steps 1 and 2 of the DNAeasy protocol were replaced by the following: The algal pellet was re-suspended in 500 μl of 2X CTAB buffer (100 mM Tris pH 8.0, 1.4 M NaCl, 25 mM EDTA, 2% (w/v)

CTAB, 0.2% v/v 2-mercaptoethanol). The re-suspended cells were ground within the microcentrifuge tubes with plastic pestles. 4 µl of 10 mg/ml Proteinase K was added to each tube and mix by inversion. The tubes were incubated at 65°C for ≥ 2 hours and were inverted every 30 minutes. The extractions then carried forward from step 3 of the DNAeasy protocol. DNA yield was verified on 1% UltraPure (Gibco BRL, Inc.) agarose TAE gels stained with ethidium bromide.

Nucleic acid extractions done by the Rowan and Powers method also contained substantial amounts of usable RNA. To purify the RNA for Reverse Transcription PCR (described below), 5 µl of the total nucleic acid preps were diluted into 95 µl of 1X DNAse Buffer and incubated with DNAse I (Ambion, Inc.) at 37°C for 1 hour. The RNA was phenol/chloroform extracted, ethanol precipitated, desalted, air dried and re-suspended in RNAse-free H_2O. RNA yield from this purification was checked on 1% agarose TBE gels stained with ethidium bromide.

Primer Design

Primer used for standard PCR, Reverse Transcription PCR, Quantitative Real Time PCR and for sequencing are listed in Table 1.1. Published PCP gene sequences *Symbiodinium* from *Acropora formosa* (Norris and Miller 1994, GenBank accession no. L13613), *A. carterae* (Sharples et al. 1996, Z50792, and Z50793) and *L. polyedra* (Le et al. 1997,U93077) were aligned with Clustal X (Thompson et al. 1997) to identify conserved regions and for initial design of primer set U325/L537. Thereafter, sequence from derived clones was used to design further primers. Primer selection for standard and reverse transcription

PCR was optimized with Oligo 4.0 (National Biosciences, Inc.). Primers for quantitative real-time PCR were designed with Primer Express (Applied Biosystems, Inc.).

Table 1.1 Primers used for amplification and sequencing.

Numbers indicate location of 5' end of primer relative to the *Symbiodinium* 203 coding region. Those in parentheses are outside of the coding and variable due to indels in the spacers.

Upper Primers	Lower Primers	Primer Sequence (5' - 3')
U325		AAGAATTCGAAGGACGCAGCAGAAGC
	L537	CAGAATTCCTTCATGTACGCTGGCAC
U448		TCGGTCCCCAAAGCAAAGGTCA
	L423	CATTCACGGCATCCCAGTCAGC
U935		CCGTGCCCAAGTCCACTGTCA
	L(-67)	GCAGGATGATTGGGATGAGT
U(1540)		GAGCCGAACACATCCAGCAG
	L24	AGCTTTCCTTGCTCCACGCAC
U3		GGTGCGTGGAGCAAGGAAAG
	L913	GCATTCACAGCTTCGTAGT
U(-28)		TCCGGCCCACTTTTAGTTTT
	L(1180)	TTTTCCCATTTGTTTCAGAG
SYBRf938		TGCCCAAGTCCACTGTCATG
	SYBRr1006	TGTTGGTCACGGTGGAATCA

PCR Conditions, Identification and Purification of Products

Standard PCR conditions (Palumbi, 1996) were generally used with modification to the 10X PCR buffer (200 mM Tris-HCl pH 8.8, 100 mM KCl, 100 mM $(NH_4)_2SO_4$, 20 mM $MgSO_4$, 1% Triton X-100 and 1 mg/ml BSA). 50 µl reactions were carried out in MJ Research, Inc PTC-100 and Mini-Cycler thermocyclers. Thermal cycle profiles were adjusted to accommodate the

annealing temperature of the primer sets and the length of the expected PCR product (Table 1.2). Each round of PCR included a negative control to check for contamination of reagents. PCR reactions were generally run out on 2% agarose TAE gels stained with ethidium bromide. When multiple bands were present in a given reaction, Southern hybridization (Sambrook, 1989) was used to identify PCP gene amplification products. PCR products were transferred to Hybond-N+ nylon membranes (Amersham, Inc.). Probes were created with DECAprime II random priming DNA labeling kits (Ambion, Inc.). Gel excised PCR products of interest were purified with QIAEX II gel extraction kits (Qiagen, Inc.).

Table 1.2 Primer Sets & Thermoprofiles.

Steps are expressed as time in minutes at a temperature in degrees C or as cycles to repeat.

Reaction	Primer(s)	step 1	step 2	step 3	step 4	step 5	step 6	step 7
PCR	U325/L537	3:00/94C	1:00/50C	:45/72C	:45/94C	GO TO 2 30X	:45/60C	:45/72C
PCR	U448/L423	3:00/94C	1:00/55C	2:15/72C	1:00/94C	GO TO 2 30X	1:00/55C	3:00/72C
PCR	U(-28)/L(1180)	3:00/94C	:45/50C	1:15/72C	1:00/94C	GO TO 2 30X	:45/ 50C	3:00/72C
RT	OLIGOdT	5:00/65C	1:00/30C	30:00/30C to 65C	5:00/98C	5:00/ 5C		
cDNA SYNTH	U3/L913	:30/94C	:30/55C	1:00/72C	GO TO 2 30X	4:00/72C		
CYCLE SEQ	VARIOUS	2:00/95C	:30/95C	:30/42C	1:00/70C	GO TO 2 30X	5:00/4C	
AUTO SEQ	VARIOUS	:10/96C	:05/50C	4:00/60C	GO TO 1 24X			
Q RT PCR	SYBRf938/SYBRr1006	5:00/94C	:20/60C	:20/72C	:20/94C	GO TO 2 40X	:20/60C	:20/72C

Reverse Transcription PCR (RT-PCR)

A BcaBEST RNA PCR Kit Ver. 1.1 (TaKaRa, Inc.) was used for Reverse Transcription and amplification of cDNA. The oligo dT primer included in the kit was used for synthesis of single stranded cDNA following a thermoprofile (including a 30-minute gradual rise in temperature from 30°C to 65°C) suggested by the manufacturer (Table 1.2). Subsequent amplification of double stranded cDNA was done with the U3/L913 primer set using "A Method" from the kit manual. Each round of RT-PCR included controls to check for contamination.

Cloning & Screening Plasmid Libraries

The PCR product from U325/L537 was blunt cloned into the pBluescript II SK vector (Stratagene, Inc.), and the ligation was used to transform chemically competent *E. coli*. All other PCR products were cloned into pCR 2.1 vectors with TA Cloning kits (Invitrogen, Inc.) and were used to transform INVαF' cells included in the kits according to the manufacturers protocol. White transformants growing on L-broth agar plates containing ampicillin and X-Gal were picked for replica plating and PCR screening of the plasmid libraries. To screen the libraries, colony bits were re-suspended into 50 µl ddH_2O in 1.5 ml tubes by vortexing for 30 seconds. Tubes were heated to 99°C for 5 minutes to lyse the cells and to denature DNAses and then were placed on ice. Lysates were spun at 10,000 g for 2 minutes to pellet bacterial debris. 10.825 µl of lysate from each tube was combined with 14.125 µl of PCR cocktail consisting of 6.25 µl ddH_2O, 5 µl of 5 mM dNTPs, 2.5 µl of standard 10X PCR buffer, 1.5 µl 25 mM $MgCl_2$,

0.625 µl of each 20 µM primer and 0.125 µl of *Taq*. Amplifications were done with themoprofiles listed above (Table 1.2). Cells from positive clones were picked from replica plates grown overnight in 3 ml cultures of L-broth containing ampicillin. Clones were purified with QIAamp DNA Mini kits (Qiagen, Inc.). PCR products from these mini preps were gel excised, purified as described above and eluted in ddH_2O for sequencing.

Sequencing & Sequence Analysis

All published sequences for this project were sequenced at least twice and most were done in both the forward and reverse directions. The sequencing thermoprofiles used are described in Table 1.2. The clone from the U325/L537 PCR product was sequenced with a *fmol* DNA Cycle Sequencing System kit (Promega, Inc.), electrophoresed on conventional 6% polyacrylamide gels and visualized on autoradiographs. Automated sequencing was done for all other clones using BigDyeV2 Terminator kits (ABI, Inc.). Half-reaction mixes consisted of a maximum of 3µl of template (adjusted to include 20-40 ng), 1 µl of 10 µM primer, 1 µl 5X Sequencing Buffer, 2 µl BigDyeV2 Terminator mix and the balance with ddH_2O. Automated sequencing reactions were purified through Sephadex G-50 Medium (Sigma, Inc.) columns and vacuum dried before re-suspensions in loading dye. Samples were run out in 6% ThermoPage acrylamide gels (Amerisco, Inc), and data was collected on a Perkin Elmer ABI PRISM 377 DNA Sequencer. Sequence contigs were assembled with Seqman (DNA Star, Inc). Translation of predicted proteins and pI estimation with done with Gene Runner v3.04 (Hastings Software Inc.) and with Edit Seq (DNA Star Inc.).

GenBank database searches for similar nucleotide and amino acid sequences were done using the BLAST algorithm (Altschul et al. 1990). Nucleotide sequences of clones used in the analyses within this chapter were submitted to GenBank and assigned the accession numbers listed in Table 1.3.

Table 1.3 *Symbiodinium* 203 PCP clone accession numbers

Species	Clone Name	Accession Number
Symbiodinium 203	203clone41	AY149122
"	203clone79	AY149123
"	203clone80	AY149124
"	203clone81	AY149125
"	203clone82	AY149126
"	203clone83	AY149127
"	203clone84	AY149128
"	203clone85	AY149129
	203clone87	AY149130
"	203clone89	AY149131
"	203clone90	AY149132
"	203cDNAclone34	AY149133
"	203cDNAclone39	AY149134
"	203cDNAclone42	AY149135
"	203cDNAclone50	AY149136
"	203cDNAclone51	AY149137
"	203cDNAclone55	AY149138
"	203cDNAclone59	AY149139

Nucleotide Sequence Divergence

K-Estimator 5.5 (Cameron, 1995; 1999) was used to calculate of the number of nonsynonymous (Ka) and synonymous (Ks) nucleotide substitutions per site for all paired comparison of complete coding regions. No comparison restrictions were introduced. 1095 bp sites within 365 codons were analyzed for

each comparison. The Kimura-2p method was used to correct the number of substitution hit per site. Ka and Ks values were generated from separate analyses, and then Ka/Ks was calculated for each pair of compared sequences. DNA coding regions were aligned with Clustal X (Thompson et al, 1997), MegaAlign (DNA Star, Inc.). Nexus files were generated with MegaAlign and with MacClade v4.0 (Maddison and Maddison, 2000). Execution of the phylogenetic analysis was done with PAUP* 4.0 (Swofford, 1998). Neighbor Joining analysis settings used were with DNA distances corrected by the Kimura-2p model. A Neighbor Joining tree was generated to visually depict the nucleotide sequence divergence.

PCR Recombination and Fidelity Controls

2 µl of each mini-prep from three unique clones were diluted 1:100 with Tris HCl buffer pH 8.3. Pair-wise combinations of the diluted clones (1 µl/ea) were used as templates for 50 µl PCR reactions that were run under conditions described above. Amplification products from each reaction were TA cloned. Three sets of 16 sub-clones were isolated, sequenced with U448 and with L423. Each was then identified as being either one of the original templates or a recombinant. The location of the recombination was roughly determined and the observable recombination frequency was calculated as:

$$\text{Number of recombinant clones} \div \text{total number of clones} \times 100\% = \text{observable recombination frequency} \quad (1.1)$$

In addition, 1 µl of 1:100 diluted complete cds clone was amplified in 5 separate PCR reactions under the same conditions as above. A single TA clone was isolated from each reaction and sequenced with U(-28), U448, L913 and L(1180) to check for nucleotide substitutions introduced by the PCR process.

PCP Gene Copy Number and Genome Size Estimation

The number of PCP genes/ genome for *Symbiodinium* 203 was calculated by the overall equation:

$$\frac{\text{PCP genes} \pm \sigma_{genes}}{\text{pg genomic DNA}} \times \frac{\text{pg genomic DNA} \pm \sigma_{\text{pg}}}{\text{genome}} = \frac{\text{PCP genes} \pm \sigma_{\text{genes}}}{\text{genome}} \quad (1.2)$$

where σ equals the standard deviation.

The $\frac{\text{PCP genes} \pm \sigma_{genes}}{\text{pg genomic DNA}}$ term in equation 1.2 was determined by quantitative real-time PCR with a Perkin Elmer ABI 7700. Data was analyzed with ABI Sequence Detection Systems 1.7 software. Amplifications of a 69 bp PCP gene segment were compared between known amounts of *Symbiodinium* 203 genomic DNA and a linearized PCP clone. Concentrations of genomic and plasmid DNA were quantified in triplicate on a Beckman Coulter DU640 spectrophotometer. 1.492 µg of the clone was digested with 2 units Not I (New England Biolabs, Inc.) at 37ºC for 1 hour to cut the pUC19 vector at a single site but not the insert. The reaction was terminated by heat inactivation at 65ºC for 20 minutes. The linearized plasmid DNA was used to create a dilution series that

provided the templates for the amplification standard curve. Replica 50 µl reactions were set up for each of the standards and for negative controls. Triplicate 50 µl reactions were set up for the *Symbiodinium* 203 genomic samples. Reagents were assembled in master mix and were distributed so that each reaction contained 4 µl of 5 mM dNTPs, 5 µl of 10X PCR buffer (100 mM Tris HCl pH 8.3, 500 mM KCl, 30 mM $MgCl_2$, 10% glycerol, 0.1% Tween 20 and 1:4000 dilution of SYBR®Green I (Molecular Probes, Inc.), 6 µl of 25 mM $MgCl_2$, 0.5 µl (2 units) of SuperTaq (Ambion, Inc.), 1 µl of 5 µM SYBRf938 primer, 1 µl of 5 µM SYBRr1006 primer, 27.5 µl ddH_2O and 5 µl DNA template (= 10,000 pg of genomic DNA, and 1492000 to 14.92 fg in 10-fold dilutions of plasmid DNA). In order to compare PCP copy number to ng of genomic DNA, PCP gene copies/ fg of PCP clone were converted as follows:

$$\frac{(\text{plasmid} + \text{insert})}{5108\ \text{bp}} \times \frac{1\ \text{mole bp}}{660\ \text{g}} \times \frac{6.02 \times 10^{23}}{\text{mole bp}} \times \frac{1\ \text{g}}{10^{15}\ \text{fg}} = \frac{179(\text{plasmids} + \text{inserts})}{\text{fg}} \qquad (1.3)$$

The $\frac{\text{pg genomic DNA} \pm \sigma_{\text{pg}}}{\text{genome}}$ term in equation 1.2 was estimated by comparing the average genome sizes of *Symbiodinium* 203 and *L. polyedra* by flow cytometry using a Beckton Dickenson FACS Calibur. The instrument was equipped with a 15 mW Argon laser producing excitation at 488 nm. A modified version of the Veldhuis et al. 1997 protocol for fixing and staining the cells was

used. An equal volume of algal cells suspended in F/2 was combined with 4% paraformaldehyde and let stand at room temperature for 15 minutes with 0.33 µl of 1% Triton X-100 and 1 µl of 1µg/ml RNAse A were added per ml of fixed cells. The solutions were mixed by gentle inversion several times and kept at room temperature for 20 minutes. Cells were pelleted by certifugation at 3000 rpm at 4ºC in a Sorvall Ultrafuge. After the liquid was carefully decanted, the fixed cells were re-suspended in 500 µl of TE. 1 µl 1% Triton X-100 and 50 µl of 20X PicoGreen (Molecular Probes, Inc.) were added and mixed by inversion. The cells were allowed to stain at room for at least one hour prior to flow cytometry (Figure 1.6). Fluorescence of PicoGreen stained DNA was measured through the FACS FL1 filter at 530 ± 30 nm. To calculate pg genomic DNA for *Symbiodinium* 203:

$$\frac{\text{mean fluorescence} \pm \sigma_{\text{fluorescence}} \text{ for } \textit{Symbiodinium } 203}{\text{mean fluorescence} \pm \sigma_{\text{fluorescence}} \text{ for } \textit{L. polyedra}} \times \frac{200 \text{ pg DNA}}{\text{genome}} \text{ for } \textit{L. polyedra} \qquad (1.4)$$

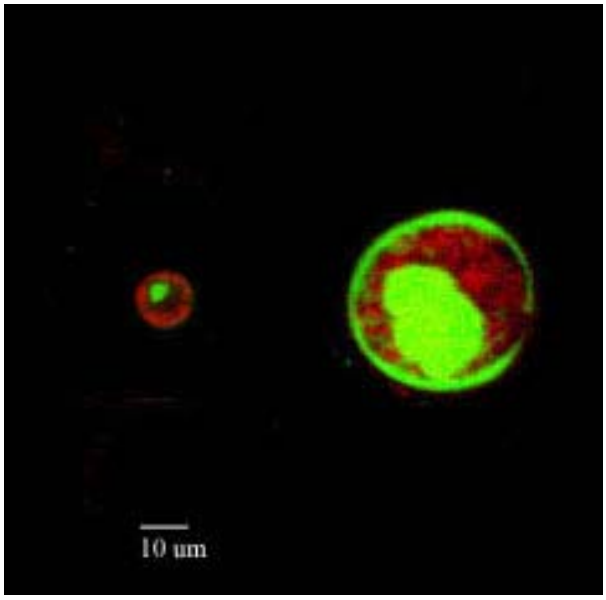


Figure 1.6 Superimposed confocal images of *Symbiodinium* 203 (left) and *L. polyedra* (right) cells at their proportional sizes showing fluorescence of PicoGreen stained nuclear DNA and red autofluorescence of chloroplasts.

Within the simple mathematic operations above, the following equations were used to accurately account for the propagation of indeterminate errors (Melissinos, 1966). Let A through D stand for means of independent data sets or resultants from previous calculations.

For the division $\frac{A \pm \sigma_A}{B \pm \sigma_B}$:

$$\frac{\sigma_{\frac{A}{B}}}{\frac{A}{B}} = \sqrt{\left(\frac{\sigma A}{A}\right)^2 + \left(\frac{\sigma B}{B}\right)^2}$$

$$\sigma_{\frac{A}{B}} = \frac{A}{B} \times \sqrt{\left(\frac{\sigma A}{A}\right)^2 + \left(\frac{\sigma B}{B}\right)^2}$$

The resulting quotient will be:

$$\frac{A}{B} \pm \frac{A}{B} \times \sqrt{\left(\frac{\sigma A}{A}\right)^2 + \left(\frac{\sigma B}{B}\right)^2} \tag{1.5}$$

And for the multiplication $(C \pm \sigma_C) \times (D \pm \sigma_D)$:

$$\frac{\sigma_{CD}}{CD} = \sqrt{\left(\frac{\sigma C}{C}\right)^2 + \left(\frac{\sigma D}{D}\right)^2}$$

$$\sigma_{CD} = CD \times \sqrt{\left(\frac{\sigma C}{C}\right)^2 + \left(\frac{\sigma D}{D}\right)^2}$$

The resulting product will be:

$$CD \times CD \times \sqrt{\left(\frac{\sigma C}{C}\right)^2 + \left(\frac{\sigma D}{D}\right)^2} \tag{1.6}$$

Amino Acid Substitution Modeling

Predicted amino acid substitutions were mapped onto the crystal structure for the *A. carterae* PCP trimer 1PPR using Swiss-PDB Viewer v3.7(b2) (Glaxo Wellcome, Inc.) as follows: Pair wise amino acid alignments between *Symbiodinium* 203 and *A. carterae* PCPs were created to identify conserved and variable sites. Mutations were individually introduced in the 1PPR structure at each of the fixed and polymorphic sites. As substitutions were made, rotomer conformations were optimized by the software. Distances were then calculated between polymorphic sites and either molecules in neighboring monomers or chromophores within the same monomer. A single layer pdb was rendered reflecting all changes. 3D protein structure images were rendered using POV-Ray v3.1 (POV-Ray Team) and could only be created from single layer pdb files.

RESULTS

Organization and Diversity of Genomic Coding and Spacer Regions

The cloning strategy for this project is summarized in Figure 1.7. The long PCP gene sequences *Symbiodinium* from *Acropora formosa* (GenBank accession no. L13613), *A. carterae* (Z50792, and Z50793) and *L. polyedra* (U93077) are highly divergent and consequently did not produce large, conserved blocks that could be easily targeted for amplification. The alignment suggested that an approximately 200 bp region in the 5' half of the long PCP genes was moderately

conserved. Primers U325 and L537 based on sequence L13613 successfully amplified a 212 bp fragment from *Symbiodinium* 203 (Figure 1.7a). The sequence of small cloned fragment was positively identified by a BLAST search as partial PCP gene.

When this research began, only long PCP genes from the free-living dinoflagellates *A. carterae* and *L. polyedra* were shown to exist in tandem arrays (Sharples et al. 1996; Le et al. 1997). Although it was not known at the time whether *Symbiodinium* 203 had long or short PCP genes, it was reasonable to assume that either variety would be arranged in the same manner.

Sequence data from the 212 bp fragment was used for generation of outward facing primers U448 and L423 to amplify between adjacent gene copies. Amplification products of various sizes were present in the reaction. A Southern Blot was created and probed with the small fragment. There was substantial hybridization to a 1.9 kb band which was then excised and purified from another gel. Direct sequencing of the 1.9 kb PCR product effectively failed, because the reads contained many unresolved bases. This was an important clue that multiple templates were present in the sequencing reaction. The 1.9 kb fragments were TA cloned, and a single clone was sequenced completely by primer walking (see Appendix 1.1). This sequence confirmed that the PCP genes from *Symbiodinium* 203 are arranged in tandem arrays like *A. carterae* and *L. polyedra*. Open reading frames at the 5' and 3' ends of this clone both translate to PCP and flank a 903 bp spacer region.

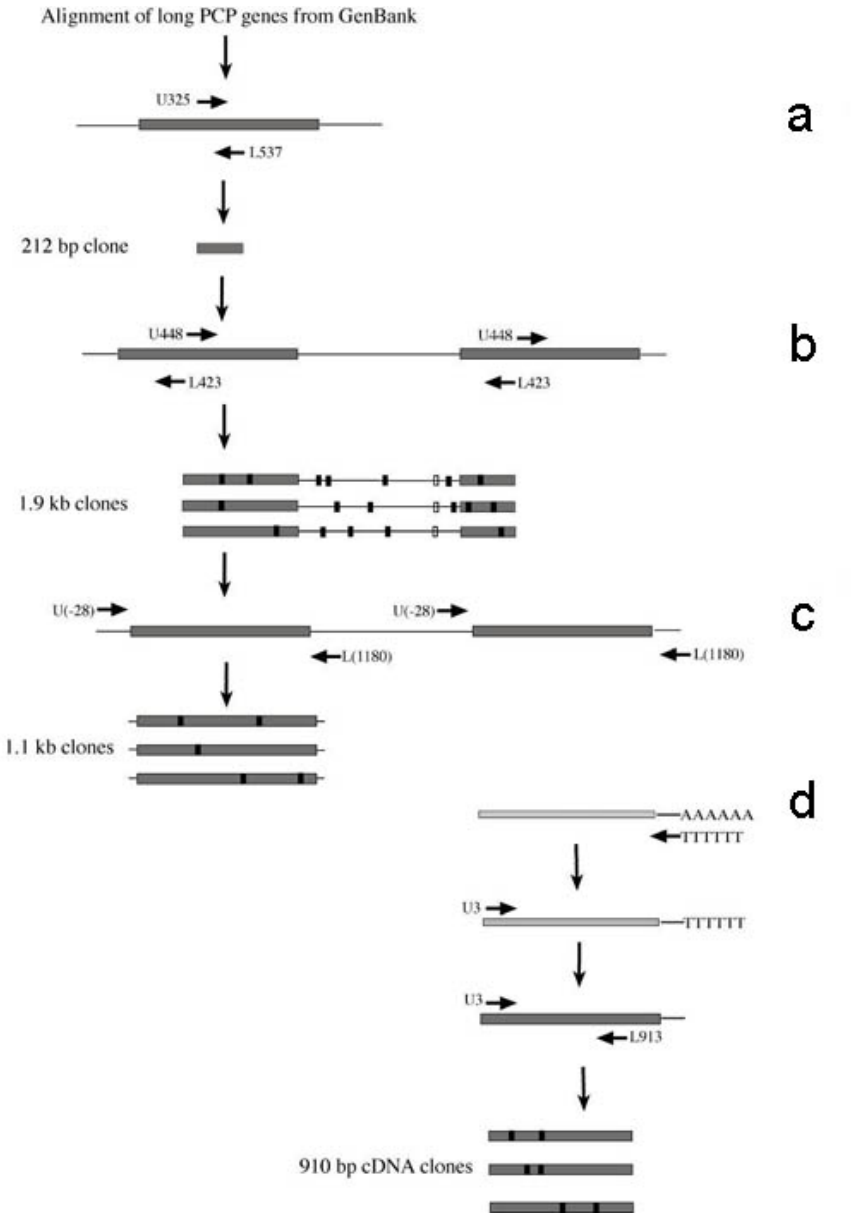


Figure 1.7 Cloning strategies. Horizontal bars represent coding regions. Horizontal lines represent spacer. Smaller black vertical bars represent hypothetical locations of substitutions, Open vertical bars represent putative promoter; **(a)** original clone; **(b)** clones between adjacent coding regions; **(c)** clones of complete coding regions; and **(d)** cDNA clone

Sequencing additional 1.9 kb clones revealed that there were numerous substitutions distributed throughout the coding and spacer regions (complete nucleotide sequence data for all tandem array clones not shown). As expected, the spacers contained the majority of the substitutions including many insertions and deletions. Despite the highly variable nature of the spacers, each was found to contain a conserved 13 bp sequence CTTGAATGCAGAA approximately -201 to -188 bp upstream of the start codon. Figure 1.8 shows that 9 of the 13 bases in this sequence are identical to and in the same relative location as the promoter previously identified from the *L. polyedra* luciferase and PCP spacers (Li and Hastings. 1998). This conserved sequence is very likely to be the *Symbiodinium* 203 promoter.

```
                        (-228)    (-218)    (-208)    (-198)    (-188)    (-178)
                           .         .         .         .         .         .

L poly promoter                                          CGTGAACGCAGTG
203_41             GTGCTTGTCC----------GTTTGCAATCTG---TCTTGAATGCAGAATCGAGACAGCA
203_43             GGGCTTGTCC----------GTTTGCAATCTGTCTTCTTGAATGCAGAATCGAGAAAGCA
203_44             GTGCTTGTCC----------GTTTGCAATCTGTCTTCTTGAATGCAGAATCCAGACAGCA
203_46             GTGCTTGTCC----------GTTTGCAATCTG---TCTTGAATGCAGAATCGAGACAGCA
203_49             GTGCTTGTCC----------GTTTGCGATCTGTCTTCTTGAATGCAGAATCGAGACAGCA
203_50             GTGCTTGTCC----------GTTGGCAATCTGTCTTCTTGAATGCAGAACCGAGACAGCA
203_52             GTGTTTGTCC----------GTTTGCAATCTGTCTTCTTGAATGCAGAATCCAGACAGCA
203_54             GTGCTTGTCCGTTTATGTCCGTTTGCAATCTGTCTTCTTGAATGCAGAATCGAGACAGCA
203_55             GTGCTTGTCC----------GTTTGCAATCTGTCTTCTTGAATGCAGAATCGAGACAGCA
                   * * ******          ************   ** **** ****  ** *** ****
```

Figure 1.8 Alignment of conserved upstream region of *Symbiodinium* 203 spacers vs. the *L. polyedra* promoter (both in bold).

Amplification of *Symbiodinium* 203 genomic DNA with inward facing primers U(-28) and L(1180) produced a 1.1 kb band. After gel extraction and

purification, direct sequencing resulted in the same pattern of numerous un-resolved bases seen with the 1.9 kb PCR products and again suggested that multiple templates were present. TA cloning produced clones 79-90. Clones 79 and 81-90 contained a 1098 bp complete long PCP coding sequence (Figure 1.7c). The open reading frame in clone 80 had a stop 27 bases earlier than other. All coding sequences except those within clones 86 and 88 were distinct from each other in terms of their nucleotide substitutions (97.8-99.5% identical). Of the published PCP gene sequences in GenBank, the *Symbiodinium* 203 coding sequences were most similar to the coding sequence of *Symbiodinium sp.* L13613 (92.6-94.5% identical). An alignment of genomic and partial cDNA clones (described below) is presented in Appendix 1.2. The majority of nucleotide substitutions between the complete 1098 bp coding regions were in synonymous codon positions resulting in Ka/Ks <1 (Table 1.4). Only the comparison of clones 79 and 82 had a Ka/Ks >1.

Table 1.4 Ka/Ks values for paired comparisons of *Symbiodinium* 203 1095 bp coding regions

clone 79	clone 81	clone 82	clone 83	clone 84	clone 85	clone 87	clone 89	
0.385								clone 81
1.115	0.373							clone 82
0.326	0.526	0.375						clone 83
0.294	0.0222	0.26	0.203					clone 84
0.276	0.208	0.233	0.178	0.276				clone 85
0.363	0.248	0.399	0.213	0.0113	0.038			clone 87
0.324	0.274	0.266	0.218	0.333	0.474	0.111		clone 89
0.306	0.404	0.399	0.335	0.186	0.162	0.22	0.196	clone 90

Low PCR Recombination Frequency and High PCR Fidelity

To control for possible generation of recombinant DNA sequences by amplifying genes in tandem (Bradley and Hillis, 1997), a re-amplification and sub-cloning experiment was conducted using pairs of distinctive clones as the templates. All other PCR conditions were the same as when genomic *Symbiodinium* 203 was amplified. The results are presented in Table 1.5. Sixteen sub-clones from each of the amplifications were partially sequenced to identify them as either one of the original templates within the PCR reaction or as recombinant. The first 500 bp of sequence from both the U448 (5') and L423 (3') ends were used for the identification. Four of the 48 sub-clones had different U448 and L423 identities. For each of these recombinants, the flanking regions sequenced completely matched one or other of the templates suggesting that recombination had likely occurred in the spacer. This may be attributable to secondary structure formation within the spacer regions of the clones during the PCR reactions. One of the sub-clones was excluded as a contaminant. The observable recombination frequency was estimated from equation 1 as:

$$4 \text{ recombinants} \div 47 \text{ total sub-clones} \times 100\% = 8.5\%$$

Table 1.5 PCR Recombination Control

	Identity of 1st 500bp from U448		Identity of 1st 500bp from L423		
Sub-clone	CLONE 41	CLONE 52	CLONE 41	CLONE 52	**RECOMBINANT**
A1		*		*	
A2	*		*		
A3		*	*		*
A4		*		*	
A5	*		*		
A6	*		*		
A7	*		*		
A8		*		*	
A9	*		*		
A10	*		*		
A11		*		*	
A12		*	*		*
A13		*		*	
A14	*		*		
A15	*		*		
A16		*		*	
Sub-clone	CLONE 41	CLONE 54	CLONE 41	CLONE 54	**RECOMBINANT**
B1		*		*	
B2	*		*		
B3		*		*	
B4	*			*	*
B5		*		*	
B6	*		*		
B7	*		*		
B8		*		*	
B9	*		*		
B10	*		*		
B11		*		*	
B12		*		*	
B13	*		*		
B14		*		*	
B15	*		*		
B16	*		*		
Sub-clone	CLONE 52	CLONE 54	CLONE 52	CLONE 54	**RECOMBINANT**
C1	*		*		
C2		*		*	
C3		*		*	
C4		*		*	
C5	*		*		
C6	*		*		
C7		*		*	
C8		*		*	
C9		*		*	
C10		*	*		*
C11	*		*		
C12	not PCP	not PCP	not PCP	not PCP	
C13	*		*		
C14		*		*	
C15		*		*	
C16		*		*	

The rate at which *Taq* error introduced substitutions in our PCR reactions was also estimated. A single complete cds clone was re-amplified in five separate PCR reactions and a single sub-clone from each reaction was completely sequenced in both directions. There were no changes introduced in to any of the 1198 base sub-clones.

Diversity Also Expressed at mRNA Level

The surprising diversity found in the genomic clones was also present in partial cDNA clones. Poly-A mRNA from *Symbiodinium* 203 was reverse transcribed with oligo dT primers and then double-stranded cDNA was amplified with the U3/L913 primer set (Figure 1.7d). A single band with the expected size was seen after gel electrophoresis, and this was excised and purified. As with the previous genomic amplifications, directed sequencing failed because of apparent multiple temples and individual cloned RT-PCR products had diversified nucleotide sequences. Eight positive cDNA clones were sequenced and none were identical to the overlapping regions in the complete coding sequence clones. Five of these cDNAs were unique (Appendix 1.2). The absence of any insertions in the complete coding sequences compared to the cDNAs is evidence that *Symbiodinium* 203 PCP genes, like *L. polyedra* (Le et al. 1997) and *H. pygmaea* (Hiller et al. 2001) are intron-less. The overall diversity of genomic and cDNA nucleotide sequences is graphically depicted in the Neighbor Joining tree in Figure 1.9.

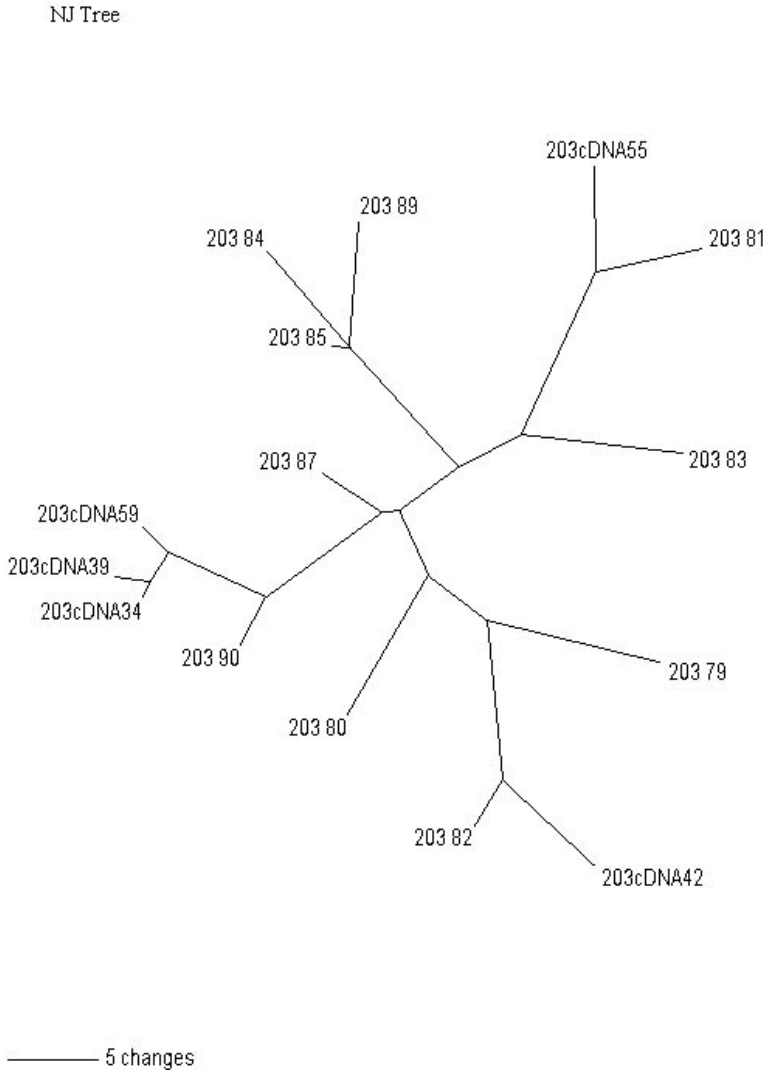


Figure 1.9 Neighboring Joining tree of *Symbiodinium* 203 genomic and cDNA nucleotide sequences.

Symbiodinium 203 Gene Family Size

Quantitative Real-Time PCR and flow cytometry were used to estimate the number of PCP genes in the *Symbiodinium* 203 genome. The curves in Figure 1.10 represent the changing quantities of 68 bp SYBRf938/SYBRr1006 PCR products as amplified from 10 ng of *Symbiodinium* 203 genomic DNA verses the dilution series of a single compete cds clone. Figure 1.11 is the standard curve generated by comparing the quantity of amplicons in each reaction at the threshold value in the log phase of amplification suggested by the ABI 377. There was a 0.996 correlation coefficient between all points (unknowns and standards) on the curve. The mean amplification value for three replicates of 10 ng of genomic DNA was 667.60 with a standard deviation of 51.89. This indicates that 10 ng of *Symbiodinium* 203 genomic DNA contains about the same number of PCP gene copies as 667.60 fg of clone 79. From equation 3, each fg of clone 79 has 179 PCP gene copies. Therefore 10 ng of genomic DNA contains 119500 ± 9288 PCP genes. In other words, there are roughly 12 ± 0.9 PCP genes per pg of genomic DNA.

The genome size of *Symbiodinium* 203 was compared to *L. polyedra* by flow cytometry to estimate the DNA content per nucleus. Table 1.6 shows the comparison of relative mean fluorescence of PicoGreen strained DNA from at least 3000 cells of each species.

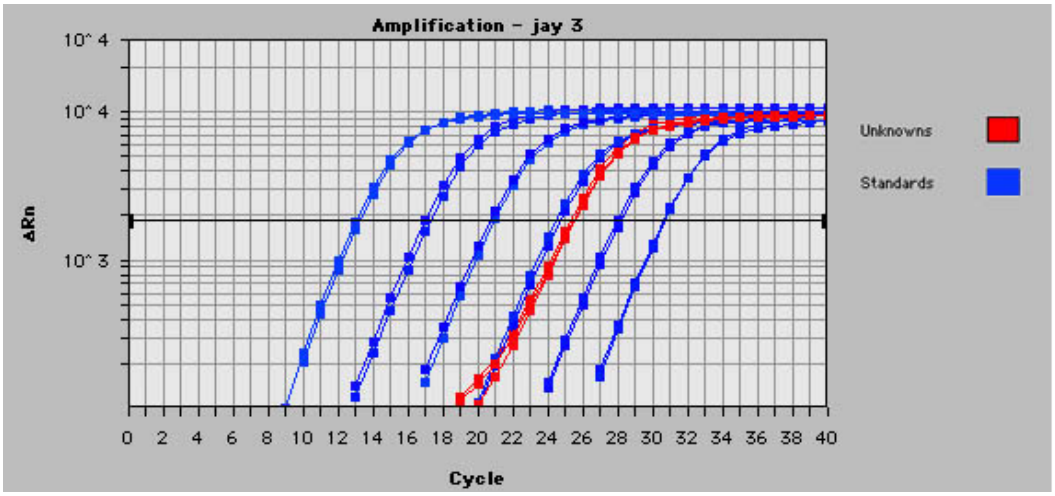


Figure 1.10 Quantitative Real-Time PCR amplification curves for *Symbiodinium* 203 genomic DNA vs. dilution series of cloned PCP cds.

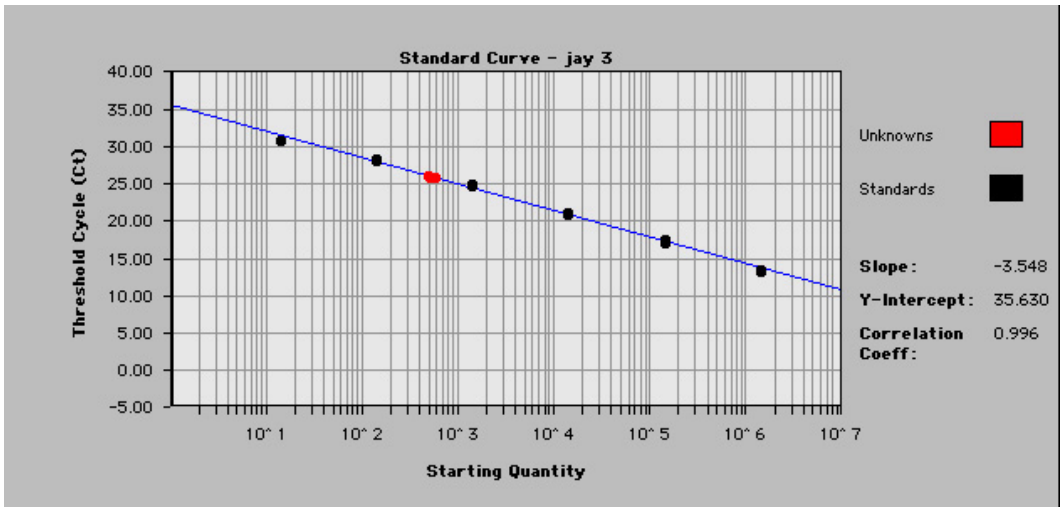


Figure 1.11 Standard curve comparing PCP gene copies in *Symbiodinium* 203 genomic DNA vs. dilution series of clones PCP cds.

Table1.6 Relative fluorescence of DNA stained with PicoGreen

Sample name	Parameter	Mean	CV (%)	Std. Dev
Symbiodinium 203	FL1-H	1.92	16.73	0.32
L. polyedra	FL1-H	152.09	26.47	40.26

Using equations 1.4 and 1.5:

$$\frac{1.92 \pm .32 \text{ for } \textit{Symbiodinium } 203}{152.09 \pm 40.26 \text{ for } \textit{L. polyedra}} \times \text{for } \textit{L. polyedra}$$

$$= \left(\frac{1.92}{152.09} \pm \frac{1.92}{152.09} \times \sqrt{\left(\frac{.32}{1.92}\right)^2 + \left(\frac{40.26}{152.09}\right)^2} \right) \times \frac{200 \text{ pg DNA}}{\text{genome}}$$

$$= \frac{3 \pm 1 \text{ pg DNA}}{\text{genome}} \text{ for } \textit{Symbiodinium } 203$$

Combining the quantitative real-time PCR and flow cytometry results within equations 1.2 and 1.7:

$$\frac{12 \pm 0.9 \text{ PCP genes}}{\text{pg genomic DNA}} \times \frac{3 \pm 1 \text{ pg genomic DNA}}{\text{genome}} \text{ for } \textit{Symbiodinium } 203$$

$$= 12 \times 3 \pm 12 \times 3 \times \sqrt{\left(\frac{0.9}{12}\right)^2 + \left(\frac{1}{3}\right)^2}$$

$$= \frac{36 \pm 12 \text{ PCP genes}}{\text{genome}} \text{ for } \textit{Symbiodinium } 203$$

Predicted Proteins, Isoelectric Points and Amino Acid Substitutions

Despite Ka/Ks values <1 for most of the compared *Symbiodinium* 203 coding regions, there was still considerable amino acid sequence variation. An alignment of predicted proteins from cloned complete coding regions is presented in Appendix 1.3. All clones but clone 80 (which had a stop codon 27 bases upstream of others) coded for 365 aa PCP pre-proteins with 52 aa transit peptides and 313 aa apoproteins. Pre-proteins ranged from 96.2-99.7% identical when compared to each other and 93.3-94.9% when compared to *Symbiodinium sp*. L13613. Within the *Symbiodinium* 203 transit peptides and apoproteins there were 9 and 20 polymorphic sites respectively. The 313 aa apoproteins had an average mass of 33 kDa and were the same length as those predicted from the *Symbiodinium sp*. L13613 and *A. carterae* Z50792 and Z50793 sequences. By contrast, the *L. polyedra* PCP gene sequence U93077 codes for a 375 aa pre-protein with a 59 aa transit peptide and a 316 aa apoproteins. Variability in long PCP mass is attributable to amino acids composition and also to polypeptide length.

Most of the predicted apoproteins from the clones had subtle differences in their mass, but they also varied in terms of their isoelectric points (Table 1.7). The pIs for each of the clones fell within a range of pH 5.73-6.78. The PCP isoforms reported from other dinoflagellates are not acidic for all species. Most isoforms from *A. carterae* (Haxo et al. 1976) and *Glenodinium (Heterocapsa) sp.* (Prezelin and Haxo 1976) are basic. However, there are several examples of

symbiotic dinoflagellates that produce predominantly acidic PCP isoforms with pI ranges similar to those predicted here including *S. microadriaticum, S. goreauii* (Trench and Blank, 1987) *Symbiodinium* from *Montastrea annularis* and *Symbiodinium* from *M. cavernosa* (Chang and Trench, 1982). In addition, the calculated apoprotein pI of the *Symbiodinium* from *A. formosa* sequence is 5.28.

Table 1.7 Calculated mass and isoelectric points from *Symbiodinium* 203 apoproteins

203 clone	Apoprotein mass (kDa)	Isoelectric point (pI)
79	33.38	6.03
80	31.956	6.37
81	33.01	6.78
82	32.965	6.78
83	33.115	5.74
84	33.068	6.37
85	32.983	6.03
87	32.983	6.37
89	33.012	5.73
90	32.926	5.73

The variability of calculated *Symbiodinium* 203 PCP pIs is a direct result of predicted amino acid substitutions. To investigate possible functional significance, a composite of these changes was mapped onto the *A. carterae* PCP apoproteins crystal structure 1PPR. There were 70 sites at which *Symbiodinium* 203 and *A. carterae* 1PPR sequences differed. Among these sites, 43 were fixed substitutions between *Symbiodinium* 203 and *A. carterae* 1PPR apoproteins, and 27 were polymorphic between individual *Symbiodinium* 203 clones. Table 1.8

lists the specific amino acids and types of side chains substituted at each polymorphic site. Ten of the polymorphic sites (positions 6, 24,134,137,182,222, 239,244, 253 and 287) are predicted to accommodate the presence or absence of amino acids with polar side chains.

Table 1.8 Substitutions within polymorphic sites of predicted *Symbiodinium* 203 PCP apoproteins compared to 1PPR amino acids. Sites in bold accommodate the presence or absence of amino acids with polar side chains.

	Symbiodinium 203		*A. carterae* 1PPR			*Symbiodinium* 203		*A. carterae* 1PPR	
Site #	Substitutions	Side Chain	Substitutions	Side Chain	Site #	Substitutions	Side Chain	Substitutions	Side Chain
6	Ala	nonpolar	Ala	nonpolar	172	Ala	nonpolar	Gln	uncharged polar
	Thr	uncharged polar				Val	nonpolar		
24	Asn	uncharged polar	Asn	uncharged polar	**182**	Lys	basic	Lys	basic
	Asp	acidic				Gln	uncharged polar		
39	Glu	acidic	Glu	acidic	217	Met	nonpolar	Ala	nonpolar
	Lys	basic				Val	nonpolar		
51	Gly	nonpolar	Gly	nonpolar	**222**	Ala	nonpolar	Leu	nonpolar
	Glu	acidic				Lys	basic		
96	Ile	nonpolar	Ile	nonpolar		Thr	uncharged polar		
	Val	nonpolar			**239**	Lys	basic	Thr	uncharged polar
101	Glu	acidic	Glu	acidic		Gln	uncharged polar		
	Lys	basic			**244**	Ala	nonpolar	Ala	nonpolar
103	Lys	basic	Met	nonpolar		Lys	basic		
	Arg	basic				Thr	uncharged polar		
105	Met	nonpolar	Met	nonpolar	245	Ala	nonpolar	Ala	nonpolar
	Val	nonpolar				Gly	nonpolar		
107	Val	nonpolar	Val	nonpolar		Asp	acidic		
	Ala	nonpolar			251	Asn	uncharged polar	Asn	uncharged polar
118	Lys	basic	Lys	basic		Tyr	uncharged polar		
	Ile	nonpolar			**253**	Ala	nonpolar	Ala	nonpolar
	Ala	nonpolar				Ser	uncharged polar		
121	Ala	nonpolar	Ala	nonpolar	275	Gly	nonpolar	Gly	nonpolar
	Asp	acidic				Gln	nonpolar		
	Val	nonpolar				Lys	basic		
134	Lys	basic	Lys	basic	276	Val	nonpolar	Val	nonpolar
	Asn	uncharged polar				Ile	nonpolar		
137	Gln	uncharged polar	Glu	acidic	279	Thr	uncharged polar	Thr	uncharged polar
	Lys	basic				Ser	uncharged polar		
154	Ala	nonpolar	Ala	nonpolar	**287**	Ser	uncharged polar	Ser	uncharged polar
	Arg	basic				Pro	nonpolar		

Figure 1.12 is a rendering of all 10 polar substitutions showing their spatial orientations in relation to the 8 peridinins within 1PPR. Thr118, Ser253 and Ser287 are of particular interest because of their proximities to peridinin polyene chains and furanic rings and possible influence on spectral tuning of these chromophores. The polar side chain of Thr118 is 7.19 Å from PID612 (Figure

1.13). Ser253 is 10.06 Å from PID624 (Figure 1.14). Ser287 is from 8.99 Å of PID622 and 9.15 Å from PID621 (Figure 1.15).

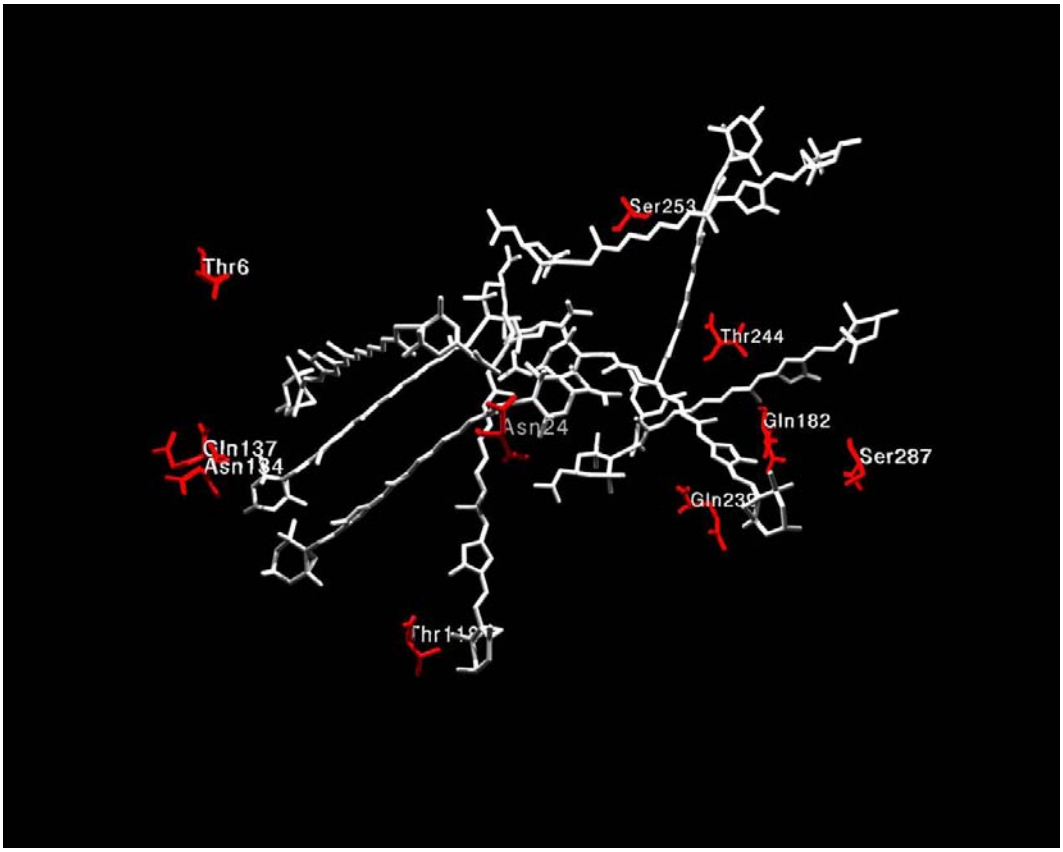


Figure 1.12 Substitutions of amino acids with polar side chains occurring at polymorphic sites within *Symbiodinium* 203 predicted apoproteins show in relation to peridinins.

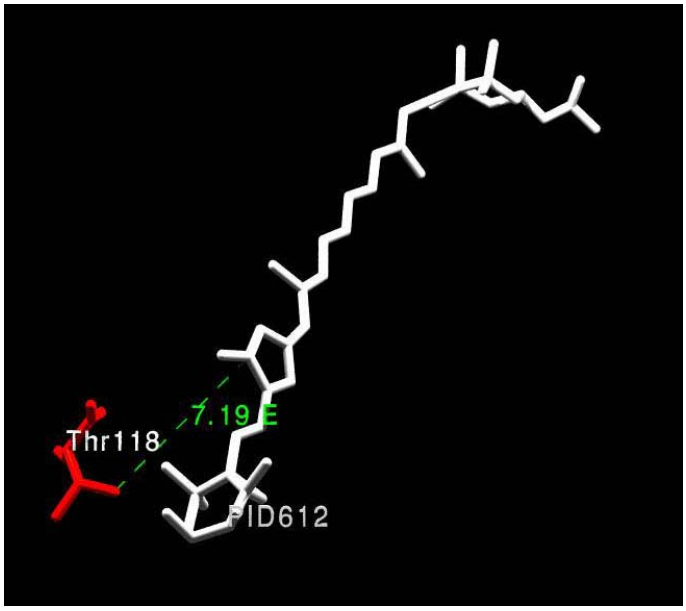


Figure 1.13 Distance (E=Å) between the polar side chain of threonine 118 and the furanic ring of peridinin 612.

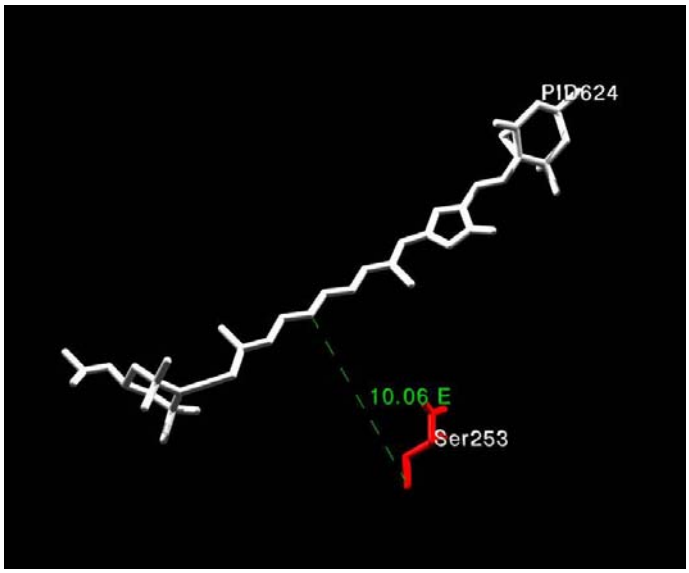


Figure 1.14 Distance (E=Å) between the polar side chain of serine 253 and the polyene chain of peridinin 624.

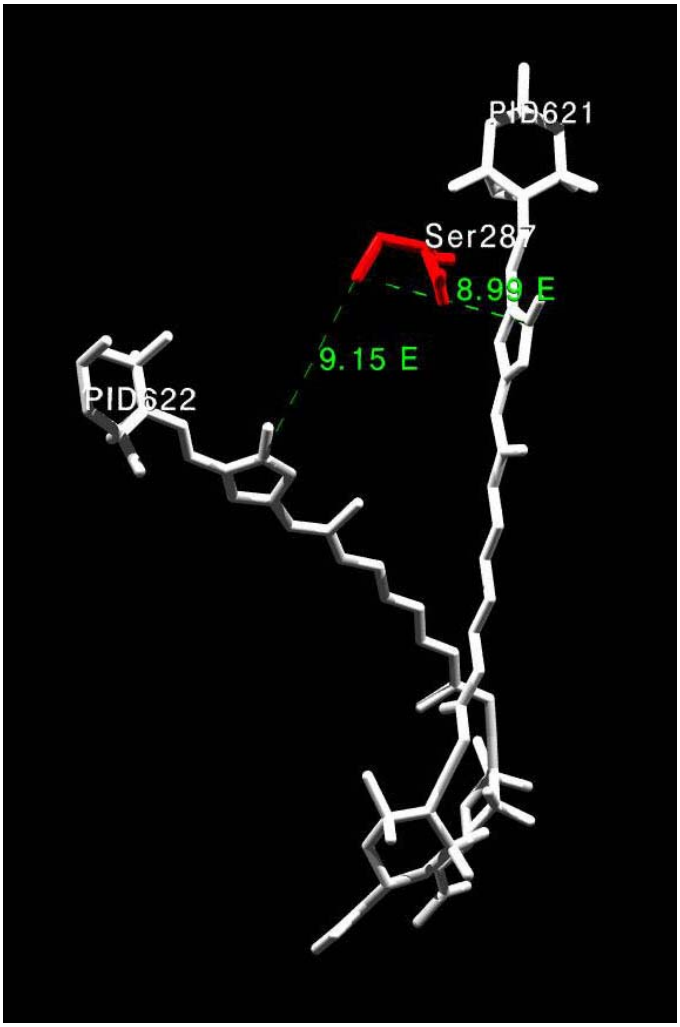


Figure 1.15 Distances (E=Å) between the polar side chain of serine 287 and the furanic rings of peridinins 621 and 622.

DISCUSSION

PCP gene organization, diversity and family size

This is the first characterization of a PCP gene family from any dinoflagellate in the genus *Symbiodinium.* Earlier descriptions of the organization of PCP genes came from free-living species A. *carterae* (Sharples et al. 1996), *L. polyedra* (Le et al. 1997) and *H. pygmaea* (Hiller et al. 2001). There are only 2 previous reports of *Symbiodinium* PCP nucleotide sequences and both were single cDNAs (Norris and Miller, 1994; Weis et al. 2002). The results of these molecular experiments demonstrate that the PCP genes of *Symbiodinium* 203 are organized essentially like those of A. *carterae* and *L. polyedra* in that the genes are arranged in tandem arrays, the coding regions code for long PCP polypeptides and the coding regions are intronless. The coding regions of *Symbiodinium* 203 are very similar in size to those from A. *carterae* and *Symbiodinium* from *A. formosa* yet shorter than *L. polyedra.* The *Symbiodinium* 203 PCP cassette has untranscribed spacers of variable sizes that are smaller than those from *L. polyedra*, but longer than *H. pygmaea.* Li and Hastings (1998) identified a likely promoter sequence from *L. polyedra* luciferase and PCP spacers. The putative promoter sequence that we found in the same relative location upstream of *Symbiodinium* 203 coding regions was 69% identical to the *L. polyedra* promoter. Dinoflagellate promoters do not fit into the common motifs used by other eukaryotes, and dinoflagellate promoters may be genus or species specific.

There is far greater diversity in the coding regions of *Symbiodinium* 203 PCP genes than previously described from any other dinoflagellate species. I have identified ten unique complete coding sequences and an additional five unique partial cDNA sequences. The majority of positive clones screened from genomic and cDNA PCR libraries were distinct however, redundant sequences were also found. For reasons of practicality, the searches were not exhaustive, so the limit on the number of unique PCP genes in *Symbiodinium* 203 may not been reached. The low Ka/Ks ratios for paired comparison between coding sequences from ten clones indicate that the majority of nucleotide substitutions occur at synonymous codon positions and that there is not a clear signal of entire coding regions being selected for within the *Symbiodinium* 203 PCP gene family. Although not shown here, the same is true when these sequences are compared to *A. carterae* or *Symbiodinium* from *A. formosa* coding sequences. This does not exclude the possibility that specific sites within PCP genes could be under selection. When PCP gene sequences are described for more species, Bayesian and likelihood analyses of PCP gene phylogenies may detect such sites.

As it became apparent that very few of coding sequences were identical, there was concern that at least part of the variation that was being observed was the result of recombination within the PCR reactions (Bradley and Hillis, 1997). When amplifying genes from within tandem arrays it is possible to generate incomplete extensions products that can act as primers and anneal to various locations in the array resulting in PCR fragments whose sequences can be different from those actually present in the genomic DNA. The results of the

recombination experiment showed that 8.5% of subclones had different 5' and 3' identities with all recombinations apparently occurring in spacer regions. Early termination of PCR extensions in the spacers may be attributable to formation of secondary structure during the annealing phase of the reactions. Conclusions to be drawn from this type of control to detect recombination are limited in a few regards. Reamplification of cloned DNA can be different than amplifying genomic regions. An increase in the number of unique templates available within each reaction could increase recombination frequency. And this type of analysis only detects recombinants that are observable by comparing coding sequence in subclones to coding sequence in the original templates. Assuming that under our PCR conditions early extension termination is an artifact of amplifying across spacers, it is possible that clone 41 is a chimera, but it is less likely that the complete coding sequence clones of 79-90 or cDNA clones 34-59 are recombinants. Another source or artificial substitution is through *Taq* polymerase errors. The *Taq* polymerase fidelity experiment did not detect any introduced changes in subclones of reamplified clone 79 further suggesting that the coding sequence differences reported here are reliable.

Le et al. (1997) reported that *L. polyedra* had a PCP gene family of roughly 5000 copies per 200 pg genome, which is one of the largest gene families for any organism. It should be noted that the *L. polyedra* genome is also unusually large compared to many other dinoflagellates and other eukaryotes. For perspective, human haploid cells typically have 3.2 pg of DNA per nucleus (Holm-Hanson, 1969). The Le et al. copy number was based on intensity of

hybridization signals of slot blots. I sought to further refine the methods of estimating PCP gene family size in this investigation. Quantitative Real-Time PCR is an extremely accurate and reliable way of determining the amount of starting template within amplification reactions and was used here to give a mean the number of PCP genes per pg of genomic DNA for *Symbiodinium* 203. Flow cytometry is now routinely used to measure various cellular parameters including DNA content and offers the advantage of being able to quickly collect data on large numbers of cells. Fluorescence is typically standardized to chick red blood cells (CRBCs), chick erythrocyte nuclei (CENs) or synthetic beads. When Veldhuis et al. (1997) quantified the DNA content of 121 strains of marine phytoplankton by flow cytometry, they pointed out that the staining of nuclei in whole cells with cell walls was not directly comparable to CRBCs. I chose not to standardize to CRBCs or CENs, but rather to make a relative comparison of the mean genome sizes of the *Symbiodinium* 203 and *L. polyedra*. *L. polyedra* cells are less than perfect standards for this purpose because they have armored theca while *Symbiodinium* 203 cells do not. In addition, the genome size of *L. polyedra* is substantially bigger than *Symbiodinium* 203. Nevertheless, comparing fluorescence of stained DNA between these two species is more valid than comparing either to the commercially available standards which stain much more readily than intact algal cells. The combined results of our quantitative real-time PCR and flow cytometry experiments show that *Symbiodinium* 203 has 36 ± 12 PCP genes per 3 ± 1 pg genome. This gene family size is much closer to the 50 PCP genes per genome for *H. pygmaea* suggested by Hiller et al. (2001) than to

L. polyedra. The PCP gene copy number estimates for all three species rely in one way or the other to the Holm-Hansen (1969) algal DNA content values based on amount of nuclear organic carbon as standards. As the DNA content of dinoflagellates becomes more accurately measured with newer techniques, the current PCP gene family sizes may be revised. Beyond the absolute differences in PCP gene family sizes, *Symbiodinium* 203 has proportionally fewer PCP genes per pg of genomic DNA than *L. polyedra.* There may be selection in symbiotic dinoflagellates toward smaller allocation of genomes to PCP gene families compared to free-living species. The test of this hypothesis awaits the characterization of PCP gene families from several more species of each type.

Low levels of concerted evolution in PCP gene family

Symbiodinium 203 has a relatively small PCP gene family that contains numerous divergent coding sequences. Despite the fact that these sequences bear greater resemblance to each other than they do to other published long PCP genes, I found little evidence that *Symbiodinium* 203 PCP genes are evolving in concert. Concerted evolution of tandemly repeated sequences such as the genes for ribosomal RNA, transfer RNA and histones is usually explained in terms of two mechanisms; continual expansion and contraction due to unequal crossing over and biased gene conversion. The relative importance of one or the other mechanism varies depending on which organisms are studied (see Hillis et al. 1991). Both rely on duplex formation between homologous loci as most often occurs in meiosis. While sexual reproduction has been reported in other dinoflagellate genera (Pfiester, 1984), there is an apparent absence of a haploid

sexual phase in *Symbiodinium* (Schoenberg and Trench, 1980a, Trench, 1993). However, Baillie et al. (2000) reported that RAPD variation in *Symbiodinium* from giant clams was comparable to patterns seen in dinoflagellates undergoing sexual recombination. Lack of heterogeneity and ease of direct sequencing of amplified large subunit ribosomal genes from various *Symbiodinium* species suggest that concerted evolution does occur within *Symbiodinium* genomes (see Wilcox, 1998). It is unclear if this is accomplished in the formation of cryptic haploid sexual cells, through mitotic crossing over or other as of yet described mechanisms particular to dinoflagellate genomes. Nevertheless, the current results strongly suggest that PCP genes are handled differently than ribosomal genes. PCP genes may contain variable sites into which mutations are introduced faster than they can be removed by homogenizing mechanisms. Another possibility is that PCP genes may reside in regions of dinoflagellate genomes where such mechanisms occur less often. The net effect is that PCP genes are evolving under little or no concerted evolution. PCP isoform diversity may be important in broadening the range of light wavelength that can be harvested by dinoflagellates for photosynthesis.

Affects of genetic diversity on predicted PCP apoproteins

While there are no reports of empirically determined PCP isoforms for *Symbiodinium* 203, translations of coding sequences cloned in this project demonstrate that there is sufficient genetic diversity to account for a suite of PCP apoprotein pIs comparable to those found in several other *Symbiodinium* species. This does not rule out the possibility that post-translational modification of PCP

polypeptides could still occur. The results suggest that post-translational modifications are not necessary to explain the multiple PCP isoforms. This point could be tested by additional PCR based characterizing PCP coding sequences of species for which PCP pIs are known.

The potential for examining functionally significant differences between predicted PCP apoproteins was greatly enhanced by mapping amino acid substitutions on to the *A. carterae* 1PPR crystal structure. Collectively, the *Symbiodinium* 203 polypeptides differed from *A. carterae* 1PPR at 70 out of 312 sites within each monomer. The 43 fixed substitutions were distributed through all domains of 1PPR including regions near adjacent monomers, the hydrophilic exterior and chromophores. The majority of these changes that faced hydrophobic interior of monomers did not affect the polarity of this environment. However, they could be important in giving *Symbiodinium* 203 PCP holoproteins a different conformation than 1PPR. Ten of the 27 polymorphic sites hold amino acids with or without polar side chains. The apoproteins from clones 79 and 83 contain polar substitutions with OH groups in positions that could influence the spectral tuning of nearby peridinins. Predicting the direction and magnitude of overall changes to the spectroscopic properties of holoproteins is beyond the scope of this investigation. Distances calculated by Swiss-PDB Viewer v3.7(b2) between polar side chains and peridinins could shift for any individual apoprotein if only substitutions from a single clone were introduced. Rendering of new tertiary structures using more sophisticated software could also improve accuracy of their physical arrangement.

CONCLUSION

Several points regarding the PCP gene family of *Symbiodinium* 203 have been established in this chapter. *Symbiodinium* 203 has long PCP genes organized like those of *A. carterae* and *L. polyedra*, but with a putative promoter that is different from *L. polyedra*. There are at least 14 distinct coding regions out of 36 ± 12 PCP genes in this family. Diversity of *Symbiodinium* 203's PCP gene family appears to be consequence of low levels of concerted evolution and acts as a primary source of variability in PCP isoforms. Amino acid substitutions in *Symbiodinium* 203's PCP apoproteins result in shifts of isoelectric points and probably influence light harvesting of holoproteins. Heterogeneity in dinoflagellate PCP gene families may provide a selective advantage as means of broadening the range of wavelengths of light that can be captured for photosynthesis.

CHAPTER TWO

Analyses of PCP genes and predicted proteins from *S. pilosum, Symbiodinium sp*. from *Dichocoenia stokesii, S. pulchrorum,* and *S. kawagutii*: Diverse PCP gene families occur in all major clades of *Symbiodinium* and in both size classes of the gene

ABSTRACT

Multiple PCP coding regions were cloned and sequenced from *S. pilosum* (ls rDNA clade A)*, Symbiodinium sp.* from *D. stokesii* (clade B)*, S. pulchrorum* (clade B) and *S. kawagutii* (clade C). *S. pilosum* has small PCP genes while the other species in this study have the large variety. In each case, the pattern of PCP gene diversity was similar to that reported for *Symbiodinium* 203 in Chapter One. These results demonstrate that diverse PCP gene families are a common feature of *Symbiodinium* species dinoflagellates, and that both size classes of PCP genes appear to be evolving under low levels of concerted evolution. PCP pIs predicted from *S. pilosum* and *S. kawagutii* complete coding sequences match data previously published from analyses of their apoproteins. This reinforces a conclusion from Chapter One that genetic diversity is the primary source of PCP isoform variation. Both *S. pilosum* and *S. kawagutii* PCP apoproteins contain polymorphic sites that are likely to influence the spectral tuning of peridinins.

INTRODUCTION

Chapter One established the structure, organization and size of the PCP gene family of *Symbiodinium* 203. This gene family was shown to be highly

diversified and capable of expressing distinct 33 kDa apoproteins with a range of pIs comparable to those previously reported for several other species of *Symbiodinium*. These results agree with previous reports that PCP gene families are probably diversified (Chang and Trench, 1984; Triplett et al. 1993; Sharples et al. 1996; Hiller et al. 2001) rather than highly conserved (Le et al. 1997). Furthermore, they support the hypothesis that genetic diversity, rather than post-translational modification, is the primary source of PCP isoform variation. The results from Chapter One also revealed possible functional significance of isoform variation by demonstrating that substitutions of amino acids with polar side chain may effect spectral tuning of peridinins within PCP holoproteins (see Hofmann et al. 1996). By expressing multiple isoforms each absorbing light at subtly different maxima, dinoflagellates effectively harvest light from a broader range of light wavelengths than they would otherwise do with only a single PCP isoform. It is unclear if the diversity at the protein level is adaptive. It is also unknown if it being maintained by positive selection or if it is simply a tolerable byproduct of genetic polymorphism.

Groundwork on *Symbiodinium* 203 provided the basis for the study of PCP genes and their predicted proteins in other *Symbiodinium* species. PCP gene sequences described here for four additional species will then be combined with *Symbiodinium* 203 sequences and previously published data for a comparative analysis of the evolution of PCP genes in Chapter Three. The current investigation builds on the detailed characterization of *Symbiodinium* 203 PCP genes to specifically address the following questions about PCP genes in other

symbiotic dinoflagellates. Do *Symbiodinium* species from each of the major ribosomal clades have diversified PCP gene families? Does this diversification extend to small and large PCP genes? Do predicted PCP pIs for *Symbiodinium* species match results of isoelectric focusing experiments on apoproteins from those same organisms (see Trench and Blank, 1987)? Do the predicted PCP apoproteins of other *Symbiodinium* species besides 203 also contain polymorphic sites that could effect the spectral tuning of peridinins?

To answer these questions, analyses of PCP genes and predicted proteins from *S. pilosum, Symbiodinium sp.* from *D. stokesii, S. pulchrorum,* and *S. kawagutii* are presented here. This group of symbiotic dinoflagellates was selected because it contains representatives from ls rDNA clades A, B, and C with either large or small PCP genes, and species for which the actual isoelectric forms were already known.

MATERIALS AND METHODS

Algal Cultures

The four species of *Symbiodinium* used in this investigation are described in Table 2.1. Robert K. Trench (University of California, Santa Barbara) donated three starter cultures of *S. pilosum, S. pulchrorum and S. kawagutii.* These are endosymbionts from a Hawaiian stony coral, a Hawaiian anemone, and a Jamaican zoanthid respectively.

I isolated *Symbiodinium sp.* from the colonies of the stony coral *Dichocoenia stokesii* collected at Lee Stocking Island, Bahamas. My cultures 28,

32 and 35 were established from separate *D. stokesii* colonies located at depths of 2 m, 15 m and 34 m. Algal cells were extracted by modification of standard methods. A WaterPic was used to remove tissue from coral skeletons. Slurries were filtered through sterile gauze to remove large debris and algal cells were pelleted by centrifugation. The cells were resuspended in distilled water to lyse ciliate contaminates, pelleted again and resuspended in 10 ml of Provasoli's Enriched Seawater plus antibiotics. Cultures were grown for two days under broad-spectrum fluorescent lights on a 12:12 light to dark cycle at 27°C. Motile cells from these cultures were plated on to agar plates containing PES and antibiotics and were kept at the same lighting and temperature conditions. Algal colonies were picked from these plates after two weeks and used to establish unialgal cultures in 10 ml of 0.45 μm filtered Gulliard's F/2. Isolates 28, 32 and 35 were indistinguishable from each other in terms of behavior in culture, external morphology and ls rDNA sequence (data not shown). They are presumably the same dinoflagellate species.

All *Symbiodinium* cultures were subsequently maintained under conditions described above and were serially transferred every 3-4 weeks in 10 ml volumes of filtered Gulliard's F/2.

Table 2.1 Algal cultures

Symbiodinium Culture	Host	Isolate	ls rDNA clade	PCP apoprotein size
S. pilosum	*Zoanthus sociatus* (Zoantharia)	RK Trench #185	A	17kDa
Symbiodinium sp.	*Dichocoenia stokesii* (Scleractinaria)	JR Reichman #28, 32& 35	B	undescribed
S. pulchrorum	*Aiptasia pulchella* (Actinaria)	RK Trench #8	B	35 kDa
S. kawgutii	*Montipora verrucosa* (Scleractinaria)	RK Trench #135	C	35 kDa

DNA Extractions

Genomic DNA was extracted from pelleted log growth phase algal cells using the modification to the Qiagen DNAeasy kit protocol described in Chapter One in which steps one and two were replaced with CTAB resuspension and grinding followed by proteinase K digestion. Extraction yields were verified on ethidium bromide stained agarose gels. Genomic DNA was resuspended in Qiagen EB (0.01 M Tris buffer without EDTA).

PCR Primers and Conditions

The primers used to amplify and sequence PCP genes from each of the four *Symbiodinium* species are listed in Table 2.2. The sPCP-F1/sPCP-R3 primer sequences were provided by V.M. Weis (Weis et al. 2002). The U(448)/L(423) and U(-28)/L(1180) primers set were used in Chapter One on *Symbiodinium* 203. PCR reactions mixes contained the same components in the same concentrations used to amplify PCP genes from *Symbiodinium* 203 in Chapter One. Other primers were designed with Oligo 4.0 (National Biosciences, Inc.) based preliminary PCP gene sequences of the species in this study. Figure 2.1 shows relative primer locations.

Table 2.2 PCR and sequencing primers

Upper/forward and Lower/reverse primers are listed sequentially in the pairs that were most often used for amplifications. Numbers in parentheses represent arbitriary location within contigs. sPCP-F1 and sPCP-R3 primers were originally designed for *S. muscatinei* (Weis et al.,2002).

Upper Primers	Lower Primers	Primer Sequence (5' - 3')	Genomic DNA Applied to
sPCP-F1		CGCTTCAAAGACCATGGATG	*S. pilosum*
	sPCP-R3	TAACGCTGGGATGCTTTGAC	"
U(112)		AAGGTTAAACTTGGCGAACG	"
	L(36)	AGCTTACAAGGCACTCATGG	"
U(758)		GGCTTGAGGTCAGACACATCCAGGC	"
	L(189)	GCTTGCCGCACACGTTTAGAAAATG	"
U(448)		TCGGTCCCCAAAGCAAAGGTCA	*S. sp.* & *S. pulchrorum*
	L(423)	CATTCACGGCATCCCAGTCAGC	"
U(182)		GGCATGATGTGGCTAGACGA	"
	L(1315)	TGTATGCAATTAAAAAAGTATG	"
U(-28)		TCCGGCCCACTTTTAGTTTT	*S. kawagutii*
	L(1180)	TTTTCCCATTTGTTTCAGAG	"

Cloning & Screening Plasmid Libraries

PCR products of interest were gel excised, purified with QIAEX II kits (Qiagen, Inc.) and cloned into pCR2.1 vectors with TA Cloning kits (Invitrogen, Inc.). Blue/white screens were performed on transformant INVαF' bacterial colonies. PCR screens of all plasmid libraries were done using the same protocol described on Chapter 1. Positive PCP gene clones were verified by separate amplification and sequencing of inserts.

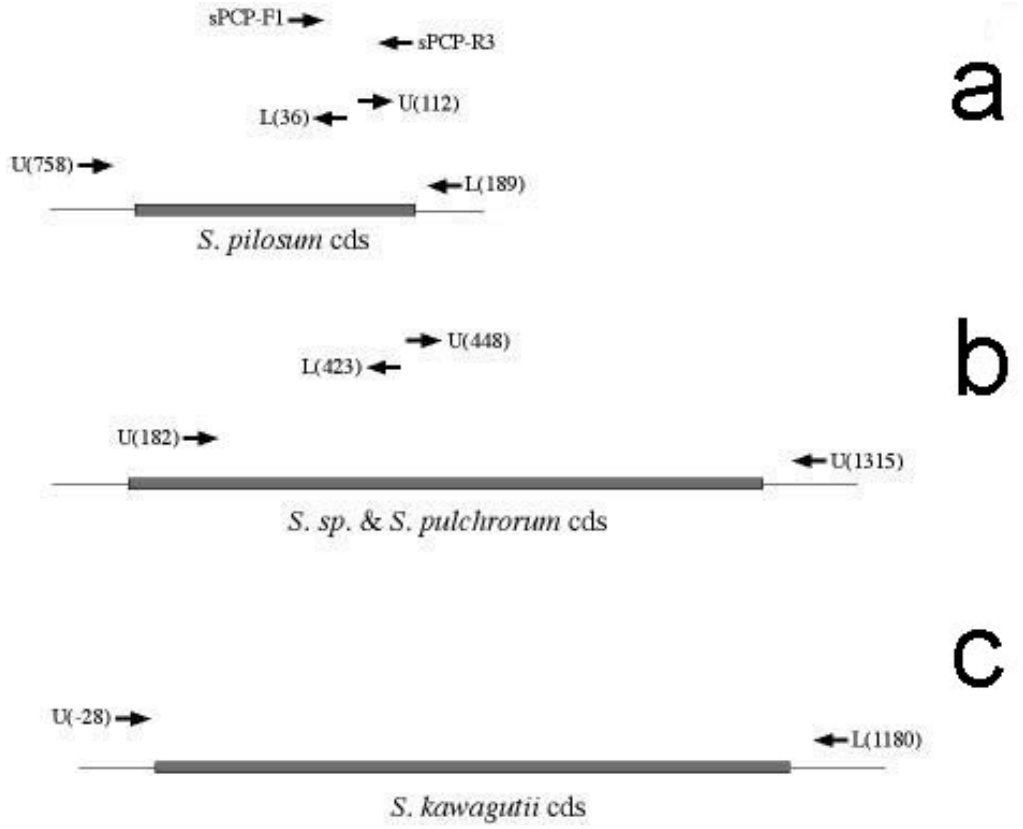


Figure 2.1 Primer map**; (a)** *S. pilosum* amplifications sequentially consisted of a 3' segment (sPCP-F1/sPCP-R3), the 3' and 5' region from 2 adjacent coding genes (U(112)/L(36)) and the entire cds (U(758)/L(l89); **(b)** For *Symbiodinium sp.* from *D. stokesii* and *S. pulchrorum*, outward facing primers (U(448)/L(423)) amplified between adjacent genes that lead to inward facing primers (U(182)/L(1315)) producing the majority of the cds; **(c)** Complete *S. kawagutii* cds were amplified with *Symbiodinium* 203 primers (U(-28)/L(1180)).

Sequencing & Sequence Analysis

The nucleotide sequences for all clones were determined by automated sequencing with BigDye V2 Terminator mix (Applera, Inc.) using the same half reaction volumes described in Chapter One. Sequence data were collected on either Perkin Elmer ABI PRISM 377 or 3100 DNA Sequencers. Sequence contigs were assembled with Seqman (DNA Star, Inc). All sequences were verified by sequencing twice and in most cases by sequencing in both directions. Sequences of clones were submitted to GenBank (accession numbers listed in Table 2.3). Translation of predicted proteins and pI estimation with done with Edit Seq (DNA Star Inc.).

Amino Acid Substitution Modeling

The predicted 16 kDa PCP apoprotein structure for *S. pilosum* was rendered by submitting the mature polypeptide sequence for a single clone to the Swiss Model server (Peitsch and Guex, 1997) with coordinates of the *A. carterae* 1PPR (monomer M) C-terminate domain as the template. The structure was imported into Swiss-PDB Viewer v3.7(b2) (Glaxo Wellcome, Inc.) where chromophores were added and merged into a single pdb layer. Polymorphic sites were mapped and distances from polar side chain of amino acids in polymorphic sites were measured in relation to furanic rings and polyene chains of the nearest peridinin molecules as was previously done with *Symbiodinium* 203.

The *S. kawagutii* 35 kDa predicted apoprotein was rendered using the methods applied to *Symbiodinium* 203 in Chapter One. Likewise, mapped polymorphic sites of *S. kawagutii* PCPs were converted to a single layer pdb file containing the substitutions prior to rendering graphics. For both *S. pilosum* and *S. kawagutii* 3D protein structure graphics were rendered using POV- Ray v3.1 (POV-Ray Team).

Table 2.3 PCP clone accession numbers for *S. pilosum*, *Symbiodinium sp*. from *D. stokesii*, *S. pulchrorum* and *S. kawagutii*.

Species	Clone Name	Accession Number
S. pilosum	185clone23	AY149140
"	185clone25	AY149141
"	185clone28	AY149142
"	185clone30	AY149143
"	185clone32	AY149144
"	185clone33	AY149145
"	185clone34	AY149146
"	185clone35	AY149147
Symbiodinium sp. from *D. stokesii*	28clone1	AY149148
"	28clone2	AY149149
"	28clone5	AY149150
"	32clone1	AY149151
"	35clone3	AY149152
"	35clone42	AY149153
"	35clone43	AY149154
"	35clone44	AY149155
"	35clone46	AY149156
"	35clone47	AY149157
"	35clone49	AY149158
S. pulchrorum	8clone1	AY149159
"	8clone2	AY149160
"	8clone3	AY149161
"	8clone4	AY149162
S. kawgutii	135clone37	AY149163
"	135clone38	AY149164
"	135clone41	AY149165
"	135clone42	AY149166
"	135clone43	AY149167
"	135clone49	AY149168
"	135clone50	AY149169
"	135clone52	AY149170
"	135clone54	AY149171
"	135clone60	AY149172

RESULTS

Multiple PCP coding regions cloned from each Symbiodinium species

Primers that were previously designed for *Symbiodinium* 203 did not amplify PCP genes from *S. pilosum*. This is not surprising and is attributable to two things. First, *S. pilosum* has small PCP genes unlike *Symbiodinium* 203. Hiller et al. (2001) and Weis et al. (2002) demonstrated that small PCPs were most homologous to the C-terminal end of the large PCPs from *A. carterae*, *L. polyedra* and *Symbiodinium sp.* from *A. formosa*. The same homology applies to small and large PCP genes nucleotide sequences. The outward facing U(448)/L(423) primer set is nested in the N-terminal end of the *Symbiodinium* 203 coding region, and is not compatible with *S. pilosum*. Second, the *Symbiodinium* 203 U(-28)/L(1180) primers were designed to flank the coding region in positions just upstream and downstream in untranscribed spacers. Spacers evolve more rapidly than coding regions and may be highly divergent from species to species. While these primers were not useful on *S. pilosum*, the general cloning strategy used for *Symbiodinium* 203 was successfully applied here again. *S. pilosum* PCP coding regions were cloned by initially amplifying a small 3' segment using the sPCP-F1/sPCP-R3 primer set designed by VM Weis. The sequences from two clones of this region (data not shown) were used to design outward facing primers U(112)/L(36). This set amplified fragments approximately 1.2 kb in length between adjacent PCP coding regions containing roughly 500 bp spacers sequences and demonstrated the size of the *S. pilosum*

small PCP gene cassette. Sequences from four of these clones (data not shown) were used to design inward facing primers U(758)/L(189) that amplified entire coding regions. As with *Symbiodinium* 203, multiple coding regions were cloned and sequenced. After screening 20 complete cds clones, 8 unique sequences (clones 23,25,28,30 and 32-35) were found. The nucleotide sequences for these clones are aligned in Appendix 2.1.

Comparison of the unique *S. pilosum* PCP gene clones to other sequences in GenBank via BLAST search indicated that they were most similar to the *S. muscatinei* PCP gene (AF425735). While *S. pilosum* small PCP genes are roughly half the size of the 1098 bp versions found in *Symbiodinium* 203, *S. pilosum* PCP genes contain proportionally fewer polymorphic sites. Across the 615 bp cds there were 36 polymorphic sites compared to 83 sites in *Symbiodinium* 203. The nucleotide sequences of 8 unique *S. pilosum* coding regions also shared from 96.6% to 99.7% identity with each other.

The U(448)/L(423) primers did amplify 2.3 kb fragments between adjacent PCP coding regions from both *Symbiodinium sp*. from *D. stokesii* and *S. pulchrorum*, compared to 1.9 kb in *Symbiodinium* 203. Govind et al. (1990) reported that *S. pulchrorum* expressed 35 kDa PCP apoproteins, so it is reasonable to assume that that it's coding regions are approximately the same size (1.1 kb) as other large PCP genes including those from *Symbiodinium* 203. Both clade B species in this study had larger untranscribed spacers (1.2kb) than either *Symbiodinium* 203 or *S. pilosum*. As before, the objective of beginning with outward facing primers was to identify the 5' and 3' ends of the genes and to

design flanking inward facing primers that would amplify entire coding regions. Although a functional primer downstream of the stop codon was designed, L(1315), none of the primers intended to anneal upstream of the start codon worked reliably for amplifying and sequencing. These difficulties appear to be the result of false priming sites and formation of primer dimers. The U(182)/L(1315) combination amplified the majority of the coding regions from *Symbiodinium sp*. from *D. stokesii* and was the source of 28clone5 through 35clone49. *Symbiodinium sp*. 28clones1&2 and all *S. pulchrorum* clones were amplified with U(448)/L(1315). Eleven unique partial PCP gene fragments were cloned from *Symbiodinium sp*. from *D. stokesii* and four from *S. pulchrorum*. An alignment of these nucleotide sequences is presented in Appendix 2.2. Because each of the clones from clade B species contained partial cds, the total number of polymorphic sites could not be determined. Nevertheless, the same pattern of genetic diversity found in the PCP gene families of *Symbiodinium* 203 and *S. pilosum* was clearly present in *Symbiodinium sp*. from *D. stokesii* and *S. pulchrorum* with eleven and four distinct sequences respectively.

Complete coding regions from *S. kawagutii* were amplified with little difficulty using the *Symbiodinium* 203 flanking primers U(-28)/L(1180). *S. kawagutii* is a large subunit rDNA clade C species and is more closely related to *Symbiodinium* 203 than the other species in this report. Of the 20 clones screened, 10 unique sequences were detected, 135clones 37, 38, 41, 42, 43, 49, 50, 52, 54 and 60. An alignment of these nucleotide sequences is present in Appendix 2.3. Nine coding regions were 1098 bp long as were those from *Symbiodinium* 203

and *Symbiodinium sp* from *A. formosa* (Norris and Miller, 1994; accession L13615). The coding region in 135clone42 was 6 bp longer. Minor differences in the lengths of PCP coding regions within the same species have been previously been noted for *A. carterae* (Sharples et al. 1996) and for *Symbiodinium* 203 in Chapter One. Although generally the same size as *Symbiodinium* 203 PCP genes, the *S. kawagutii* genes were substantially more polymorphic with 138 variable nucleotide sites. The coding regions shared from 95.7% to 98.4% identity with each other.

Predicted amino acid sequences, isoelectric focusing points and protein modeling for S. pilosum PCPs

The predicted PCP amino acid sequences for *S. pilosum* are shown in as an alignment in Appendix 2.4. The small PCPs from *S. pilosum* are structured essentially like those of *S. muscatinei* and *H. pygmaea*. Each *S. pilosum* clone codes for a 21 kDa PCP preprotein consisting of a 54 aa transit peptide plus a 150 aa 16 kDa mature apoprotein. The *S. pilosum* transit peptides are 6 aa longer than those identified for the *S. muscatinei* PCP (Weis et al. 2002) and 3 aa longer than those from *H. pygmaea* (Hiller et al., 2001), yet the transit peptides from all three species share common elements. Each contains an N-terminal positively charged sequence followed by two hydrophobic domains that are separated by a another positively charged residue containing the general motif (F/H)V(+/-P)GP(+/-G)R. In *S. pilosum*, this region is conserved as FVPGPGR. The transit peptides all end at a predicted cleavage point with the motif A(Y/H)A. *S. pilosum* and *S. muscatinei* end with AYA while *H. pygmaea* ends with AHA.

S. pilosum, like *S. muscatinei* and *H. pygmaea*, produces 150 aa apoproteins that show sequence homology with the C-terminal domains of the large PCP apoproteins in GenBank. All apoproteins had isoelectric focusing points in the basic range from pH 7.28 to 8.77 (see Table 2.4).

The range and specific pIs values calculated here correspond with data previously published by Trench and Blank (1987) on *S. pilosum* PCPs (Figure 2.2). This strongly supports the hypothesis that genetic diversity among PCP genes contributes to the generation of PCP isoforms.

Table 2.4 Calculated mass and isoelectric points from *S. pilosum* apoproteins

S. pilosum clone	Apoprotein mass (kDa)	Isoelectric point (pI)
23	16.056	7.28
25	16.01	8.77
28	16.056	7.28
30	16.075	8.27
31	16.056	7.28
33	16.066	7.28
34	16.07	7.28
35	16	7.28

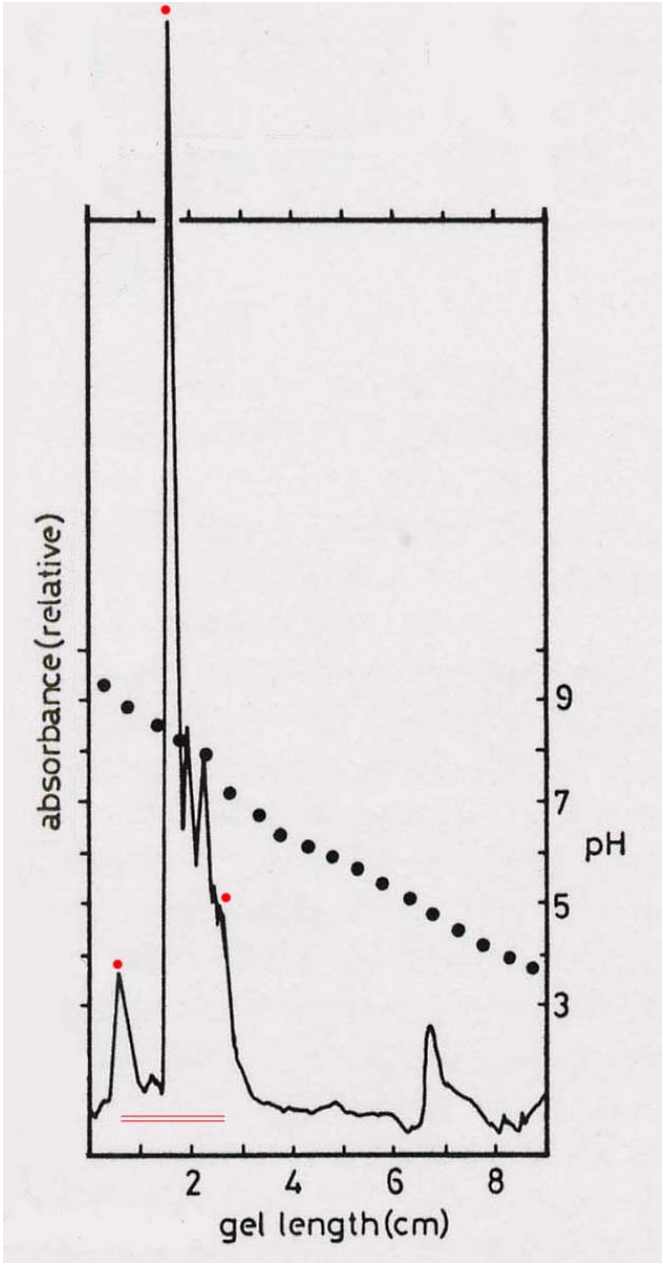


Figure 2.2 Isoelectric focusing patterns from *S. pilosum*. The red double line indicates the range of calculated pIs for S. pilosum apoproteins, and the red dots show matches between the calculated and actual pIs (After figure 1d, Trench and Blank, 1987).

As with the apoproteins of *Symbiodinium* 203 in Chapter One, the differences in pI for *S. pilosum* PCPs are the direct result of amino acid substitution at polymorphic sites. The predicted apoproteins from *S. pilosum* are polymorphic at nine sites. Three of these sites (positions 33, 100, and 113) accommodate the presence or absence of amino acids with polar side chains (Table 2.5). As with the nucleotide sequences, amino acid polymorphism among predicted *S. pilosum* PCPs was proportionally less than in *Symbiodinium* 203.

Table 2.5 Substitutions within polymorphic sites of predicted *S. pilosum* PCP apoproteins. Bold signifies polarity polymorphic sites.

Site #	Substitutions	Side Chain	Site #	Substitutions	Side Chain
11	Lys	basic	92	His	basic
	Pro	nonpolar		Arg	basic
23	Trp	nonpolar	**100**	Ser	uncharged polar
	Arg	basic		Pro	basic
29	Ala	nonpolar	**113**	Asn	uncharged polar
	Gly	nonpolar		Lys	basic
33	Thr	uncharged polar	145	Val	nonpolar
	Ala	nonpolar		Ile	nonpolar
58	Leu	nonpolar			
	Pro	nonpolar			

In Chapter One, substitutions at polarity polymorphic sites of *Symbiodinium* 203 PCPs were shown to potentially influence the spectral tuning of peridinins. This was accomplished by mapping the amino acids substitutions into the *A. carterae* 1PPR crystal structure and measuring the distance between polar side chains and specific regions of peridinins. However, mapping *S. pilosum* PCP substitutions is more complicated because no crystal structure for a

small PCP is currently available. To get a reasonable approximation of *S. pilosum* PCP conformation, the C-terminal half of 1PPR was used as a template and a structure was generated through the Swiss-Prot server as described by Hiller et al. (2001). Previous characterization of the 16 kDa PCPs from *S. pilosum* by Govind et al. (1990) indicated that they existed only as monomers and not as hexadimers suggested by Hiller et al. (2001) for *H. pygmaea* PCPs. The first approximation monomer is shown in Figure 2.3a. After refining the structure by the addition of chromophores and rotomer optimization (Figure 2.3b), amino acid substitutions were mapped and the locations of the 3 polarity polymorphic sites were determined in relation to four peridinins (Figure 2.4).

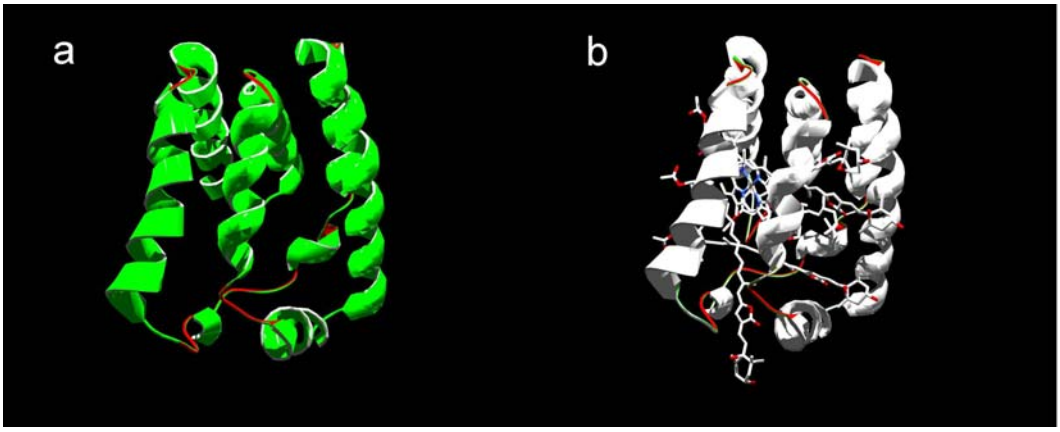


Figure 2.3 *S. pilosum* 16 kDa PCP monomer; **(a)** first approximation apoprotein; **(b)** refined holoprotein.

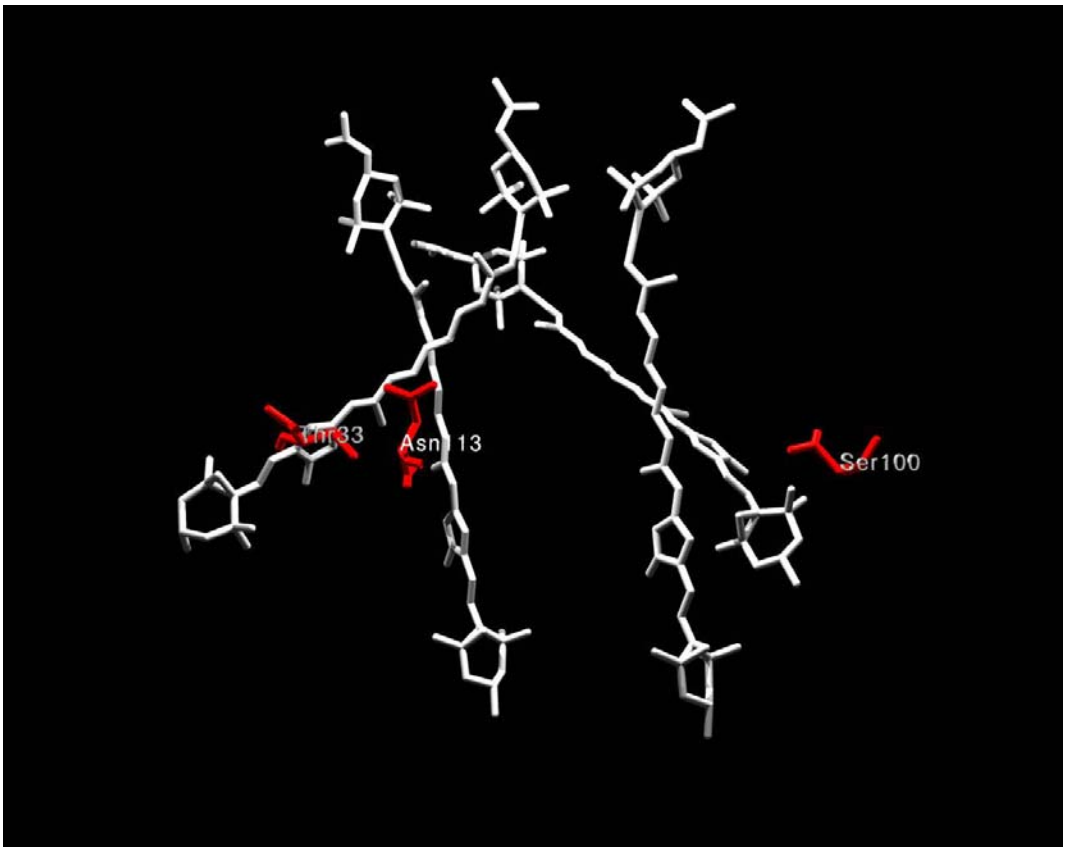


Figure 2.4 Substitutions of amino acids with polar side chains occurring at polymorphic sites within *S. pilosum* predicted PCP apoproteins shown in relation to peridinins.

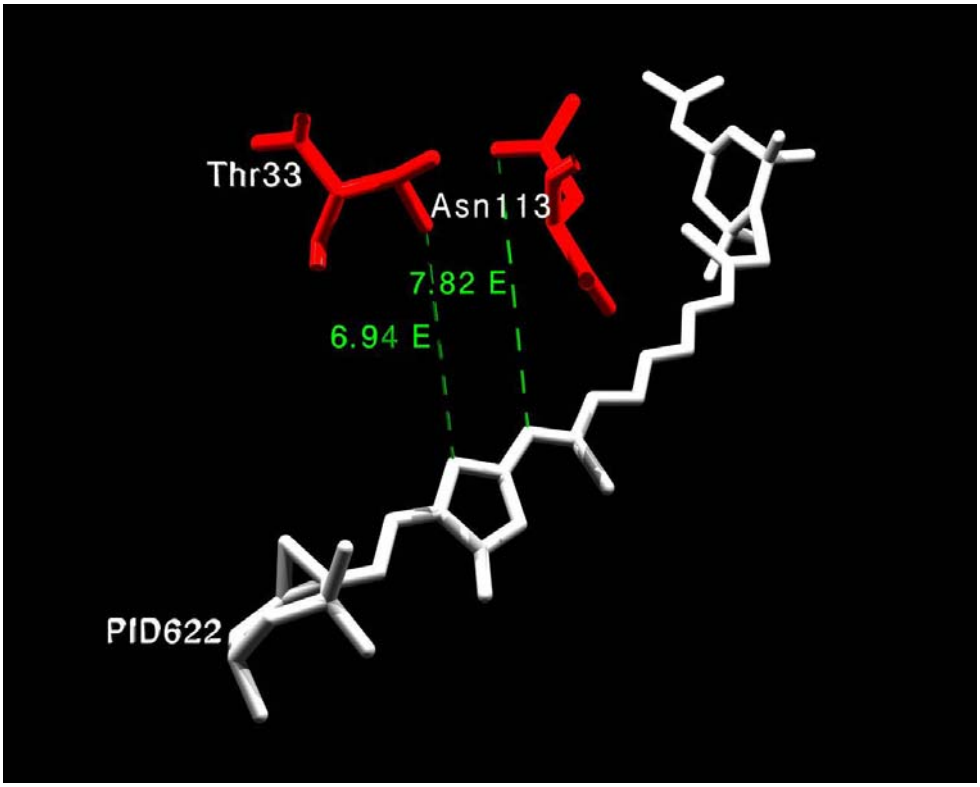


Figure 2.5 Distances (E=Å) from the polar side chains of threonine 33 and asparagine 133 to the furanic ring and polyene chain of peridinin 622.

Predicated amino acid sequences, isoelectric focusing points and protein modeling for S. kawagutii PCPs

Ten unique *S. kawagutii* PCP preproteins sequences are aligned in Appendix 2.5. All polypeptides but that predicted from a single clone are 365 aa long and like the majority from *Symbiodinium* 203 contain a 52 aa 12 kDa transit peptide followed by a 313 aa 33 kDa mature apoprotein. The apoprotein from 135clone42 differs slightly in that it is 2 aa longer on the C-terminal end. The

transit polypeptides from *S. kawagutii* PCPs are arranged like those from *Symbiodinium* 203 and *Symbiodinium sp*. from *A. formosa* (Norris and Miller, 1994). The positively charged central domain has a F(V/A)PGPR motif while the cleavage site is denoted by an AFA sequence.

The apoproteins from *S. kawagutii* were most similar in terms of amino acid sequence to those described in Chapter One for *Symbiodinium* 203, but the calculated isoelectric focusing points for *S. kawagutii* PCPs were not restricted to solely acid or basic forms. To the contrary, they span from pH 5.5 to 8.38 (see Table 2.6).

Table 2.6 Calculated mass and isoelectric points for *S. kawagutii* apoproteins

135 clone	Apoprotein mass (kDa)	Isoelectric point (pI)
37	32.9378	6.04
38	33.012	7.41
41	32.8408	8.1
42	33.057	7.46
43	32.1997	5.73
49	33.0351	7.97
50	32.9777	5.5
52	32.994	8.38
54	33.0941	8.11
60	33.0158	6.51

Just as there is correlation between predicted and previously measured pIs for *S. pilosum* apoproteins, the same is true for *S. kawagutii.* Figure 2.6 shows that the range of *S. kawagutii* PCP pIs encompasses many of the isoforms reported by Trench and Blank (1987), and that there are six matches between the data sets.

This relationship reinforces the idea that genetic diversity of PCP gene families is the underlying source of isoform generation.

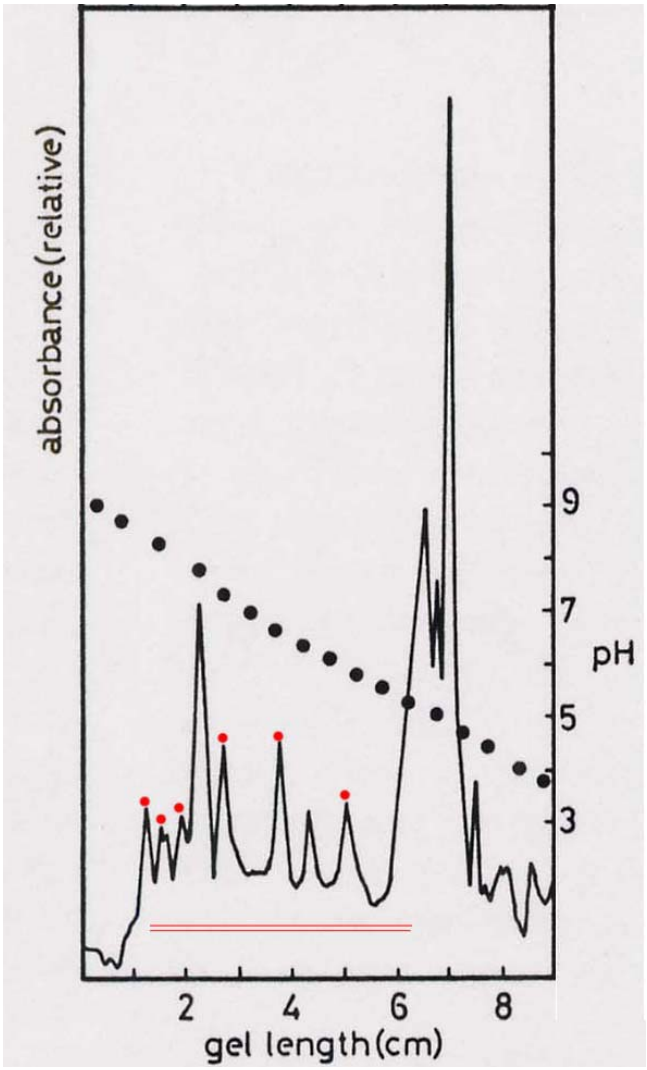


Figure 2.6 Isoelectric focusing patterns from *S. kawagutii*. The red double line indicates the range of calculated pIs for *S. kawagutii* apoproteins, and the red dots show matches between the calculated and actual pIs (After figure 1c, Trench and Blank, 1987).

The *S. kawagutii* PCPs are significantly more diversified than the PCPs from either *Symbiodinium* 203 or *S. pilosum*. Among the *S. kawagutii* clones, there are 9 polymorphic sites in the transit peptides and 55 in the apoproteins. Of the differences in the apoproteins, 22 sites (positions 15, 16, 24, 36, 70, 89, 130, 150, 154, 155, 162, 169, 178, 199, 215, 235, 238, 239, 280, 287 and 288) are predicted to accommodate either the presence or absence of amino acids with polar side chains. There are an additional four polymorphic sites that hold various polar amino acids. Specific amino acid substitutions and types of side chains are listed in Table 2.7.

Mapping substitutions into the *A. carterae* 1PPR crystal structure revealed that there were 94 out of 313 sites at which the *S. kawagutii* and *A. carterae* PCPs differed. Beyond the 55 polymorphic sites, there were 39 fixed changes present in each *S. kawagutii* apoprotein. A composite structure containing all of the polar substitutions within polymorphic sites was used to show the positions of these amino acids in relation to the 8 peridinins present in large PCP (see Figure 2.7). As with protein modeling presented for *Symbiodinium* 203 and *S. pilosum*, polar side chains within 10 Å of furanic ring and polyene chains of peridinins are considered likely to affect the specific wavelength of blue-green light harvested by the accessory pigment (see Hofmann et al. 1996). Eight amino acids predicted from *S. kawagutii* PCP apoproteins (positions 70, 89, 150, 169, 178, 280, 287 and 288) meet this proximity criterion and are rendered in Figures 2.8-10.

Table 2.7 Substitutions within polymorphic sites of predicted *S. kawagutii* PCP apoproteins. Sites in bold accommodate the presence or absence of amino acids with polar side chains.

Site #	Substitutions	Side Chain	Site #	Substitutions	Side Chain	Site #	Substitutions	Side Chain
9	Lys	basic	118	Ala	nonpolar	**202**	Glu	acidic
	Arg	basic		Lys	basic		Gln	uncharged polar
10	Lys	basic	**130**	Pro	nonpolar		Ala	nonpolar
	Pro	nonpolar		Ser	uncharged polar	211	Ile	nonpolar
15	Tyr	uncharged polar	138	Ala	nonpolar		Val	nonpolar
	Phe	nonpolar		Gly	nonpolar	**215**	Ala	nonpolar
16	Ser	uncharged polar	142	Phe	nonpolar		Ser	uncharged polar
	Pro	nonpolar		Ile	nonpolar	226	Val	nonpolar
24	Asn	uncharged polar	146	Val	nonpolar		Ala	nonpolar
	Asp	acidic		Asp	acidic	233	Phe	nonpolar
	Tyr	uncharged polar	**150**	Gln	uncharged polar		Ile	nonpolar
28	Phe	nonpolar		His	basic	**235**	Ser	uncharged polar
	Leu	nonpolar	152	Thr	uncharged polar		Cys	nonpolar
36	Gln	uncharged polar		Ser	uncharged polar	**238**	Ala	nonpolar
	His	basic	153	Thr	uncharged polar		Thr	uncharged polar
39	Glu	acidic		Ser	uncharged polar	**239**	Lys	basic
	Lys	basic	**154**	Ala	nonpolar		Asn	uncharged polar
50	Val	nonpolar		Thr	uncharged polar	246	Asp	acidic
	Met	nonpolar	**155**	Arg	basic		His	basic
57	Lys	basic		Ser	uncharged polar	275	Gly	nonpolar
	Met	basic	160	Val	nonpolar		Lys	basic
61	Lys	basic		Ala	nonpolar	276	Ile	nonpolar
	Arg	basic	**162**	Ser	uncharged polar		Val	nonpolar
	Val	nonpolar		Phe	nonpolar	**280**	Thr	uncharged polar
64	Ala	nonpolar	166	Ile	nonpolar		Pro	nonpolar
	Gly	nonpolar		Val	nonpolar	**287**	Ser	uncharged polar
70	Thr	uncharged polar	168	Glu	acidic		Pro	nonpolar
	Ile	nonpolar		Gly	nonpolar		Lys	basic
	Val	nonpolar		Val	nonpolar	**288**	Lys	basic
	Phe	nonpolar	**169**	Ala	nonpolar		Asn	uncharged polar
81	Ser	uncharged polar		Ser	uncharged polar	290	Ser	uncharged polar
	Thr	uncharged polar	170	Ala	nonpolar		Asn	uncharged polar
85	Trp	nonpolar		Val	nonpolar	295	Met	nonpolar
	Cys	nonpolar	**178**	Tyr	uncharged polar		Ile	nonpolar
89	Asn	uncharged polar		Phe	nonpolar		Val	nonpolar
	Asp	acidic	185	Asp	acidic	305	Lys	basic
102	Ala	nonpolar		His	basic		Arg	basic
	Glu	acidic	**199**	Thr	uncharged polar	310	Ala	nonpolar
111	Val	nonpolar		Pro	nonpolar		Gly	nonpolar
	Ala	nonpolar						

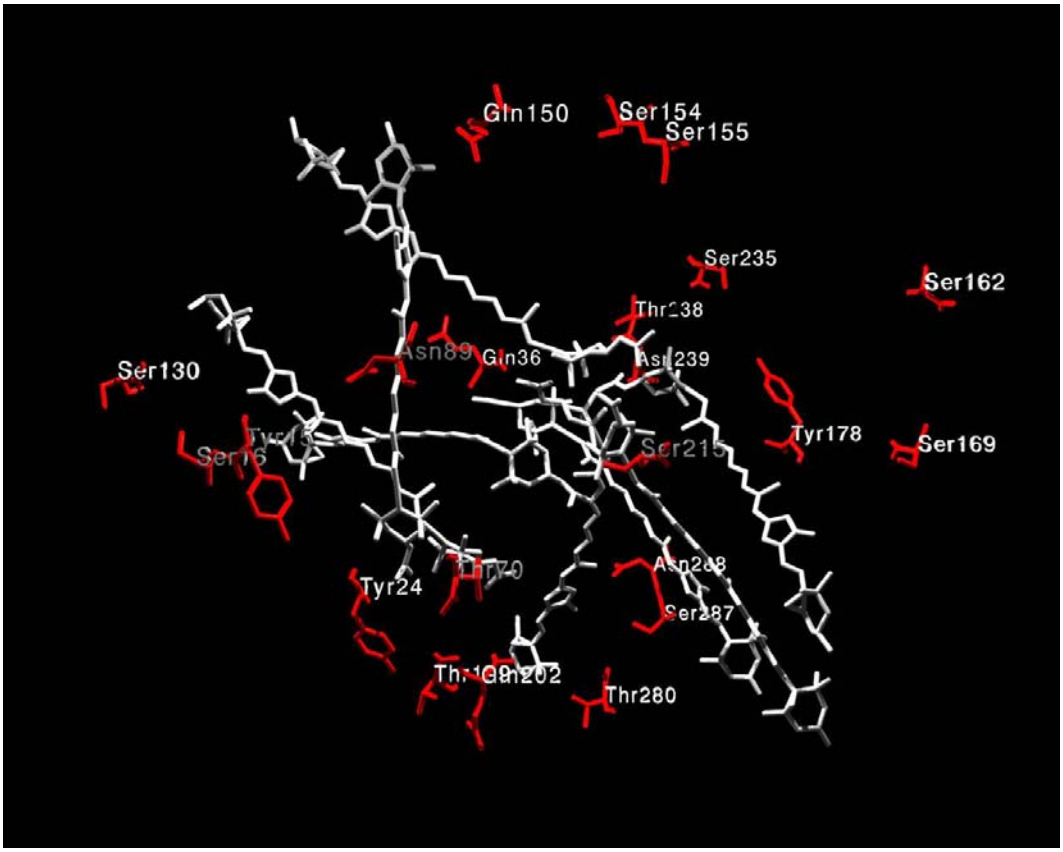


Figure 2.7 Twenty two polar amino acids occurring in polymorphic sites of *S. kawagutii* predicted PCP apoproteins located relative to eight peridinin chromophores.

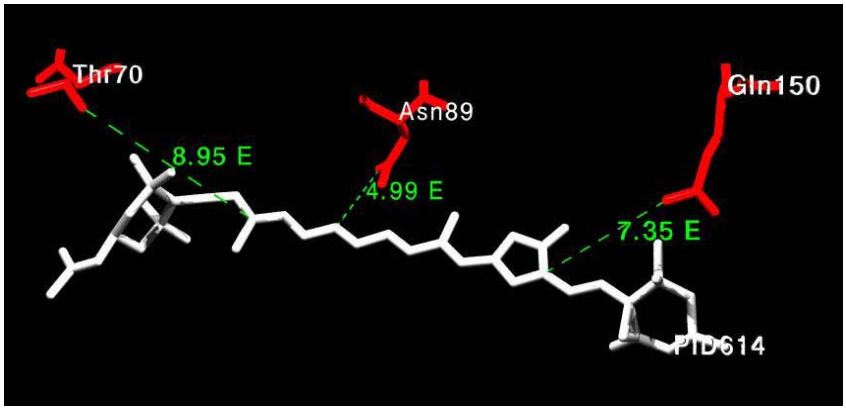


Figure 2.8 Distances (E=Å) from the polar side chains of threonine 70, asparagine 89 and glutamine 150 to the furanic ring and polyene chain of peridinin 614.

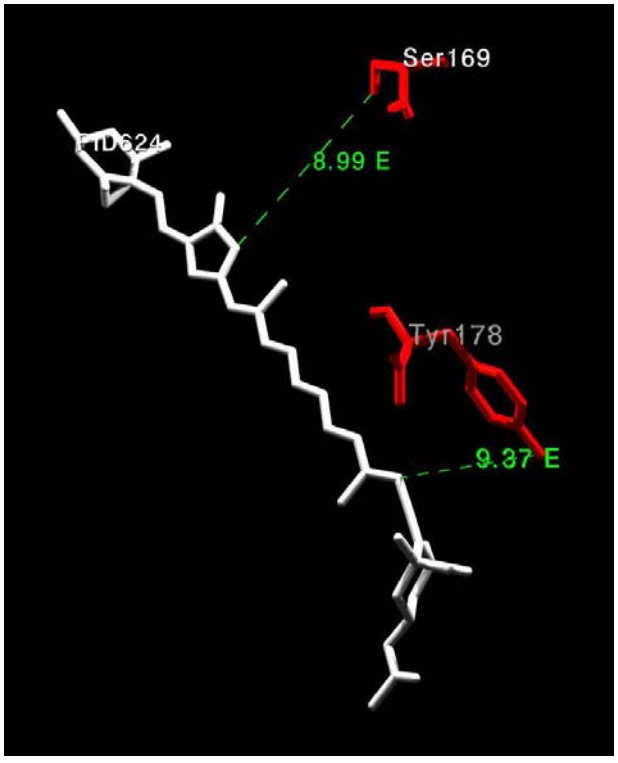


Figure 2.9 Distances (E=Å) from the polar side chains of serine 169 and tyrosine 178 to the furanic ring and polyene chain of peridinin 624.

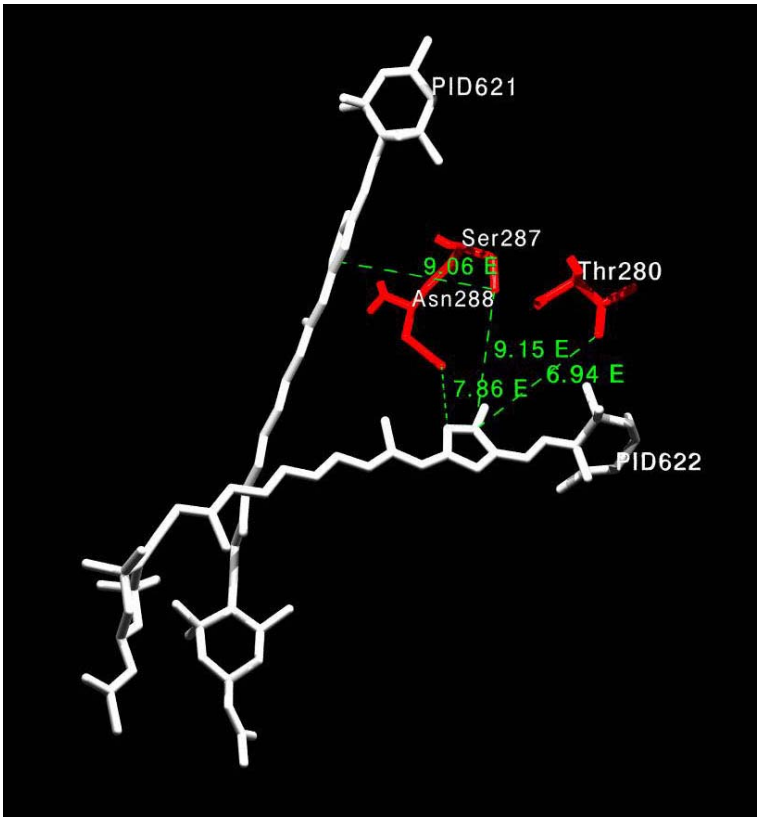


Figure 2.10 Distances (E=Å) from the polar side chains of threonine 280, serine 287 and asparagine 288 to the furanic ring and polyene chain of peridinins 621 and 622.

DISCUSSION

Diverse PCP gene families found in Symbiodinium sp. from each major rDNA clade

Evidence of genetic variability within individual dinoflagellate PCP gene families began to emerge in 1993 when Triplett et al. cloned 2 distinct small PCP transcripts from *H. pygmaea.* Despite reading frame problems with their data that were noticed after publication, their general multi-gene premise was correct. Sharples et al. (1996) reported two versions of large PCPs in *A. carterae* after cloning transcripts and genomic coding regions. Le et al. (1997) concluded that the large PCP gene family of *L. (Gonyaulax) polyedra* was highly conserved as judged by Southern hybridization, yet they acknowledged the possibility that differences could exist between individual gene copies. More recently, Hiller et al. (2001) indicated that up to five different PCP genes were found in *H. pygmaea*. The only previous data on PCP genes from *Symbiodinium* prior to Chapter One were single sequence reports (see Norris and Miller, 1994; Weis et al. 2002).

The results from Chapter One demonstrated several important points regarding PCP genes. This was the first detailed characterization of a PCP gene family from a *Symbiodinium* species. The results showed that the PCP gene family in *Symbiodinium* 203 is highly diversified, is apparently subjected to little or no homogenization by concerted evolution, contains sufficient non-synonymous substitutions to code for multiple PCP isoforms, and that its predicted apoproteins have polymorphic sites that could affect the tuning of peridinins.

One of the objectives of the present report is to determine if the diversity of *Symbiodinium* 203 PCP gene family is common to other *Symbiodinium* species or if *Symbiodinium* 203 is anomalous. Multiple PCP coding regions were found for each of the additional species tested here (*S. pilosum, Symbiodinium sp*. from *D. stokesii, S. pulchrorum,* and *S. kawagutii*). Together with *Symbiodinium* 203 there are now examples of diversified PCP gene families from species in each major *Symbiodinium* rDNA clade; 1 from A, 2 from B and 2 from C. This set of taxa includes endosymbionts from various Pacific and Caribbean invertebrate hosts, and it is reasonable to assume that the pattern of genetic diversity may extend to all *Symbiodinium* species.

Both small and large PCP genes from Symbiodinium occur in diversified families

Beyond showing that diversified PCP gene families are found in species from *Symbiodinium* clades A, B, and C, the present results also demonstrate that both size classes of PCP genes are diversified as well. This agrees with previous reports of multiple small and large PCP coding regions from *H. pygmaea* and *A. carterae* respectively (Triplett et al, 1993; Sharples et al., 1996; and Hiller et al., 2001). This widespread diversity should be considered within the context of dinoflagellate and PCP gene evolution. Free-living dinoflagellates are generally considered to be ancestral to symbiotic species. In addition, the axis of symmetry in large PCP genes and the nucleotide sequence homology between small and large PCP genes (Norris and Miller, 1994; Hiller et al., 2001) support the

hypothesis that large PCP genes evolved from the duplication and fusion of small PCP genes. An implication of these points taken together is that PCP gene polymorphism is probably ancient. This characteristic of PCP genes may predate the divergence of *Symbiodinium* from free-living dinoflagellates and the evolution of species with large genes from those with small genes. However, it is also possible that evolutionary history of PCP genes is more complicated. Large PCP genes and endosymbioses may both have evolved more than once. It is also possible that certain small PCP genes have been formed via truncation of large genes. The observable differences in level of diversification between species suggest that various factors affecting PCP gene diversification (selection, drift, etc.) are still at work. Although polymorphism among PCP genes has probably existed for millions of years, it is not clear if it offers a selective advantage and is being maintained. In general, PCP gene families appear to have evolved under little to no concerted evolution, unlike dinoflagellates ribosomal gene families that are more uniform and easily sequenced directly from PCR amplifications (see Rowan and Powers, 1991a & b; 1992, Rowan and Knowlton, 1995; Rowan et al. 1997; Wilcox, 1998).

Calculated pIs from predicted PCP apoproteins match measured values

From some of the earliest characterizations of PCPs, it was noted that these dinoflagellate light harvesting complexes were normally expressed in multiple isoforms distinguished by their isoelectric focusing points. While it was recognized that genetic diversity might be the source of the isoforms, preliminary

experimental results pointed to post-translational modification instead (see Haxo et al. 1975; Haxo et al. 1976; Seigelman et al. 1977). Trench and Chang (1984) suggested to the contrary, that the genetics were likely responsible. It was not until the tools of molecular biology were applied to the cloning and sequencing of PCP transcripts and genomic coding regions that evidence supporting the genetic hypothesis was produced from experiments on free-living dinoflagellates (Triplett et al. 1993; Sharples et al. 1996; Hiller et al. 2001).

The predicted isoelectric focusing points for *S. pilosum* and *S. kawagutii* presented in this chapter reasonably match the Trench and Blank (1987) data on the actual apoproteins for these species. For both species, the range of predicted pIs overlap with the majority of dominant PCP isoforms expressed. However, the fit is not perfect in either case. This may be attributable to the fact that the searches for unique coding regions were not exhaustive. Each species may contain many more versions of PCP genes than were detected here. While results from this and the previous chapter substantiate that genetic diversity is common to PCP gene families in *Symbiodinium*, post-translational modification may also be contributing to isoform diversity. It now seems very unlikely that dinoflagellates only express a single PCP preprotein that is then converted to multiple isoforms.

Polarity polymorphic sites of PCP apoproteins likely to influence tuning of peridinins

Not all PCPs capture exactly the same wavelengths of light, but they generally overlap. Different dinoflagellates species can contain PCPs that have unique spectroscopic absorbance maxima (Prezelin and Haxo, 1976; Song et al.

1976), and individual PCP isoforms within a single species can have different absorption and fluorescence maxima (Iglesias-Prieto et al. 1991). The protein modeling results from *Symbiodinium* 203 (Chapter One), *S. pilosum* and *S. kawagutii* provide the likely structural basis for these observations. The predicted PCP apoproteins for all three species contain polarity polymorphic sites that may house amino acids with or without polar side chains. In each case there is a subset of these polymorphic sites that have side chains within 10Å of furanic rings and polyene chains of peridinins potentially affecting spectral tuning.

Although, a consistent signature of polarity polymorphic sites is present in each group of predicted PCP apoproteins, there is variation between species in terms of the relative proportion of these sites (*Symbiodinium* 203, 3%; *S. pilosum*, 2%; and *S. kawagutii*, 7%) and their specific locations. *Symbiodinium* 203 and *S. kawagutii* have three polarity polymorphic sites in common and only one (position 287) is close to a peridinin in both cases. Neither species has sites of this nature that correspond to those in *S. pilosum.* Despite the apparent functional significance of presence or absence of polar residues in areas that differentially tune the peridinins of PCPs, it is entirely possible that the substitutions within PCP genes occur at random locations and are only weakly filtered out by gene conversion. Depending on their host association and dominant habitat, different species of dinoflagellates may be able to tolerate more or less variability of PCP light harvesting characteristics that are simply the by products of genetic polymorphism. On the other hand there may be regions of PCPs that are under selection that are not yet revealed by the methods applied here or there may be

purifying selection for a range spectral absorption maxima. Phylogenetic analyses of the now expanded PCP gene sequences available will determine if the nucleotide substitutions are non-random and if they are, then sites under selection could also potentially be detected.

CONCLUSION

The first PCP coding sequences from *S. pilosum*, *Symbiodinium sp*. from *D. stokesii*, *S. pulchrorum* and *S. kawagutii* were presented in this chapter. Multiple coding sequences were cloned from each species. I established that diverse PCP gene families occur in all major clades of *Symbiodinium* and in both size classes of the gene. As with *Symbiodinium* 203 in Chapter One, these PCP gene families do not appear to have been homogenized through concerted evolution. The predicted PCP apoproteins from *S. pilosum* and *S. kawagutii* have calculated isoelectric focusing points that generally match values previously measured for these species, which supports the hypothesis that genetic polymorphism is the primary source generating differences in PCP isoforms, however, post-translational modification cannot be ruled out as another possible mechanism. Protein modeling based on *A. carterae* PCP produced a putative tertiary structure for *S. pilosum* apoproteins and was used to identify polymorphic sites in *S. pilosum* and *S. kawagutii* PCPs that could affect spectral tuning of peridinins. Additional phylogenetic work is needed to determine if amino acids sites within PCP are under selection.

CHAPTER THREE

Phylogenetic analyses and evolution of dinoflagellate PCP gene families: diversity within families is consistent across lineages, but individual sites are most likely to be under purifying rather than positive selection.

ABSTRACT

Photosynthetic dinoflagellates use multiple PCP isoforms to harvest blue-green light from aquatic habitats. Genetic differences among copies in PCP gene families account for expression of many of these isoforms, but it is unclear if selective pressures are maintaining the polymorphism. This is the first phylogenetic study of the evolution of PCP gene families with the aim of measuring selection at the codon level. Nucleotide substitution model likelihood ratio tests (LRTs) on PCP coding sequences from seven symbiotic and three free living dinoflagellate species suggests that PCP gene sequence changes are most closely fit by the GTR+I+G model. Consensus and highest posterior probability trees from Bayesian analyses show that topologies of PCP gene and large subunit (ls) rDNA trees are generally similar. As expected, PCP sequence data indicate that symbiotic species evolved from free-living ancestors and that the small PCP gene state consistently appears to be ancestral to large PCP gene state. There is also not a clear correlation between PCP gene size class and ribosomal clade. However, there are inter-digitations between large PCP sequences from clade B

and C species. These results suggest that PCP gene polymorphism is ancient and preceded the evolution of large PCP genes by fusion of small genes, the divergence of the genus *Symbiodinium* from free-living dinoflagellates and speciation within *Symbiodinium*. They also suggest that small and large PCP genes may have previously been present or may still be present in species from each *Symbiodinium* clade. Several PAML CODONML models were tested on a single 17 sequence tree with the highest posterior probability. LRTs indicate that that data is best fit by Model M3 (discrete ωs). There was no indication of codons under positive selection. To the contrary, purifying selection is affecting PCP genes. It is likely that PCP polymorphism is adaptive and that there is purifying selection for isoforms within a range of spectral absorption maxima.

INTRODUCTION

The evolution of PCPs has been a prominent adaptation for photosynthetic dinoflagellates. Free-living and symbiotic dinoflagellates utilize their unique PCP light harvesting complexes to capture blue-green light with remarkable efficiency from aquatic habitats. Regardless of whether it has small or large PCP genes, an individual species will express a stable combination of apoprotein isoforms in an inversely proportional response to light intensity (Chang and Trench, 1982; 1984; Trench and Blank, 1987). The source of PCP isoform diversity has been a matter of speculation that was addressed in the previously sections of this dissertation. In Chapters One and Two, I established that diverse PCP genes are common to several *Symbiodinium* species and that genetic polymorphism can account for the

majority of isoform pI variation through amino acid substitutions. Some of these changes can potentially affect absorption maxima of chromophores within PCP complexes and provide a structural mechanism for subtle absorption maxima differences reported for isoforms from *S. microadriaticum* (Iglesias-Prieto et al. 1991). Despite the genetic explanation for these PCP properties, it is not clear that PCP isoform polymorphisms are adaptive to the present environments that dinoflagellates live in. If they are adaptive, then positive selection of sites within PCP genes may be maintaining the polymorphism, or there may be purifying selection for ranges of isoforms.

Sequence data on PCP genes have been slowly accumulating for almost a decade. While one phylogenetic study has shown how soluble and membrane-bound PCPs are related to light harvesting complexes of other algae and higher plants (Durnford et al. 1999), no investigation to date has concentrated at the evolution of PCP genes between dinoflagellates species or among gene families. There are now examples of multi-version PCP gene families from two free-living species, *H. pygmaea* (Triplett et al. 1993; Hiller et al. 2001), and *A. carterae* (Sharples et al. 1996) and five symbiotic species, *Symbiodinium* 203, *S. pilosum*, *Symbiodinium sp*. from *D. stokesii*, *S. pulchrorum* and *S. kawagutii* (Chapter One and Two). There are additional single sequence characterizations from free-living *L. polyedra* and endosymbionts *Symbiodinium* from *A. formosa* and *S. muscatinei.* The total number of PCP nucleotide sequences available for phylogenetic analyses was increased over six fold by the results from Chapters One & Two.

Comparative methods need to be applied to this collection of sequence to answer the following basic questions regarding PCP evolution. Does the evolution of PCP genes between species simply reflect the divergence of species as indicated by ribosomal sequences or does the phylogenetic reconstruction of PCP gene evolution show a more complex pattern? Are there codon sites that are under positive selection (ω (= d_N/d_S = Ka/Ks) >1)? If so, are they associated with spectral tuning of chromophores or some other domains? Are PCP genes subject to purifying selection (ω < 1)? This chapter seeks to answer these questions through the use of combined Bayesian analysis and maximum likelihood (ML) approaches.

MATERIALS AND METHODS

DNA sequences and alignments

Fifty eight PCP coding sequences from Genbank were used in this project, and their accession numbers are shown in Table 3.1. Sequences were aligned with MegaAlign (DNA Star, Inc.) and imported into Se-Al Sequence Alignment Editor v2.0 (Copyright 1996-2002, Andrew Rambaut). Within Se-Al, the nucleotide sequences were translated and manual adjustments were made to the amino acid alignment. The final alignment was back-translated (see Appendix 3.1) and exported as a Nexus file.

Table 3.1 Accession numbers of PCP gene sequences used for phylogenetic analyses

Species	Accession Numbers
A. carterae	AF298220, Z50792 & Z50793
H. pygmaea	AJ298192 & AJ298193
L. polyedra	U93077
Symbiodinium sp . from *A. formosa*	L13613
Symbiodinium sp . from *H. hippopus*	AY149123 - AY149139
S. pilosum	AY149140 - AY 149147
Symbiodinium sp . from *D. stokesii*	AY149148 - AY149158
S. pulchrorum	AY149159 - AY149162
S. kawgutii	AY149163 - AY149172

Nucleotide substitution model testing

The Nexus file described above was executed in PAUP* 4.0 (Swofford, 1998). With the optimality criterion set to maximum parsimony, a heuristic search of 10,000 trees was conducted and the best tree was retained. Likelihood scores for this tree were successively generated under the JC69, K2P, HKY85, GTR, GTR+I and GTR+I+G series of nucleotide substitution models. The PAUP* setting were as follows: JC69 (one substitution type, nucleotide frequencies equal, no invariable sites and equal rates at variable sites), K2P (two substitution types, transition / transversion ratio estimated by ML, nucleotide frequencies equal, no invariable sites and equal rates at variable sites), KHY85 (two HKY85 variant substitution types, transition / transversion ratio estimated by ML, nucleotide frequencies estimated by ML, no invariable sites and equal rates at variable sites),

GTR (six substitution types, rate matrix estimated by ML, nucleotide frequencies estimated by ML, no invariable sites and equal rates at variable sites), GTR+I (six substitution types, rate matrix estimated by ML, nucleotide frequencies estimated by ML, proportion of invariable sites estimated and equal rates at variable sites) and GTR+I+G (six substitution types, rate matrix estimated by ML, nucleotide frequencies estimated by ML, proportion of invariable sites estimated, gamma distribution of variable sites with shape parameter estimated and four rate classes). Pairs of nested models (a model in the series vs. the next more general model) were compared using a standard LRT in which twice the difference of negative log likelihood was compared to a χ^2 distribution at the degrees of freedom resulting from the difference in the number of free parameters of the models under comparison. In each comparison, the more specific model represents the null hypothesis and the next more general model is the alternative. The "p" statistic reported is probability of observing the LRT result if the null hypothesis was true.

Bayesian inference of PCP gene phylogeny

The same Nexus file was executed in MrBayes 2.01 and 3.0 (Huelsenbeck and Ronquist, in press) with the parameters of the likelihood model setting (lset) corresponding to GTR+G+I. The prior probabilities settings (prset) for the rate maxima were adjusted to give a negative exponential curve (exp (.2)). Using a negative exponential curve more accurately parameterizes the substitution rate matrix (personal communication, Derrick Zwickl, University of Texas at Austin).

In two MrBayes runs, the Markov Chain Monte Carlo (MCMC) settings included 500,000 generations with trees sampled every 100 generations with 4 MCMC chains. The temperature of the three "heated chains" was left at the default (.2). To check for convergence an additional 4,000,000 generation run was executed with the same settings. For all runs, -ln likelihood scores from the output files were graphically post "burnin" to determine which trees should be excluded from the calculation of consensus trees. Trees that were retained were sorted by in order to identify the tree with highest posterior probability.

An additional 4,000,000 generation MrBayes run (same settings and post-run analyses) was executed on a 17 PCP gene sequence subset of the original nexus. The subset nexus file was generated by removing sequences from the alignment within PAUP*. The subset contained a single sequence from each free-living species, the two *Symbiodinium* sequences published prior to Chapter One and pairs of sequences from each of the species in Chapters One and Two that were most divergent in terms of their "p" distance as identified by PAUP*. This subset was generated to reduce to occurrence of near zero branch lengths in the highest posterior probability tree.

Testing PAML CODONML models and detecting selection.

The 17 sequence Nexus file and the corresponding tree with the highest posterior probability were loaded in to PAML 3.13 (Yang, 1997). The data was analyzed with six codon substitution models from CODONML; M0 (single ω estimated for all sites), M1 ($\omega_0 = 0$ and $\omega_1 = 1$), M2 ($\omega_0 = 0$, $\omega_1 = 1$ and ω_2

estimated from the data), M3 (three discrete ωs all of which could be >1), M7 (ten discrete ωs picked from gamma distribution) and M8 (ten discrete ωs picked from gamma distribution plus one estimated from the data that can be >1). LRTs were performed on the likelihood scores from nested pairs of models to determine which models best fit the data. The ω values and probabilities generated by the best fitting models were compared for indications of selective pressure on PCP genes and to check for sites under positive selection (see Yang et al. 2000; Yang and Bielawski, 2000).

RESULTS

Bayesian phylogenetic analysis of PCP genes with GTR+I+G

Based on the likelihood ratios of six standards nucleotide substitution models against a parsimony tree of PCP gene sequences, each more general set of parameters had a significantly better fit than the previous model (Table 3.2). Consequently, GTR+I+G was specified as the likelihood model for all MrBayes runs.

Table 3.2 Substitution model likelihood ratio test results for PCP genes

Model	- ℓ	LRT	-2 Δℓ	df	p
JC69	9526.02663				
K2P	9417.92171	JC69 vs K2P	216.20984	1	6.06941E-49
HKY85	9392.31734	K2P vs HKY85	51.20874	3	4.41592E-11
GTR	9375.89712	HKY85 vs GRT	32.84044	5	4.04818E-06
GTR+I	9224.37258	GTR vs GTR+I	303.04908	1	7.13659E-68
GTR+I+G	9144.89927	GTR+I vs GTR+I+G	158.94662	1	1.92222E-36

Post-run analyses indicated that there was overall convergence between the likelihood scores and consensus tree from the 500,000 and 4,000,000 generation MrBayes runs (data not shown). The longer run produced over 36,000 trees with approximately the same posterior probabilities. The consensus (Figure 3.1 colored by species) and highest posterior probability trees (Figure 3.2 colored by species and 3.3 colored by ribosomal clade) retained from this group both are shown below.

Some broad aspects of the topology of the PCP trees were present as expected. The tree structures are in agreement with the widely held view that species with small PCP genes gave rise to those with large PCP genes. PCP sequence data also support the hypothesis that *Symbiodinium* species have evolved from free-living ancestors. *H. pygmaea* small PCP genes are more basal than either *S. muscatinei* or *S. pilosum*. *A. carterae* and *L. polyedra* PCP genes are more basal than all other large PCP genes from *Symbiodinium* species. In addition, high levels of similarity between individual copies of the same PCP gene family (see Chapters One and Two) resulted in lack of resolution and polytomies in the consensus tree. This is also reflected in the short branch lengths near terminal nodes of the highest posterior probability tree. Finally, PCP size classes are not strictly correlated with ribosomal clades as demonstrated by the PCP genes from clade B species *S. muscatinei*, *Symbiodinium* from *D. stokesii* and *S. pulrchrorum.*

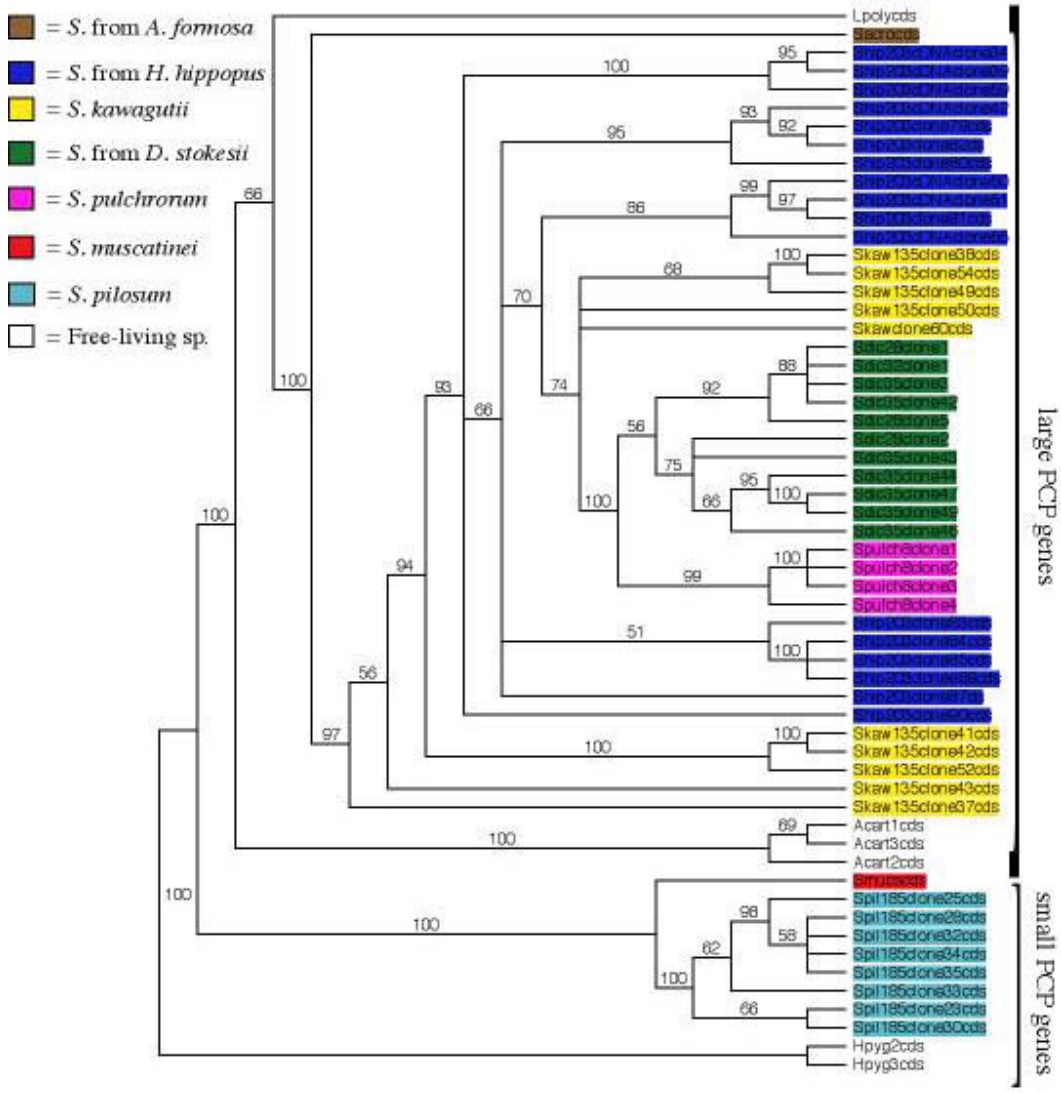


Figure 3.1 Consensus (50% majority rule) of 36,000 trees for complete PCP gene data set, colored by species. Numbers are posterior probability support.

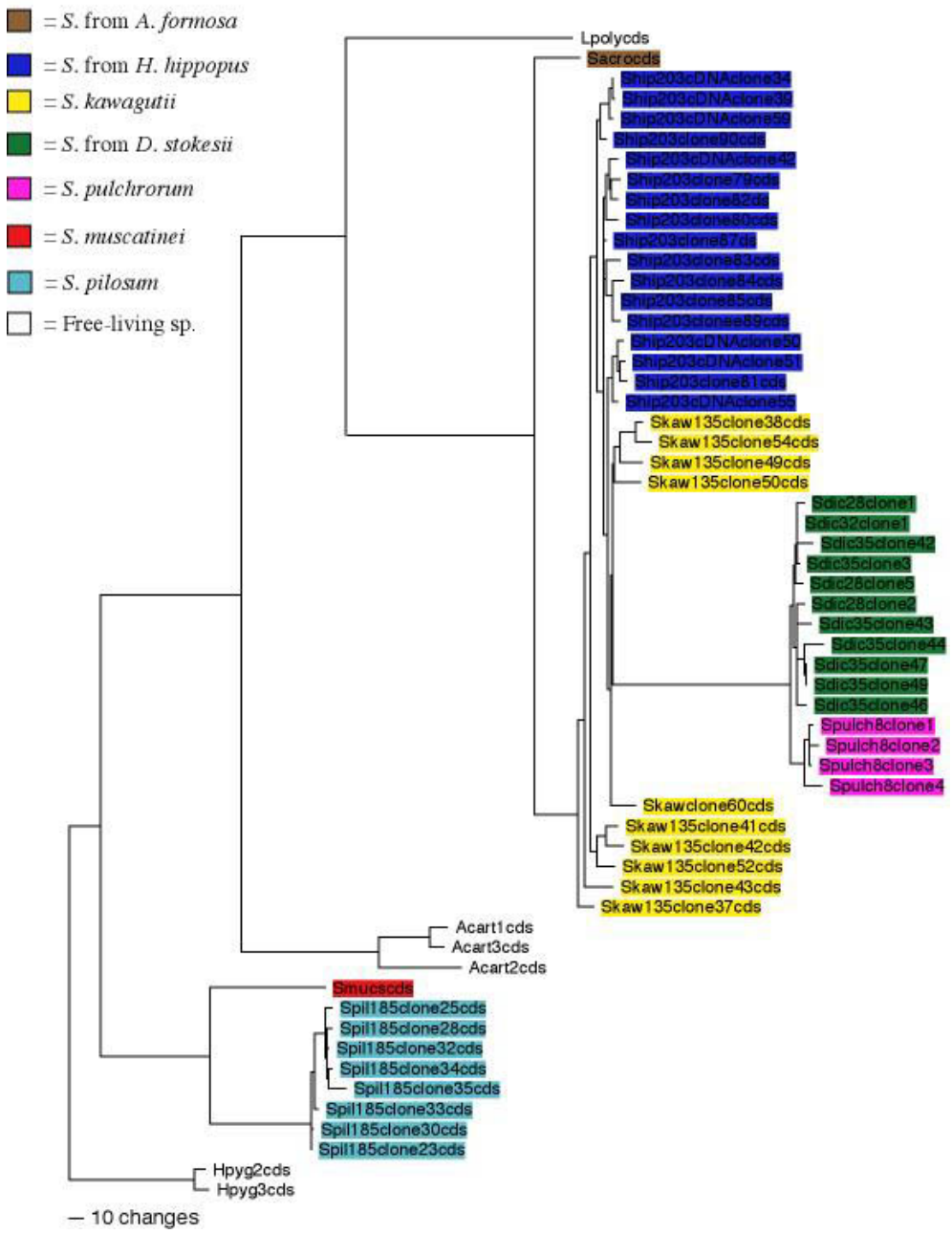


Figure 3.2 Highest posterior probability tree for the complete PCP gene data set, colored by species.

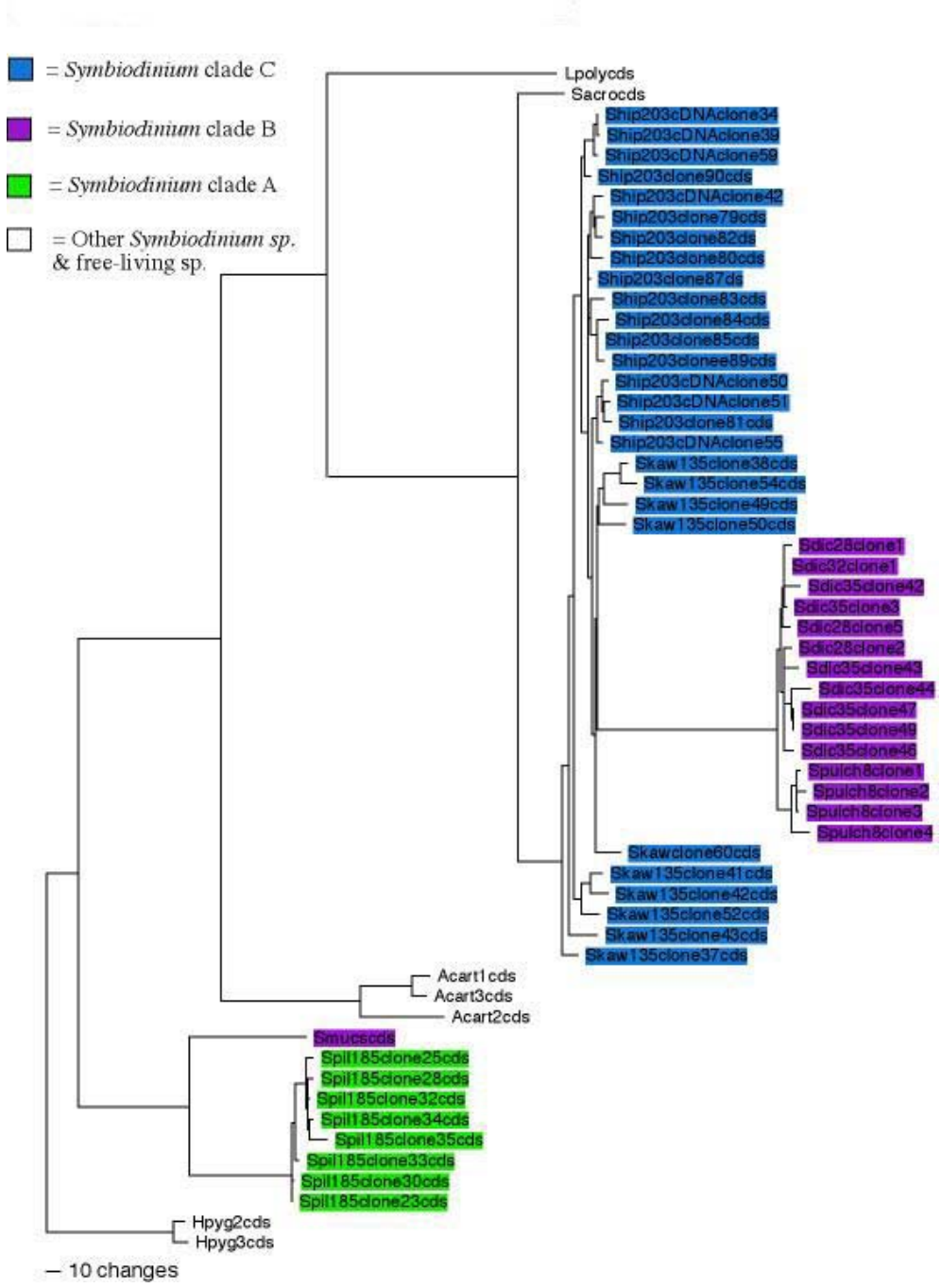


Figure 3.3 Highest posterior probability tree for the complete PCP gene set colored by large subunit ribosomal clade of the species.

By contrast, there were also unanticipated aspects to these trees. Figures 3.1 and 3.2 suggest that there were interdigitations of branches between the large PCP genes from different *Symbiodinium* species. Furthermore, Figure 3.3 shows that large PCP genes from a clade C species (*S. kawagutii*) could be more basal to those from clade B species (*Symbiodinium* from *D. stokesii* and *S. pulrchrorum*). A summary of the results presented in the cladograms in Figure 3.4.

The highest posterior probability tree from Bayesian analysis on the subset of most divergent PCP sequences is shown (Figure 3.5). This tree retains most of the same characteristics of the larger tree in Figures 3.2 and 3.3. This tree and its corresponding Nexus files were analyzed with PAML CODONML substitution models.

PAML CODONML analyses

The output files from PAML CODONML include likelihood scores, ω values (nonsynonymous to synonymous substitution ratios) and p scores (probability of sites having a corresponding ω value). As with the nucleotide substitution models in PAUP*, it was necessary to conduct LRTs on PAML codon substitution models to determine which model best fits the data. This was done prior to interpretation of estimated parameters. Table 3.3 show the comparisons made in these tests and probabilities where significant differences between models exited. M3 had the best fit. M2 and M3 fit the data better than either M0 or M1. M0 had a smaller negative log-likelihood score than M1, so an LRT was not appropriate. Neither M7 nor M8 fit significantly better than M3,

and there was also no significant difference between the likelihoods of M7 and M8.

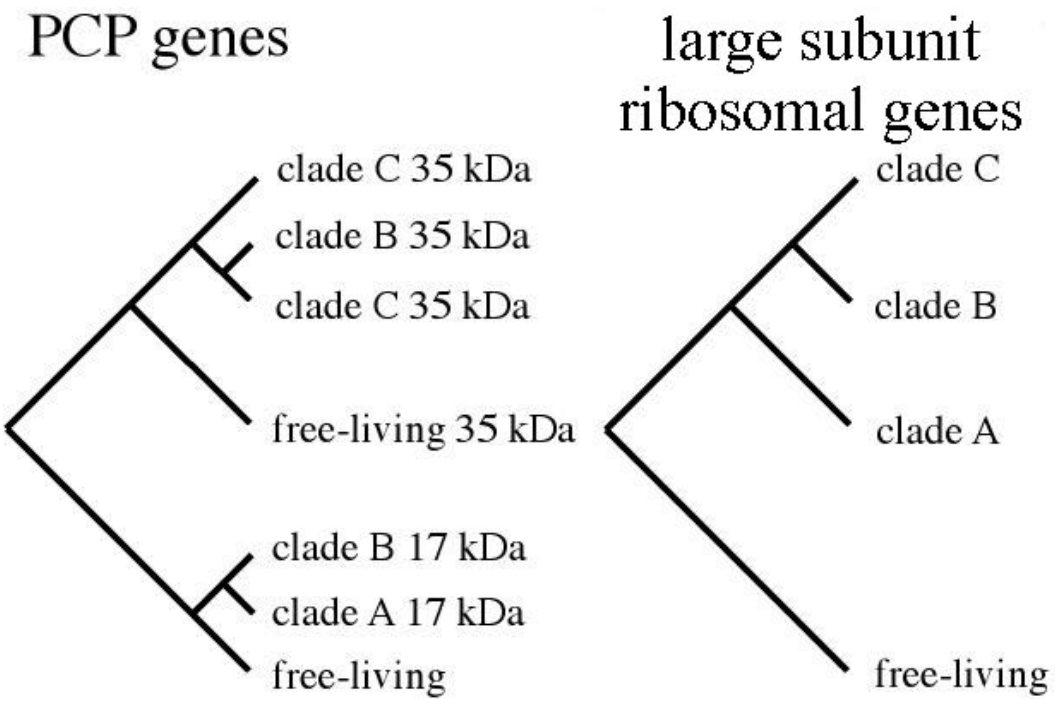


Figure 3.4 Cladograms comparing the evolution of PCP and large subunit ribosomal genes. Clade B 35 kDa PCPs and clade B small PCP are from different species. (Ribosomal tree was redrawn from Wilcox, 1998).

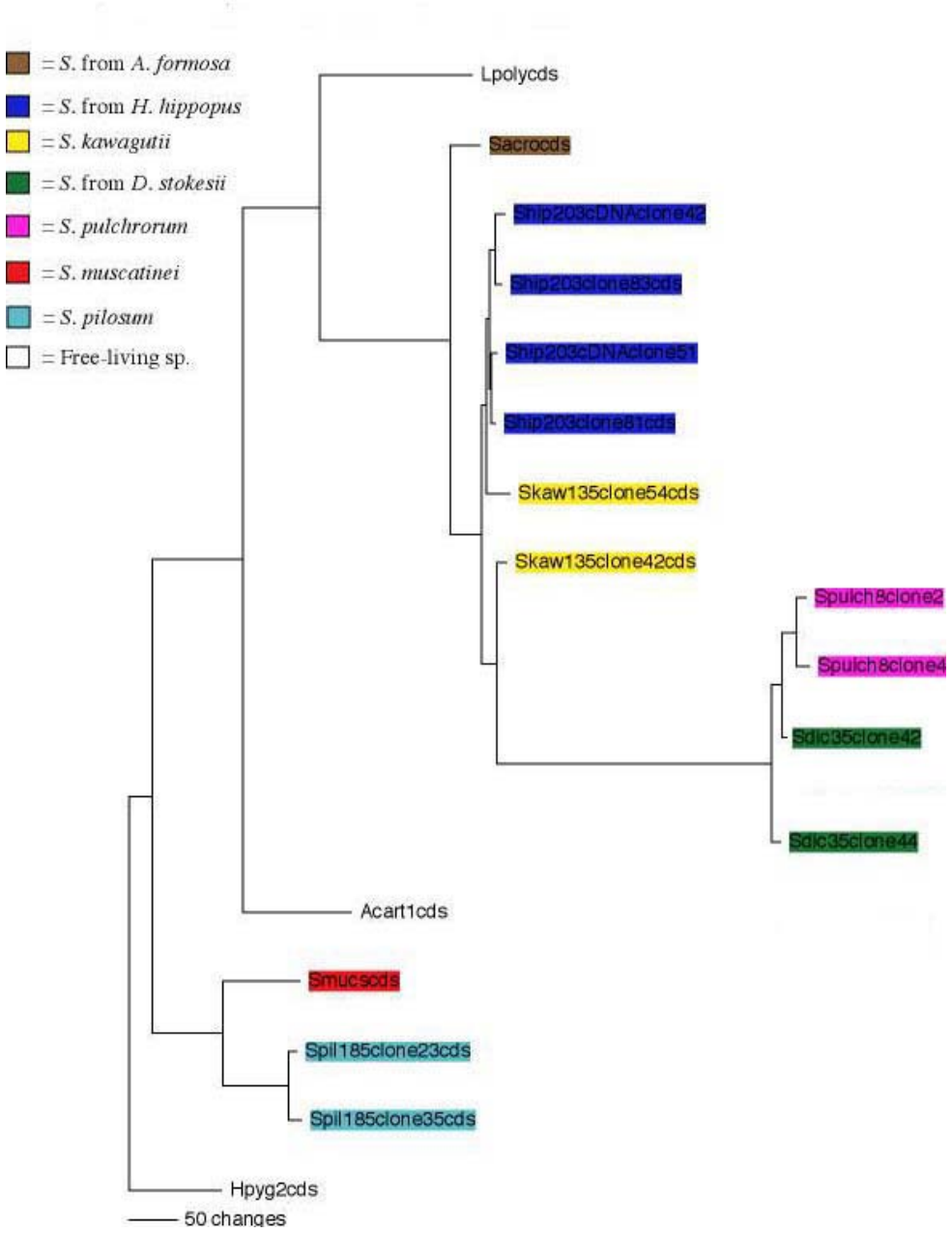


Figure 3.5 Highest posterior probability tree for PCP gene subset colored by species.

Table 3.3 PAML CONDONML model likelihood ratio test results

Model Code	- ℓ	LRT	-2 Δℓ	df	p
M0 (one-ratio)	6328.168135	M0 vs M2	187.592692	2	1.83977E-41
M1 (neutral)	6493.227752	M0 vs M3	212.324186	4	8.40292E-45
M2 (selection)	6234.371789	M1 vs M2	517.711926	2	3.8044E-113
M3 (discrete)	6222.006042	M1 vs M3	542.44342	4	4.414E-116
		M2 vs M3	24.731494	2	4.26211E-06
M7 (beta)	6224.635905				
M8 (beta&ω)	6224.066493				

The estimated parameters for all six codon substitution models are listed in Table 3.4. PAML CODONML can conduct paired comparisons between models based on the estimated parameters in order to identify the presence of any codons under positive selection (M2 vs. M3) and to more accurately estimate the strength of selection (M7 vs. M8). No sites were identified as being under positive selection. Probabilities reported for M2 and M3 suggest that most codons have ωs < 0.2 and are under purifying selection instead.

Table 3.4 PAML CODONML estimated parameters for PCP genes

Model Code	d_N/d_S for all branches	Parameters		
M0	0.1025	ω = 0.1025		
M1	0.5875	p0 = 0.41252	p1 = 0.58748	
		ω0 = 0.00	ω1 = 1.00	
M2	0.1467	p0 = 0.28871	p1 = 0.06443	p2 = 0.64685
		ω0 = 0.00	ω1 = 1.00	ω2 = 0.12718
M3	0.1303	p0 = 0.54959	p1 = 0.40577	p2 = 0.04464
		ω0 = 0.02218	ω1 = 0.19529	ω2 = 0.87118
M7	0.1238	p = 0.53740	q = 3.68372	
M8	0.1244	p = 0.72473	q = 3.84597	p1 = 0.22793
				ω1 = 0.01947

DISCUSSION

PCP gene diversity hypotheses

The novel evolution of PCP light harvesting complexes was a substantial adaptive success for photosynthetic dinoflagellates. The vast majority of extant photosynthetic dinoflagellate species express PCPs. It is now well established that small and large PCP genes from free-living and symbiotic dinoflagellates occur in diversified families and that this genetic polymorphism is the primary basis for expression of multiple PCP isoforms (Triplett et al. 1993; Sharples et al. 1996; Hiller et al. 2001; Chapters One and Two). Given the prevalence of PCP gene heterogeneity among divergent lineages of photosynthetic dinoflagellates, it is easy to assume that this characteristic has been maintained over evolutionary time because it too is adaptive, perhaps in terms of broadening light harvesting range.

Selective pressures could lead to adaptive polymorphism in at least two scenarios. First, there is positive selection for individual PCP isoforms with distinct spectral properties, and consequently corresponding codon sites within PCP genes have $\omega > 1$. Second, there is purifying selection for isoforms that have spectral properties within a specific range of absorption maxima (all codons with $\omega < 1$). While it seems unlikely that PCP gene polymorphism would be so wide spread without some adaptive functional significance, it is possible that there is purifying selection for isoforms with a specific absorption maximum, and that minor variants of that isoform are tolerated (also codons with $\omega < 1$).

Heterogeneity among PCP genes is ancient

The main objective of the present research was to use Bayesian phylogenetic reconstruction and maximum likelihood methods to determine the selective pressures that exist within PCP genes at the codon level and to thereby narrow the explanation of PCP polymorphism. The consensus tree, phylograms and cladograms presented in this report (see Figures 3.1 – 3.5) revealed that the evolution of PCP genes within families and between species is not simply a reflection of divergence between taxa, but rather that the history has been more complex and interesting. For example, PCP polymorphism appears to be ancient. Its presence in lineages of small and large PCP genes and in both free-living and *Symbiodinium* species suggests that the polymorphism predates the beginning of these divergences. This is also in agreement with the inter-digitations of PCP gene families among more closely related *Symbiodinium* species. If differences in PCP genes were not instrumental in speciation and the polymorphism was present as speciation events occurred, then overlap between gene families of recently divergent species would be expected. In addition functional constraints on PCP polypeptides may contribute to keeping sequences between species similar as in major histocompatibility complex genes. The nesting of large PCP genes from clade B species within those from clade C species is curious. More accurate placement of clade B large PCP genes may come by obtaining complete coding region sequences instead of partials.

PCP size class not correlated to ribosomal clade.

The lack of strict correlation between PCP gene size and ribosomal clade (see Figure 3.3 and 3.4) agrees with a similar observation about PCPs size classes and ITS clades (LaJeunesse, 2001). These results suggest that first fusions of small PCP genes to produce the large variety probably preceded the initial split of *Symbiodinium* from free-living species. Figure 3.4 demonstrates that different clade B species have either small or large PCP genes. It is very likely that all *Symbiodinium* clades once contained both size classes of PCP genes. In fact, C is the only major clade in which all characterized apoproteins are of one size class; 35 kDa (Govind et al. 1990; LaJeunesse, 2001). Clade C and clade A species may have lost their small and large PCP genes, respectively. On the other hand, it is also possible that these genes are still present yet were not detected by methods used in Chapter Two. In most instances, species specific primers had to be designed for each new PCP gene family. Characterization of PCP gene sequences from additional clade A and C species may demonstrate that both size classes of genes are still present.

It would be interesting to learn if any single species possesses both small and large PCP genes as this could represent an intermediate state in the evolution of PCP gene size classes. There are examples of clade A species such as *S. microadriaticum* that express small and large apoproteins. However, the large polypeptides from these species are homodimers and may come from the same small PCP genes that code for their small apoproteins. Some of the small

apoproteins may then undergo post-translational modification to allow dimerization (personal communication, Robert Trench, UC Santa Barbara).

PCP gene codons under purifying selection rather than positive selection

The results from the two PAML CODONML models which best fit the data, M2 and M3, were in agreement and were conclusive. They both indicate that it is most likely for codons to be conserved and have ω ratios estimated from the data at < 0.2. Neither model assigned codons to any ω class with a value higher than 1. Furthermore, M3 estimated ωs and probabilities suggest that codons from the PCP gene compared here have 55% probability being highly conserved with a $\omega < 0.02$, 41% probability of have $\omega < 0.2$ and only 4% probability of experiencing the highest ratio of 0.87. The consensus after comparison of the results from both models is that no PCP codon sites are under positive selection. Instead of having sites under positive selection, codons within PCP genes are subject to purifying selection like most other duplicated genes (Wagner, 2002). Although nonsynonymous substitutions exist between copies, the locations are not consistent within families or across species. This may be an important reflection of selective pressures in different habitats occupied by different species.

Because the tertiary structure of PCP polypeptides has a direct effect on the physical arrangement of chromophores and the efficiency of energy transfer from peridinin to chlorophyll *a*, there would clearly be constraints on the type and amount of amino acid substitutions that could occur without disrupting the light

harvesting function of the holoproteins. Purifying selection has the affect of removing deleterious mutations while leaving behind both nonsynonymous and synonymous substitutions of marginal to zero functional consequence. If PCP genes families were also subjected to highly efficient homogenization, then differences between copies would fade rapidly. However, there can be numerous differences at the nucleotide level between copies, and the distinctive pI properties of expressed isoforms are stable. This suggests that PCP genes are not being thoroughly homogenized by biased gene conversion or other mechanisms. Why PCP genes should be homogenized to a lesser degree than ribosomal genes (or any other gene family) is still unclear. However, it may be related to where PCP genes reside within dinoflagellate chromosomes. Their locations have yet to be mapped.

Suggested experiments to test for adaptation through purifying selection

If PCPs are adapted to harvest specific light wavelengths, then the PCPs from dinoflagellate species living under different light environments should also have absorption maxima that correspond to available light in their specific habitat. For instance, the expectation would be that dinoflagellates living in clear seawater dominated by blue light would have PCP isoforms with different absorption maxima than those from dinoflagellates living in turbid freshwater in which most of the ambient light is green.

It should also be possible to determine if PCPs are under purifying selection to harvesting a range of wavelengths rather than a single wavelength.

The 3.6 nm range of absorption maxima reported by Iglesias-Prieto et al. (1991) for *S. microadriaticum* may be well suited to the fluctuations in the light field most often experienced by this organism. Endosymbiotic dinoflagellates of hosts that live in the shallow water with highly variable light environment should have PCP isoforms with a relatively broad range of absorption maxima, while those from deep water with generally uniform lighting should have a more narrow range of PCP variability.

Furthermore, if PCP isoform polymorphism is adapted to variable light, then a physiological experiment designed to measure the rate of saturation of photosystem II should demonstrate more rapid saturation under fluctuating light wavelengths than when exposed to a single wavelength.

CONCLUSION AND FUTURE DIRECTIONS

This chapter has presented the first phylogenetic analyses of the evolution of dinoflagellate PCP gene families. There has been special emphasis on incorporating new sequence data generated for five *Symbiodinium* species in Chapters One and Two of this dissertation. Together, the results of this dissertation have demonstrated several main points. PCP genes from *Symbiodinium* species occur in diversified families. Genetic heterogeneity is the primary source of PCP isoform variation. Amino acid substitutions within isoforms may affect spectral tuning of peridinins. PCP polymorphism is ancient and present in major lineages of photosynthetic dinoflagellates. Codon sites within PCP genes are evolving under purifying selection and are subjected to net

reduced levels of concerted evolution. The polymorphism likely to be adaptive, but additional experiments are required to substantiate this hypothesis.

It is possible that PCP genes contain sites that are under positive selection that were not detected here. Supplementary work is needed to characterize the PCP genes families from other species. The inclusion PCP gene sequences from additional dinoflagellate species in future phylogenetic analyses will provide a clearer picture of PCP gene evolution. Where possible the use of complete coding sequences will also improve the accuracy and resolution of PCP gene phylogenies and determination of selective pressures affecting PCP evolution.

In situ hybridization experiments could be used to determine the distribution of PCP genes among dinoflagellate chromosomes. PCP genes could potentially be used as molecular markers to identify certain chromosomes or to gauge changes in ploidy level between species. The location of PCP genes on dinoflagellate chromosomes may also be relevant to the levels of concerted evolution that the genes experience.

Finally, studies on the functional significance of amino acid substitution in PCPs will be enhanced by the determination of PCP crystal structures from many more species. The protein modeling presented here was dependent on the single structure that is currently available. Combined comparisons of nucleotide sequences, amino acid sequences and crystal structures from various dinoflagellate species may demonstrate if and how PCPs have been specialized for light harvesting in particular habitats.

Appendices

APPENDIX 1.1 *SYMBIODINIUM* 203 CLONE 41 TANDEM REPEAT. UPSTREAM AND DOWNSTREAM CODING SEQUENCES ARE UNDERLINED. THE PUTATIVE PROMOTER WITHIN THE SPACER IS IN BOLD.

TGGCCGTTTACGATTCAGTGAAAGGCATCACGGACCCCAAAGTGCCAGCTTACATGAAGTCCTTGG
TGAACGGGCCCGATGCTGAGAACGCCTACAAAGCGTTCCTGGAATTCAAGGATGTTGTTGCAAAGA
ACCAGGTGACCACCGCCAGTGCCCCTGCAGCTGTGCCTTCTGGGGACAAGATTGGTGTAGCTGCAA
AAGCGTTGTCCGATGCATCCTATCCTTTCATCAAGGACATCGATTGGCTGTCCGACATTTACCTGA
AGCCGCTGCCCGGCAAGACTGCCCCGGAGACCCTGAAAGCCATTGACAAGATGATCGTGATGGGTG
CCAAGATGGATGGGAACCTCTTGAAGGCAGCAGCAGAGGCACACCACAAGGCCATTGGCAGCATTG
ATGCCAAGGGTGTGACGTCCGCGGCCGACTACGAAGCTGTGAATGCAGCATTGGGGCGCTTGGTGG
CATTCCGTGCCCAAGTCCACTGTCATTGGACGTGTACAATTCCATGGCCAAAGTCGTTGGTTCCAC
CGTGACCAACAACATGTTCTCGAAGGTGAATCCATTGGATGCAGTGGCTGCCGCCAAGGGTTTCTA
CACCTTCAAGGATGTTGTGGAGGCTTCCCAGCGCTGAAGGTGAAGACCATCGGCATGAAGTTTCGA
CATTGCTTTGATGTGCGTTGACCCTTTTTTTAATCTTGTATAGCACGAGCCTCTGAAAATGGAAAA
AGAAGCTCACAATTTGCTTTGTCACCAATGTATTGCCCGCTGCAGCTCCACTGTTTTCCAGATCTT
TCGATATCTTTCAAGACATTTTAAAAGTCATAGCCATGCCTCGATTCAACTAGAGGCGCCAAAGCC
TTGCTCCAACTCCCAACTCACCGACTTGGAGCCGAACACATCCAGCAGTTGAAGCCTCAGCATGAC
ACGATATGACATGATATGACACGATTCCTCCATGTTCTCGCTTTGTTGCTCGTGTACCAATCGAAC
ACCACAGAAGCATCATGAGATGGTGATGAAATGGAAGTGGCCGGTGGCGAGAAACCACCCTGCGTC
ACGTTGAATCTGATCAGTGATTGCATTGTTTTGTTTTGATATGGTAAGCACCCCTTGATAGCTCTT
AGGACTAGATGGACTAGGATCTCACATAGATGTCCAGTGACATAGATTCTGTGTTGATTTCTTGCT
ACAAGATACAGTACGTCTCTGGCAGTGCTTGTCCGTTTGCAATCTGT**CTTGAATGCAGAAT**CGAGA
CAGCAAGAGGTAGGATTTTCGTGGATTTTGGTAACTCAATGATGATAGCGATATGGCCCGGGCCCG
GCCCTGCTTATTTGGTTTATCTGTACTCATCCCAATCGTCCTGCGGTACTTGGTTCAAGTTAGTTT
CAGAGGCTATTTTTTGAGCAACTCCGGCCCACTTTTAGTTTTTGGTGGCCATGGTGCGTGGAGCAA
GGAAAGCTGTTGCTGTCGGAGTTGCGGTGGCTGTTGCCTGCAGTGTCCAGCAGCATTTGAATTTTG
TGCCTGGGCCTCGGCATGCTGCTCCAGTGGCCGCAGCAGCAGCCAGCATGATGATGGCTCCCGCGG
CTTTTGCTGATGAGATCGGCGATGCTGCAAAGAAGCTTGGAGATGCTTCCTACTCTTTTGCCAAGG
AAGTGGATTGGAGCAATGGAATTTTCCTCCAGGCCCCTGGCAAGTTTCAGCCCTTGAAGGCGTTGA
AAGCGATTGACAAGATGATCGAAATGGGGGCAGCTGCCGATCCCAAGCTTCTCAAAGAGGCAGCAG
AAGCACATCACAAGGCCATCGGGAGCATCAGCGGGCCAAATGGTGTGACTTCGCGT

APPENDIX 1.2 ALIGNMENT OF CLONED *SYMBIODINIUM* 203 PCP GENE NUCLEOTIDE SEQUENCES (START AND STOP CODONS ARE UNDERLINED).

```
                    10        20        30        40        50        60
                    .         .         .         .         .         .

203_79       GTGGCCATGGTGCGTGGAGCAAGGAAAGCTGTTGTTGTCGGAGTTGCGGTGGCTGTTGCC    60
203_80       ................................C.......................    60
203_81       ................................C.......................    60
203_82       ---.............................C.......................    57
203_83       ................................C.......................    60
203_84       ................................C....A..................    60
203_85       ................................C.......................    60
203_87       ................................C.......................    60
203_89       .C..............................C.......................    60
203_90       ................................C.......................    60
203cDNA34    ------------------------------....C............A.........    30
203cDNA39    ------------------------------....C............A.........    30
203cDNA42    ------------------------------....C.......................    30
203cDNA55    ------------------------------....C.......................    30
203cDNA59    ------------------------------...TC............A.........    30

                    70        80        90        100       110       120
                    .         .         .         .         .         .

203_79       TGCAGTGTCCAGCAGCATTTGAATTTTGTGCCTGGGCCTCGGCATGCTGCTCCAGTGGCC    120
203_80       ....T.T...G.....................................................    120
203_81       ................CC.....................................................    120
203_82       ................C....C.................................................    117
203_83       ..............CC.....................................................T    120
203_84       .....................................................................T    120
203_85       .....................................................................T    120
203_87       .....................................................................    120
203_89       .....................................................................T    120
203_90       .....................................................................    120
203cDNA34    ............A..........................................................    90
203cDNA39    ............A..........................................................    90
203cDNA42    .....................................................................    90
203cDNA55    ........A........CC.....................................................    90
203cDNA59    ............A..........................................................    90

                    130       140       150       160       170       180
                    .         .         .         .         .         .

203_79       GCAGCAACAGCCAGCATGATGATGGCTCCTGCGGCTTTTGCTGATGAGATTGGCGATACT    180
203_80       .....................................................G..    180
203_81       ......G.......................A.................C......G..    180
203_82       ......G.....................................................G..    177
203_83       ......G..A.................C...................C......G..    180
203_84       ......G.....................................C......G..    180
203_85       ......G.....................................C......G..    180
203_87       ......G.....................................G..    180
203_89       ......G.....................................C......G..    180
203_90       ......G.....................................G..    180
203cDNA34    ......G.....................................G..    150
203cDNA39    ......G.....................................G..    150
203cDNA42    .....................................G..    150
203cDNA55    ......G.....................................T...G..    150
203cDNA59    ......G.....................................G..    150
```

```
                  190       200       210       220       230       240
                  .         .         .         .         .         .

203_79       GCAAAGAAGCTTGGAGATGCTTCCTACTCTTTTGCCAAGGAAGTGGATTGGAACAATGGA    240
203_80       ............................................................    240
203_81       ............................................................    240
203_82       ............................................................    237
203_83       ............................................................    240
203_84       ............G...............................................    240
203_85       ............................................................    240
203_87       ............................................................    240
203_89       ...................................................G........    240
203_90       ............................................................    240
203cDNA34    ............................................................    210
203cDNA39    ............................................................    210
203cDNA42    ............................................................    210
203cDNA55    ............................................................    210
203cDNA59    ............................................................    210

                  250       260       270       280       290       300
                  .         .         .         .         .         .

203_79       ATTTTCCTCCAGGCCCCTGGCAAGTTTCAGCCCTTGGAGGCGTTGAAAGCAATTGACAAG    300
203_80       ............................................................    300
203_81       .................................A.............G.........    300
203_82       ............................................................    297
203_83       ............................................................    300
203_84       ............................................................    300
203_85       ............................................................    300
203_87       ............................................................    300
203_89       ............................................................    300
203_90       ...................................................G........    300
203cDNA34    ............................................................    270
203cDNA39    ............................................................    270
203cDNA42    ............................................................    270
203cDNA55    .................................A..........................    270
203cDNA59    ............................................................    270

                  310       320       330       340       350       360
                  .         .         .         .         .         .

203_79       ATGATCGAAATGGGGGCAGCCGCCGATCCCAAGCTTCTCAAAGAGGCAGCAGAAGCACAT    360
203_80       ............................................................    360
203_81       ............................................................    360
203_82       ............................................................    357
203_83       ............A...............................................    360
203_84       ............................................................    360
203_85       ............................................................    360
203_87       ............................................................    360
203_89       ............................................................    360
203_90       ..................T.........................................    360
203cDNA34    ........................C...................................    330
203cDNA39    ........................C...................................    330
203cDNA42    ............................................................    330
203cDNA55    ............................................................    330
203cDNA59    ............................................................    330
```

```
                  370       380       390       400       410       420
                  .         .         .         .         .         .

203_79       CACAAGGCCATCGGGAGCATCAGCGGACCAAATGGTGTGACTTCGCGTGCTGACTGGGAT    420
203_80       ........................G.................................    420
203_81       ..........T.............G.................................    420
203_82       ........................G.................................    417
203_83       ..........T.............G.................................    420
203_84       ...........................................................    420
203_85       ...........................................................    420
203_87       ........................G.................................    420
203_89       ...........................................................    420
203_90       ..........T.............G.................................    420
203cDNA34    ..........T.............G.................................    390
203cDNA39    ..........T.............G.................................    390
203cDNA42    ..........T.............G.................................    390
203cDNA55    ........................G.................................    390
203cDNA59    ..........T.............G.................................    390

                  430       440       450       460       470       480
                  .         .         .         .         .         .

203_79       GCCGTGAATGCAGCCCTGGGCCGCGTAGTCGCTTCGGTCCCCAAAGCAAAGGTCATGGCC    480
203_80       ................T..........................................    480
203_81       ................T......................G...................    480
203_82       ................T..........................................    477
203_83       ................T..........................................    480
203_84       ................T..........................................    480
203_85       ................T..........................................    480
203_87       ................T..........................................    480
203_89       ................T..........................................    480
203_90       ................T..........................................    480
203cDNA34    ................T.........................G.G..........    450
203cDNA39    ................T.........................G.G..........    450
203cDNA42    ............T..T.........A.........................G.....    450
203cDNA55    ................T..........................................    450
203cDNA59    ................T.........................G............    450

                  490       500       510       520       530       540
                  .         .         .         .         .         .

203_79       GCTTACGATTCAGTGAAAGACATCACGGACCCCACAGTGCCAGCTTACATGAAGTCCTTG    540
203_80       .T..........................................................    540
203_81       .T.............................A............................    540
203_82       ...........................................................    537
203_83       .T.....................................A....................    540
203_84       .T..........................................................    540
203_85       .T..........................................................    540
203_87       .T.........G................................................    540
203_89       .T..........................................................    540
203_90       .T..........................................................    540
203cDNA34    .T.....................................T....................    510
203cDNA39    .T.....................................T....................    510
203cDNA42    .T...........................G..............................    510
203cDNA55    .T.............................A..............................    510
203cDNA59    .T.....................................T....................    510
```

```
                  550       560       570       580       590       600
                  .         .         .         .         .         .

203_79       GTGAACGGGCCCGATGCTGAGAAGGCCTACCAAGCATTCCTGGAATTCAAGGATGTTGTT    600
203_80       ......................C......A.G.................A.........    600
203_81       ......................C......A............................C    600
203_82       ............................A.............................    597
203_83       ..............................................................    600
203_84       ..............................................................    600
203_85       ..............................................................    600
203_87       ..............................................................    600
203_89       ..............................................................    600
203_90       ..............................................................    600
203cDNA34    ..............................................................    570
203cDNA39    ..............................................................    570
203cDNA42    ............................A.............................    570
203cDNA55    ...........T..................................................    570
203cDNA59    ..............................................................    570

                  610       620       630       640       650       660
                  .         .         .         .         .         .

203_79       GCAAAGAACCAGGTGACCACCGCCAGTGCTCCTGCAGTTGTGCCTTCTGGGGACAAGATT    660
203_80       ..............................................................    660
203_81       ..............................................................    660
203_82       ..............................................................    657
203_83       ..............................................................    660
203_84       ....................CG........................................    660
203_85       ..............................................................    660
203_87       ..............................................................    660
203_89       ..G...........................................................    660
203_90       ..............................................................    660
203cDNA34    ..............................................................    630
203cDNA39    ..............................................................    630
203cDNA42    ..............................................................    630
203cDNA55    ..............................................................    630
203cDNA59    ..............................................................    630

                  670       680       690       700       710       720
                  .         .         .         .         .         .

203_79       GGTGTAGCTGCAAAAGCTTTGTCCGATGCATCCTATCCTTTCATCAAGGACATCGATTGG    720
203_80       ................G.....T.....................................    720
203_81       ................G.....................................    720
203_82       ..............................................................    717
203_83       ................G.....................................    720
203_84       ................G.........................A.............    720
203_85       ................G.....................................    720
203_87       ................G.....................................    720
203_89       ..............TG.....................................    720
203_90       ................G.........................C.............    720
203cDNA34    ................G.........................A.............    690
203cDNA39    ................G.........................A.............    690
203cDNA42    ..............................................................    690
203cDNA55    ................G.....................................    690
203cDNA59    ................G.........................A.............    690
```

```
                  730       740       750       760       770       780
                  .         .         .         .         .         .

203_79       CTGTCCGACATTTATCTGAAGCCGCTGCCCGGCAAGACTGCCCCAGAGACCCTGAAAGCC    780
203_80       ............................................................    780
203_81       ..............C.............................................    780
203_82       ............................................................    777
203_83       ............................................................    780
203_84       ............................................................    780
203_85       ............................................................    780
203_87       ............................................................    780
203_89       ..............C..............T...............................    780
203_90       ............................................................    780
203cDNA34    ............................................................    750
203cDNA39    ............................................................    750
203cDNA42    ............................................................    750
203cDNA55    ..............C.............................................    750
203cDNA59    ............................................................    750

                  790       800       810       820       830       840
                  .         .         .         .         .         .

203_79       ATTGACAAGATGATCGTGATGGGCGCCAAGATGGATGGAAACCTCTTGAAGGCAGCAGCA    840
203_80       ............................................................    840
203_81       ............................................................    840
203_82       ............................................................    837
203_83       ............................................................    840
203_84       .......................T....................................    840
203_85       .......................T....................................    840
203_87       ............................................................    840
203_89       .......................T....................................    840
203_90       ............................................................    840
203cDNA34    ............................................................    810
203cDNA39    ............................................................    810
203cDNA42    ............................G...............................    810
203cDNA55    ............................................................    810
203cDNA59    ............................................................    810

                  850       860       870       880       890       900
                  .         .         .         .         .         .

203_79       GAGGCACACCACAAGGCCATTGGCAGCATTGATGCCAAGGGTGTGACGTCCGCGGCCGAC    900
203_80       ............................................................    900
203_81       ............................................................    900
203_82       ............................................................    897
203_83       ............................................................    900
203_84       ..................................C.......C.....T..A......    900
203_85       ..................................C.......C.....T.........    900
203_87       ............................................................    900
203_89       ..................................C.......C.....T.........    900
203_90       ............................................................    900
203cDNA34    ............................................................--    868
203cDNA39    ..........................................................AA..G..--    868
203cDNA42    ............................................................--    868
203cDNA55    ..........................................................A.....--    868
203cDNA59    ..........................................................A...A..--    868
```

```
                  910       920       930       940       950       960
                  .         .         .         .         .         .

203_79       TACGAAGCTGTGTATGCAGCATTGGGGCGCTTGGTGGCATCCGTGCCCAAGTCCACTGTC    960
203_80       ............A.............................................    960
203_81       ............A.............................................    960
203_82       ............A.............................................    957
203_83       ............A.....T.......................................    960
203_84       ............A.............................................    960
203_85       ............A.............................................    960
203_87       ............A.............................................    960
203_89       ............A.............................................    960
203_90       ............A.............................................G    960
203cDNA34    ------------------------------------------------------------    868
203cDNA39    ------------------------------------------------------------    868
203cDNA42    ------------------------------------------------------------    868
203cDNA55    ------------------------------------------------------------    868
203cDNA59    ------------------------------------------------------------    868

                  970       980       990       1000      1010      1020
                  .         .         .         .         .         .

203_79       ATGGACGTGTACAATTCCATGGCCCAAGTCGTTGATTCCACCGTGACCAACAACATGTTC   1020
203_80       ........................A.....................................   1020
203_81       ........................A...........A.........................   1020
203_82       ........................A.....................................   1017
203_83       ........................A.....................................   1020
203_84       ........................A.....................................   1020
203_85       ........................A.....................................   1020
203_87       ........................A.....................................   1020
203_89       ........................A.....................................   1020
203_90       ......................TGGCA...................................   1020
203cDNA34    ------------------------------------------------------------    868
203cDNA39    ------------------------------------------------------------    868
203cDNA42    ------------------------------------------------------------    868
203cDNA55    ------------------------------------------------------------    868
203cDNA59    ------------------------------------------------------------    868

                  1030      1040      1050      1060      1070      1080
                  .         .         .         .         .         .

203_79       TCGAAGGTGAATCCATTGGATGCGATGGCTGCCGCCAAGGGTTTCTACACCTTCAAGGAT   1080
203_80       ........................A.............................T.....   1080
203_81       ........................A.....................T...............   1080
203_82       C..............C.............................................   1077
203_83       ........................A..............A...........T.........   1080
203_84       ........................T.....................................   1080
203_85       ........................A.....................................   1080
203_87       ........................A.....................T.....T.........   1080
203_89       ........................A.....................T...............   1080
203_90       ........................A..............A......................   1080
203cDNA34    ------------------------------------------------------------    868
203cDNA39    ------------------------------------------------------------    868
203cDNA42    ------------------------------------------------------------    868
203cDNA55    ------------------------------------------------------------    868
203cDNA59    ------------------------------------------------------------    868
```

```
                    1090      1100      1110      1120      1130      1140
                    .         .         .         .         .         .

203_79       GTTGTGGAGGCTTCCCAGCGCTGAAGGTGAAGACCATCAGCATGAAGTTTCGACCTTGCT   1140
203_80       ............................................................   1140
203_81       ...........................................C....A.....   1140
203_82       ............................................................   1137
203_83       ............................................................   1140
203_84       ...........................................A.....   1140
203_85       ...........................................A.....   1140
203_87       ............................................................   1140
203_89       ...........................................A.....   1140
203_90       ...........................................A.....   1140
203cDNA34    ------------------------------------------------------------    868
203cDNA39    ------------------------------------------------------------    868
203cDNA42    ------------------------------------------------------------    868
203cDNA55    ------------------------------------------------------------    868
203cDNA59    ------------------------------------------------------------    868

                    1150      1160      1170      1180      1190      1200
                    .         .         .         .         .         .

203_79       TTGACATGCTTTGACACTTTTTTA-AATCTTGTATCTTGTATAGCAGTATGTCACGG---   1196
203_80       ........................-..........................A...G.A---   1196
203_81       ....T....G.....C.......TT.......................------.GA.----   1190
203_82       ........................-.......C.................TCAGGA.TT-   1196
203_83       ........................-.............................-..-----C   1193
203_84       ........................-.......CG.............------GA.CTTT   1193
203_85       ..........C.............-.......CG.............------GA-CTTT   1192
203_87       ............A...........-.........................A...GACTT-   1199
203_89       ........................-.......CG............C------GA-C---   1189
203_90       ........................-.............................--------TC   1191
203cDNA34    ------------------------------------------------------------    868
203cDNA39    ------------------------------------------------------------    868
203cDNA42    ------------------------------------------------------------    868
203cDNA55    ------------------------------------------------------------    868
203cDNA59    ------------------------------------------------------------    868



203_79       GC    1198
203_80       ..    1198
203_81       -.    1191
203_82       AA    1198
203_83       .-    1194
203_84       AA    1195
203_85       AA    1194
203_87       A-    1200
203_89       --    1189
203_90       A-    1192
203cDNA34    --    868
203cDNA39    --    868
203cDNA42    --    868
203cDNA55    --    868
203cDNA59    --    868
```

APPENDIX 1.3 ALIGNMENT OF PREDICTED AMINO ACID SEQUENCES FROM *SYMBIODINIUM* 203 CLONES. TRANSIT PEPTIDE IS UNDERLINED. NUMBERS IN SQUARE PARENTHASES INDICATE LOCATIONS WITHIN APOPROTEINS

```
                        10        20        30        40        50        60[8]
                        .         .         .         .         .         .

203 79aa        MVRGARKAVVVGVAVAVACSVQQHLNFVPGPRHAAPVAAATASMMMAPAAFADEIGDTAK    60
203 80aa        .........A.........IFR.................................A..    60
203 81aa        .........A.............................A.................A..    60
203 82aa        .........A.............S.L.............A.................A..    60
203 83aa        .........A.............................AT................A..    60
203 84aa        .........A.R...........................A.................A..    60
203 85aa        .........A.............................A.................A..    60
203 87aa        .........A.............................A.................A..    60
203 89aa        .........A.............................A.................A..    60
203 90aa        .........A.............................A.................A..    60
203cDNA 34aa    --------.A.............................A.................A..    52
203cDNA 39aa    --------.A.............................A.................A..    52
203cDNA 42aa    --------.A.............................................A..    52
203cDNA 55aa    --------.A.............................A.................A..    52
203cDNA 59aa    --------.S.............................A.................A..    52

                        70[18]    80[28]    90[38]    100[48]   110[58]   120[68]
                        .         .         .         .         .         .

203 79aa        KLGDASYSFAKEVDWNNGIFLQAPGKFQPLEALKAIDKMIEMGAAADPKLLKEAAEAHHK    120
203 80aa        ............................................................    120
203 81aa        ...........................K................................    120
203 82aa        ............................................................    120
203 83aa        ........................................E...................    120
203 84aa        ............................................................    120
203 85aa        ............................................................    120
203 87aa        ............................................................    120
203 89aa        ..............D.............................................    120
203 90aa        ............................................................    120
203cDNA 34aa    ............................................................    112
203cDNA 39aa    ............................................................    112
203cDNA 42aa    ............................................................    112
203cDNA 55aa    ...........................K................................    112
203cDNA 59aa    ............................................................    112

                        130[78]   140[88]   150[98]   160[108]  170[118]  180[128]
                        .         .         .         .         .         .

203 79aa        AIGSISGPNGVTSRADWDAVNAALGRVVASVPKAKVMAAYDSVKDITDPTVPAYMKSLVN    180
203 80aa        ....................................V.....................    180
203 81aa        ............................E.....V..........K..........    180
203 82aa        ............................................................    180
203 83aa        ....................................V.............D.......    180
203 84aa        ....................................V.....................    180
203 85aa        ....................................V.....................    180
203 87aa        ....................................V.....................    180
203 89aa        ....................................V.....................    180
203 90aa        ....................................V.....................    180
203cDNA 34aa    ................................R...V.............V.......    172
203cDNA 39aa    ................................R...V.............V.......    172
203cDNA 42aa    ..................................I........V.V..........A.........    172
203cDNA 55aa    ....................................V..........K..........    172
203cDNA 59aa    ....................................V.............V.......    172
```

```
                    190[138]  200[148]  210[158]  220[168]  230[178]  240[188]
                    .         .         .         .         .         .

203 79aa        GPDAEKAYQAFLEFKDVVAKNQVTTASAPAVVPSGDKIGVAAKALSDASYPFIKDIDWLS    240
203 80aa        .....N..K.....................................................    240
203 81aa        .....N..K.....................................................    240
203 82aa        ........K.....................................................    240
203 83aa        .....................................................................    240
203 84aa        ........................R.....................................    240
203 85aa        .....................................................................    240
203 87aa        .....................................................................    240
203 89aa        ...........................................V.........................    240
203 90aa        ...........................................................Q.......    240
203cDNA 34aa    .....................................................................    232
203cDNA 39aa    .....................................................................    232
203cDNA 42aa    ........K.....................................................    232
203cDNA 55aa    .....................................................................    232
203cDNA 59aa    .....................................................................    232

                    250[198]  260[208]  270[218]  280[228]  290[238]  300[248]
                    .         .         .         .         .         .

203 79aa        DIYLKPLPGKTAPETLKAIDKMIVMGAKMDGNLLKAAAEAHHKAIGSIDAKGVTSAADYE    300
203 80aa        .....................................................................    300
203 81aa        .....................................................................    300
203 82aa        .....................................................................    300
203 83aa        .....................................................................    300
203 84aa        ...............................................Q.....................    300
203 85aa        ...............................................Q.....................    300
203 87aa        .....................................................................    300
203 89aa        ...............................................Q.....................    300
203 90aa        .....................................................................    300
203cDNA 34aa    .....................................................------------    289
203cDNA 39aa    .....................................................KG---------    289
203cDNA 42aa    ........................V.............................---------    289
203cDNA 55aa    .....................................................T.---------    289
203cDNA 59aa    .....................................................TD---------    289

                    310[258]  320[268]  330[278]  340[288]  350[298]  360[308]
                    .         .         .         .         .         .

203 79aa        AVYAALGRLVASVPKSTVMDVYNSMAQVVDSTVTNNMFSKVNPLDAMAAAKGFYTFKDVV    360
203 80aa        ..N.....................K.............................----    356
203 81aa        ..N.....................K...T.........................    360
203 82aa        ..N.....................K..........P..................    360
203 83aa        ..N.S...................K.............................    360
203 84aa        ..N.....................K.............................    360
203 85aa        ..N.....................K.............................    360
203 87aa        ..N.....................K.............................    360
203 89aa        ..N.....................K.............................    360
203 90aa        ..N.....................GI............................    360
203cDNA 34aa    ------------------------------------------------------------    289
203cDNA 39aa    ------------------------------------------------------------    289
203cDNA 42aa    ------------------------------------------------------------    289
203cDNA 55aa    ------------------------------------------------------------    289
203cDNA 59aa    ------------------------------------------------------------    289
```

```
203 79aa        EASQR    365
203 80aa        -----    356
203 81aa        .....    365
203 82aa        .....    365
203 83aa        .....    365
203 84aa        .....    365
203 85aa        .....    365
203 87aa        .....    365
203 89aa        .....    365
203 90aa        .....    365
203cDNA 34aa    -----    289
203cDNA 39aa    -----    289
203cDNA 42aa    -----    289
203cDNA 55aa    -----    289
203cDNA 59aa    -----    289
```

APPENDIX 2.1 ALIGNMENT OF CLONED *S. PILOSUM* PCP GENE NUCLEOTIDE SEQUENCES (START AND STOP CODONS ARE UNDERLINED).

```
                    10        20        30        40        50        60
                    .         .         .         .         .         .

185clone23    GTGGGTCCCAG-TGCATTTTT-AGCCATGGCAAAGACGGCTCGCGCGGTTTCAGTTCTGC    58
185clone25    ......T....-........G.-.....................................    58
185clone28    ...........G-.........T.................................T...    59
185clone30    ...........-.........-.....................................    58
185clone32    ...........G.........-...................A.................    59
185clone33    ...........-.........-.....................................    58
185clone34    .........--AG........-.T...................................    57
185clone35    C........--.G........-.T...................................    57

                    70        80        90        100       110       120
                    .         .         .         .         .         .

185clone23    TTGCGGGAGTTTGCCTCTGCTGCCTCCCGGCCTTCGTGCCGGGGCCTGGTCGCTTTGCTC    118
185clone25    ............................................C.........T..    118
185clone28    ............................................C.............    119
185clone30    ............................................C.............    118
185clone32    ............................................C.............    119
185clone33    ............................................C.............    118
185clone34    ..............................T.............C.............    117
185clone35    ............................................C.............    117

                    130       140       150       160       170       180
                    .         .         .         .         .         .

185clone23    GCAGTGTGGCCCCTGCAGCTGTTGGGGCTGGGGCGCTGGGCATGCTTGGCGCAGCTCCTG    178
185clone25    ................................................................    178
185clone28    ....C.....A.....................................................    179
185clone30    ................................................................    178
185clone32    ....C.....A.....................................................    179
185clone33    ................................................................    178
185clone34    ....C.....A.....................................................    177
185clone35    ....C..A..A.....................................................    177

                    190       200       210       220       230       240
                    .         .         .         .         .         .

185clone23    CCTACGCCGACAAGATCGATGATGCCGCCAAAGTCCTCTCAGAGAAGTCCTATCCTTTCC    238
185clone25    ................................................................    238
185clone28    ................................................................    239
185clone30    ................................................................    238
185clone32    ................................................................    239
185clone33    ................................................................    238
185clone34    ................................................................    237
185clone35    ................................C.....................T.    237

                    250       260       270       280       290       300
                    .         .         .         .         .         .

185clone23    TGAAGGAGATCGATTGGACCTCGGACGTCTATGCCAAGCTCCCCACGCAGCCTCCTCTGA    298
185clone25    ....A.........C..................G.....................    298
185clone28    ....A...........................................................    299
185clone30    ................................................................    298
185clone32    ....A...........................................................    299
185clone33    ................................................................    298
185clone34    ................................................................    297
185clone35    ....A..A.....................G.........G.............    297
```

```
                    310       320       330       340       350       360
                    .         .         .         .         .         .

185clone23    AGGTGATGACAGCCATTGACACAATGCTGAAGATGGGTGCTGCCATGGATCCAGCTGCTC    358
185clone25    ..............C.............................................    358
185clone28    ..............C.............................................    359
185clone30    ............................................................    358
185clone32    ..............C.............................................    359
185clone33    ............................................................    358
185clone34    ..............C.............................................    357
185clone35    ............A..C.............................................    357

                    370       380       390       400       410       420
                    .         .         .         .         .         .

185clone23    TTAAGACGGGTGTTCTGGCCCACAGCCAGGCGATTGCCAACATGGATTCCAAAGGCGTGG    418
185clone25    C.............T.....................................T....    418
185clone28    ............................................................    419
185clone30    ............................................................    418
185clone32    ............................................................    419
185clone33    .......C....................................................    418
185clone34    ............................................................    417
185clone35    .......C.....................C.....................T....    417

                    430       440       450       460       470       480
                    .         .         .         .         .         .

185clone23    CGACGCTGGCCGATTACACCGCCATCAACTCAGCGATTGGTCACATGATCTCTTCAGTGC    478
185clone25    ....C.........T.........................T.................    478
185clone28    ....C............G.....................T.................    479
185clone30    ..........................................G.................    478
185clone32    ....C.........T.........................T.................    479
185clone33    ............................................................    478
185clone34    ....C.........T................A........T.................    477
185clone35    ..............T................A.........................    477

                    490       500       510       520       530       540
                    .         .         .         .         .         .

185clone23    CCGCCTCGAAGACCATGGATGTTTACAATGCGTTCGCCAAGTTTAACCTTGGCTCAGACG    538
185clone25    ...........................................G........G....    538
185clone28    ............................................................    539
185clone30    ............................................................    538
185clone32    ............................................................    539
185clone33    .....C......................................................    538
185clone34    ..........................................................G....    537
185clone35    ..........................................T........G....    537

                    550       560       570       580       590       600
                    .         .         .         .         .         .

185clone23    TTGGCCCCTACATGATGAGCAAGGTGAATGCCGCAGATGCCCAGGCAGCTTACAAGGCAC    598
185clone25    .......T.....................T..........................    598
185clone28    .......T.....................T..........................    599
185clone30    ............................T..........................    598
185clone32    .......T.....................T..........................    599
185clone33    .......T.....................T..........................    598
185clone34    .......T.....................T..........................    597
185clone35    .......T.....................T..........................    597
```

```
                    610       620       630       640       650       660
                    .         .         .         .         .         .

185clone23    TCATGGACTTCAAGGATGTTGTCAAAGCCTCCCAGCGCTGAGCCCCAAACTCAGAAACCT    658
185clone25    ..............................................A...........T..    658
185clone28    ..............................................G.......TAG    659
185clone30    ............................................................    658
185clone32    ..............................................G.......TAG    659
185clone33    ............................................................    658
185clone34    ...................A......................A...........T..    657
185clone35    ............................................................T..    657

                     670
                     .

185clone23    TCTTGAAACAA    669
185clone25    ...........    669
185clone28    .....C....-    669
185clone30    ..........-    668
185clone32    ...........    670
185clone33    ...........    669
185clone34    ...-......-    666
185clone35    ..........-    667
```

APPENDIX 2.2 ALIGNMENT OF CLONED *SYMBIODINIUM SP. FROM D. STOKESII AND S. PULCHRORUM* PCP GENE NUCLEOTIDE SEQUENCES

```
                    10        20        30        40        50        60
                    .         .         .         .         .         .

28clone5     ATGATCCAGATGGGAGCCGCCGCTGATCCTGAATTGTTGAAGAAGGCAGCTGATGCTCAC    60
28clone1     ------------------------------------------------------------    0
28clone2     ------------------------------------------------------........    8
32clone1     ......G.....................................................    60
35clone3     ......G.....................................................    60
35clone42    ......G.....................................................    60
35clone43    ......G.....................................................    60
35clone44    ......G.....................................................    60
35clone46    ......G.....................................................    60
35clone47    ......G.....................................................    60
35clone49    ......G.....................................................    60
8clone1      ------------------------------------------------------------    0
8clone2      ------------------------------------------------------------    0
8clone3      ------------------------------------------------------------    0
8clone4      ------------------------------------------------------------    0

                    70        80        90        100       110       120
                    .         .         .         .         .         .

28clone5     CACAAAGCCATCGGCAGCATTAGTGGACCAAATGGAGTAACTTCTCGTGCTGACTGGGAT    120
28clone1     ------------------------------------------------------------    0
28clone2     ............................................................    68
32clone1     ............................................................    120
35clone3     ............................................................    120
35clone42    ............................................................    120
35clone43    ...........................................T................    120
35clone44    ...........................................T................    120
35clone46    ............................................................    120
35clone47    ............................................................    120
35clone49    ............................................................    120
8clone1      ------------------------------------------------------------    0
8clone2      ------------------------------------------------------------    0
8clone3      ------------------------------------------------------------    0
8clone4      ------------------------------------------------------------    0

                    130       140       150       160       170       180
                    .         .         .         .         .         .

28clone5     GCAGTGAATGCGGCCCTTGGCCGTGTGGTCGCTTCTGTCCCAAAGCAAAAGGTCATGGAC    180
28clone1     ------------------------------------------------------.....    5
28clone2     ............................................................    128
32clone1     ............................................................    180
35clone3     ............................................................    180
35clone42    ...A........................................................    180
35clone43    ...AA..................A..A.................................    180
35clone44    ...CA......A...........A......T..............................    180
35clone46    ..............T.............................................    180
35clone47    ............A...............................................    180
35clone49    ...........A................................................    180
8clone1      ------------------------------------------------------------...    3
8clone2      ------------------------------------------------------------...    3
8clone3      ------------------------------------------------------------...    3
8clone4      ---------------------------------------------------------TGCC..    6
```

```
                  190       200       210       220       230       240
                  .         .         .         .         .         .

28clone5     GTCTATAATGCAGTCAAAGACATCACCGACCCCAAGGTGCCAGCTTACATGAAGTCCTTG    240
28clone1     ......G.....................................................    65
28clone2     ......G.....................................................    188
32clone1     ......G.....................................................    240
35clone3     ......G.....................................................    240
35clone42    ......G.............G.......................................    240
35clone43    ......G.....................................................    240
35clone44    ......G.....T..............................T................    240
35clone46    ......G.....................................................    240
35clone47    ......G.....................................................    240
35clone49    ......G.....................................................    240
8clone1      ......G.....................................................    63
8clone2      ......G..............................C......................    63
8clone3      ......G.....................................................    63
8clone4      ......G..............................C......................    66

                  250       260       270       280       290       300
                  .         .         .         .         .         .

28clone5     GTGAACGGTGCAGATGCGGAGAAAGCCTACCAAGGTTTCTTGGAGTTCAAGGATGCTGCG    300
28clone1     ..........................................................T..T.    125
28clone2     ....G.............G.....................................T..T.    248
32clone1     ..........................................................T..T.    300
35clone3     ..........................................................T..T.    300
35clone42    ..A.......................................................T..T.    300
35clone43    ..........................................................T..T.    300
35clone44    ..........................................................T..T.    300
35clone46    ..........................................................T..G.    300
35clone47    ..........................................................T..T.    300
35clone49    ..........................................................T..T.    300
8clone1      ..........................................................T..T.    123
8clone2      ...........G..............................................T..T.    123
8clone3      ..........................................................T..T.    123
8clone4      ................A.........................................T..TC    126

                  310       320       330       340       350       360
                  .         .         .         .         .         .

28clone5     GCCGCAAACCAGGTGACCACCGCTAGCGCTGCTGCCACGGTGCCTACTGGAGACAAGATT    360
28clone1     ..........T................................C..............    185
28clone2     ...........................................................    308
32clone1     ..........................................C..............    360
35clone3     ..........................................C..............    360
35clone42    ..........................................C....A..........    360
35clone43    ...........................................................    360
35clone44    ...........T..................T............................    360
35clone46    ......................A....................................    360
35clone47    ...........................................................    360
35clone49    ...........................................................    360
8clone1      ....................................T........A............    183
8clone2      ....................................T........A............    183
8clone3      ....................................T........A............    183
8clone4      .....G.............T.................A........A............    186
```

```
                  370       380       390       400       410       420
                  .         .         .         .         .         .

28clone5     GGCACAGCTGCTAAAGCTTTGTCCGATGCATCATATCCATTCATCAAGGACATCGATTGG    420
28clone1     ..............................................................    245
28clone2     ..............................................................    368
32clone1     ..............................................................    420
35clone3     ..............................................................    420
35clone42    ....................................A.........................    420
35clone43    ................................T.........A.....A......        420
35clone44    ..................A...........................................    420
35clone46    ..................A...........................................    420
35clone47    ..................A...........................................    420
35clone49    ..................A...........................................    420
8clone1      .....G........................................................    243
8clone2      ...G.G........................................................    243
8clone3      .....G........................................................    243
8clone4      .....G......................C..C.....................          246

                  430       440       450       460       470       480
                  .         .         .         .         .         .

28clone5     CTGTCAGACATTTACCTGAAGCCATTGCCTGGCAAGACTGCCCCAGAGACCTTGACAGCC    480
28clone1     ............................G.............................    305
28clone2     .........G.....................................A....          428
32clone1     ..............................................................    480
35clone3     ..............................................................    480
35clone42    ..........................................G...........        480
35clone43    .........G.....................................A....          480
35clone44    .........G.....................................A....          480
35clone46    .........G.....................................A....          480
35clone47    .........G.....................................A....          480
35clone49    .........G.....................................A....          480
8clone1      ..............................................A....          303
8clone2      ..........................................G....C...A...T        303
8clone3      ..............................................A....          303
8clone4      ..............................................CA....          306

                  490       500       510       520       530       540
                  .         .         .         .         .         .

28clone5     ATCGACAAGATGATCGTCATGGGAAGCAAAATGGATGGCAACCTCTTGAAGGCAGCCGCT    540
28clone1     ....................................G.........................    365
28clone2     .................................C............................    488
32clone1     ..............................................................    540
35clone3     ..............................................................    540
35clone42    ............................C.................              540
35clone43    .................................C............................    540
35clone44    ..............................................................    540
35clone46    ..............................................................    540
35clone47    ..............................................................    540
35clone49    ..............................................................    540
8clone1      ...............G..............................                  363
8clone2      ...............G..............................                  363
8clone3      ...............G..............................                  363
8clone4      ..............................................................    366
```

```
                  550       560       570       580       590       600
                  .         .         .         .         .         .

28clone5     GAGGCACACCACAAGGCCATTGGCAGCATTGATGCCAAGGGTGTCACATCTCCAGCAGAC    600
28clone1     ..........................C.............................    425
28clone2     ..............................................................    548
32clone1     ..............................................................    600
35clone3     ..............................................................    600
35clone42    ............T.............................G...........G...    600
35clone43    ..............................................................    600
35clone44    ..............................................................    600
35clone46    ..............................................................    600
35clone47    ..............................................................    600
35clone49    ..............................................................    600
8clone1      ........................T.....................................    423
8clone2      ..............................................................    423
8clone3      ..............................................................    423
8clone4      ...........................................T..........A....    426

                  610       620       630       640       650       660
                  .         .         .         .         .         .

28clone5     TACGAAGCGGTGAACGCAGCTTTGGGCCGTATTGTGGCGTCGGTGCCAAAACAAACCGTC    660
28clone1     ..............................................................    485
28clone2     ...........................................A..................    608
32clone1     ..............................................................    660
35clone3     ..............................................................    660
35clone42    ..............................................................    660
35clone43    ..............................................................    660
35clone44    ........A.............................A.....................    660
35clone46    ..............................................................    660
35clone47    ..............................................................    660
35clone49    ..............................................................    660
8clone1      ..............................................G...........    483
8clone2      ..............................................................    483
8clone3      ..............................................................    483
8clone4      ..............................................................    486

                  670       680       690       700       710       720
                  .         .         .         .         .         .

28clone5     ATGGATGTGTACAACTCAATGGCAAAGATTGTGGACCCAAGCGTGACCAACAACATGTTC    720
28clone1     ..............................................................    545
28clone2     ..............................................................    668
32clone1     ..............................................................    720
35clone3     ..............................................................    720
35clone42    ................................T.............................    720
35clone43    ............................G.................................    720
35clone44    ................................T.............................    720
35clone46    .C........................G.....................................    720
35clone47    ............................G.....................................    720
35clone49    ............................G.....................................    720
8clone1      ............G........................T.........................    543
8clone2      ............G........................T.........................    543
8clone3      ............G........................T.........................    543
8clone4      ...........................................T..................    546
```

```
                  730       740       750       760       770       780
                  .         .         .         .         .         .

28clone5     TCTAAGGTCAACCCCTTGGATGCATTGTCTGCTGCCAAGGGCTTTTACACCTTCAAGGAT    780
28clone1     ..............................-------------------------    579
28clone2     ............................................................    728
32clone1     ............................................................    780
35clone3     ...............................................T............    780
35clone42    ............................................................    780
35clone43    ............................C...............................    780
35clone44    ............................G...................T............    780
35clone46    ............................C...............................    780
35clone47    ..C................................G............T............    780
35clone49    ..C................................G............T............    780
8clone1      ..G.C...................C.............A....-----------------    586
8clone2      ..G.C...................C.............A...........A.....----    599
8clone3      ..G.C...................C.............A...........A.....----    599
8clone4      ..G.C.........T.......................----------------------    584

                       790       800
                       .         .

28clone5     GTTGTGGAAGCTGTCCAGCGTTAA    804
28clone1     ------------------------    579
28clone2     ........................    752
32clone1     ................A..C...    804
35clone3     ...................C..G    804
35clone42    ...................C...    804
35clone43    ...................C...    804
35clone44    ...................C..G    804
35clone46    ...................C...    804
35clone47    ...................C..G    804
35clone49    ...................C..G    804
8clone1      ------------------------    586
8clone2      ------------------------    599
8clone3      ------------------------    599
8clone4      ------------------------    584
```

APPENDIX 2.3 ALIGNMENT OF CLONED *S. KAWAGUTII* PCP GENE NUCLEOTIDE SEQUENCES (START AND STOP CODONS ARE UNDERLINED).

```
                    10        20        30        40        50        60
                    .         .         .         .         .         .

135clone37    -CGGTGGCCATGGTGCGTGGAGCAAGGAAAGCTGTTGCTGTCGGAGTTGCGGTGGCTGTT    59
135clone38    -T.....................................................C...    59
135clone41    -T.....................................................    59
135clone42    G.....................................................    60
135clone43    -T......-.....................................................    58
135clone49    -.....................................................    59
135clone50    -T.-.....................................................T........    58
135clone52    -T.....................................C.....................    59
135clone54    -T.....................................................    59
135clone60    ---------.....................................................    51

                    70        80        90        100       110       120
                    .         .         .         .         .         .

135clone37    GACTGCAGTGTCCAGCAGCACTTGAATTTTGTGCCTGGGCCTCGGCGTGCTGCTCCAGTG    119
135clone38    .C..............A.............T.............A......C......    119
135clone41    .C...........T.............................A.............    119
135clone42    .C.................................A.......A.............    120
135clone43    .C.............A.........................A.............    118
135clone49    .C.........................C......A.......A.............    119
135clone50    .C.........................................A.............    118
135clone52    .C.........................................A.............    119
135clone54    .C.............A.............T.............A......C......    119
135clone60    .C.................................A.......A.............    111

                    130       140       150       160       170       180
                    .         .         .         .         .         .

135clone37    GCCGCAGCAGCAGCCAGCATGATGATGGCTCCTGCGGCTTTTGCTGATGAGATCGGCGAT    179
135clone38    .....................................................    179
135clone41    .....................................................T..T...    179
135clone42    .....................................................T......    180
135clone43    ........................A.............................    178
135clone49    .....................................................T......    179
135clone50    ..................G..............................T......    178
135clone52    .....................................................T......    179
135clone54    ..................A..................................    179
135clone60    .....................................................T......    171

                    190       200       210       220       230       240
                    .         .         .         .         .         .

135clone37    GCTGCAAAGAAGCTTGGAGATGCTTCCTACTCTTTTGCCAAGGAAGTGGATTGGAACAAT    239
135clone38    .....................................................    239
135clone41    .....................................................    239
135clone42    ........................T.........................G.....    240
135clone43    .....................................................    238
135clone49    ..................................T..............T.....    239
135clone50    .....................................................    238
135clone52    ..........G..........................................    239
135clone54    ............C........................................    239
135clone60    .....G...........G............C.....................    231
```

```
                    250       260       270       280       290       300
                    .         .         .         .         .         .

135clone37    GGAATTTTCCTCCAGGCCCCTGGCAAGTTTCAGCCCTTGGAGGCGTTGAAAGCAATTGAC    299
135clone38    ....................................A...................    299
135clone41    ....................................A...................    299
135clone42    ................................C..G...A...................    300
135clone43    ......C.........T.....................A...................    298
135clone49    ..............T.....................A...................    299
135clone50    ......C.................................................    298
135clone52    ....................................A...................    299
135clone54    ....................................A...................    299
135clone60    .....................................A..................    291

                    310       320       330       340       350       360
                    .         .         .         .         .         .

135clone37    AAGATGATCGAAGTGGGGGCAGCCGCCGATCCCAAGCTTCTCAAAGAGGCAGCAGAAGCA    359
135clone38    ............A.....................T.........................    359
135clone41    ............A.....................................................    359
135clone42    ............A.....................................................G    360
135clone43    ............A.....................................................    358
135clone49    ............A.....................................................    359
135clone50    ............A.....................................G.......    358
135clone52    ............A.....................................................    359
135clone54    ............A.............................G.................    359
135clone60    ............A.....................................................    351

                    370       380       390       400       410       420
                    .         .         .         .         .         .

135clone37    CATCACAAGGCCACTGGGAGCATCAGCGGGCCAAATGGTGTGACTACGCGTGCTGACTGG    419
135clone38    .............................................T..............    419
135clone41    ............T.....................C.........T..............    419
135clone42    ............GT..............................T.T............    420
135clone43    ...........TTC...............................T..............    418
135clone49    ............TC..............................T.............T    419
135clone50    ...........TTC.....................C.........T.A............    418
135clone52    ............TC....................C.........T....C.........    419
135clone54    ............TC..A.............................T..............    419
135clone60    ...........T.T...............................T..............    411

                    430       440       450       460       470       480
                    .         .         .         .         .         .

135clone37    GATGCCGTGAATGCAGCCCTTGGCCGCGTAGTCGCTTCGGTCCCCAAAGCAAAGGTCATG    479
135clone38    ................................................................    479
135clone41    ................................................................    479
135clone42    ................................................................    480
135clone43    ........AG..........................A.....................    478
135clone49    ....................................A.....................    479
135clone50    ................................................A..........    478
135clone52    ................................................................    479
135clone54    ................................................................    479
135clone60    ................................................................    471

                    490       500       510       520       530       540
                    .         .         .         .         .         .

135clone37    GCCGTTTACGATTCAGTGAAAGACATCACGGACCCCGCAGTGCCAGCTTACATGAAGTCT    539
135clone38    ....................................AA.....................C    539
135clone41    ....................................AA.....................C    539
135clone42    ....................................AA.....................C    540
135clone43    ................................................................C    538
135clone49    ................C...................AA.....................C    539
135clone50    ....................................AA.....................C    538
135clone52    ....................................AA.....................C    539
135clone54    ....................................AA.....................C    539
135clone60    ....................................AA.....................C    531
```

```
                    550       560       570       580       590       600
                    .         .         .         .         .         .

135clone37    TTGGTGAACGGGCCCGATGCTGAGAAGGCCTACCAAGCATTCCTGGAATTCAAGGATGTT    599
135clone38    ........................................T.................    599
135clone41    ..........T..T.....................G.....................    599
135clone42    ..........T..T.....................G.....................    600
135clone43    .....................................T.........A......    598
135clone49    ............................................................    599
135clone50    ...........T................................A...........    598
135clone52    .....................................G......................    599
135clone54    .....................................T.................    599
135clone60    ............................................................    591

                    610       620       630       640       650       660
                    .         .         .         .         .         .

135clone37    GTTGCAAAGAACCAGGTGACCACCGCCAGGGCTCCTGCAGTTGTGCCTTCTGGAGACAAG    659
135clone38    ..........................T.......................G......    659
135clone41    ................G..G......T............C.........G......    659
135clone42    ...................G......T.....................G......    660
135clone43    ..........................T..........C..........G......    658
135clone49    .....T.................A....T.....................G......    659
135clone50    ......................T.....T.....................G......    658
135clone52    ..C................G......T.....................G......    659
135clone54    ............C.............T................T...G......    659
135clone60    .A........................T.....................G......    651

                    670       680       690       700       710       720
                    .         .         .         .         .         .

135clone37    ATTGGTGAAGCTGCAAAAGCGTTGTCCGATGCATCCTATCCTTTCATCAAGGACATCGAT    719
135clone38    G......T..............................A.................    719
135clone41    .......T.................................................T...    719
135clone42    .......T.................................................T...    720
135clone43    ................G.........................................    718
135clone49    .......T..............................A.................    719
135clone50    ...................................T...A........A.........    718
135clone52    .......T.....T.............................................    719
135clone54    G....A.TGT..............................A.................    719
135clone60    .......G..............................A.............C..    711

                    730       740       750       760       770       780
                    .         .         .         .         .         .

135clone37    TGGCTGTCCGACATCTACCTGAAGCCGCTGCCCGGCAAGACTGCCCCAGAGACCCTGAAA    779
135clone38    ........................................G..............    779
135clone41    ........G..............................C........C.........    779
135clone42    ........G.....................................C.........    780
135clone43    ........G.....T.............................................    778
135clone49    ........G................A.................G..............    779
135clone50    ........G.....................................................    778
135clone52    ........G.....................................C.........    779
135clone54    ........................................G..............    779
135clone60    ........G.....T........A.....................................    771

                    790       800       810       820       830       840
                    .         .         .         .         .         .

135clone37    GCCATTGACAAGATGATCGTGATGGGCGCCAAGATGGATGGTAACCTCTTGAAGGCAGCA    839
135clone38    ........................TT.............A.................    839
135clone41    ........................................A.................    839
135clone42    ........................................A.................    840
135clone43    ..............................A.........A.................    838
135clone49    ........................................A.................    839
135clone50    ........................................A.................    838
135clone52    .....C..................................A.................    839
135clone54    ..............G..........TT.............A..T..............    839
135clone60    ........................................A.................    831
```

```
                  850       860       870       880       890       900
                  .         .         .         .         .         .

135clone37    GCAGAGGCACACCACAAGGCCATTGGCAGCATTGATGCCAAGGGTGTGACGTCCGCGGCC    899
135clone38    ......................C.........................A.........    899
135clone41    ......................C.....................................    899
135clone42    ......................C.....................................    900
135clone43    ......................C.................C..............C...    898
135clone49    .T....................C............A..................A...    899
135clone50    ........................................C.................    898
135clone52    ....................T.....T.............................    899
135clone54    ......................C.........................A.........    899
135clone60    ........................................C.................    891

                  910       920       930       940       950       960
                  .         .         .         .         .         .

135clone37    GACTACGAAGCTGTGAATGCAGCATTGGGGCGCTTGGTGGCATCCGTGCCCAAGTCCACT    959
135clone38    ............................................................    959
135clone41    ............................................................A   959
135clone42    ............................................................A   960
135clone43    ..........................................................A...   958
135clone49    ............................................................    959
135clone50    ............................................................    958
135clone52    C...........................................................    959
135clone54    ............................................................    959
135clone60    ............................................................    951

                  970       980       990       1000      1010      1020
                  .         .         .         .         .         .

135clone37    GTCATGGACGTGTACAATTCCATGGCTGGCATCGTTGATTCCACCGTGACCAACAACATG    1019
135clone38    ........................GAAAG.............................    1019
135clone41    ............................G...........C.................    1019
135clone42    ............................................................    1020
135clone43    ..........................................T.................    1018
135clone49    ........................CAAAG.............................    1019
135clone50    ........................CAAAG.............................    1018
135clone52    ..........T................................G..........T...    1019
135clone54    ........................GAAAG.............................    1019
135clone60    ........................CAAAG.............................    1011

                  1030      1040      1050      1060      1070      1080
                  .         .         .         .         .         .

135clone37    TTTTCGAAGGTGAGCCCATTGGATGCAATGGCTGCCGCCAAGGGTTTTTACACCTTCAAG    1079
135clone38    ..CC.........AT.............................................    1079
135clone41    ..C..........AT..............T...............................    1079
135clone42    ..C..........AT............G.................................    1080
135clone43    ..C..........AT.............................................    1078
135clone49    ..C..........AT....................T........................    1079
135clone50    ..C.....T....AT..............T..C.........................A    1078
135clone52    ..C..........AT.............................................    1079
135clone54    ..C..........AT.............................................    1079
135clone60    ..C.T........AT...C...........................C..........G.    1071

                  1090      1100      1110      1120      1130      1140
                  .         .         .         .         .         .

135clone37    GATGTTGTGGAGGCTTCCCAGCGCTGAAGGTGAAGACCATCAGCAGGAAGTTTCGACATT    1139
135clone38    ..............................................T.............    1139
135clone41    ..............................................T............A    1139
135clone42    ..........A..............G..................T.............    1140
135clone43    ................T...........................C.............    1138
135clone49    ..............................................T.............    1139
135clone50    ..............................................T.............    1138
135clone52    ..............................................T.............    1139
135clone54    ..............................................T.............    1139
135clone60    ............G...T............................T.............    1131
```

```
                    1150      1160      1170      1180      1190      1200
                    .         .         .         .         .         .

135clone37    GCTTT-GACATGCTTTAACAGAATTTTAAATCTTGCATCTGCCTTTGCAATTGCATTGCC    1198
135clone38    .....-...........CA.CTT....G.......T.....------------------    1178
135clone41    .....-.............ATT....-----.--A..AA------------------    1171
135clone42    .....-..........G..CCTT....-----..AA..A.------------------    1174
135clone43    .....-.............CTT...A....G..----------------------------    1171
135clone49    .....-.............CTT...----------------------------------    1164
135clone50    .....-...........G..CTT...-........T.....------------------    1176
135clone52    .....T..............-CT....-----..AA.....------------------    1173
135clone54    .....-.............CTT............------------------------    1173
135clone60    ......A............CTT............T.....------------------    1170

                    1210      1220      1230      1240      1250      1260
                    .         .         .         .         .         .

135clone37    ATCTTTGTAATTGCATTGTCATGCAGCGTCGCAGAGTGAGTGAGATTTGTTTAAGATATT    1258
135clone38    -------...-------------------------------------------------    1182
135clone41    -------.-------------------------------------------------    1172
135clone42    -------.-------------------------------------------------    1175
135clone43    -------------------------------------------------------------    1171
135clone49    -------------------------------------------------------------    1164
135clone50    -------.-------------------------------------------------    1177
135clone52    -------.-------------------------------------------------    1174
135clone54    -------------------------------------------------------------    1173
135clone60    -------.-------------------------------------------------    1171

                    1270      1280
                    .         .

135clone37    GTATCTTGTATAGCAAAAGC    1278
135clone38    ....................    1202
135clone41    ....................    1192
135clone42    ....................    1195
135clone43    ...A................    1191
135clone49    -A.............----    1179
135clone50    -------.........-----    1185
135clone52    ....................    1194
135clone54    -...........-------    1185
135clone60    -------...C---------    1175
```

APPENDIX 2.4 ALIGNMENT OF PREDICTED AMINO ACID SEQUENCES FROM *S. PILOSUM* CLONES. TRANSIT PEPTIDE IS UNDERLINED. NUMBERS IN SQUARE PARENTHASES INDICATE LOCATIONS WITHIN APOPROTEINS

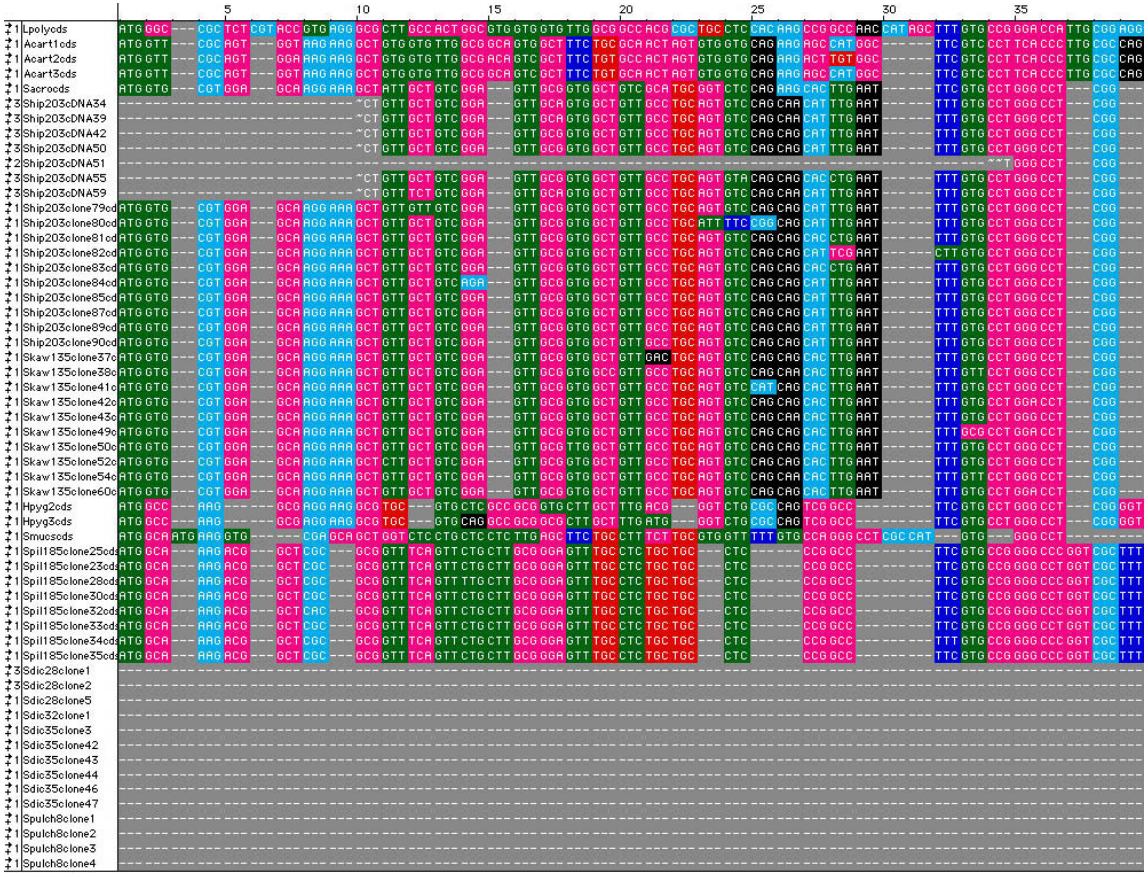
```
                    10        20        30        40        50        60[6]
                    .         .         .         .         .         .

185clone23aa    MAKTARAVSVLLAGVCLCCLPAFVPGPGRFARSVAPAAVGAGALGMLGAAPAYADKIDDA    60
185clone25aa    ............................V............................    60
185clone28aa    ................................................................    60
185clone30aa    ................................................................    60
185clone31aa    .....H..........................................................    60
185clone33aa    ................................................................    60
185clone34aa    ................................................................    60
185clone35aa    ................................................................    60

                    70[16]    80[26]    90[36]    100[46]   110[56]   120[66]
                    .         .         .         .         .         .

185clone23aa    AKVLSEKSYPFLKEIDWTSDVYAKLPTQPPLKVMTAIDTMLKMGAAMDPAALKTGVLAHS    120
185clone25aa    ................R.....G.............................P........    120
185clone28aa    ................................................................    120
185clone30aa    ................................................................    120
185clone31aa    ................................................................    120
185clone33aa    ................................................................    120
185clone34aa    ................................................................    120
185clone35aa    ...P.................G...A..............................    120

                    130[76]   140[86]   150[96]   160[106]  170[116]  180[126]
                    .         .         .         .         .         .

185clone23aa    QAIANMDSKGVATLADYTAINSAIGHMISSVPASKTMDVYNAFAKFNLGSDVGPYMMSKV    180
185clone25aa    ............................................K.............    180
185clone28aa    ................................................................    180
185clone30aa    ........................R.......................................    180
185clone31aa    ................................................................    180
185clone33aa    ................................P.......................    180
185clone34aa    ................................................................    180
185clone35aa    ................................................................    180

                    190[136]  200[146]
                    .         .

185clone23aa    NAADAQAAYKALMDFKDVVKASQR.    205
185clone25aa    .........................    205
185clone28aa    .........................    205
185clone30aa    .........................    205
185clone31aa    .........................    205
185clone33aa    .........................    205
185clone34aa    ..................I......    205
185clone35aa    .........................    205
```

APPENDIX 2.5 ALIGNMENT OF PREDICTED AMINO ACID SEQUENCES FROM *S. KAWAGUTII* CLONES. TRANSIT PEPTIDE IS UNDERLINED. NUMBERS IN SQUARE PARENTHASES INDICATE LOCATIONS WITHIN APOPROTEINS

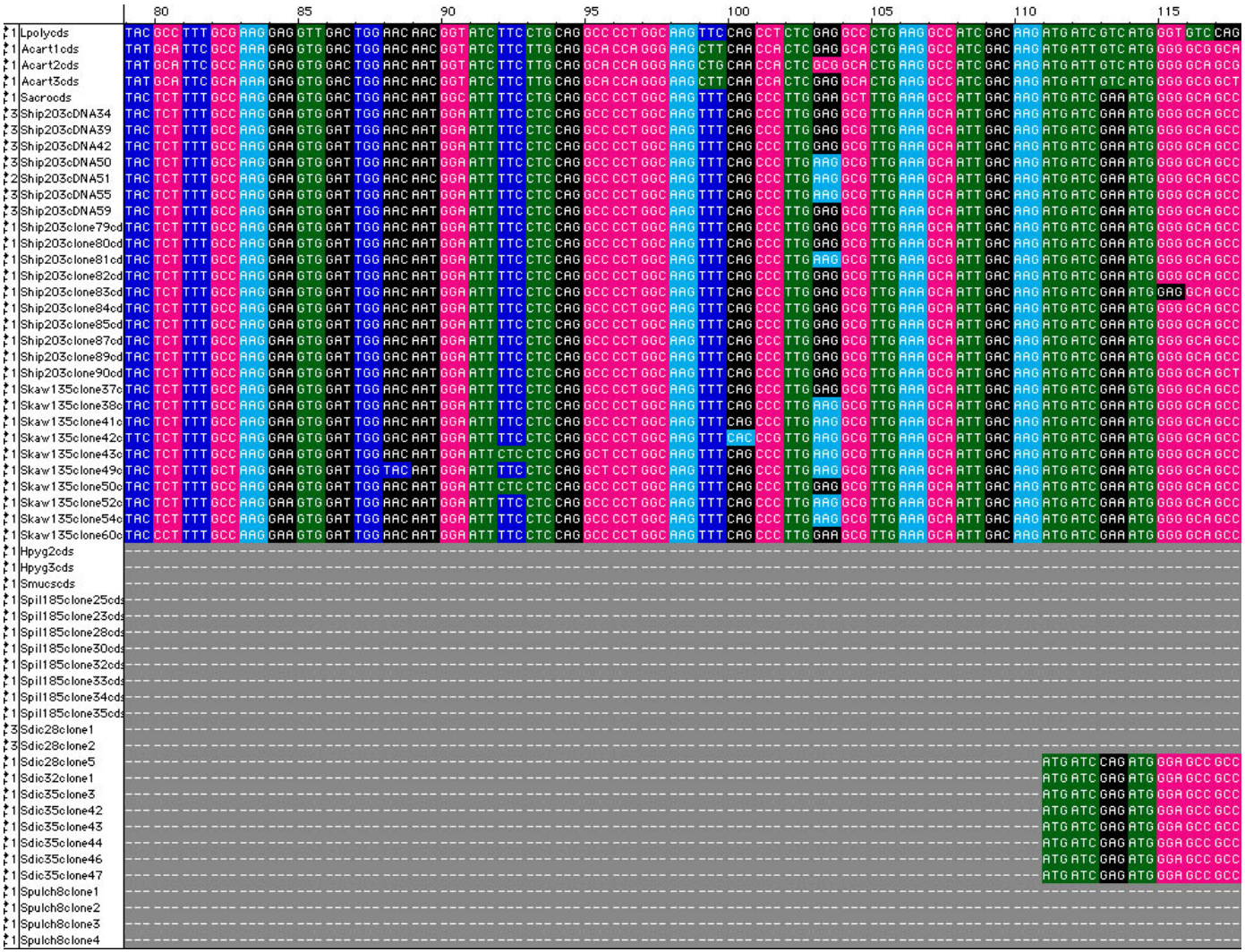
```
                        10        20        30        40        50        60
                        .         .         .         .         .         .

135clone37aa    MVRGARKAVAVGVAVAVDCSVQQHLNFVPGPRRAAPVAAAAASMMMAPAAFADEIGDAAK    60
135clone38aa    ................A.............H..........................    60
135clone41aa    ................A...H.........H..........................    60
135clone42aa    ................A.............H..........................    60
135clone43aa    ................A.............H...........I..............    60
135clone49aa    ................A.........A....H..........................    60
135clone50aa    ............L..A.............H..........V...............    60
135clone52aa    ........L........A.............H..........................    60
135clone54aa    ................A.............H.........I................    60
135clone60aa    ................A.............H..........................    60

                        70        80        90        100       110       120
                        .         .         .         .         .         .

135clone37aa    KLGDASYSFAKEVDWNNGIFLQAPGKFQPLEALKAIDKMIEVGAAADPKLLKEAAEAHHK    120
135clone38aa    ............................K.........M......M...........    120
135clone41aa    ............................K.........M.................    120
135clone42aa    ......F........D...........H..K.........M.................    120
135clone43aa    ..................L.........K.........M.................    120
135clone49aa    ..............Y.............K.........M.................    120
135clone50aa    ..................L.....................M............G.....    120
135clone52aa    R...........................K.........M.................    120
135clone54aa    .P..........................K.........M.........R........    120
135clone60aa    .......P..................................M.................    120

                        130       140       150       160       170       180
                        .         .         .         .         .         .

135clone37aa    ATGSISGPNGVTTRADWDAVNAALGRVVASVPKAKVMAVYDSVKDITDPAVPAYMKSLVN    180
135clone38aa    ............S.....................................K.........    180
135clone41aa    .I.........S.....................................K.........    180
135clone42aa    .V.........S.....................................K.........    180
135clone43aa    .F.........S.......D.....................................    180
135clone49aa    .I.........S...C.........................A......K.........    180
135clone50aa    .F.........S....................E................K.........    180
135clone52aa    .I.........S.....................................K.........    180
135clone54aa    .I.........S.....................................K.........    180
135clone60aa    .I.........S.....................................K.........    180

                        190       200       210       220       230       240
                        .         .         .         .         .         .

135clone37aa    GPDAEKAYQAFLEFKDVVAKNQVTTARAPAVVPSGDKIGEAAKALSDASYPFIKDIDWLS    240
135clone38aa    ..........................S.........V.V.....................    240
135clone41aa    .........G.............SS.S....A.......V.....................    240
135clone42aa    .........G.............S.S...........V.....................    240
135clone43aa    ..........................S...............................    240
135clone49aa    ........................TS...........V.....................    240
135clone50aa    .S..........I...........S.....................F...........    240
135clone52aa    .........G.............S.S...........V.V...................    240
135clone54aa    ....................H....S......F...V.VS...................    240
135clone60aa    .................D........S............G................H...    240
```

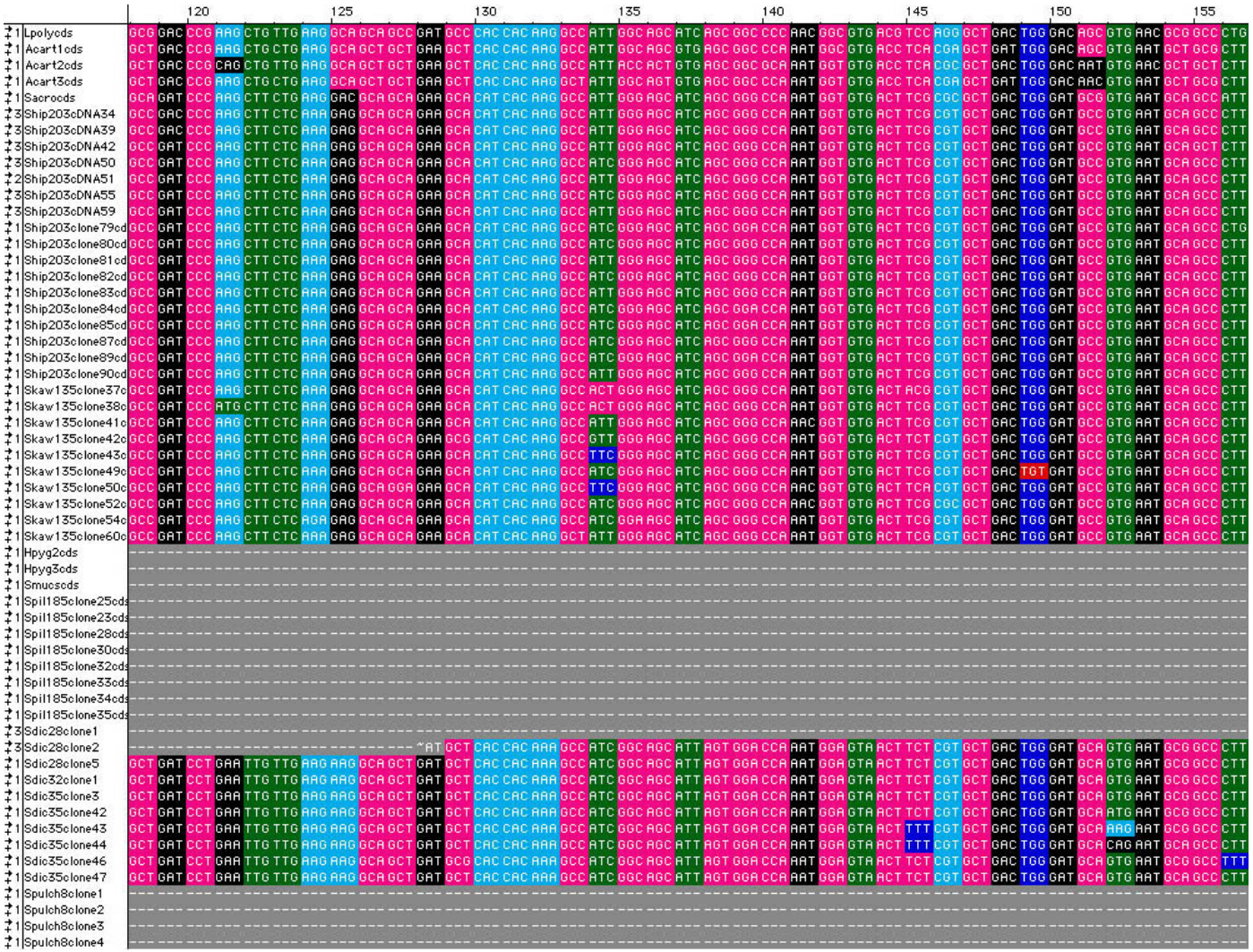
```
                         250       260       270       280       290       300
                         .         .         .         .         .         .

135clone37aa    DIYLKPLPGKTAPETLKAIDKMIVMGAKMDGNLLKAAAEAHHKAIGSIDAKGVTSAADYE    300
135clone38aa    ..........................S.................................    300
135clone41aa    ..........P..Q.............................................    300
135clone42aa    ............Q.............................................    300
135clone43aa    ...................................................N.....    300
135clone49aa    ...................................V...........T.........    300
135clone50aa    ...................................................N.....    300
135clone52aa    ............A.............................F.C.........H..    300
135clone54aa    ......................V...S.................................    300
135clone60aa    ...................................................N.....    300

                         310       320       330       340       350       360
                         .         .         .         .         .         .

135clone37aa    AVNAALGRLVASVPKSTVMDVYNSMAGIVDSTVTNNMFSKVSPLDAMAAAKGFYTFKDVV    360
135clone38aa    ........................KV..........P..N...................    360
135clone41aa    ........................V...P.........N....I.............    360
135clone42aa    ....................................N....V.............    360
135clone43aa    ....................................N...................    360
135clone49aa    ........................KV............N...................    360
135clone50aa    ........................KV...........N.N....I.............    360
135clone52aa    ....................................N...................    360
135clone54aa    ........................KV............N...................    360
135clone60aa    ........................KV..........L..N..............R...    360



135clone37aa    EASQR.--    366
135clone38aa    .....--    366
135clone41aa    .....--    366
135clone42aa    .....WR.    368
135clone43aa    .....--    366
135clone49aa    .....--    366
135clone50aa    .....--    366
135clone52aa    .....--    366
135clone54aa    .....--    366
135clone60aa    .G...--    366
```

APPENDIX 3.1 ALIGNMENT OF PCP NUCLEOTIDE SEQUENCES COLORED BY AMINO ACID TRANSLATION

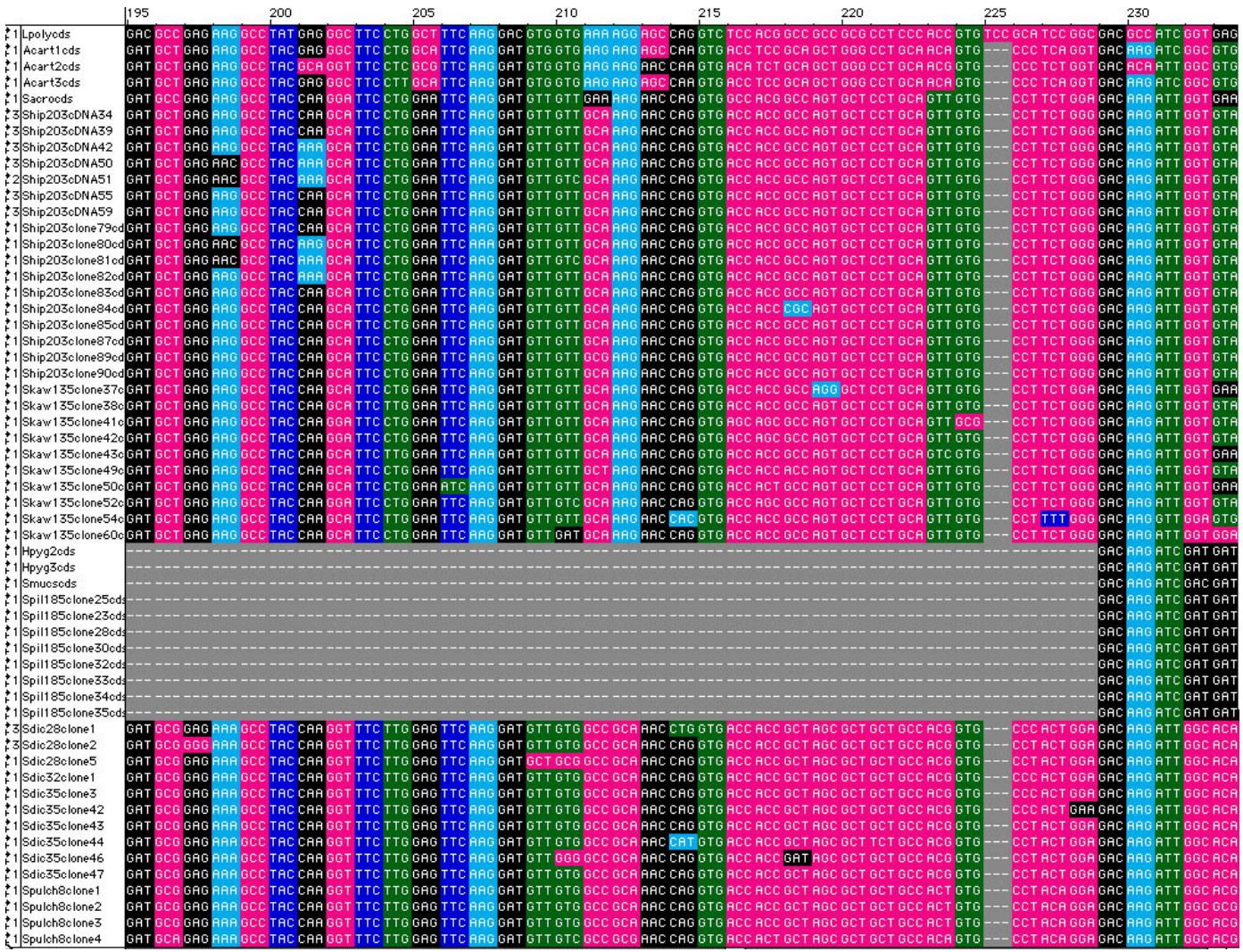


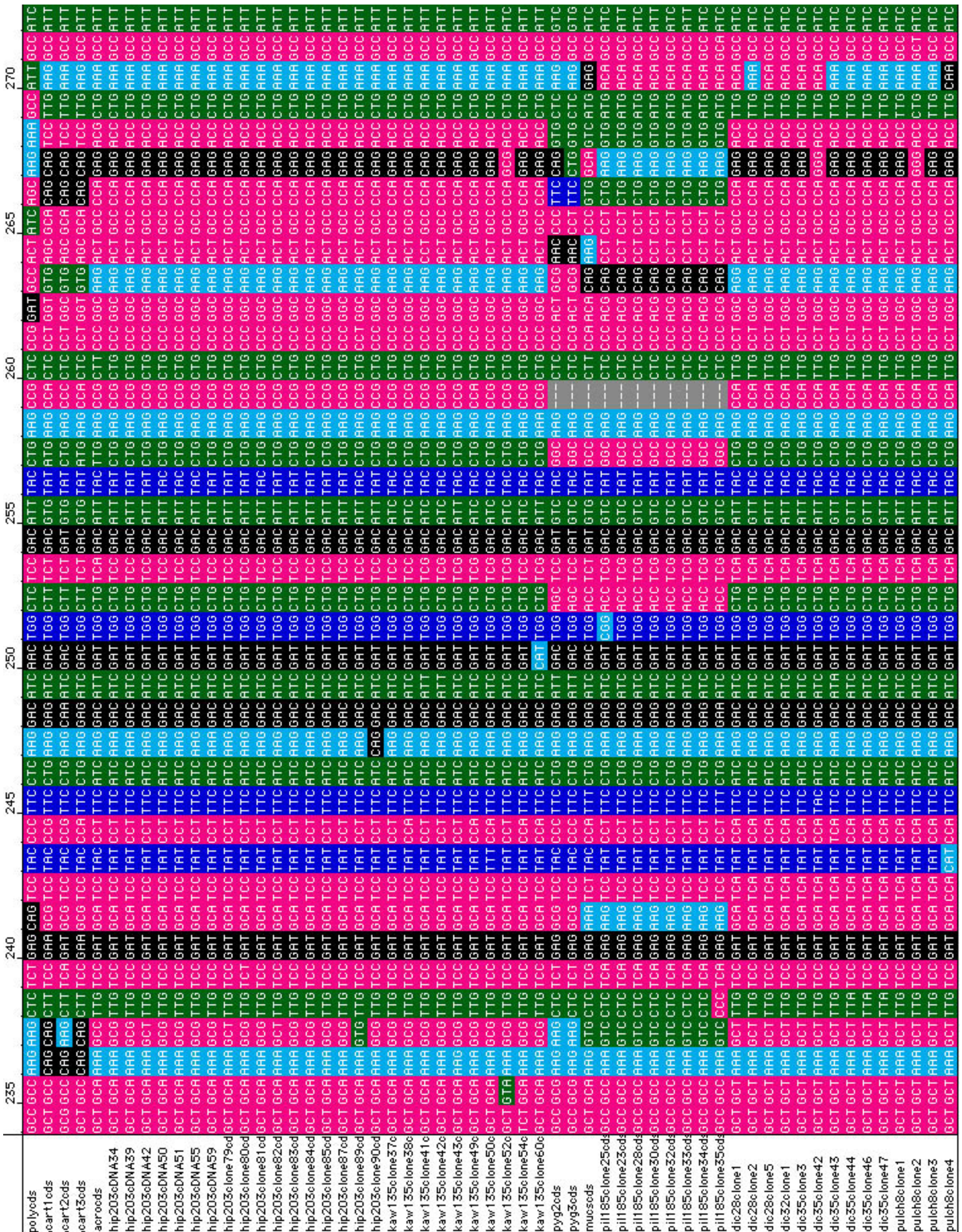

40
45
50
55
60
65
70
75
Lpolycds
Acart1cds
Acart2cds
Acart3cds
Sacrocds
Ship203cDNA34
Ship203cDNA39
Ship203cDNA42
Ship203cDNA50
Ship203cDNA51
Ship203cDNA55
Ship203cDNA59
Ship203clone79cd
Ship203clone80cd
Ship203clone81cd
Ship203clone82cd
Ship203clone83cd
Ship203clone84cd
Ship203clone85cd
Ship203clone87cd
Ship203clone89cd
Ship203clone90cd
Skaw135clone37c
Skaw135clone38c
Skaw135clone41c
Skaw135clone42c
Skaw135clone43c
Skaw135clone49c
Skaw135clone50c
Skaw135clone52c
Skaw135clone54c
Skaw135clone60c
Hpyg2cds
Hpyg3cds
Smuscds
Spil185clone25cd
Spil185clone23cd
Spil185clone28cd
Spil185clone30cd
Spil185clone32cd
Spil185clone33cd
Spil185clone34cd
Spil185clone35cd
Sdic28clone1
Sdic28clone2
Sdic28clone5
Sdic32clone1
Sdic35clone3
Sdic35clone42
Sdic35clone46
Sdic35clone47
Spulch8clone1
Spulch8clone2
Spulch8clone3
Spulch8clone4

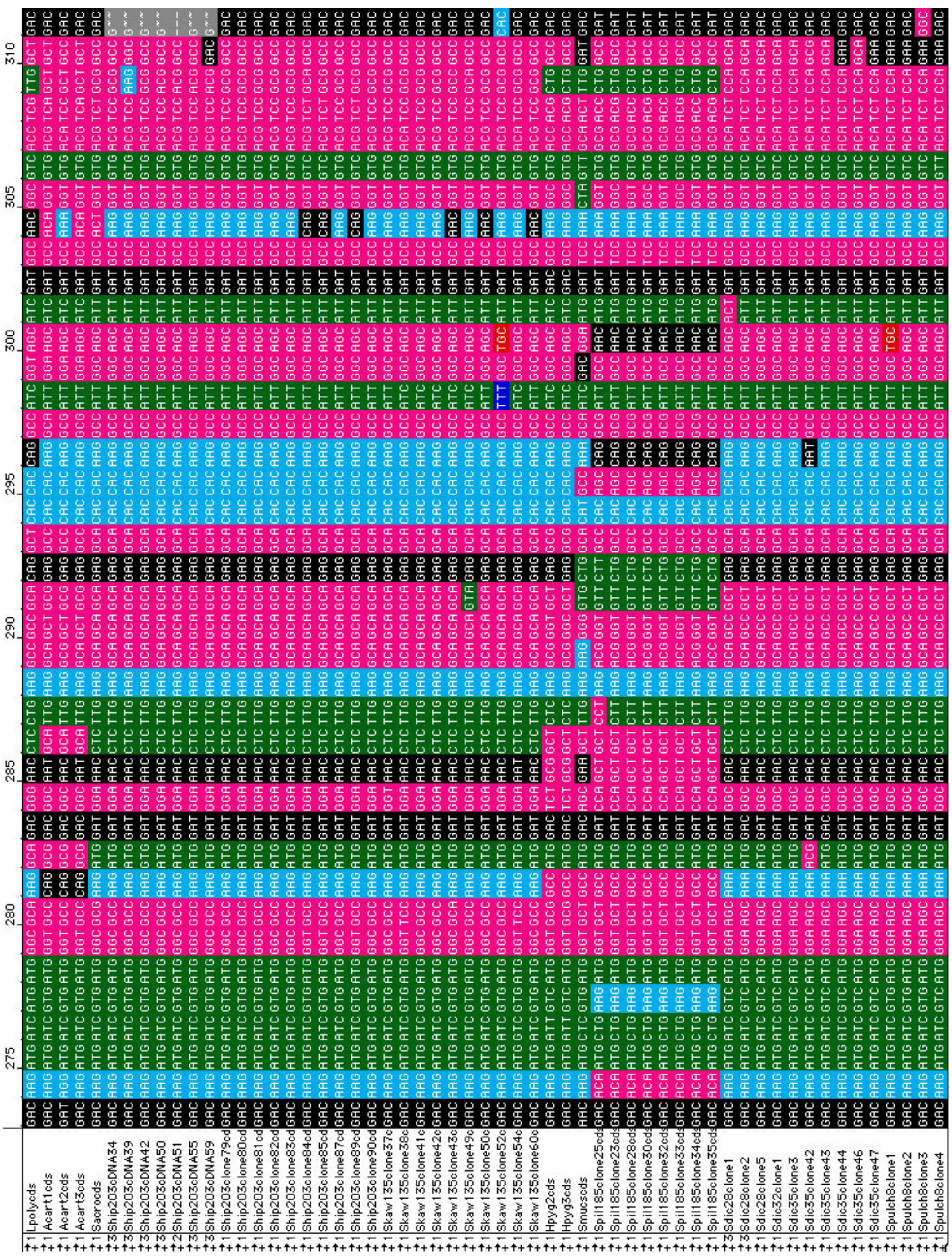

80
85
90
95
100
105
110
115
Lpolycds
Acart1cds
Acart2cds
Acart3cds
Sacrocds
Ship203cDNA34
Ship203cDNA39
Ship203cDNA42
Ship203cDNA50
Ship203cDNA51
Ship203cDNA55
Ship203cDNA59
Ship203clone79cd
Ship203clone80cd
Ship203clone81cd
Ship203clone82cd
Ship203clone83cd
Ship203clone84cd
Ship203clone85cd
Ship203clone87cd
Ship203clone89cd
Ship203clone90cd
Skaw135clone37c
Skaw135clone38c
Skaw135clone41c
Skaw135clone42c
Skaw135clone43c
Skaw135clone49c
Skaw135clone50c
Skaw135clone52c
Skaw135clone54c
Skaw135clone60c
Hpyg2cds
Hpyg3cds
Smucscds
Spil185clone25cds
Spil185clone23cds
Spil185clone28cds
Spil185clone30cds
Spil185clone32cds
Spil185clone33cds
Spil185clone34cds
Spil185clone35cds
Sdic28clone1
Sdic28clone2
Sdic28clone5
Sdic32clone1
Sdic35clone3
Sdic35clone42
Sdic35clone43
Sdic35clone44
Sdic35clone46
Sdic35clone47
Spulch8clone1
Spulch8clone2
Spulch8clone3
Spulch8clone4

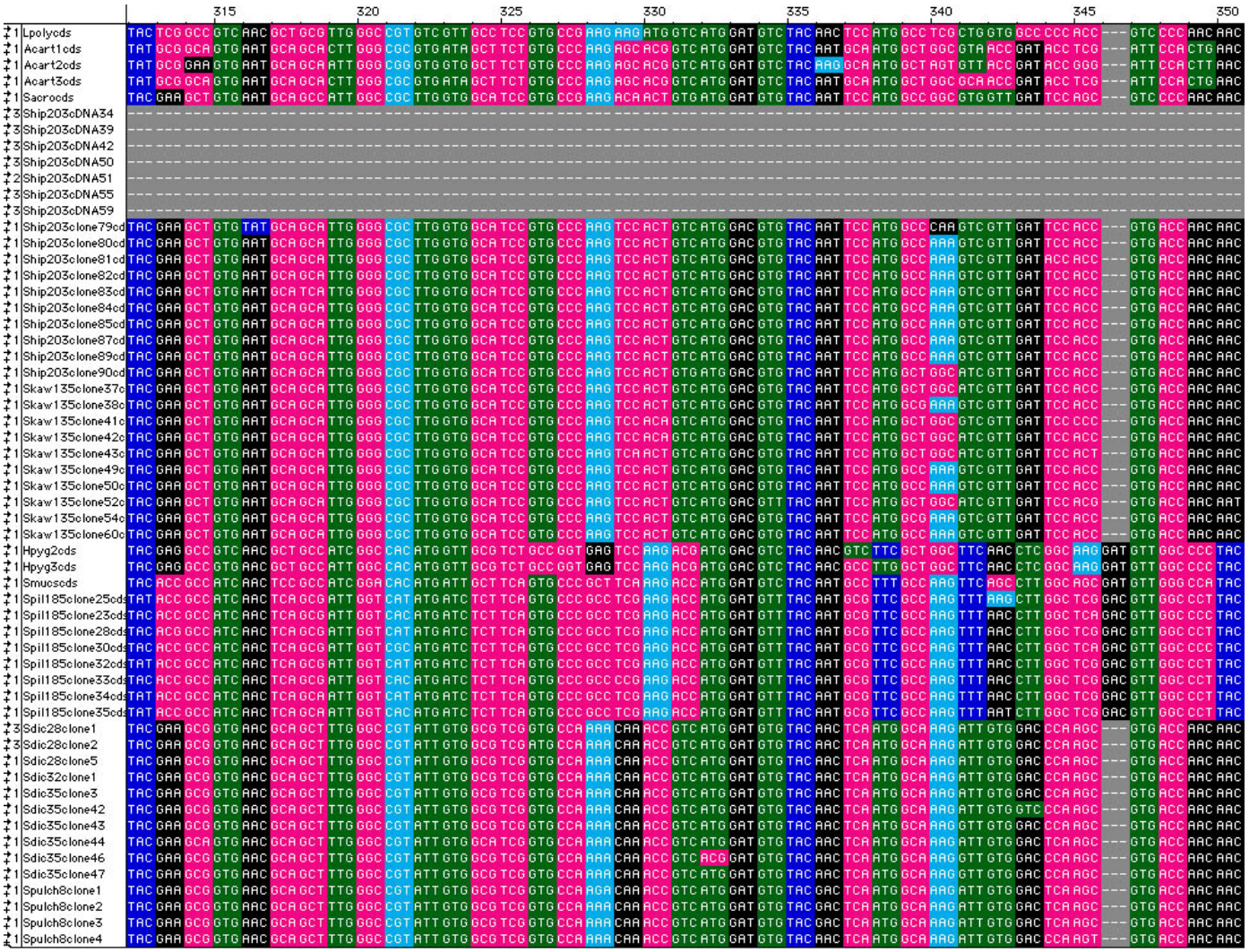

120
125
130
135
140
145
150
155
Lpolycds
Acart1cds
Acart2cds
Acart3cds
Sacrocds
Ship203cDNA34
Ship203cDNA39
Ship203cDNA42
Ship203cDNA50
Ship203cDNA51
Ship203cDNA55
Ship203cDNA59
Ship203clone79cd
Ship203clone80cd
Ship203clone81cd
Ship203clone82cd
Ship203clone83cd
Ship203clone84cd
Ship203clone85cd
Ship203clone87cd
Ship203clone89cd
Ship203clone90cd
Skaw135clone37c
Skaw135clone38c
Skaw135clone41c
Skaw135clone42c
Skaw135clone43c
Skaw135clone49c
Skaw135clone50c
Skaw135clone52c
Skaw135clone54c
Skaw135clone60c
Hpyg2cds
Hpyg3cds
Smucscds
Spil185clone25cds
Spil185clone23cds
Spil185clone28cds
Spil185clone30cds
Spil185clone32cds
Spil185clone33cds
Spil185clone34cds
Spil185clone35cds
Sdic28clone1
Sdic28clone2
Sdic28clone5
Sdic32clone1
Sdic35clone3
Sdic35clone42
Sdic35clone43
Sdic35clone44
Sdic35clone46
Sdic35clone47
Spulch8clone1
Spulch8clone2
Spulch8clone3
Spulch8clone4


160
165
170
175
180
185
190
Lpolycds
Acart1cds
Acart2cds
Acart3cds
Sacrocds
Ship203cDNA34
Ship203cDNA39
Ship203cDNA42
Ship203cDNA50
Ship203cDNA51
Ship203cDNA55
Ship203cDNA59
Ship203clone79cd
Ship203clone80cd
Ship203clone81cd
Ship203clone82cd
Ship203clone83cd
Ship203clone84cd
Ship203clone85cd
Ship203clone87cd
Ship203clone89cd
Ship203clone90cd
Skaw135clone37c
Skaw135clone38c
Skaw135clone41c
Skaw135clone42c
Skaw135clone43c
Skaw135clone49c
Skaw135clone50c
Skaw135clone52c
Skaw135clone54c
Skaw135clone60c
Hpyg2cds
Hpyg3cds
Smucsods
Spil185clone25cds
Spil185clone23cds
Spil185clone28cds
Spil185clone30cds
Spil185clone32cds
Spil185clone33cds
Spil185clone34cds
Spil185clone35cds
Sdic28clone1
Sdic28clone2
Sdic28clone5
Sdic32clone1
Sdic35clone3
Sdic35clone42
Sdic35clone43
Sdic35clone44
Sdic35clone46
Sdic35clone47
Spulch8clone1
Spulch8clone2
Spulch8clone3
Spulch8clone4


195
200
205
210
215
220
225
230
Lpolycds
Acart1cds
Acart2cds
Acart3cds
Sacrocds
Ship203cDNA34
Ship203cDNA39
Ship203cDNA42
Ship203cDNA50
Ship203cDNA51
Ship203cDNA55
Ship203cDNA59
Ship203clone79cd
Ship203clone80cd
Ship203clone81cd
Ship203clone82cd
Ship203clone83cd
Ship203clone84cd
Ship203clone85cd
Ship203clone87cd
Ship203clone89cd
Ship203clone90cd
Skaw135clone37c
Skaw135clone38c
Skaw135clone41c
Skaw135clone42c
Skaw135clone43c
Skaw135clone49c
Skaw135clone50c
Skaw135clone52c
Skaw135clone54c
Skaw135clone60c
Hpyg2cds
Hpyg3cds
Smucscds
Spil185clone25cds
Spil185clone23cds
Spil185clone28cds
Spil185clone30cds
Spil185clone32cds
Spil185clone33cds
Spil185clone34cds
Spil185clone35cds
Sdic28clone1
Sdic28clone2
Sdic28clone5
Sdic32clone1
Sdic35clone3
Sdic35clone42
Sdic35clone43
Sdic35clone44
Sdic35clone46
Sdic35clone47
Spulch8clone1
Spulch8clone2
Spulch8clone3
Spulch8clone4






315
320
325
330
335
340
345
350
Lpolycds
Acart1cds
Acart2cds
Acart3cds
Sacrocds
Ship203cDNA34
Ship203cDNA39
Ship203cDNA42
Ship203cDNA50
Ship203cDNA51
Ship203cDNA55
Ship203cDNA59
Ship203clone79cd
Ship203clone80cd
Ship203clone81cd
Ship203clone82cd
Ship203clone83cd
Ship203clone84cd
Ship203clone85cd
Ship203clone87cd
Ship203clone89cd
Ship203clone90cd
Skaw135clone37c
Skaw135clone38c
Skaw135clone41c
Skaw135clone42c
Skaw135clone43c
Skaw135clone49c
Skaw135clone50c
Skaw135clone52c
Skaw135clone54c
Skaw135clone60c
Hpyg2cds
Hpyg3cds
Smucsods
Spil185clone25cd
Spil185clone23cd
Spil185clone28cd
Spil185clone30cd
Spil185clone32cd
Spil185clone33cd
Spil185clone34cd
Spil185clone35cd
Sdic28clone1
Sdic28clone2
Sdic28clone5
Sdic32clone1
Sdic35clone3
Sdic35clone42
Sdic35clone43
Sdic35clone44
Sdic35clone46
Sdic35clone47
Spulch8clone1
Spulch8clone2
Spulch8clone3
Spulch8clone4


350
355
360
365
370
375
380
Lpolycds
Acart1cds
Acart2cds
Acart3cds
Sacrocds
Ship203cDNA34
Ship203cDNA39
Ship203cDNA42
Ship203cDNA50
Ship203cDNA51
Ship203cDNA55
Ship203cDNA59
Ship203clone79cd
Ship203clone80cd
Ship203clone81cd
Ship203clone82cd
Ship203clone83cd
Ship203clone84cd
Ship203clone85cd
Ship203clone87cd
Ship203clone89cd
Ship203clone90cd
Skaw135clone37c
Skaw135clone38c
Skaw135clone41c
Skaw135clone42c
Skaw135clone43c
Skaw135clone49c
Skaw135clone50c
Skaw135clone52c
Skaw135clone54c
Skaw135clone60c
Hpyg2cds
Hpyg3cds
Smucscds
Spil185clone25cds
Spil185clone23cds
Spil185clone28cds
Spil185clone30cds
Spil185clone32cds
Spil185clone33cds
Spil185clone34cds
Spil185clone35cds
Sdic28clone1
Sdic28clone2
Sdic28clone5
Sdic32clone1
Sdic35clone3
Sdic35clone42
Sdic35clone43
Sdic35clone44
Sdic35clone46
Sdic35clone47
Spulch8clone1
Spulch8clone2
Spulch8clone3
Spulch8clone4

References

Altschul SF, Gish W, Miller W, Myers EW and Lipman DJ (1990). Basic logical alignment search tool. *J Mol Biol* 215:403-10.

Baillie BK, Belda-Baillie CA, Silvestre V, Sison M, Gomez AV, Gomez ED and Monje V (2000). Genetic variation in Symbiodinium isolates from giant clams based on random-amplified-polymorphic DNA (RAPD) patterns. *Marine Biology* 136 (2000) 5, 829-836

Banaszak AT, Iglesias-Prieto R, and Trench RK (1993). *Scrippsiella velellae* sp. nov. (Peridinales) and *Gleodinium viscum* sp. nov. (phytodiniales), dinoflagellate symbionts of two hydrozoans (Cnidaria). *J Phycol* 29:517-528.

Bold HC and Wynne MJ, (1985). Introduction to the Algae: Structure and Reproduction. Second Edition. Prentice-Hall, Inc.: Englewood Cliffs, p.1-720. ISBN 0-13-477746-8 [Chapter 8, "Division Pyrrhophyta"]

Bradley RD and Hillis DM (1997). Recombinant DNA sequences generated by PCR amplification . *Mol Biol Evol* 14(5): 592-593.

Chang SS, Prezelin BB, and Trench RK (1983). Mechanisms of photoadaptation in three strains of the symbiotic dinoflagellate *Symbiodinium microadriaticum*. *Mar. Biol.* 76, 219-229.

Chang SS and Trench RK (1982). Peridinin-chlorolphyll *a* proteins from the symbiotic dinoflagellate *Symbiodinium (Gymnodinium) microadriaticum*, Freudenthal. *Proc R Soc Lond* 215:191-210.

Chang SS and Trench RK (1984). The Isoelectric Forms, Quaternary Structure and Amino Acid Composition of Peridinin-Chlorophyll a-Proteins from the Symbiotic Dinoflagellate Symbiodinium microadriaticum Freudenthal. *Proc R Soc Lond B*, 222 (1227):259-271.

Comeron JM (1999) K-Estimator: Calculation of the number of nucleotide substitutions per site and the confidence intervals. *Bioinformatics* 15: 763-764

Comeron JM (1995) A method for estimating the numbers of synonymous and nonsynonymous substitutions per site. *J Mol Evol*. 41: 1152-1159

Damjanovic A, Ritz T and Schulten K (2000). Excitation transfer in the peridinin-chlorophyll-protein of *Amphidinium carterae*. *Biophys J*. 79(4):1695-705.

Durnford DG, Deane JA, Tan S, McFadden GI, Gantt E, and Green BR (1999). A Phylogenetic Assessment of the Eukaryotic Light-Harvesting Antenna Proteins, with Implications for Plastid Evolution. *J Mol Evol* 48:59-68.

Falkowski PG and Raven JA (1997). Aquatic photosynthesis. Malden, Mass.: Blackwell Science.

Govind NS, Roman SJ, Iglesias-Prieto R, Trench RK, Triplett EL, and Prezelin BB (1990). An Analysis of the Light-Harvesting Peridinin-Chlorophyll a-Proteins from Dinoflagellates by Immunoblotting Techniques. *Proc R Soc Lond B* 240(1297): 187-195.

Grossman AR, Bhaya D, Apt KE and Kehoe DM (1995). Light-harvesting complexes in oxygenic photosynthesis: diversity, control, and evolution. *Annu Rev Genet*. 29:231-88.

Haidak DJ, Mathews CK and Sweeney BM (1966). Pigment Protein Complex from *Gonyaulax*. *Science* 152 (3719) 212-213.

Haxo FT, Kycia JH, Somers GF, Bennett A, and Siegelman HW (1976). Peridinin-chlorolphyll *a* proteins of the dinoflagellate *Amphidinium carterae* (Plymoth 450). *Plant Physiol* 57:297-303.

Hiller RG, Wrench PM, Gooley AP, Shoebridge G and Breton J. (1993). The major intrinsic light-harvesting protein of *Amphidinium*: characterization and relation to other light-harvesting proteins. *Photochem Photobiol* 57(1):125-31

Hiller RG, Crossley LG, Wrench PM, Santucci N and Hofmann E (2001). The 15-kDa forms of the apo-peridinin-chlorophyll a protein (PCP) in dinoflagellates show high identity with the apo-32 kDa PCP forms, and have similar N-terminal leaders and gene arrangements. *Mol Genet Genomics* 266 (2): 254-259.

Hillis DM, Moritz C, Porter CA and Baker RJ (1991). Evidence for biased gene conversion in concerted evolution of ribosomal DNA. *Science* 251(4991):308-10.

Hofmann E, Wrench PM, Sharples FP, Hiller RG, Welte W and Diederichs K (1996). Structural basis of light harvesting by carotenoids: Peridinin-chlorophyll-protein from *Amphidinium carterae*. *Science* 272 (5269): 1788-1791.

Holm-Hansen, O (1969). Algae: Amounts of DNA and Organic Carbon in Single Cells. *Science* 163 (3862) 87-88.

Huelsenbeck JP and Ronquist FR (in press). MRBAYES:Bayesian inference of phylogeny. *Biometrics*

Huelsenbeck JP, Ronquist F, Nielsen R and Bollback JP (2001). Bayesian inference of phylogeny and its impact on evolutionary biology. *Science* 294 (5550): 2310-2314.

Iglesias-Prieto R, Govind NS, and Trench RK (1991). Apoprotein Composition and Spectroscopic Characterization of the Water-Soluble Peridinin--Chlorophyll a--Proteins from Three Symbiotic Dinoflagellates. *Proc R Soc Lond B* 246 (1317): 275-283.

Iglesias-Prieto R, Govind NS, and Trench RK (1993). Isolation and characterization of three membrane-bound chlorophyll-protein complexes from four dinoflagellates. *Phil Trans R Soc* (Ser B) 340: 381-392.

Iglesias-Prieto R and Trench RK (1994). Acclimation and adaptation to irradiance in symbiotic dinoflagellates. I. responces of the photosynthetic unit to changes in photon flux density. *Mar Ecol Prog Ser* 113:163-175.

Iglesias-Prieto R and Trench RK (1997). Acclimation and adaptation to irradiance in symbiotic dinoflagellates. II. Response of chlorophyll-protein complexes to different photon-flux densities. *Mar Biol* 130 (1): 23-33.

Jovine RVM, Triplett EL, Nelson NB and Prezelin BB (1992). Quantification of chromophore pigments, apoprotein abundance and isoelectric variants of peridinin-chlorophyll a-protein complexes (PCPs) in the dinoflagellate Heterocapsa *pygmaea* grown under variable light conditions. *Plant Cell Physiol* 33 (6): 733-741.

LaJeunesse TC (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 (5): 866-880

LaJeunesse TC and Trench RK, (2000). Biogeography of Two Species of *Symbiodinium* (Freudenthal) Inhabiting the Intertidal Sea Anemone *Anthopleura elegantissima* (Brandt). *Biol Bull* 199:126-30.

Larkum T (1996). How dinoflagellates make light work with peridinin. *Trends Plant Sci* 1 (8): 247-248

Larkum, T and Howe CJ (1997). Molecular aspects of light-harvesting processes in algae. *Advances in Botanical Research* 27:257-330

Le QH, Markovic P, Hastings JW, Jovine KVM and Morse D (1997). Structure and organization of the peridinin-chlorophyll a-binding protein gene in *Gonyaulax polyedra. Mol. Gen Genet* 255 (6) 595-604.

Lee, DH Mittag, M Sczekan, S Morse, D and Hastings JW (1993). Molecular cloning and genomic organization of a gene for luciferin- binding protein from the dinoflagellate *Gonyaulax polyedra*. *J Biol Chem* 268: 8842-8850.

Li L, and Hastings W (1998). The structure and organization of the luciferase gene in the photosynthetic dinoflagellate *Gonyaulax polyedra. Plant Molecular Biology* 36 (2): 275-284.

Loeblich AR III. (1976). Dinoflagellate evolution: Speciation and evidence. *J Protozool* 23(1):13-28.

Loeblich AR III. (1982). Dinophyceae. Pages 101-115 in *Synopsis and Classification of Living Organisms.* McGraw-Hill Book Company, Inc..

Loeblich AR III. (1987). Dinoflagellate evolution. Pp. 482-522 *in* "Dinoflagellates" DL Spector ed.. San Diego: Academic Press.

Machabee' S, Wall L and Morse D. (1994). Expression and genomic organization of a dinoflagellate gene family. *Plant Molecular Biology* 25: 23-31.

Maddison DR and Maddison WP. (2000). MacClade 4: Analysis of phylogeny and character evolution. Version 4.0. Sinauer Associates, Sunderland, Massachusetts.

McNally KL, Govind NS, Thome PE, and Trench RK (1994). Small-subunit ribosomal DNA sequence analyses and a reconstruction of the inferred phylogeny amoung symbiotic dinoflagelltes (Pyrrophyta). *J Phycol* 30:316-329.

Melissinos AC (1966). Errors and their propagation. Pages 467-479 in A.C. Melissinos ed. "Experiments in Modern Physics". Academic Press, Orlando.

Moffat AS (1996). Form Follows Function When Plants Harvest Light. *Science* 272 (5269) pp. 1743-1744.

Moldowan JM, and Talyzina NM (1998). Biogeochemical Evidence for Dinoflagellate Ancestors in the Early Cambrian. *Science* 281 pp. 1168-1170.

Nielsen R and Yang ZH (1998). Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* 148 (3): 929-936.

Norris BJ and Miller DJ (1994). Nucleotide sequence of a cDNA encoding the precursor of the peridinin-chlorophyll *a*-binding protein from the dinoflagellate *Symbiodinium* sp.. *Plant Mol Biol* 24:673-677.

Ogata T, Kodama M, Nomura S, Kobayashi M, Nozawa T, Katoh T and Mimuro M. (1994). A novel peridinin-chlorophyll a protein (PCP) from the marine dinoflagellate *Alexandrium cohorticula*: a high pigment content and plural spectral forms of peridinin and chlorophyll a. *FEBS Lett*. 356(2-3):367-71

Palumbi SR (1996). Nucleic Acids II: The Polymerase Chain Reaction. *In* "Molecular Systematics, 2nd. Ed." (D. M. Hillis, B. K. Mable and C. Moritz, Eds), pp. 205-247. Sinauer, Sutherland, Massachusetts.

Pfiester LA (1984). Sexual Reproduction. Pp. 181-199 *in* " Dinoflagellates" DL Spector ed.. San Diego: Academic Press.

Pinto E, Catalani LH, Lopes NP, Di Mascio P and Colepicolo P. (2000). Peridinin as the major biological carotenoid quencher of singlet oxygen in marine algae *Gonyaulax polyedra*. *Biochem Biophys Res Commun*; 268(2):496-500.

Prezelin BB (1976). The role of peridinin-chlorolphyll *a*- proteins in the photosynthetic light adaptation of the marine dinoflagellate, *Glenodinium sp.*. *Planta (Berl.)* 130:225-233.

Prezelin BB, and Haxo FT (1976). Purification and characterization of peridinin-chlorophyll *a*-proteins from the marine dinoflagellates *Gleodinium sp.* and *Gonyaulax polyedra. Planta (Berl)*. 128:133-141.

Rizzo PJ (1991). The enigma of the dinoflagellate chromosome. *J Protozool* 38(3):246-252.

Roman JS, Govind NS, Triplett EL and Prezelin BB (1988). Light regulation of peridinin-chlorophyll a-protein (PCP) complexes in the dinoflagellate *Glenodinium sp. Plant Physiol* 88: 594-599.

Rowan R and Knowlton N (1995). Intraspecific and ecological zonation in coral-algal symbiosis. *Pro Nat Acad Sci* 92:2850-2853.

Rowan R, Knowlton N, Baker A. and Jara J. (1997). Landscape ecology of algal symbionts creates variation in episodes of coral bleaching. *Nature* 338: 265-269.

Rowan R and Powers DA (1991a). A Molecular Genetic Classification of Zooxanthellae and the Evolution of Animal-Algal Symbioses. *Science*. 251(4999). 1348-1351.

Rowan R and Powers DA (1991b). Molecular genetic identification of symbiotic dinoflagellates (zooxanthallae). *Mar Ecol Prog Ser* 71:65-73.

Rowan R and Powers DA (1992). Ribosomal RNA sequences and the diversity of symbiotic dinoflagellates (zooxanthallae). *Pro Nat Acad Sci* 89:3639-3643.

Rowan R, Whitney SM, Fowler A and Yellowlees D (1996). Rubisco in marine symbiotic dinoflagellates: form II enzymes in eukaryotic oxygenic phototrophs encoded by a nuclear multigene family. *Plant Cell*.8: 539-553.

Sadler LA, McNally KL, Govind NA, Brunk CF, and Trench RK (1992). The nucleotide sequence of the small subunit ribosomal RNA gene from *Symbiodinium pilosum*, a symbiotic dinoflagellate. *Curr Genet* 21:409-416.

Sambrook J, Fritsc, EF and Maniatis T (1989). *In* "Molecular cloning: a laboratory manual". Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory.

Scharf SJ (1990). Cloning with PCR. Pp. 84-91 *in* "PCR protocols : a guide to methods and applications". MA innis, DH Gefland, JJ Sninski and TJ White, eds. San Diego: Academic Press.

Schoenberg DA and Trench RK (1976). Specificity of symbioses between marine cnidarians and zooxanthallae. Pp. 423-432 *in* G. O. Macke ed., *Coelenterate Ecology and Behavior.* Plenum Publishing, N.Y..

Schoenberg DA and Trench RK (1980a). Genetic variation in *Symbiodinium (Gymnodinium) microadriaticum* Freudenthal, and specificity with its symbiosis with marine invertebrates. I. Isoenzyme and soluable protein patterns of axenic cultures of *Symbiodinium microadriaticum. Proc R Soc Lond,* 207:405-427.

Schoenberg DA and Trench RK (1980b). Genetic variation in *Symbiodinium (Gymnodinium) microadriaticum* Freudenthal, and specificity with its symbiosis with marine invertebrates. II. Morphological variation in *Symbiodinium microadriaticum. Proc R Soc Lond,* 207:429-444.

Schoenberg DA and Trench RK (1980c). Genetic variation in *Symbiodinium (Gymnodinium) microadriaticum* Freudenthal, and specificity with its symbiosis with marine invertebrates. III. Specificity and infectivity of *Symbiodinium microadriaticum. Proc R Soc Lond,* 207:429-444.

Sharples FP, Wrench PM, Ou K and Hiller RG (1996). Two distinct forms of the peridinin-chlorophyll a-protein from *Amphidinium carterae*. *Biochim Biophys Acta* 12;1276(2):117-23.

Siegelman HW, and Kycia JH, and Haxo FT (1977). Peridinin-chlorophyll *a*-proteins of the dinoflagellate algae. *Broohaven Symp Biol* no. 28, pp. 162-169.

Song PS, Koka P, Prezelin BB and Haxo FT (1976). Molecular topology of the photosynthetic light-harvesting pigment complex, peridinin-chlorophyll a-protein, from marine dinoflagellates. *Biochemistry*15(20):4422-7.

Spector, D (1984). Dinoflagellate Nuclei. Pp. 107-147 *in* " Dinoflagellates" DL Spector ed.. San Diego: Academic Press.

Suzuki Y and Gojobori T (1999). A method for detecting positive selection at single amino acid sites. *Mol Biol Evol* 16: 1315-1328.

Swofford DL 1998. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

ten Lohuis MR, and Miller DJ (1998). Light-Regulated Transcription of Genes Encoding Peridinin Chlorophyll a Proteins and the Major Intrinsic Light-Harvesting Complex Proteins in the Dinoflagellate *Amphidinium carterae* Hulburt (Dinophycae). Changes in Cytosine Methylation Accompany Photoadaptation. *Plant Physiol* 117: 189-196.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F and Higgins DG (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25(24):4876-82.

Trench RK (1981). Cellular and molecular interactions in symbioses between dinoflagellates and marine invertebrates. *Pure and Appl Chem,* 53:819-835.

Trench RK (1993). Microalgal-invertebrate symbioses: a review. *Ectocytobiosis & Cell Res,* 9:135-175.

Trench RK and Blank RJ (1987). *Symbiodinium microadriaticum* Freudenthal, *S. goreauii* sp. nov., *S. kawagutii* sp. nov. and *S. pilosum* sp. nov.: Gymnodinioid dinflagellates of marine invertebrates. *J Phycol* 23:469-481.

Triplett EL, Jovine RVM, Govind NS, Roman SJ, Chang SS, & Prezelin BB (1993). Characterization of two full-length cDNA sequences encoding for apoproteins of peridinin-chlorophyll *a*-proptein (PCP) complexes. *Mol Mar Biol Biotechnol*. 2(4):246-254.

Wagner, A (2002). Selection and gene duplication: a view from the genome. *Genome Biology* 2002 3(5): reviews1012.1-1012.3

Weis VM, Verde EA, and Reynolds WS (2002). Characterization of a short form peridinin-chlorophyll-protein (PCP) cDNA and protein from the symbiotic dinoflagellate *Symbiodinium* muscatinei (Dinophyceae) from the sea anemone *Anthopleura elegantissima* (Cnidaria). *J Phycol* 38 (1), 157-163.

Wilcox TP (1998). Large-subunit ribosomal RNA systematics of symbiotic dinoflagellates: Morphology does not recapitulate phylogeny. *Mol Phylogenet Evol* 10 (3): 436-448.

Yang Z (1997). PAML: a program package for phylogenetic analysis by maximum likelihood *Cabios* 13:555-556.

Yang Z and Bielawski JP (2000). Statistical methods for detecting molecular adaptation. *Tree* 15(12): 496-503.

Yang Z, Nielsen R, Goldman N, and Pedersen AK (2000). Codon-Substitution Models for Heterogeneous Selection Pressure at Amino Acid Sites. *Genetics* 155: 431-499.

Veldhuis MJ, Cucci TL and Sieracki ME (1997). Cellular DNA content of marine phytoplankton using two new fluorochromes: taxonomic and ecological implications. *J Phycol* 33: 527-541.

VITA

Jay R. Reichman was born on January 20, 1960 in Odessa, Texas to parents Ann and Roy. Jay was raised in Dallas, Texas where he attended Lamplighter School, St. Marks School of Texas and Jesuit College Preparatory School. He received a B.S. from Texas A & M in 1982. Jay met his future wife Melada in 1983. Jay worked as a purchasing agent and as an operations manager in the private sector in Austin, Texas prior to entering graduate school. While in graduate school at UT Austin, Jay served as a teaching assistant in the Department of Zoology for courses in genetics, oceanography and limnology, ecology, and human physiology. Jay and Melada were married in 1994. In 1998, he was promoted to Lecturer in the Department of Kinesiology and Health Education where he supervised the University of Texas Underwater Science and Scuba Diving Program. Jay and Melada's daughter Lauren was born was born in 2001. Jay will be doing postdoctoral research with the US Environmental Protection Agency's Western Ecology Division in Corvallis, Oregon.

Permanent address: Unknown at this time.

This dissertation was typed by the author.