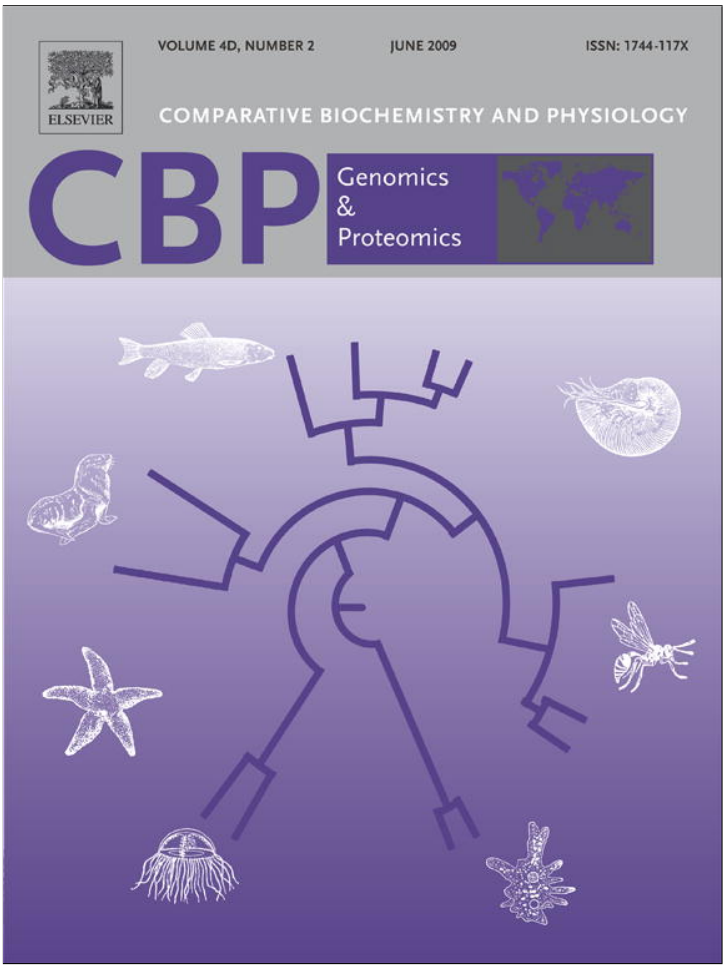

Comparative Biochemistry and Physiology, Part D 4 (2009) 67–74

Contents lists available at ScienceDirect

Comparative Biochemistry and Physiology, Part D

journal homepage: www.elsevier.com/locate/cbpd

Evolutionary analysis of orthologous cDNA sequences from cultured and symbiotic dinoflagellate symbionts of reef-building corals (Dinophyceae: *Symbiodinium*)

Christian R. Voolstra [a], Shinichi Sunagawa [a], Jodi A. Schwarz [b], Mary Alice Coffroth [c], Dave Yellowlees [d], William Leggat [d], Mónica Medina [a,*]

[a] School of Natural Sciences, University of California, Merced, P.O. Box 2039, Merced, CA 95344, USA
[b] Department of Biology, Vassar College, 124 Raymond Avenue Box 731, Poughkeepsie, NY 12604, USA
[c] Graduate Program in Evolution, Ecology and Behavior and Department of Geology, State University of New York at Buffalo, 447 Hochstetter Hall, Buffalo, NY 14260, USA
[d] Comparative Genomics Centre, School of Pharmacy and Molecular Sciences and ARC Centre of Excellence for Coral Reef Studies, James Cook University, Townsville, Qld, 4811, Australia

ARTICLE INFO

Article history:
Received 1 September 2008
Received in revised form 13 November 2008
Accepted 16 November 2008
Available online 6 December 2008

Keywords:
cDNA library
Dinoflagellate
Ortholog comparison
Symbiosis
Zooxanthellae

ABSTRACT

Dinoflagellates are ubiquitous marine and freshwater protists. The endosymbiotic relationship between dinoflagellates of the genus *Symbiodinium* (also known as zooxanthellae) and corals forms the basis of coral reefs. We constructed and analyzed a cDNA library from a cultured *Symbiodinium* species clade A (CassKB8). The majority of annotated ESTs from the *Symbiodinium* sp. CassKB8 library cover metabolic genes. Most of those belong to either carbohydrate or energy metabolism. In addition, components of extracellular signal transduction pathways and genes that play a role in cell–cell communication were identified. In a subsequent analysis, we determined all orthologous cDNA sequences between this library (1,484 unique sequences) and a library from a *Symbiodinium* species clade C (C3) (3,336 unique sequences) that was isolated directly from its symbiotic host. A set of 115 orthologs were identified between *Symbiodinium* sp. CassKB8 and *Symbiodinium* sp. C3. These orthologs were subdivided into three groups that show different characteristics and functions: conserved across eukaryotes (CE), dinoflagellate-specific (DS) and *Symbiodinium*-specific (SS). Orthologs conserved across eukaryotes are mainly comprised of housekeeping genes, photosynthesis-related transcripts and metabolic proteins, whereas the function for most of the dinoflagellate-specific orthologs remains unknown. A d*N*/d*S* analysis identified the highest ratio in a *Symbiodinium*-specific ortholog and evidence for positive selection in a dinoflagellate-specific gene. Evolution of genes and pathways in different dinoflagellates seems to be affected by different lifestyles, and a symbiotic lifestyle may affect population structure and strength of selection. This study is the first evolutionary comparative analysis of orthologs from two coral dinoflagellate symbionts.

1. Introduction

Dinoflagellates are omnipresent marine and freshwater protists. As free-living photosynthetic plankton, they account for much of the primary productivity of oceans and lakes. As photosynthetic symbionts, they provide essential nutrients to reef-building corals and other marine invertebrates (Muscatine et al., 1981). Phylogenetically, they have been placed in the super group Chromalveolata (Parfrey et al., 2006). This grouping was introduced as a parsimonious, albeit controversial, explanation for the presence of plastids of red algal origin in photosynthetic members of the Alveolata and Chromista (Cavalier-Smith, 1999). Whereas the monophyly of Alveolata (e.g. ciliates, dinoflagellates, and apicomplexa) is well supported, the kingdom Chromista was created to unite diverse microbial lineages with red algal plastids (Cavalier-Smith, 1998; Cavalier-Smith, 2002). Dinoflagellates are adapted to a wide variety of environments as reflected by a tremendous diversity in form and nutrition (e.g. autotrophy versus heterotrophy) (Graham and Wilcox, 2000). Some remarkable features of dinoflagellate biology are their unique genome structure and gene regulation. The nuclear genomes of these algae lack nucleosomes, have permanently condensed chromosomes, and are estimated to be up to 3,000–215,000 Mb in size (the haploid human genome is ~3,000 Mb) (Spector, 1984). The reasons for this large amount of cellular DNA are not known. Two mutually exclusive theories attempting to explain this large amount of cellular DNA can be summarized as the 'adaptive' versus the 'junk' DNA theories (Petrov, 2001). Studies of DNA reassociation kinetics suggest that dinoflagellate genomes are complex and non-repetitive (Allen et al., 1975; Hinnebusch et al., 1980). Due to their size, there is currently no sequenced dinoflagellate genome. However, dinoflagellate transcript sequences or Expressed Sequence Tags (ESTs) are available. Currently, NCBI holds 89,436 EST sequences of *Dinophyceae* (as of January 2008).

* Corresponding author. Tel.: +1 209 228 7863; fax: +1 209 228 4053.
E-mail address: mmedina@ucmerced.edu (M. Medina).

1744-117X/$ – see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.cbd.2008.11.001

The vast majority of ESTs (>85,000) stem from studies of six different dinoflagellates – *Lingulodinium polyedrum*, *Alexandrium tamarense*, *Amphidinium carterae*, *Karenia brevis*, *Karlodinium micrum*, and *Heterocapsa triquetra* (Bachvaroff et al., 2004; Hackett et al., 2004b, 2005; Tanikawa et al., 2004; Patron et al., 2005, 2006; Yoon et al., 2005) – none of those belong to the same phylogenetic branch as *Symbiodinium* (Suessiales).

The majority of symbiotic dinoflagellates are members of the diverse genus *Symbiodinium* (also called zooxanthellae) (Rowan, 1998). They live in a symbiotic lifestyle with a broad range of marine animals and other unicellular eukaryotes (Trench, 1993). The hosts of symbiotic dinoflagellates include foraminifera, radiolarians, flatworms, anemones, jellyfish, and mollusks. The best-studied relationship, however, is between zooxanthellae and reef-building corals. Zooxanthellae contribute to their hosts' nutrition and provide energy in the form of amino acids and simple sugars that may sustain high rates of calcification, while the host provides a sheltered, light-rich environment, as well as inorganic nitrogen and carbon (Yellowlees et al., 2008). Molecular systematics applied to *Symbiodinium* spp. divide the group into eight highly divergent subgeneric lineages, or clades (A–H) (Rowan and Powers, 1991; Santos et al., 2002; Baker 2003; Pochon et al., 2006). Within each clade, numerous closely related types or species exhibit distinctive host taxa, geographic, and/or environmental distributions (Baker, 2003). The first radiation event within the genus *Symbiodinium* probably occurred around 50 MYA (Pochon et al., 2006), whereas the major diversification of extant *Symbiodinium* lineages started around 10–15 MYA (John et al., 2003; Pochon et al., 2006). At present, we have little or no understanding of the physiological benefits to a coral of hosting different zooxanthellae strains, and why the majority of corals acquire their symbionts *de novo* each generation from the environment (Szmant 1986; Coffroth et al., 2006). Kitano and Oda (2006) devised a principle of 'self-extending symbiosis' as a mechanism under which evolvable robust systems continue to extend their system boundary by the integration of foreign biological entities (genes, microorganisms, etc.) to enhance the adaptive capacity of the system. As 'loose integration' is more adaptive than 'tight integration', *de novo* acquisition of dinoflagellate symbionts would provide the coral host with a flexible and adaptive solution to counteract environmental perturbations.

In this study, we report the construction and analysis of a newly generated EST dataset from cultured *Symbiodinium* sp. CassKB8. Orthologs between this dataset and an EST library of *Symbiodinium* sp. C3 (Leggat et al., 2007) were determined and analyzed. This is only the second report of ESTs from *Symbiodinium* (Leggat et al., 2007) and the first time that a comparison between orthologs from two different strains has been conducted. Annotation of the newly created library identified mainly metabolic genes, but also genes belonging to important signaling pathways and cell communication genes. In a subsequent analysis, all identified orthologs from both EST libraries were analyzed in terms of d*N*/d*S*. The set of orthologs that is conserved across eukaryotes is mainly comprised of housekeeping genes and genes related to metabolism and photosynthesis. In contrast, the putative functions for most of dinoflagellate- and *Symbiodinium*-specific genes remain unknown. Based on our results the characterization of dinoflagellate-specific genes may identify candidates that are responsible for the vast diversity of dinoflagellate lifestyles, while *Symbiodinium*-specific genes may be related to symbiosis-specific functions. Our analyses suggest that a symbiotic lifestyle may be linked to an increase in evolutionary rates of some genes. As such, our study provides insights into how a given lifestyle might affect the evolution of an organism.

2. Materials and methods

2.1. Isolation of RNA and cDNA library construction

The dataset used in this analysis includes the original ESTs from the normalized cDNA library published in Leggat et al. (2007) and additional ESTs from the same library. The species is *Symbiodinium* sp. C3 *sensu* LaJeunesse (2002) which was isolated from the blue morph of the coral *Acropora aspera* (Dove, 2004) from a reef flat at Heron Island (23.44°S, 151.91°E). This clade (C3) is widely distributed in corals from the Indo-Pacific and Caribbean oceans (LaJeunesse et al., 2003). In the case of *Symbiodinium* sp. CassKB8, we sampled material for construction of cDNA libraries from cells in culture. *Symbiodinium* sp. CassKB8 was originally isolated from *Cassiopea* sp. in Kaneohe Bay, Hawaii. This strain has been in culture for over 25 years. Although the name is similar to the KB symbionts by Apprill and Gates (2007), the similarity is coincidental. Whereas all *Symbiodinium* species characterized in the above mentioned study belong to clade C, CassKB8 belongs to clade A. *Symbiodinium* sp. CassKB8 is A1 *sensu* LaJeunesse (2001). CassKB8 cultures were reared at 28 °C under a 12/12 hour light: dark cycle. We isolated total RNA using Qiazol reagent (Qiagen) and vortexing with glass beads to break open the cells. To remove residual phenol or other contaminants, the RNA was purified using RNeasy clean up kit (Qiagen). Total RNA was quantified using a Nanodrop spectrophotometer, and RNA quality was assessed using an Agilent Bioanalyzer. We used the Clontech SMART cDNA Library Construction Kit with the pDNR-lib vector to construct non-normalized cDNA libraries. The cDNA was PCR-amplified with the Advantage 2 PCR kit using the SMART 5′ PCR III primer and CDS III/3′ PCR primer, and using between 18 and 26 cycles (depending on the starting amount of RNA). To minimize cloning of incomplete or degraded transcripts, we preferentially selected cDNA >500 bp by first passing the SfiI-digested cDNA over CHROMA SPIN-400 columns, and then in some cases, cutting out a >500 bp smear from a 1.1% agarose gel. The size-selected cDNA was ligated to the pDNR-lib vector. Electrocompetent cells were transformed with the vector, grown overnight in liquid suspension and then plated onto Teknova LB agar plates with 30 μg ml^{-1} chloramphenicol. Colonies were picked into 384 well plates using a QBot robot (Genetix), and sequenced from both 5′ and 3′ ends on ABI 3730 Sequencers at the Joint Genome Institute (JGI).

2.2. EST assembly

The chromatogram files of both libraries were base-called using phred (Ewing et al., 1998) and assembled after vector-screening with the cross-match tool of the phred/phrap/consed package (http://www.phrap.org/phredphrapconsed.html). The vector-screened ESTs were assembled by CAP3 (Huang and Madan, 1999). The contigs and singlets (not assembled ESTs) of each library were concatenated to multi-FASTA files, which were subsequently used for sequence annotation and reciprocal BLAST analysis. For *Symbiodinium* sp. C3, sequences are accessible under Taxonomy ID 154562 (GenBank accession no. EH035872 - EH058216 and FE63988 – FE866467). For *Symbiodinium* sp. CassKB8, EST sequences are accessible under Taxonomy ID 230985 (GenBank accession no. FE537410-FE540062). Assembled sequences are accessible through our EST database at http://sequoia.ucmerced.edu/SymBioSys/index.php, and orthologs detailed in this particular study are available as additional files (Additional File 2, Additional File 3). There is a potential for contamination with coral genes in the EST library of *Symbiodinium* sp. C3 given the way in which the library was constructed. A small number of contigs (192) were identified as cnidarian contaminants with a BLASTx *e*-value of $<10^{-5}$. Those were removed prior to further analyses.

2.3. EST annotation

All assembled unigenes greater than 100 bp from the *Symbiodinium* sp. CassKB8 library were uploaded to http://www.genome.jp/kaas-bin/kaas_main?mode=est_s and annotated with the KEGG Automatic Annotation Server (KAAS) (Moriya et al., 2007). A single-directional best hit assignment method, together with a bit score threshold of greater than 40 was chosen as suggested for EST libraries.

Annotation maps were downloaded and parsed to yield the numbers of annotated sequences for a given term within the KEGG orthology (Table 1). KEGG associates genes to pathways in a way that identical genes can appear in different pathways. However, genes that share the same annotation will only be counted once.

2.4. Identification of orthologs by Best Reciprocal BLAST Hit (BRBH) approach

We followed a Best Reciprocal BLAST Hit (BRBH) approach to isolate orthologs from both EST libraries. Initially, the method was developed as a shortcut to identify orthologs between genomes (Tatusov et al., 1997, 2003), but is assumed to work equally well for EST sequences (Telford, 2007). Briefly, all-against-all BLASTs of both libraries were conducted to get a library-specific best hit (BeT). A BeT is the EST sequence in a target library which is most similar to a given sequence from the query library. Provided the two genes are true orthologs, the closest significant match in the initial BLAST should select the original reference gene as the top hit in a reciprocal BLAST (Telford, 2007). The underlying premise is that orthologs are more similar to each other than they are to any other sequence from the source libraries. To identify putative orthologs from our set of EST libraries, we used tBLASTx with a bitscore cutoff of 40 for any given alignment.

To determine the fraction of genes that are conserved among eukaryotes, the orthologs from both libraries were blasted against the NCBI nr nucleotide and protein sequence databases with BLASTn and BLASTx. All orthologs that gave hits with an *e*-value $<10^{-5}$ were assigned to the conserved in eukaryotes category (CE). All other sequences were searched against the complete set of dinoflagellate ESTs from NCBI (as of December 2007) by tBLASTx and were assigned to the dinoflagellate-specific category (DS) if they showed *e*-values $<10^{-5}$. The portion of genes that gave no hit to dinoflagellates other than *Symbiodinium* were assigned to the *Symbiodinium*-specific category (SS). As there is a chance that orthologs that have no matches to the NCBI nr nucleotide and protein databases might still have homologs in species other than dinoflagellates, a tBLASTx with an *e*-value cutoff of $<10^{-5}$ against the full 'est_others' database (NCBI) was carried out for all those sequences.

For the comparison of the CE orthologs we used the *Plasmodium falciparum* proteome (most complete apicomplexan genome). All CE orthologs were blasted against nr and *P. falciparum* in a BLASTx search with an *e*-value cutoff of 10^{-10}. Orthologs were considered only if they reported the same first BLAST hit to the target database.

Table 1
Number of unigenes and distribution of KEGG orthologies for the EST library of *Symbiodinium* sp. CassKB8.

KEGG orthologies	# of UniSeqs		%
total # of annotated UniSeqs >100 bp	1,484		
# of annotated UniSeqs with BLAST bit scores >40	155		10.4%
Metabolism	**82**		**52.9%**
Carbohydrate metabolism	23		28.0%
Energy metabolism	17		20.7%
Lipid metabolism	5		6.1%
Nucleotide metabolism	5		6.1%
Amino acid metabolism	12		14.6%
Metabolism of other amino acids	3		3.7%
Metabolism of cofactors and vitamins	5		6.1%
Biosynthesis of secondary metabolites	3		3.7%
Xenobiotics biodegradation and metabolism	9		11.0%
Genetic information processing	**21**		**13.5%**
Translation	13		61.9%
Folding, sorting and degradation	8		38.1%
Environmental information processing	**16**		**10.3%**
Signal transduction	14		87.5%
MAPK signaling pathway		2	
Wnt signaling pathway		4	
Notch signaling pathway		1	
TGF-beta signaling pathway		2	
ECM-receptor interaction		5	
Cellular processes	**28**		**18.1%**
Cell motility	2		7.1%
Cell growth and death	7		25.0%
Cell communication	10		35.7%
Focal adhesion		3	
Adherens junction		2	
Tight junction		3	
Gap junction		2	
Endocrine system	6		21.4%
Immune system	1		3.6%
Nervous system	2		7.1%
Human diseases	**8**		**5.2%**
Neurodegenerative disorders	3		37.5%
Metabolic disorders	1		12.5%
Cancers	4		50.0%

2.5. Detection of positive selection

All sequences were aligned using MUSCLE (Edgar, 2004). To determine evolutionary relationships, nucleotide Maximum Likelihood (ML) trees were constructed on the basis of amino acid-aligned EST sequences using PHYLIP (http://evolution.genetics.washington.edu/phylip.html). Trees were drawn with TreeView (Page, 1996). We tested for evidence of positive selection by comparing the nonsynonymous substitution rate (d*N*) to the synonymous substitution rate (d*S*). Overall d*N*/d*S* values were calculated with the codeml program from PAML (Model M0) (Yang, 2007). Site-specific ML models were used to detect positive selection on specific amino acids. Site models treat the d*N*/d*S* ratio for any site (codon) in the gene as a random variable from a statistical distribution, thus allowing d*N*/d*S* to vary among codons (Nielsen and Yang, 1998; Yang et al., 2000). Positive selection is defined as the presence of some codons at which d*N*/d*S*>1. We implemented models M7 (beta) and M8 (beta&ω) (Nielsen and Yang, 1998; Wong et al., 2004; Yang et al., 2005) with the codeml program in PAML (Yang 2007). This test compares the likelihood of the data under a model in which d*N*/d*S* among sites is constrained between 0 and 1 (M7), and against a model in which an additional category of sites with d*N*/d*S*>1 is allowed (M8). Data analyses and computer simulations showed that these pairs of site models are well suited to detect positively selected sites, even for small phylogenies (Anisimova et al., 2001, 2002, 2003). A likelihood-ratio test (LRT) was used to compare selection models. The test statistic $-2\Delta\ln L$ follows a χ^2 distribution with critical values to be 5.99 and 9.21 at 5% and 1% ($df=2$). When the LRT was significant, a Bayes Empirical Bayes (BEB) procedure was used to identify amino acid under positive selection (Yang et al., 2005). All sites with a posterior probability of >0.5 of being in the positively selected class were reported. Only those codons that were represented in all species were chosen for analysis. The F3x4 model of codon frequencies was used. As for the identification of orthologs between *Symbiodinium* sp. C3 and *Symbiodinium* sp. CassKB8, a tBLASTx bitscore cutoff of 40 was used for the alignment between a given DS ortholog and the corresponding hit in *A. tamarense*.

3. Results

3.1. EST library analysis

A total of 2,304 clones were sequenced from 5′ and 3′ ends giving rise to 4,608 sequence reads that were assembled into 1,738 (1129 contigs and 609 singlets) unique sequences (UniSeqs). The number of UniSeqs probably represents an overestimation of the true number of unique genes, as some of them may represent non-overlapping sequences originating from the same gene. 1,484 UniSeqs had an insert greater than 100 bp and were subsequently subjected to annotation with the KEGG Automatic Annotation Server (KAAS) (Moriya et al., 2007) (Table 1). 155 sequences (10.4%) were annotated

Table 2
Comparison of Top 20 BLASTx hits of orthologs to *P. falciparum* and nr database.

Top 20 C3 and CassKB8 EST BLASTx hits to the proteins of the apicomplexan *P. falciparum*			
Ortholog	C3 *e*-value	CassKB8 *e*-value	Protein description
CE_ortholog1	0	0	Actin
CE_ortholog2	0	E −98	Ribonucleoside-diphosphate reductase small chain
CE_ortholog29	0	E −81	ATP synthase beta chain
CE_ortholog6	0	E −52	alpha tubulin
CE_ortholog5	E −72	E −80	14-3-3 protein homologue
CE_ortholog8	E −88	E −55	Adenosylhomocysteinase
CE_ortholog18	0	E −35	Elongation factor 2
CE_ortholog12	E −68	E −66	ADP-ribosylation factor
CE_ortholog16	E −80	E −47	Cytochrome oxidase subunit 1
CE_ortholog19	E −89	E −31	Glyceraldehyde-3-phosphate dehydrogenase
CE_ortholog25	E −73	E −38	Cytochrome b
CE_ortholog13	E −53	E −49	Cyclophilin
CE_ortholog67	E −56	E −30	Ribosomal protein family L5
CE_ortholog47	E −42	E −39	Splicesome-associated protein
CE_ortholog39	E −24	E −34	Protein phosphatase
CE_ortholog11	E −30	E −25	Proliferating cell nuclear antigen
CE_ortholog9	E −25	E −25	Nitrate transporter
CE_ortholog82	E −36	E −13	T-complex protein beta subunit
CE_ortholog22	E −26	E −20	RNA-binding protein mei2 homologue
CE_ortholog45	E −28	E −15	Glycine–tRNA ligase
Top 20 C3 and CassKB8 EST BLASTx hits to the GenBank nr database			
Ortholog	C3 *e*-value	CassKB8 *e*-value	Protein description
CE_ortholog1	0	0	Actin
CE_ortholog2	0	0	Ribonucleotide Reductase, R2/beta subunit
CE_ortholog3	0	0	Peridinin chlorophyll-a binding protein apoprotein precursor
CE_ortholog7	0	E −91	Plastid C1 class II fructose bisphosphate aldolase
CE_ortholog29	0	E −88	F1 ATP synthase beta subunit
CE_ortholog25	0	E −79	Cytochrome b
CE_ortholog16	0	E −72	Cytochrome oxidase subunit 1
CE_ortholog11	E −92	E −79	Proliferating cell nuclear antigen
CE_ortholog8	E −97	E −66	Adenosylhomocysteinase
CE_ortholog6	0	E −6	alpha tubulin
CE_ortholog19	0	E −58	Glyceraldehyde-3-phosphate dehydrogenase
CE_ortholog13	E −82	E −73	Cyclophilin-like protein
CE_ortholog12	E −78	E −73	ADP-ribosylation factor
CE_ortholog46	E −91	E −5	Homolog to co-chaperone p23
CE_ortholog18	0	E −38	Elongation factor 2
CE_ortholog40	0	E −31	Elongation factor 1, alpha-like protein
CE_ortholog52	0	E −21	Glyceraldehyde-3-phosphate dehydrogenase
CE_ortholog56	0	E −19	Conserved hypothetical plastid protein
CE_ortholog55	E −94	E −19	Oxygen evolving enhancer 1 precursor (photosystem II)
CE_ortholog43	E −69	E −42	small GTPase Rab1

for *Symbiodinium* sp. CassKB8. Approximately half of all annotated genes were involved in metabolism in this library (82 genes). Within this category the majority of genes were either assorted to 'carbohydrate metabolism' (23 out of 82 genes), 'energy metabolism' (17 out of 82 genes), or 'amino acid metabolism' (12 out of 82 genes) (Table 1). In addition, we found genes belonging to a number of important signal transduction pathways present in the library (e.g. MAPK signaling pathway, Wnt signaling pathway, Notch signaling pathway, TGF-beta signaling pathway). Furthermore, a number of genes were assorted to 'cell communication'. This category included genes associated with focal adhesion, adherence junctions, tight junctions, and gap junctions (e.g. collagen, type I/II/III/V/XI, alpha; actin beta/gamma 1; Wiskott–Aldrich syndrome protein; protein phosphatase 2; tubulin alpha) that mediate contacts and communication between cells.

3.2. Ortholog comparison

We conducted a tBLASTx-based Best Reciprocal BLAST Hit (BRBH) approach to identify orthologs between our *Symbiodinium* sp. CassKB8 (CassKB8) library and another EST library from *Symbiodinium* sp. C3 (C3). For C3, a collection of 5,664 ESTs comprised of single-pass 5′ reads was assembled into 3,466 UniSeqs (903 contigs and 2,563 singlets). The sequences used here are an extension of those described in Leggat *et al.* (2007). We identified 132 potential orthologs from both EST libraries. After manual inspection of the alignments, 115 sequences were selected for further analysis. We subdivided our analysis between the two *Symbiodinium* strains into three groups of genes: CE (conserved across eukaryotes), DS (dinoflagellate-specific), and SS (*Symbiodinium*-specific). 82% of the 115 putative orthologs were CE, 15% were DS, and 3% were SS.

The closest species to dinoflagellates that are represented by sequenced genomes belong to the apicomplexa (e.g. *P. falciparum*). Although dinoflagellates and apicomplexans are sister taxa within the alveolates (Bhattacharya et al., 2004), apicomplexans lead an exclusively parasitic lifestyle. The set of 95 CE orthologs was blasted against the nr protein database and the full set of *P. falciparum* proteins in order to assess coherence. When blasted against the GenBank protein database (nr), 44 out of the 95 orthologs had the highest similarity to the same protein (*e*-values $< 10^{-10}$). In contrast, only 24 of the 95 orthologs had the same best BLAST hit in the *P. falciparum* protein sequence database (*e*-value $< 10^{-10}$). The most highly conserved proteins between these organisms were housekeeping genes such as actin, α-tubulin, and glyceraldehyde-3-phosphate dehydrogenase (Table 2). Nevertheless, there were also differences. *P. falciparum*, not surprisingly, appears to have lost most of the genes related to photosynthesis among other metabolic genes. Thus, despite their close phylogenetic relationship, there are substantial differences in gene content between dinoflagellates and apicomplexans.

We were interested to find out whether the symbiotic lifestyle might have affected the evolution of some of the dinoflagellate-specific genes in C3 and CassKB8. If this is the case, selective constraints on a gene should be similar between C3 and CassKB8 and different from other non-symbiotic dinoflagellate taxa. Pairwise d*N*/d*S* ratios for C3 and CassKB8 were compared to the d*N*/d*S* ratios

Table 3
Evolutionary analysis of DS orthologs (dinoflagellate-specific genes).

Ortholog	C3 CassKB8 alignment bit score	*A. tamarense* EST	tBLASTx *e*-value C3	tBLASTx *e*-value CassKB8	LC	d*N*/d*S* C3 CassKB8	d*N*/d*S* C3 CassKB8 *A. tamarense* (M0)	logL M7	logL M8	*p* (M7 vs M8)
DS_ortholog2	**177**	**CF947354**	**1.00E −11**	**7.00E −11**	**108**	**0.003**	**0.006**	**−507.5611**	**−503.9808**	**0.0279**
DS_ortholog4	174	CF948464	1.00E −51	9.00E −30	100	0.007	0.002	−785.1313	−782.9563	ns
DS_ortholog5	125	CK784543	4.00E −16	5.00E −25	67	0.124	0.057	−572.2892	−570.8564	ns
DS_ortholog6	116	CK433542	2.00E −11	2.00E −12	44	0.026	0.036	−336.8355	−336.5720	ns
DS_ortholog7	108	CK785584	2.00E −05	4.00E −12	84	0.004	0.003	−845.2768	−845.2768	ns
DS_ortholog9	89	CK784533	3.00E −33	7.00E −19	36	0.020	0.016	−239.6259	−239.6259	ns
DS_ortholog10	88	CK786644	4.00E −10	4.00E −08	51	0.030	0.036	−440.6544	−440.6544	ns
DS_ortholog12	70	CK785754	2.00E −15	5.00E −10	52	0.329	0.123	−541.6122	−541.6122	ns
DS_ortholog13	65	CX769605	4.00E −04	4.00E −04	65	0.174	0.008	−704.6121	−702.2552	0.0947
				AV d*N*/d*S*		**0.080**	**0.032**			

Note. LC, length of sequence in codons; *p*, *p* value of LRT comparing models listed, significant values (0.05 level) are bold; ns, not significant.

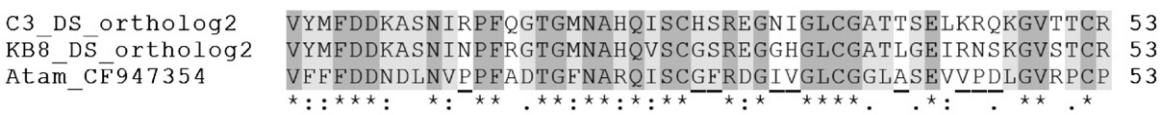

Fig. 1. Alignment of DS_ortholog 2 from *Symbiodinium* sp. C3, *Symbiodinium* sp. CassKB8, and *A. tamarense*. Underlined amino acids present sites of positive selection with a posterior probability of >0.5 as determined by Bayes Empirical Bayes analysis. *A. tamarense* accession number is given next to species acronym. Atam: *A. tamarense*, C3: *Symbiodinium* sp. C3, KB8: *Symbiodinium* sp. CassKB8. The symbols "*", ":", "." denote the degree of conservation observed in each column. "*": residues are identical, ":" conservative substitution, "." semi-conservative substitution.

that were obtained from the alignment of C3 and CassKB8 to the homologous sequence of *Alexandrium tamarense*. This organism is an autotrophic dinoflagellate from a different order, the Gonyaulacales. Of the 17 DS genes, one sequence was a putative rRNA-coding gene as indicated by multiple stop codons in the BLAST alignments and the annotation of homologous sequences from other dinoflagellates. This gene was excluded from subsequent analyses. Seven genes had no homolog in *A. tamarense*. Therefore, a total of nine genes were analyzed in the DS group (Table 3). On average, d*N*/d*S* values between C3 and CassKB8, and between the latter and *A. tamarense* were much smaller than 1. The mean d*N*/d*S* rate of substitution was 0.080 for C3 and CassKB8, indicating low average divergence between different *Symbiodinium* species. In the comparison of *Symbiodinium* sp. C3 and CassKB8 to *A. tamarense* the mean d*N*/d*S* ratio was 0.032. Thus, it was generally lower between different dinoflagellate taxa than between species within the same taxon. Six out of the nine genes showed higher d*N*/d*S* ratios between *Symbiodinium* strains than the latter to *A. tamarense*. Site-specific analyses identified one gene with significant signatures of Darwinian selection (DS_ortholog2, $\chi^2 = 7.16$, $p < 0.05$), and one gene that barely missed significance (DS_ortholog13, $\chi^2 = 4.71$, $p < 0.1$). Nine adaptively evolving sites were identified for DS_ortholog2 (position 12, 27, 28, 32, 33, 40, 44, 45, 46; average d*N*/d*S* at these sites: 4.37) (Fig. 1). An InterProScan analysis did not match with any known protein motif, and the function of this ortholog remains to be elucidated.

DS_ortholog9 seems to be highly conserved across dinoflagellates as this gene gave significant tBLASTx hits to 5 other dinoflagellate species besides *A. tamarense* (CK784533), namely *Amphidinium carterae* (CF065838), *Karlodinium micrum* (EC159040), *Lingulodinium polyedrum* (CD809735), *Pfiesteria piscicida* (DQ864906), and *Heterocapsa triquetra* (DT384419). Fig. 2 depicts the full coding sequence (cds) as indicated by stop codons upstream and downstream of the aligned sequences. This gene codes for a small protein that is highly conserved throughout dinoflagellates (average d*N*/d*S* = 0.009). An InterProScan revealed that the protein has a Zinc finger, C2H2-type domain in position 30 to 52. As the structural motif occupies about a third of the whole protein sequence, other structural features need to be organized in a compact way.

Three out of a total of 20 orthologs that were found in dinoflagellates appear to be *Symbiodinium*-specific. SS_ortholog1 covers an ORF of 163 amino acids (Fig. 3a). The ortholog showed a high degree of conservation between C3 and CassKB8 at the amino acid level (95%). Nevertheless, the pairwise d*N*/d*S* ratio was 0.520, indicating accelerated divergence in this protein between both species. An InterProScan did not indicate any conserved domains. SS_ortholog2 contained an ORF of 56 amino acids that was represented in both EST libraries (Fig. 3b). This portion covered the 3′ end of the protein as indicated by a 3′ stop codon and a poly(A) tail in both sequences. The pairwise d*N*/d*S* ratio was 0.189, which is higher than the average of the orthologs that are conserved across dinoflagellates (Table 3). A search for conserved domains was inconclusive. SS_ortholog3 shared a stretch of 30 amino acids in both *Symbiodinium* strains. Interestingly, CassKB8 contained a 3′ end that is not shared by C3 as indicated by a stop codon (Fig. 3c). The pairwise d*N*/d*S* was low (0.041), which suggests that the sequence is under negative selection. An InterProScan did not reveal any homologies to known protein motifs.

4. Discussion

4.1. EST library analysis

The results of the KEGG analysis show that the majority of the annotated genes identified from the CassKB8 EST dataset are metabolic genes. The genes that were assorted to signal transduction pathways belong to cell surface receptor-linked signal transduction pathways. This may suggest that *Symbiodinium* in culture express a broad range of receptors to sense and explore their environment. Furthermore, we found a number of genes that have functions related to cell–cell contacts and communication, specifically genes playing a role in focal adhesions, adherens junctions, tight junctions, and gap junctions. All these processes manage contacts and communication between cells. The identification of genes that play a role in cell–cell adhesion and cell–cell communication highlight the possibility that *Symbiodinium* in culture communicate and associate with each other. It is known that zooxanthellae in culture can occur as coccoid non-motile cells or non-motile gymnodinoid cells (Freudenthal, 1962; Trench, 1981), but there has been little attempt to relate these forms to differences in aggregation behavior. Nevertheless, most of the genes that were associated to cell–cell contacts and communication serve other functions as well (e.g. parts of cytoskeletal network, signal pathways).

4.2. Ortholog comparison

Using a BRBH approach, around 8% of the sequences in our EST library were identified as orthologs. Given the distribution of orthologs

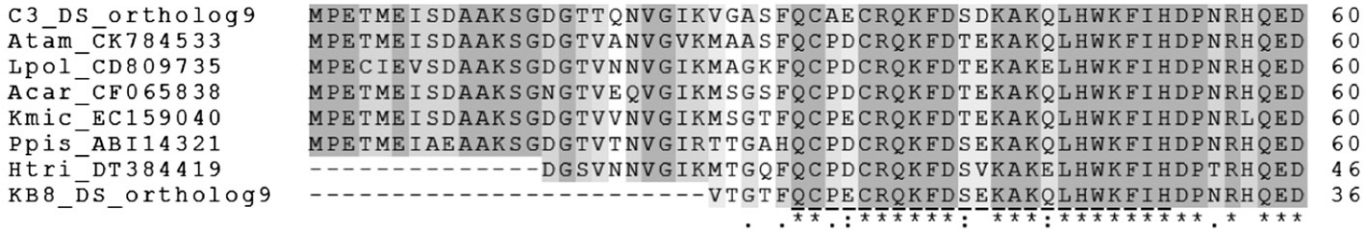

Fig. 2. Alignment of full cds of DS_ortholog9 from *Symbiodinium* sp. C3, *Symbiodinium* sp. CassKB8, *Alexandrium tamarense*, *Amphidinium carterae*, *Karlodinium micrum*, *Lingulodinium polyedrum*, *Heterocapsa triquetra*, and *Pfiesteria piscicida*. Underlined amino acids present a Zinc finger, C2H2-type domain that is conserved across all proteins. Accession numbers are given next to species acronym. Atam: *A. tamarense*, C3: *Symbiodinium* sp. C3, KB8: *Symbiodinium* sp. CassKB8, Acar: *Amphidinium carterae*, Kmic: *Karlodinium micrum*, Lpol: *Lingulodinium polyedrum*, Htri: *Heterocapsa triquetra*, Ppis: *Pfiesteria piscicida*. The symbols "*", ":", "." denote the degree of conservation observed in each column. "*": residues are identical, ":" conservative substitution, "." semi-conservative substitution.

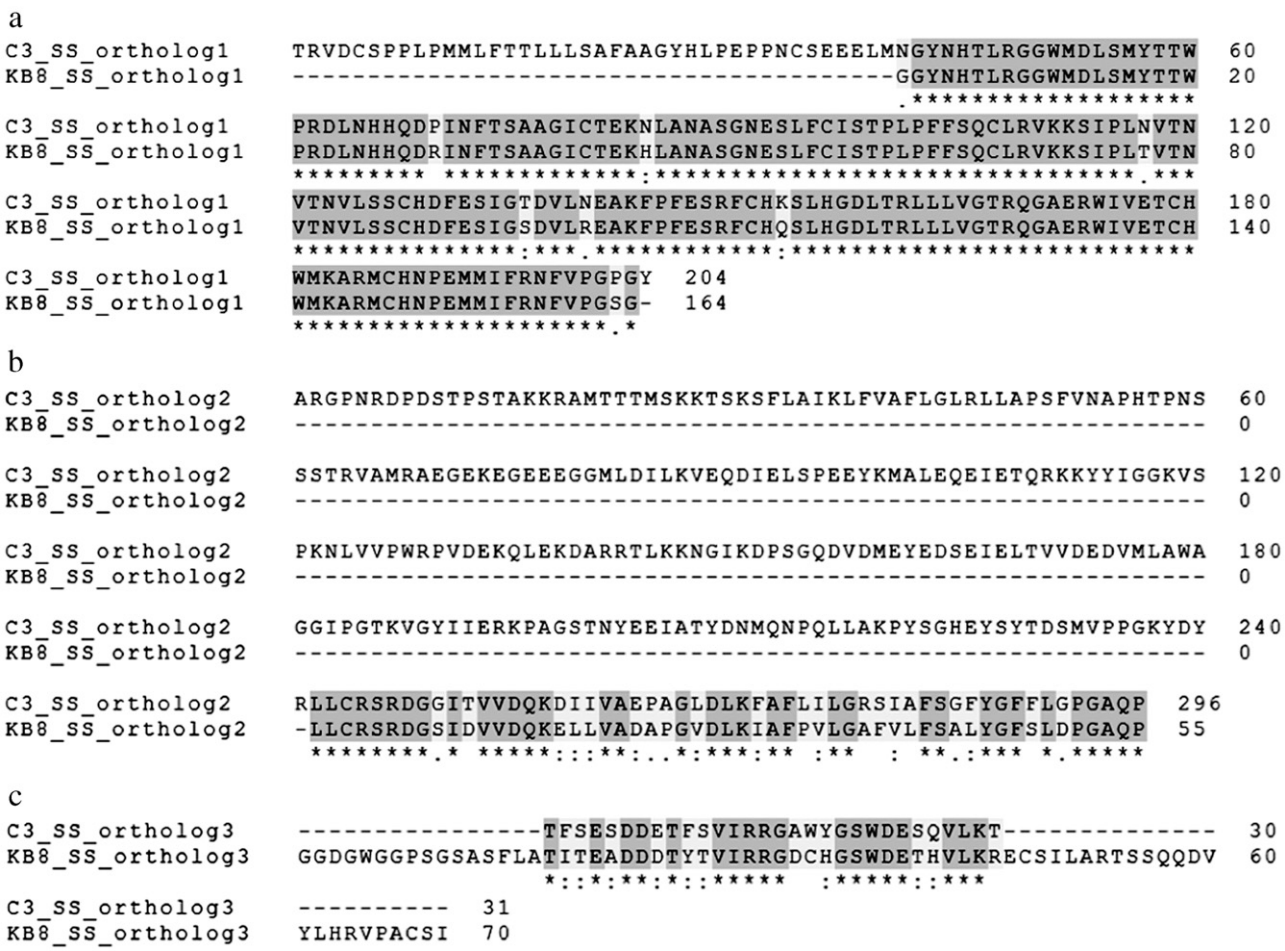

Fig. 3. Alignment of the three *Symbiodinium*-specific orthologs. Longest ORFs for given species are shown in alignment. a) SS_ortholog1; b) SS_ortholog2; c) SS_ortholog3; C3: *Symbiodinium* sp. C3, KB8: *Symbiodinium* sp. CassKB8. The symbols "*", ":", "." denote the degree of conservation observed in each column. "*": residues are identical, ":" conservative substitution, "." semi-conservative substitution.

for the three categories, our data suggest that 18% of a *Symbiodinium* transcriptome consist of DS genes (3% of which are SS), while the majority of genes fall in the CE category (82%). The difference in the number of hits with an *e*-value $<10^{-10}$ for the CE orthologs in *P. falciparum* (24/88) in comparison to the NCBI nr database (44/88) emphasizes that there are likely to be substantial differences in gene content between dinoflagellates and apicomplexans. Although other apicomplexan genomes were available (e.g. *Toxoplasma gondii*), *P. falciparum* provided the most comprehensive annotation of proteins, which is why we chose this species for the comparison. With the forecast of a *Symbiodinium* genome becoming available, further studies should enable the identification of genes that are crucial to a lifestyle inside the cells of other eukaryotes, or alternatively for the entry into the host, by unearthing those genes that are exclusively found and conserved between *Symbiodinium* spp. and one (*P. falciparum*) or more of the apicomplexans.

There is a possibility that we were not able to identify the true ortholog from *A. tamarense* for some of the DS genes. However, as orthologs from C3 and CassKB8 had to identify the same sequence in *A. tamarense*, the chances of identifying a false positive were significantly lowered. In any case, the nature of the approach itself made it impossible to rule out that there were other potential homologs that are the true orthologs. We tried to minimize the source of error by choosing the phylogenetically least distant taxon (Hackett et al., 2004a) that is covered by a large collection of ESTs. Other dinoflagellate species with similar sized EST libraries (e.g. *Karenia* sp., *Karlodinium* sp.) had poorer overall alignments and BLAST *e*-values. Furthermore, phylogenetic distance and differences in ecological niche were the same by selecting the same outgroup for all DS orthologs analyzed. We found evidence for positive selection in one of the nine DS genes. d*N*/d*S* ratios were on average lower between different dinoflagellate taxa than between species of *Symbiodinium*. It is unlikely that the observed pattern is a systematic artifact as there was no significant correlation between d*N*/d*S* ratios and 1) the quality of the ortholog alignment, 2) the number of codons that were subjected to analysis, or 3) the quality of the alignment of the *Symbiodinium* orthologs to the *A. tamarense* homolog (Pearson Product Moment Correlation: all $p>0.05$). However, we cannot rule out that demographic/population genetic factors are affecting the d*N*/d*S* ratio across genes. There is a significant correlation between the d*N*/d*S* ratios from different *Symbiodinium* species and from the latter to *A. tamarense* (Pearson Product Moment Correlation: correlation coefficient >0.8, $p<0.01$). This shows that, although d*N*/d*S* ratios are consistently higher between both *Symbiodinium* species, selective constraints for a given protein are connected across lineages (Supplementary Fig. 1). The pattern we see is based on a small number of genes, and CassKB8 and C3 belong to two extremely divergent clades of *Symbiodinium*. Thus, there is a possibility that d*S* values are underestimated, which in consequence would lead to increased d*N*/d*S* ratios. Nevertheless, d*N*/d*S* ratios between CassKB8 and C3 were low on average (Table 3). Hence, it is unlikely that this is a strong confounding factor.

It is interesting that we find the highest d*N*/d*S* in a SS gene. Since there are only three of them, we might speculate that the portion of genes with high d*N*/d*S* in the SS group is high. The evolution of proteins within *Symbiodinium* seems to be elevated and we speculate that the symbiotic lifestyle might affect population structure and strength of selection (e.g. genes might need to coevolve with their coral host, or

evolve as a consequence of competition among different symbiont species). Lutzoni and Pagel (1997) demonstrate a highly significant association between mutualism and increased rates of evolution in nuclear ribosomal DNA in fungi, and suggest that the increased rate of evolution after the adoption of a mutualistic lifestyle is generalized across the genome of mutualists. When Castillo-Davis et al. (2004) compared positively selected genes between two nematodes (*Caenor-* and *Caenorhabditis briggsae*) and two vertebrates (*Homo sapiens* and *Mus musculus*), they found that fast evolving proteins were distributed heterogeneously among functional classes in both lineages, but that all genes were to a large extent components of passive or active coevolving systems, either within or between genomes. As *Symbiodinium* species live as symbionts, they need to coevolve with their coral host. This could ultimately lead to accelerated protein divergence. Prominent examples that coevolution leads to accelerated evolution are 1) the genetic "arms race" between genomes of host and pathogen (Fitch et al., 1991; Gray and Gill, 1993), or 2) the sexual selection between or among sexes (Singh and Kulathinal, 2000; Wyckoff et al., 2000; Swanson and Vacquier, 2002; Torgerson et al., 2002). If there is faster evolutionary change in *Symbiodinium* spp. because of coevolution due to a symbiotic lifestyle, then the pattern should be especially visible in genes that play a role in symbiosis. Indeed, one of the three SS orthologs shows a high d*N*/d*S* ratio (SS_ortholog1 d*N*/d*S* = 0.520) despite a high degree of amino acid conservation (95%). Given the increased evolutionary rate of this protein at the codon level but the conservation on the amino acid sequence, we might hypothesize that this gene is coevolving passively and plays a role in symbiosis. Passive coevolution has been experimentally demonstrated for transcription factor binding sites in fruit flies, where regulatory mutations in *cis*-acting DNA-binding sites are compensated by coevolved differences at nearby sites (Ludwig et al., 2000; Shaw et al., 2002). SS_ortholog2 shows a d*N*/d*S* ratio that is higher than average among the DS genes, which further indicates that the symbiotic lifestyle of *Symbiodinium* spp. could affect evolution of genes. SS_ortholog3 in contrast has an overall low d*N*/d*S* ratio (0.041), but is not found outside of *Symbiodinium*. Highly conserved lineage-specific genes are commonly referred to as orphan genes. It has been hypothesized that slow-evolving orphan genes may be candidate genes for lineage-specific adaptations that are under strong purifying selection after experiencing a history of adaptive evolution (Domazet-Loso and Tautz, 2003).

We will soon have more sequences for different *Symbiodinium* species. New data will enable us to extend these analyses to examine 1) if genes with elevated d*N*/d*S* among species are genes important for symbiosis function, and 2) if the accelerated evolution of genes may be caused by divergence in function through positive selection or relaxed selection or both. Scientific interest in dinoflagellates has risen recently because of the many unique biochemical and genomic characteristics found in these organisms in addition to the many ecological roles they have. Arguably the most important is the role these organisms play in the health of coral reefs. A comprehensive knowledge of the biology of *Symbiodinium,* including host-symbiont mutualism, serial endosymbiosis and potential horizontal gene transfer in the coral-algae symbioses is important if we are to understand the functional genomics of this biological system and how it evolved. A deeper understanding of genes and principles of symbiosis may provide us with a novel concept and connected countermeasures to determine what crucial efforts have to be met in order to sustain thriving coral reefs.

Acknowledgements

We would like to thank Andres Aguilar for helpful comments on the d*N*/d*S* analyses, Pilar Francino, Miriam Barlow and Michael DeSalvo for helpful comments on the manuscript. We would also like to thank Robert A. Kinzie III for initial donation of the *Symbiodinium* sp. CassKB8 culture and the Aquarium of Niagara for seawater. This study was supported through NSF awards BE-GEN 0313708 and IOS 0644438 (MM), OCE 0424996 (MAC) and through the ARC Centre of Excellence for Coral Reef Studies (DY, WL).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.cbd.2008.11.001.

References

Allen, J.R., Roberts, M., Loeblich III, A.R., Klotz, L.C., 1975. Characterization of the DNA from the dinoflagellate *Crypthecodinium cohnii* and implications for nuclear organization. Cell 6, 161–169.
Anisimova, M., Bielawski, J.P., Yang, Z., 2001. Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. Mol. Biol. Evol. 18, 1585–1592.
Anisimova, M., Bielawski, J.P., Yang, Z., 2002. Accuracy and power of Bayes prediction of amino acid sites under positive selection. Mol. Biol. Evol. 19, 950–958.
Anisimova, M., Nielsen, R., Yang, Z., 2003. Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. Genetics 164, 1229–1236.
Apprill, A.M., Gates, R.D., 2007. Recognizing diversity in coral symbiotic dinoflagellate communities. Mol. Ecol. 16, 1127–1134.
Bachvaroff, T.R., Concepcion, G.T., Rogers, C.R., Herman, E.M., Delwiche, C.F., 2004. Dinoflagellate expressed sequence tag data indicate massive transfer of chloroplast genes to the nuclear genome. Protist 155, 65–78.
Baker, A.C., 2003. Flexibility and specificity in coral-algal symbiosis: diversity, ecology, and biogeography of Symbiodinium. Ann. Rev. Ecolog. Evol. Syst. 34, 661–689.
Bhattacharya, D., Yoon, H.S., Hackett, J.D., 2004. Photosynthetic eukaryotes unite: endosymbiosis connects the dots. BioEssays 26, 50–60.
Castillo-Davis, C.I., Kondrashov, F.A., Hartl, D.L., Kulathinal, R.J., 2004. The functional genomic distribution of protein divergence in two animal phyla: coevolution, genomic conflict, and constraint. Genome Res. 14, 802–811.
Cavalier-Smith, T., 1998. A revised six-kingdom system of life. Biol. Rev. Camb. Philos. Soc. 73, 203–266.
Cavalier-Smith, T., 1999. Principles of protein and lipid targeting in secondary symbiogenesis: euglenoid, dinoflagellate, and sporozoan plastid origins and the eukaryote family tree. J. Eukaryot. Microbiol. 46, 347–366.
Cavalier-Smith, T., 2002. Chloroplast evolution: secondary symbiogenesis and multiple losses. Curr. Biol. 12, R62–64.
Coffroth, M.A., Lewis, C.F., Santos, S.R., Weaver, J.L., 2006. Environmental populations of symbiotic dinoflagellates in the genus Symbiodinium can initiate symbioses with reef cnidarians. Curr. Biol. 16, R985–R987.
Domazet-Loso, T., Tautz, D., 2003. An evolutionary analysis of orphan genes in *Drosophila*. Genome Res. 13, 2213–2219.
Dove, S., 2004. Scleractinian corals with photoprotective host pigments are hypersensitive to thermal bleaching. Mar. Ecol., Prog. Ser. 272, 99–116.
Edgar, R.C., 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5, 113.
Ewing, B., Hillier, L., Wendl, M.C., Green, P., 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. 8, 175–185.
Fitch, W.M., Leiter, J.M.E., Li, X., Palese, P., 1991. Positive Darwinian evolution in human influenza a viruses. Proc. Natl. Acad. Sci. U. S. A. 88, 4270–4274.
Freudenthal, H., 1962. Symbiodinium gen. nov. and *Symbiodinium microadriaticum* sp. nov., a zooxanthellae; taxonomy, life cycle, and morphology. I. ProtozooL 9, 45–52.
Graham, L.E., Wilcox, L.W., 2000. Algae. Prentice Hall, Upper Saddle River, N.J.
Gray, G.D., Gill, H.S., 1993. Host genes, parasites and parasitic infections. Int. J. Parasitol. 23, 485–494.
Hackett, J.D., Anderson, D.M., Erdner, D.L., Bhattacharya, D., 2004a. Dinoflagellates: a remarkable evolutionary experiment. Am. J. Bot. 91, 1523–1534.
Hackett, J.D., Yoon, H.S., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Nosenko, T., Bhattacharya, D., 2004b. Migration of the plastid genome to the nucleus in a peridinin dinoflagellate. Curr. Biol. 14, 213–218.
Hackett, J.D., Scheetz, T.E., Yoon, H.S., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Bhattacharya, D., 2005. Insights into a dinoflagellate genome through expressed sequence tag analysis. BMC Genomics 6, 80.
Hinnebusch, A.G., Klotz, L.C., Immergut, E., Loeblich III, A.R., 1980. Deoxyribonucleic acid sequence organization in the genome of the dinoflagellate *Crypthecodinium cohnii*. Biochemistry 19, 1744–1755.
Huang, X., Madan, A., 1999. CAP3: a DNA sequence assembly program. Genome Res. 9, 868–877.
John, U., Fensome, R.A., Medlin, L.K., 2003. The application of a molecular clock based on molecular sequences and the fossil record to explain biogeographic distributions within the *Alexandrium tamarense* "species complex" (Dinophyceae). Mol. Biol. Evol. 20, 1015–1027.
Kitano, H., Oda, K., 2006. Self-extending symbiosis: a mechanism for increasing robustness through evolution. Biol. Theory 1, 61–66.
LaJeunesse, T.C., 2001. Investigating the biodiversity, ecology, and phylogeny of endosymbiotic dinoflagellates in the genus Symbiodinium using the ITS region: in search of a "species" level marker. J. Phycol. 37, 866–880.
LaJeunesse, T.C., 2002. Diversity and community structure of symbiotic dinoflagellates from Caribbean coral reefs. Mar. Biol. 141, 387–400.
LaJeunesse, T.C., Loh, W.K.W., van Woesik, R., Hoegh-Guldberg, O., Schmidt, G.W., Fitt, W.K., 2003. Low symbiont diversity in southern Great Barrier Reef corals, relative to those of the Caribbean. Limnol. Oceanogr. 48, 2046–2054.

Leggat, W., Hoegh-Guldberg, O., Dove, S., Yellowlees, D., 2007. Analysis of an EST library from the Dinoflagellate (Symbiodinium sp.) symbiont of reef-building corals. J. Phycol. 43, 1010.
Ludwig, M.Z., Bergman, C., Patel, N.H., Kreitman, M., 2000. Evidence for stabilizing selection in a eukaryotic enhancer element. Nature 403, 564–567.
Lutzoni, F., Pagel, M., 1997. Accelerated evolution as a consequence of transitions to mutualism. Proc. Natl. Acad. Sci. U. S. A. 94, 11422–11427.
Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A.C., Kanehisa, M., 2007. KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. 35, W182–185.
Muscatine, L., McCloskey, L.R., Marian, R.E., 1981. Estimating the daily contribution of carbon from zooxanthellae to coral animal respiration. Limnol. Oceanogr. 26, 601–611.
Nielsen, R., Yang, Z., 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics 148, 929–936.
Page, R.D., 1996. TreeView: an application to display phylogenetic trees on personal computers. Comput. Appl. Biosci. 12, 357–358.
Parfrey, L.W., Barbero, E., Lasser, E., Dunthorn, M., Bhattacharya, D., Patterson, D.J., Katz, L.A., 2006. Evaluating support for the current classification of eukaryotic diversity. PLoS Genet. 2, e220.
Patron, N.J., Waller, R.F., Archibald, J.M., Keeling, P.J., 2005. Complex protein targeting to dinoflagellate plastids. J. Mol. Biol. 348, 1015–1024.
Patron, N.J., Waller, R.F., Keeling, P.J., 2006. A tertiary plastid uses genes from two endosymbionts. J. Mol. Biol. 357, 1373–1382.
Petrov, D.A., 2001. Evolution of genome size: new approaches to an old problem. Trends Genet. 17, 23–28.
Pochon, X., Montoya-Burgos, J.I., Stadelmann, B., Pawlowski, J., 2006. Molecular phylogeny, evolutionary rates, and divergence timing of the symbiotic dinoflagellate genus Symbiodinium. Mol. Phylogenet. Evol. 38, 20–30.
Rowan, R., 1998. Diversity and ecology of zooxanthellae on coral reefs. J. Phycol. 34, 407–417.
Rowan, R., Powers, D.A., 1991. A molecular genetic classification of zooxanthellae and the evolution of animal-algal symbioses. Science 251, 1348–1351.
Santos, S.R., Taylor, D.J., Kinzie III, R.A., Hidaka, M., Sakai, K., Coffroth, M.A., 2002. Molecular phylogeny of symbiotic dinoflagellates inferred from partial chloroplast large subunit (23S)-rDNA sequences. Mol. Phylogenet. Evol. 23, 97–111.
Shaw, P.J., Wratten, N.S., McGregor, A.P., Dover, G.A., 2002. Coevolution in bicoid-dependent promoters and the inception of regulatory incompatibilities among species of higher Diptera. Evol. Dev. 4, 265–277.
Singh, R.S., Kulathinal, R.J., 2000. Sex gene pool evolution and speciation: a new paradigm. Genes Genet. Syst. 75, 119–130.
Spector, D.L., 1984. Dinoflagellate nuclei. In: Spector, D.L. (Ed.), Dinoflagellates. Academic Press, Inc., Orlando, Florida, USA.
Swanson, W.J., Vacquier, V.D., 2002. The rapid evolution of reproductive proteins. Nat. Rev., Genet. 3, 137–144.
Szmant, A.M., 1986. Reproductive ecology of Caribbean reef corals. Coral Reefs 5, 43–53.
Tanikawa, N., Akimoto, H., Ogoh, K., Chun, W., Ohmiya, Y., 2004. Expressed sequence tag analysis of the dinoflagellate *Lingulodinium polyedrum* during dark phase. Photochem. Photobiol. 80, 31–35.
Tatusov, R.L., Koonin, E.V., Lipman, D.J., 1997. A genomic perspective on protein families. Science 278, 631–637.
Tatusov, R.L., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Kiryutin, B., Koonin, E.V., Krylov, D.M., Mazumder, R., Mekhedov, S.L., Nikolskaya, A.N., Rao, B.S., Smirnov, S., Sverdlov, A.V., Vasudevan, S., Wolf, Y.I., Yin, J.J., Natale, D.A., 2003. The COG database: an updated version includes eukaryotes. BMC Bioinformatics 4, 41.
Telford, M.J., 2007. Phylogenomics. Curr. Biol. 17, R945–946.
Torgerson, D.G., Kulathinal, R.J., Singh, R.S., 2002. Mammalian sperm proteins are rapidly evolving: evidence of positive selection in functionally diverse genes. Mol. Biol. Evol. 19, 1973–1980.
Trench, R.K., 1981. Cellular and molecular interactions in symbioses between dinoflagelates and marine invertebrates. Pure Appl. Chem. 513, 819–835.
Trench, R.K., 1993. Microalgal-invertebrate symbioses — a review. Endocytobiosis Cell Res. 9, 135–175.
Wong, W.S., Yang, Z., Goldman, N., Nielsen, R., 2004. Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. Genetics 168, 1041–1051.
Wyckoff, G.J., Wang, W., Wu, C.I., 2000. Rapid evolution of male reproductive genes in the descent of man. Nature 403, 304–309.
Yang, Z., 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24, 1586–1591.
Yang, Z., Nielsen, R., Goldman, N., Pedersen, A.M., 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155, 431–449.
Yang, Z., Wong, W.S., Nielsen, R., 2005. Bayes empirical Bayes inference of amino acid sites under positive selection. Mol. Biol. Evol. 22, 1107–1118.
Yellowlees, D., Rees, T.A., Leggat, W., 2008. Metabolic interactions between algal symbionts and invertebrate hosts. Plant Cell Environ. 31, 679–694.
Yoon, H.S., Hackett, J.D., Van Dolah, F.M., Nosenko, T., Lidie, K.L., Bhattacharya, D., 2005. Tertiary endosymbiosis driven genome evolution in dinoflagellate algae. Mol. Biol. Evol. 22, 1299–1308.