The ISME Journal (2015) 9, 844–856

www.nature.com/ismej

ORIGINAL ARTICLE

Unfolding the secrets of coral–algal symbiosis

Nedeljka Rosic[1], Edmund Yew Siang Ling[2], Chon-Kit Kenneth Chan[3], Hong Ching Lee[4], Paulina Kaniewska[1,5], David Edwards[3,6,7], Sophie Dove[1,8] and Ove Hoegh-Guldberg[1,8,9]

[1]*School of Biological Sciences, The University of Queensland, St Lucia, Queensland, Australia;* [2]*University of Queensland Centre for Clinical Research, The University of Queensland, Herston, Queensland, Australia;* [3]*School of Agriculture and Food Sciences, The University of Queensland, St Lucia, Queensland, Australia;* [4]*The Kinghorn Cancer Centre, Garvan Institute of Medical Research, Sydney, New South Wales, Australia;* [5]*Australian Institute of Marine Science, Townsville, Queensland, Australia;* [6]*School of Plant Biology, University of Western Australia, Perth, Western Australia, Australia;* [7]*Australian Centre for Plant Functional Genomics, The University of Queensland, St Lucia, Queensland, Australia;* [8]*ARC Centre of Excellence for Coral Reef Studies, The University of Queensland, St Lucia, Queensland, Australia and* [9]*Global Change Institute and ARC Centre of Excellence for Coral Reef Studies, The University of Queensland, St Lucia, Queensland, Australia*

Dinoflagellates from the genus *Symbiodinium* form a mutualistic symbiotic relationship with reef-building corals. Here we applied massively parallel Illumina sequencing to assess genetic similarity and diversity among four phylogenetically diverse dinoflagellate clades (A, B, C and D) that are commonly associated with corals. We obtained more than 30 000 predicted genes for each *Symbiodinium* clade, with a majority of the aligned transcripts corresponding to sequence data sets of symbiotic dinoflagellates and <2% of sequences having bacterial or other foreign origin. We report 1053 genes, orthologous among four *Symbiodinium* clades, that share a high level of sequence identity to known proteins from the SwissProt (SP) database. Approximately 80% of the transcripts aligning to the 1053 SP genes were unique to *Symbiodinium* species and did not align to other dinoflagellates and unrelated eukaryotic transcriptomes/genomes. Six pathways were common to all four *Symbiodinium* clades including the phosphatidylinositol signaling system and inositol phosphate metabolism pathways. The list of *Symbiodinium* transcripts common to all four clades included conserved genes such as heat shock proteins (*Hsp*70 and *Hsp*90), *calmodulin*, *actin* and *tubulin*, several ribosomal, photosynthetic and cytochrome genes and chloroplast-based heme-containing *cytochrome P450*, involved in the biosynthesis of xanthophylls. Antioxidant genes, which are important in stress responses, were also preserved, as were a number of calcium-dependent and calcium/calmodulin-dependent protein kinases that may play a role in the establishment of symbiosis. Our findings disclose new knowledge about the genetic uniqueness of symbiotic dinoflagellates and provide a list of homologous genes important for the foundation of coral–algal symbiosis.
The ISME Journal (2015) **9,** 844–856; doi:10.1038/ismej.2014.182; published online 24 October 2014

Introduction

Stony corals (Scleractinia) can form mutualistic symbioses with photosynthetic dinoflagellates of the genus *Symbiodinium* that are based on nutritional exchange and allow coral reef growth in oligotrophic marine environments (Muscatine *et al.*, 1975; Trench, 1979). These reef-building corals provide the foundation for the coral reef ecosystem and a habitat for millions of marine species. Currently, because of global climate change and multiple stress factors, coral reefs around the world are in decline (Hughes *et al.*, 2003; Hoegh-Guldberg *et al.*, 2007). Hyperthermal stress conditions are proposed as a major global factor destabilizing the coral–algal symbiosis, leading to the loss of endosymbionts (or their photopigments) that is manifested on the colony level as a loss of tissue coloration, also known as 'coral bleaching' (Douglas, 2003; Weis, 2008). As a result of bleaching, coral health may be compromised, increasing the incidence of coral mortality (Baker, 2003; McClanahan, 2004; Hoegh-Guldberg *et al.*, 2007). The susceptibility of the coral–algal symbiosis to stress and bleaching is influenced by the entire holobiont, comprising coral host and associated microbes (Rowan, 1998; Rohwer *et al.*, 2002). In particular, the loss of dinoflagellate photosymbionts (Rowan, 2004) may also affect coral

Correspondence: N Rosic, School of Biological Sciences, The University of Queensland, Research Road, St Lucia, Brisbane, Queensland 4072, Australia.
E-mail: n.rosic@uq.edu.au
Received 8 January 2014; revised 5 August 2014; accepted 25 August 2014; published online 24 October 2014

fitness and its ability to respond to other stressors (Howells *et al*., 2012).

Dinoflagellates are important microbial eukaryotes that, together with diatoms, are the leading primary producers in the oceans (Lin, 2011), although recent studies suggest that dinoflagellates of the genus *Symbiodinium* are also capable of heterotrophy (Jeong *et al*., 2012). Dinoflagellates are characterized by a very large genome (Hackett *et al*., 2004) and a number of unique features such as DNA containing 5-hydoxymethylmuracil (Rae, 1976), a lack of the usual histones (Rizzo, 1981) and transcriptional regulatory elements (Li and Hastings, 1998). Dinoflagellates contain a conserved spliced leader sequence (Zhang *et al*., 2007), and highly expressed genes with elevated copy numbers and tandem repeats (Bachvaroff and Place, 2008). A number of dinoflagellate genes have been acquired from bacteria and other eukaryotes by horizontal gene transfer or endosymbiosis, resulting in important gene innovations (Wisecaver and Hackett, 2011; Wisecaver *et al*., 2013).

A symbiotic lineage of single-cell dinoflagellate protists is divided into clades (A–I) and numerous phylogenetically distinct types (Santos *et al*., 2002; Pochon *et al*., 2004). Out of the nine clades, *Symbiodinium* clades A, B, C and D are commonly associated with corals and other metazoans (Pochon and Gates, 2010). Although symbiont abundance shows some seasonal variability (Chen *et al*., 2005), *Symbiodinium* clade C are the most common endosymbionts found in reef-building corals from the Pacific and Indian Oceans (Lesser *et al*., 2013) and are represented by many distinct species (Thornhill *et al*., 2013). *Symbiodinum* clades A and B are more abundant within Caribbean corals (LaJeunesse, 2002; Baker, 2003), whereas clade D are dominant in corals living in the warm waters of the Persian Gulf (Ghavam Mostafavi *et al*., 2007). On the Great Barrier Reef (GBR), surveys of *Symbiodinium* genetic diversity can be found on the web-based SymbioGBR database (Tonk *et al*., 2013).

Recent sequencing projects have revealed some of the unique characteristics of *Symbiodinium* related to their transcriptional regulation (Bayer *et al*., 2012) and functional differences between clades (Ladner *et al*., 2012). The latest sequencing project for *Symbiodinium minutum* (Shoguchi *et al*., 2013) estimated a genome size of 1.5 Gbp with ~42 000 predicted protein-encoding genes. In this study, we focused on protein-encoding genes from four different coral-associated *Symbiodinium* clades (A, B, C and D) using transcriptomic data generated by massively parallel Illumina sequencing (San Diego, CA, USA). Our aim was to explore the presence of unique, evolutionarily conserved genes in the symbiotic dinoflagellates that would determine their capacity for symbiosis with corals and other marine species. Here, we describe *de novo* transcriptome assemblies for the four *Symbiodinium* clades and provide gene annotation, gene ontology and pathway analyses for the predicted *Symbiodinium* genes orthologous to all four clades. Finally, we explore the possible role of conserved calcium/calmodulin-dependent protein kinases (CCaMKs) and the inositol pathway for the foundation of cnidarian–dinoflagellate symbiosis.

Materials and methods

Cultures, RNA extraction and sequencing

Cultures of *Symbiodinium* spp. used in this study were polyclonal cultures of clades A (internal transcribed spacer (ITS) type A2) isolated from the coral *Zoanthus sociatus* (Caribbean region); B (ITS type B2) from *Oculina diffusa* (Bermuda); C (ITS type C1) from the anemone *Discosoma sanctithomae* (Jamaica), all obtained from Professor Roberto Iglesias-Prieto (RSU, UNAM, Puerto Morelos, Mexico); and culture of clade D (ITS type D1) isolated from *Porites annae* (W. Pacific) by Professor Michio Hidaka (University of the Ryukyus, Japan) and was donated by Ass/Professor Scott Santos (Auburn University, Auburn, AL, USA). Cultures were grown in axenic f/2 medium (Guillard and Ryther, 1962) and maintained at a temperature of 25 °C, under a constant 12:12-h day/night period, with an irradiance of 50 μmol quanta $m^{-2} s^{-1}$ (measured using a Li-Cor flat quantum sensor, Lincoln, NE, USA). The algal cultures from different cell growth phases were combined to maximize the diversity of gene expression profiles. The algal cells were centrifuged and the resulting pellet was snap-frozen in liquid nitrogen and stored at −80 °C before RNA extraction. Total RNA was extracted from algal cells as previously described (Rosic and Hoegh-Guldberg, 2010). Briefly, this method combines the usage of Trizol reagent (Ambion Life Technologies, Austin, TX, USA) followed by the RNeasy kit (Qiagen, Hilden, Germany). The RNA quantity and integrity were analyzed using a NanoDrop ND-1000 spectrometer (Wilmington, DE, USA) and an Agilent 2100 Bioanalyzer (Santa Clara, CA, USA), RNA integrity number >6. Equal concentrations of high-quality RNA from algal cultures were used to prepare libraries (using the Illumina TruSeq RNA Sample Preparation Kit, San Diego, CA, USA) for sequencing with the Illumina GA II Sequencing System at the Australian Genome Research Facility Ltd. To avoid bacterial contamination, library construction included the purification of poly-A-containing mRNA molecules using poly-T oligo-attached magnetic beads.

Short read data assessment and quality control

Phred-like quality scores are calculated by the Illumina sequencer for sequencing data, and nucleotides (nt) with quality scores <20 were trimmed from reads together with Illumina sequencing primers and multiplex adaptors. After trimming,

reads <55 nt in length were removed. The remaining 'clean' short reads were assembled into contiguous sequences (contigs) before further analyses.

De novo short read assembly

The software applications Velvet version 1.1.04 (Zerbino and Birney, 2008) and Oases version 0.1 (Schulz *et al.*, 2012) were used for *de novo* transcriptome assembly, with a *k*-mer size of 53. Oases was used to cluster the Velvet-assembled contigs to construct transcript isoforms. Contigs with less than three times *k*-mer coverage were removed from the final assemblies.

To reduce the redundancy of the assembled contigs, each transcriptome was first aligned to itself using the Nucleotide-Nucleotide Basic Local Alignment Search Tool (BLAST) software application (BLASTn, version 2.2.27+; Altschul *et al.*, 1990). Any pair of contigs that was >99% identical over 95% of the length of the shorter contig was collapsed into a single contig by removal of the shorter contig. Additional testing to reduce redundancy was carried out and pairs of contigs that were >80% and >90% identical over 95% of the length of the shorter contig were collapsed into the longer contig. The number of redundant transcripts at the >80% and >90% identity levels was not significantly reduced, and reduction of redundant transcripts was kept at the >99% identity level (Supplementary Table S1).

Assembly assessment and annotation

To assess the quality of the assembled transcriptomes, the contigs were aligned, with BLASTn and an *E*-value of $\leqslant 10^{-5}$, to *Symbiodinium* expressed sequence tags (ESTs) from Genbank and the Joint Genome Institute databases as well as *Symbiodinium* hemoglobin and heat shock protein genes (Rosic *et al.*, 2011a, 2013). Subsequently, the reference database was extended to include non-*Symbiodinium* sequences from Genbank (bacterial, environmental, invertebrate, plant and viral nucleotide sequences; *Acropora digitifera* genome) and other databases (*A. millepora* transcriptome, Moya *et al.*, 2012; *A. hyacinthus*, *A. tenuis* and *Porites astreoides* transcriptomes, Eli Meyer and Mikhail Matz, www.bio.utexas.edu/research/matz_lab) to identify non-*Symbiodinium* homologs of the contigs that did not align to known *Symbiodinium* sequences. Furthermore, contigs >300 nt in length were aligned, using the Translated Nucleotide-Protein BLAST application (BLASTx, version 2.2.27+) and an *E*-value of $\leqslant 10^{-5}$, to the SwissProt (SP) protein (version 2011_08), TrEMBL (version 2012_07) and National Center for Biotechnology Information (NCBI) non-redundant protein sequence (nr; 13 August 2011 update) databases for the purpose of annotation. Additional BLAST comparison of individually targeted sequences was carried out against an EST database (http://sequoia.ucmerced.edu/SymBioSys/) and the coral proteome database, ZoophyteBase (Dunlap *et al.*, 2013). These databases include sequences of various cnidarian species and *Symbiodinium* strains. The online software application, VENNY (http://bioinfogp.cnb.csic.es/tools/venny/), was used to identify database sequences that were aligned to contigs from two or more clades, as well as redundancy in the database alignments of contigs within each clade.

Identification of orthologs in all four Symbiodinium transcriptomes

To identify orthologous transcripts within the *de novo*-assembled transcriptomes, we applied the Reciprocal Best BLAST Hits (RBBH) approach (Telford, 2007) and adopted a four-way reciprocal BLASTn (*E*-value $\leqslant 10^{-15}$) strategy, with all pairwise comparisons (Ness *et al.*, 2011). Criteria for identifying orthologous sequences were: minimum length of 200 nt; minimum sequence identity of 90%; and minimum alignment proportion of 80% for the shorter contig (Ness *et al.*, 2011).

Functional profile and gene ontology (GO) enrichment analyses

To determine the function of the orthologous genes, we performed BLASTx against the SP protein database using an *E*-value of $\leqslant 10^{-15}$. Using this conservative approach, we only obtained sequence alignments with well-characterized orthologous proteins existing in the SP database. The SP genes that aligned best with the orthologous contigs were considered orthologous genes and used for downstream enrichment analyses.

GO enrichment analyses and pathway analyses, as well as the identification of enriched biological themes and KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways, were performed using the database for annotation, visualization and integrated discovery (DAVID) using the *UNIPROT ACCESSION* terms from the SP genes as identifiers (Huang da *et al.*, 2009a, b). All SP-annotated genes from all four clades contributed to the background gene set for the enrichment analyses. DAVID uses the Fisher's exact test to ascertain statistically significant gene enrichment for a particular pathway, and significant processes were selected based on a corrected *P*-value <0.05. We applied the corrected *P*-value (Benjamini correction) with a cutoff of 0.05 for filtering the significantly enriched pathways and a *P*-value of 0.001 for filtering enriched GO categories (Huang da *et al.*, 2009a; Meyer *et al.*, 2009).

Taxonomic analyses

To evaluate an evolutionary origin and possible horizontal gene transfer within our transcriptome sequences, predicted transcripts from each of the analyzed samples were aligned to publicly available

data sets, including ESTs, genome and transcriptome sequences from Genbank databases (accessed March 2012; that included bacterial, environmental, invertebrate, plant and viral nucleotide sequences; *A. digitifera* genome; human genome; and *Symbiodinium* ESTs); *Symbiodinium* ESTs from the Joint Genome Institute (University of California, Merced, CA, USA); the *A. millepora* transcriptome (Meyer *et al.*, 2009; Moya *et al.*, 2012); and the *A. hyacinthus*, *A. tenuis* and *P. astreoides* transcriptomes (Eli Meyer, Mikhail Matz, *et al.* data: www.bio.utexas.edu/research/matz_lab). Database alignments were carried out using BLASTn, specifying a word size of 11 and *E*-value $\leqslant 10^{-5}$.

In addition, to compare our conserved *Symbiodinium* transcripts with other eukaryotic genomes, we performed BLASTn (*E*-value $\leqslant 10^{-5}$) to genome and transcriptome sequences of three dinoflagellates *Alexandrium minutum*, *Cyanidioschyzon merolae and Symbiodinium minutum*, as well as the coral *A. digitifera* and human genomes.

Results

De novo short read assembly

Massively parallel Illumina sequencing generated over 19 million raw reads (99 nt single reads) and a total of 1.9–2.9 Gb raw data for each of the four *Symbiodinium* transcriptomes (Table 1). After trimming to remove low-quality bases, primer and adaptor sequences, the sequence reads were used to produce Velvet/Oases assemblies resulting in over 90 000 transcripts for each clade, with an average contig length between 300 and 396 bp, N50 between 310 and 479 bp and longest contigs from 3955 to 7242 bp (Table 2). In order to estimate the number of genes expressed, we clustered highly similar contigs that had at least 99%, 90% and 80% of sequence identity over 95% of the length of the shorter contig. Our results showed that when reducing redundancy, the decrease in the numbers of predicted transcripts was <1% for each algal clade (Supplementary Table S1). Therefore, we clustered similar contigs based on $\geqslant$99% sequence identity. Frequencies of the transcript length distribution for each *Symbiodinium* clade are shown in Supplementary Figure S1. After removal of short contigs (<300 nt), we obtained between 29 846 and 46 892 nonredundant transcripts for each clade. Following BLASTx comparison of the predicted transcripts with the SP, TrEMBL and the NCBI non-redundant protein sequence databases, we successfully annotated ~40–44% of the transcripts (Table 2). Similarity between our *de novo* transcriptome assemblies and publicly available *Symbiodinium* EST sequences as well as non-*Symbiodinium* databases, based on BLASTn with cutoff *E*-values of 10^{-5}, is presented in Supplementary Table S2.

Assembly assessment and annotation

In order to validate the accuracy of the *de novo* assemblies, we compared the *Symbiodinium* (C3)

Table 1 Sequencing statistics: the number of reads and total amount of sequence generated for each clade

	Clade A	*Clade B*	*Clade C*	*Clade D*
Number of raw reads	22 312 981	25 599 586	28 944 885	19 170 898
Number of raw bases (Gb)	2.21	2.53	2.87	1.90
Number of reads after quality control (QC)	17 562 275	20 743 552	23 315 723	16 489 824
Number of bases after QC (Gb)	1.54	1.84	2.06	1.51

Values include both raw sequence data and sequence data after trimming and filtering reads with low-quality bases.

Table 2 Overview of the sequencing data, assembly, clustering and annotation statistics: clustering was based on >99% pairwise identity between transcripts predicted by Velvet/Oases

	Clade A	*Clade B*	*Clade C*	*Clade D*
No. of transcripts (contigs)	92 976	145 838	106 079	98 072
Average contig length (bp)	316	300	396	389
Longest contig (bp)	4042	3955	5915	7242
N50	338	310	479	468
No. of nonredundant transcripts $\geqslant$300 bp	29 846	42 233	46 892	42 885
No. of nonredundant transcripts $\geqslant$300 bp aligning to SP, TrEMBL and NCBI nr databases, *E*-value $\leqslant 10^{-5}$	13 134	18 436	19 047	17 093
Percentage of transcripts $\geqslant$300 bp, with alignments to SP, TrEMBL and NCBI nr databases	44.00%	43.65%	40.62%	39.86%
No. of nonredundant transcripts $\geqslant$300 nt aligning to SP, *E*-value $\leqslant 10^{-15}$	3489	4307	4811	4559

Abbreviations: NCBI, National Center for Biotechnology Information; nr, non-redundant protein sequence; SP, SwissProt.
Nonredundant transcripts $\geqslant$300 nt were used in gene annotation. Percentage of transcripts with alignments to the SP, TrEMBL and NCBI databases, using BLASTx with *E*-value $\leqslant 10^{-5}$ are presented. Additional BLASTx to the SP database, with *E*-value $\leqslant 10^{-15}$, was carried out to annotate transcripts using well-characterized genes.

Sanger sequences (Leggat *et al.*, 2007) for conserved *HSP70* and *HSP90* genes (Rosic *et al.*, 2011a) and novel polymorphic hemoglobin-like proteins (Rosic *et al.*, 2013) with the assemblies. The results of BLASTn and BLASTx alignments, with stringent *E*-values and high bit scores, confirmed the presence of these genes in the *de novo* assemblies of all four transcriptomes presented here (Supplementary Table S3). For conserved genes such as *HSP70* and *HSP90*, we observed sequence identity of 95–99% between the EST Sanger sequences for *Symbiodinium* clade C3 and our *de novo*-assembled genes from the various *Symbiodinium* clades and ITS-2 types. The hemoglobin genes showed conservation at the protein level, but a greater level of sequence variability at the nucleotide level in the range of 65–94% for Hb1 and 69–96% for Hb2, across the clades.

Functional profile and GO enrichment analyses

GO enrichment analysis highlighted enriched pathways within the KEGG database, including well-described proteins and ubiquitous biochemical pathways. We identified six pathways common in all four *Symbiodinium* clades: phosphatidylinositol signaling system, inositol phosphate metabolism, spliceosome, ribosome, endocytosis and sucrose metabolic pathways (Table 3).

Enrichment analyses of our annotated transcriptomes revealed GO categories that were significantly enriched among the four analyzed dinoflagellates. Examples of enriched GO categories common to all analyzed dinoflagellates (using the Benjamini correction *P*-value <0.05 as cutoff) included 10 biological process categories related to photosynthesis: transcription, nitrate metabolism, microtubule-based processes and phosphatidylinositol metabolic processes (Table 4 and Figure 1). In the case of the molecular function category, there was no significant enrichment found that was common to all four dinoflagellates. For the cellular component category, all four *Symbiodinium* clades showed enrichment in genes related to 29 categories that included cellular components such as chloroplast, microtubules, dynein complex, cytosol and other plastid parts (Table 5).

The number of alignments to the SP database is presented in Table 2. The records of nonredundant *Symbiodinium* transcripts from different clades that aligned to the SP database, including transcripts from two or more clades that aligned to the same SP genes, are presented in a four-way Venn diagram (Figure 2). Orthology among all *Symbiodinium* clades was inferred for the 1053 genes at the core of the four-way Venn diagram. These genes shared the same *UNIPROT ACCESSION* IDs and showed a high degree of sequence similarity among the four *Symbiodinium* clades, and included a number of multiple-copy genes (Supplementary Table S4). This list of shared *Symbiodinium* genes included conserved genes such as heat shock proteins (*Hsp*70 and *Hsp*90); housekeeping genes (HKGs) including *actin*, *calmodulin*, *tubulin*, *GAPDH* and *cyclophilin*; as well as several ribosomal, cytochrome genes with chloroplast-based heme-containing *cytochrome P450* and photosynthetic genes including *ribulose bisphosphate carboxylase* and *peridinin-chlorophyll a-binding protein* (Table 6). Common antioxidant genes important in stress responses were also identified including *thioredoxin*, *ferredoxin* and *superoxide dismutase*. Furthermore, we have identified calcium-dependent protein kinases and CCaMKs that are important in intercellular signaling (Table 6). Several CCaMK isoforms within the conserved catalytic region have been identified for *Symbiodinium* clades, and are presented in a multiple sequence alignment (ClustalX) including the representative CCaMKs from other species (Figure 3). The phylogenetic analyses of the conserved catalytic domain revealed two monophyletic groups of *Symbiodinium* CCaMKs, indicating their different evolutionary origin (Figure 3c).

Table 3 Significant pathways enriched in all four *Symbiodinium* clades with corrected *P*-value $\leqslant 0.05$ after multiple testing correction (MTC) by the Benjamini procedure

Significant pathway	*No. of genes*	*Fold enrichment*
Phosphatidylinositol signaling system	7	3.9
Inositol phosphate metabolism	7	3.4
Spliceosome	10	2.3
Ribosome	7	2.7
Endocytosis	7	2.7
Starch and sucrose metabolism	4	3.9

Table 4 Biological processes (BP) enriched in all four *Symbiodinium* clades with corrected *P*-value $\leqslant 0.05$

Annotation term	*Gene ontology (GO) ID*	*No. of genes*	*Fold enrichment*
RNA splicing	GO: 0008380	47	2.08
mRNA metabolic process	GO: 0016071	61	1.79
Microtubule-based movement	GO: 0007018	36	2.16
Microtubule-based process	GO: 0007017	50	1.86
mRNA processing	GO: 0006397	52	1.81
Nitrate metabolic process	GO: 0042126	8	6.24
Nitrate assimilation	GO: 0042128	8	6.24
Phosphatidylinositol metabolic process	GO: 0046488	8	6.24
Photosynthesis	GO: 0015979	29	2.19
Photosynthesis, light harvesting	GO: 0009765	7	6.55

Taxonomic analyses

Our taxonomic analyses revealed that up to 72% of *Symbiodinium* transcripts aligned with sequences from publicly available sequence databases, whereas 25–35% had no hits. The majority of the aligned transcripts matched the symbiotic dinoflagellate

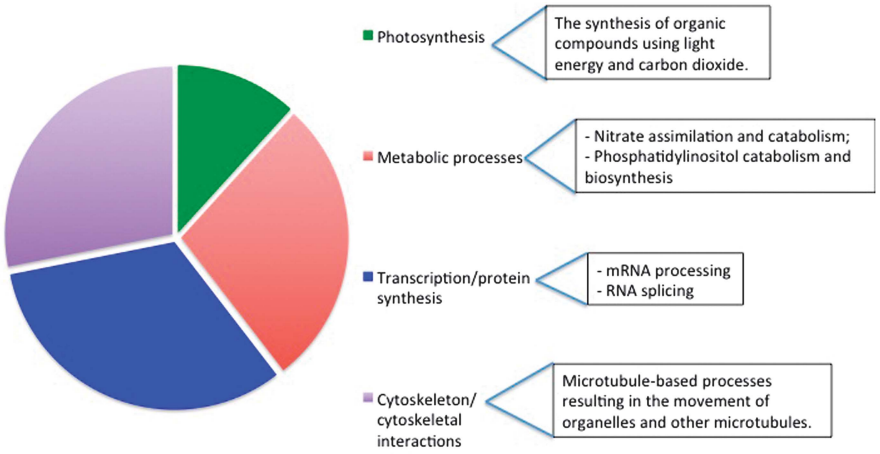

Figure 1 Biological processes (BP) enriched in all four *Symbiodinium* clades, using DAVID enrichment analyses of nonredundant transcripts ⩾300 nt with BLASTx hits to SP, *E*-value $<10^{-15}$.

Table 5 Cellular components (CC) enriched in all four *Symbiodinium* clades with corrected *P*-value ⩽0.05

Annotation term	*Gene ontology (GO) ID*	*No. of genes*	*Fold enrichment*
Chloroplast	GO: 0009507	109	1.99
Plastid	GO: 0009536	114	1.93
Axoneme	GO: 0005930	22	3.45
Dynein complex	GO: 0030286	21	3.29
Cilium	GO: 0005929	31	2.41
Chloroplast part	GO: 0044434	33	2.26
Thylakoid	GO: 0009579	36	2.13
Organelle	GO: 0043226	538	1.11
Intracellular organelle	GO: 0043229	538	1.11
Plastid part	GO: 0044435	33	2.11
Cilium axoneme	GO: 0035085	15	3.41
Microtubule	GO: 0005874	47	1.83
Photosynthetic membrane	GO: 0034357	23	2.46
Microtubule-associated complex	GO: 0005875	27	2.25
Flagellum	GO: 0019861	13	3.28
Microtubule cytoskeleton	GO: 0015630	61	1.57
Cell projection part	GO: 0044463	30	1.98
Intracellular membrane-bounded organelle	GO: 0043231	465	1.10
Chloroplast envelope	GO: 0009941	16	2.65
Membrane-bounded organelle	GO: 0043227	466	1.10
Ribonucleoprotein complex	GO: 0030529	87	1.40
Cytosolic part	GO: 0044445	34	1.80
Basolateral plasma membrane	GO: 0016323	8	4.28
Photosystem	GO: 0009521	13	2.82
Spliceosome	GO: 0005681	28	1.87
Cytosolic ribosome	GO: 0022626	28	1.86
Plastid envelope	GO: 0009526	16	2.39
Cilium part	GO: 0044441	16	2.35
Macromolecular complex	GO: 0032991	248	1.16

data sets (96.6–98%), whereas a relatively small proportion of the aligned transcripts matched bacterial (0.7–1.3%) and coral (0.2–0.7%) sequences (Supplementary Table S2).

Transcripts corresponding to the 1053 shared SP genes from each *Symbiodinium* clade were

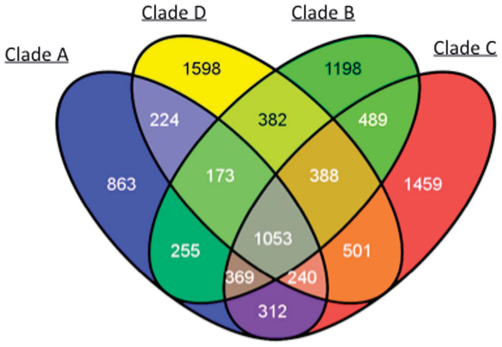

Figure 2 The Venn diagram of proposed *Symbiodinium* transcripts from each clade (⩾300 nt in size) aligned to the SP database, showing the number of genes unique to each clade as well as those shared among clades. Sequence homology was inferred when the expectation value (*E*-value) was ⩽10^{-15}.

redundant and included 2597 (clade A), 3626 (B), 2864 (C) and 2513 (D) transcripts. They were compared with five other eukaryotic genomes/transcriptomes (BLASTn *E*-value ⩽10^{-5}) resulting in 12–17% of these transcripts from each clade aligning to non-*Symbiodinium* sequences, including *Homo* sapiens and *A. digitifera* (Supplementary Table S5). More than 80% of these transcripts were aligned to *Symbiodinium* species only (our *Symbiodinium* clades and *S. minutum*) and were not found in other eukaryotes. Due to redundancy and multiple transcripts annotated with the same *UNIPROT ACCESSION* IDs, only 41% or 432 of these genes were actually unique to *Symbiodinium* when compared with other analyzed eukaryotes (Supplementary Table S6). Direct alignment of the 1053 gene sequences obtained from the SP database to the other eukaryote sequences have shown 95.6% presence (1007 out of 1053) of these SP genes in other eukaryotes and 4.4% (46 out of 1053) specific to *Symbiodinium* species (Supplementary Table S7).

Table 6 The UniProt IDs and descriptions of some of the common genes that have been found in all four *Symbiodinium* clades

UniProt ID	*Hit description*
HSP genes	
Q6TMK3	Heat shock protein 88 (heat shock 70-related protein)
P11144; P41753	Heat shock 70 kDa protein
O44001	Heat shock protein 90
O88600	Heat shock 70 kDa protein 4 (ischemia responsive 94 kDa protein)
Housekeeping genes	
A3E4D8; A8I1Q0; P41041	Calmodulin
Q6GLA2	S-adenosylmethionine mitochondrial carrier protein
P22513	Glyceraldehyde-3-phosphate dehydrogenase, glycosomal (GAPDH)
Q39613	Peptidyl-prolyl cis-trans isomerase (PPIase) (Cyclophilin)
P34791	Cyclosporin A-binding protein (Rotamase cyclophilin-4)
Q9FJX0	Cyclophilin-65
P26182; P53476	Actin
Photosynthetic genes	
Q9AW48	Photosystem II stability/assembly factor HCF136, chloroplastic
Q9XQV3	Photosystem I P700 chlorophyll a apoprotein A1 (PsaA)
P49481	Photosystem I reaction center subunit II (Photosystem I 16 kDa polypeptide)
P85194	Oxygen-evolving enhancer protein 1, chloroplastic (OEE1)
A0T0T0	Photosystem II D2 protein
Q7XY82	Photosystem II 12 kDa extrinsic protein, chloroplastic
P80483	Peridinin-chlorophyll a-binding protein 3
P55738	Caroteno-chlorophyll a-c-binding protein
Q5ENN5	Ribulose bisphosphate carboxylase, chloroplastic
Q40297	Fucoxanthin-chlorophyll a-c binding protein A, chloroplastic
Antioxidant genes	
Q59094	Superoxide dismutase [Mn]
P99029	Peroxiredoxin-5, mitochondrial
P83877	Thioredoxin-like protein 4A
Q00598	Ferredoxin—NADP reductase, cyanelle (FNR)
P10770	Ferredoxin
P41344	Ferredoxin—NADP reductase, leaf isozyme, chloroplastic
Q3IQZ9	Catalase-peroxidase (CP)
A4R606	Putative heme-binding peroxidase
Q46845	Disulfide-bond oxidoreductase YghU
Cytochrome genes	
P09437	Cytochrome *b*2, mitochondrial
Q4PBY6	Cytochrome *c* peroxidase, mitochondrial (CCP)
O23365	Cytochrome *P*450 97B3, chloroplastic
Q7XYM4	Cytochrome *b*6-f complex iron-sulfur subunit, chloroplastic
A0T0C6	Cytochrome *c*-550 (Cytochrome c550)
P29610	Cytochrome *c*1-2, heme protein, mitochondrial (Clone PC18I)
Q02766	Cytochrome *c* oxidase subunit 1
Q8M9X2	Apocytochrome *f*
P00110	Cytochrome *c*6 (Cytochrome c-553)
P31564	Cytochrome *c* biogenesis protein ccsA
Q5PQA4	NADH-cytochrome b5 reductase 2
Q4N594	Cytochrome *c*
O23066	Cytochrome *P*450 86A2
Calcium-dependent protein kinase genes	
O15865	Calcium-dependent protein kinase 2
Q63450	Calcium/calmodulin-dependent protein kinase type 1
Q7RJG2	Calcium-dependent protein kinase 4
Q7RAH3	Calcium-dependent protein kinase 1
Q9FMP5	Calcium-dependent protein kinase 17
P62343	Calcium-dependent protein kinase 1
Q9NJU9	Calcium-dependent protein kinase 3
Q7RAV5	Calcium-dependent protein kinase 3
Q8IBS5	Calcium-dependent protein kinase 4

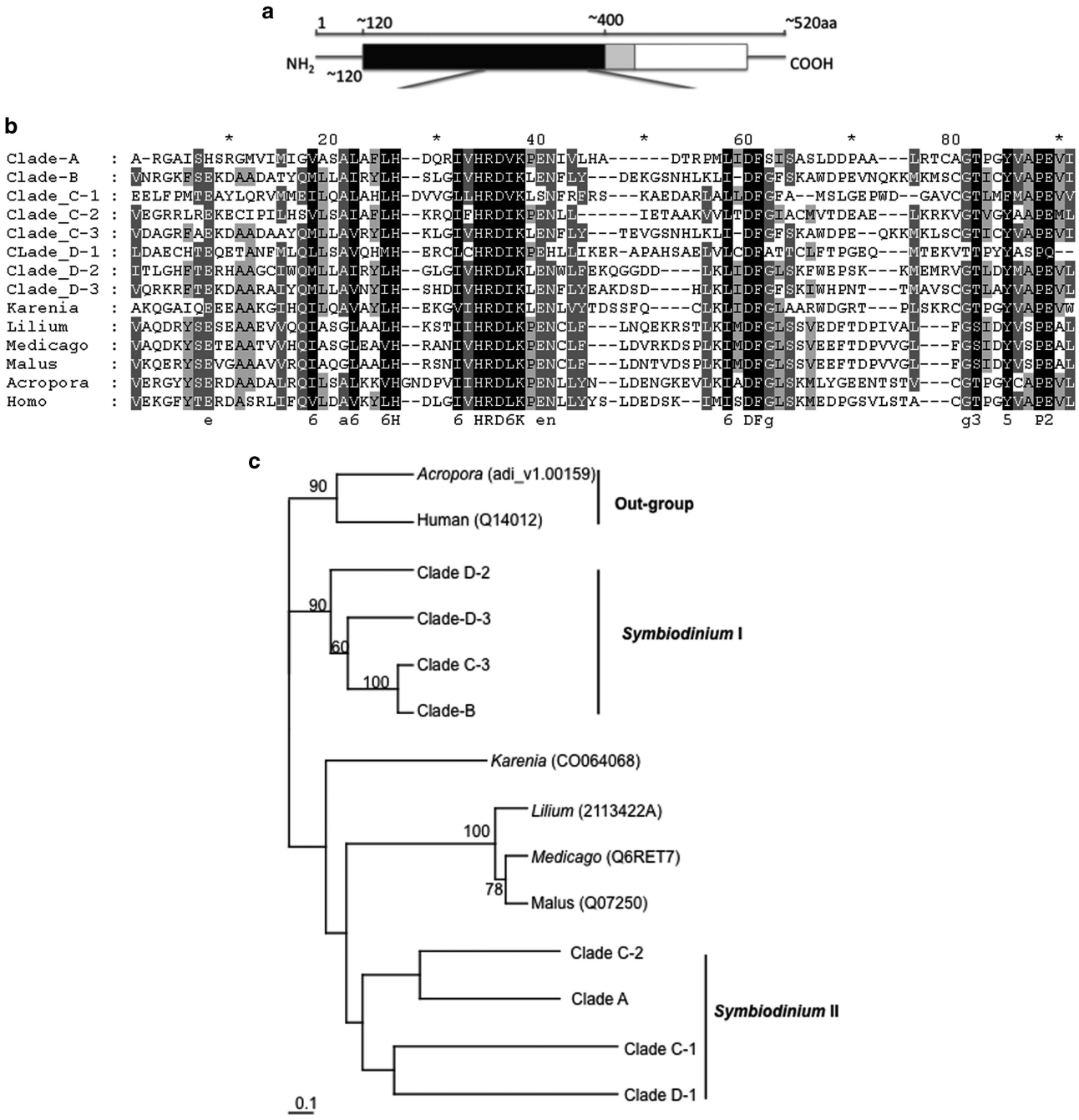

Figure 3 Diagram of the CCaMK structure (Swulius and Waxham, 2008), showing the conserved catalytic domain (black box) required for kinase activity; autoinhibitory domain (gray box) and the C-terminal regulatory domain (white box) holding Ca^{2+}-binding EF-hands (**a**). The multiple sequence alignment (MSA) of the conserved CCaMK region from four *Symbiodinium* clades including representatives of different isoforms from other species: *Karenia brevis* (CO064068); *Lilium longiflorum* (2113422A); *Medicago truncatula* (Q6RET7); CCaMK of *Malus domestica* (Q07250); *Acropora digitata* (*adi_v1.00159*); and *Homo sapiens* (Q14012) (**b**). The identical residues in all sequences are indicated by white letters with a black background (amino acids conserved in 100% of the sequences), white letters with a gray background (80% conserved) and black letters with a gray background (60% conserved). The MSA was constructed using clustalX (ftp://ftp.ebi.ac.uk/pub/software/clustalw2). Phylogenetic analyses of deduced amino acid sequences of *Symbiodinium* and other representative species were done using sequences from MSA (**c**). The phylogenetic tree was tested using a 1000-replicated bootstrap analysis (Felsenstein, 1989) and bootstrap values >50% are indicated at each node. A distance method using maximum likelihood estimates was based on the Dayhoff PAM matrix. The scale for the branch length (0.1 substitutions per site) is presented under the tree.

Discussion

The number of predicted genes for the four *Symbiodinium* clades has been estimated here to be between 30 000 and 46 000, and this is in the range of results obtained by others (Bayer *et al*., 2012; Shoguchi *et al*., 2013). Validation of the *de novo* assemblies was obtained using conserved *Symbiodinium HSP70* and *HSP90* genes (Rosic *et al*., 2011a) and polymorphic hemoglobin genes (Rosic *et al*., 2013). Heat shock proteins are molecular chaperones that are evolutionarily conserved and important for regular cellular functions involving protein folding and unfolding, degradation and transport as well as stress responses (Sørensen *et al*., 2003). In contrast, globin genes are characterized by a high level of polymorphism, although on a protein level they maintain globin-fold and a conserved Histidine residue (Royer *et al*., 2005). From our *de novo* assemblies, we recovered polymorphic hemoglobin-like genes with preserved globin domains (Supplementary Table S3).

Biological processes enriched and common to all four symbiotic dinoflagellates included photosynthesis, metabolism, transcription/translation and cytoskeletal interactions (Figure 1), as well as the phosphatidylinositol signaling system, inositol phosphate metabolism, spliceosome, ribosome, endocytosis and sucrose metabolism pathways (Table 3). In photosynthetic dinoflagellate *Alexandrium* the spliceosome pathway was also enriched, highlighting the importance of RNA-splicing mechanism in the dinoflagellates group (Zhang *et al*., 2014). These results indicate some of the fundamental processes needed for successful *Symbiodinim*–coral interaction. Pathways related to nutrient transfer, carbon concentration, nitrogen recycling, calcification, oxidative stress and cell communication have been proposed as necessary physiological adaptations required for a symbiotic lifestyle within *Symbiodinium*–holobiont systems (Weber and Medina, 2012). *Symbiodinium*, however, builds symbiotic associations with many marine species, including metazoan hosts (cnidaria, porifiera, mollusca and acoelomorpha), as well as single-celled eukaryote hosts (foraminifera, radiolarians and ciliates). Consequently, the implications and importance of these conserved biological pathways and processes for the foundation of this photosynthetic symbiosis are yet to be fully elucidated in the symbiotic and nonsymbiotic scenarios.

Finding gene orthologs among HKGs such as *actin*, *calmodulin*, *tubulin*, *GAPDH* and *cyclophilin* (Table 6) being expressed in all four clades was in some way an expected outcome as these proteins play an important role in essential cellular processes and their stable gene expression profiles are usually maintained irrespective of stress exposure (Sturzenbaum and Kille, 2001; Huggett *et al*., 2005). The *actin* gene is highly expressed and conserved in dinoflagellates (Kim *et al*., 2011). This gene and other HKGs have already been successfully applied as a reference, together with other HKGs, in several recent gene expression studies of different *Symbiodinium* clades, because of their stable expression during stress exposure (Rosic *et al*., 2011b; McGinley *et al*., 2012; Sorek and Levy, 2012; Ogawa *et al*., 2013). Furthermore, photosynthesis-related transcripts were recognized as shared among the clades and included genes such as *Ribulose-1,5-bisphosphate carboxylase oxygenase* (*Rubisco*) and *peridinin-chlorophyll a-binding protein* that are unique to dinoflagellates (Rowan *et al*., 1996; Leggat *et al*., 2011). An RBBH approach was used to find orthologs between EST libraries of two *Symbiodinium* spp. (CassKB8 and C3), revealing 132 potential orthologs that, similarly, included HKGs such as *actin* and *cyclophilin* and photosynthesis-related genes (Voolstra *et al*., 2009). Here, we have identified several antioxidant genes common to the four coral dinoflagellates (Table 6), including genes from the *thioredoxin* (*Trx*) superfamily, as well as *superoxide dismutase* (*SOD Mn*) and *catalase* (*CAT*) genes. The *Trx* genes have also been found in two *Symbiodinium* clades, A and B (Bayer *et al*., 2012), whereas SOD and CAT genes are known to be involved in the oxidative stress response and scavenging reactive oxygen species (Lesser and Shick, 1989; Lesser, 2006; Levy *et al*., 2006). Although there are different forms of SOD metalloproteins (McCord and Fridovich, 1969), the conserved SOD Mn form found in mitochondria is also common in bacteria and many eukaryotic algae, and is considered, together with Fe SOD, to be an evolutionarily ancient form of SODs (Lesser, 2006). CAT is a heme-containing enzyme that catalyzes the conversion of hydrogen peroxide to water and oxygen and, similarly, peroxidases catalyze the conversion of hydrogen peroxide to water (Lesser, 2006). The Trx enzymes in the chloroplast regulate the activity of photosynthetic enzymes via ferreodoxin-thioredoxin reductase (Arnér and Holmgren, 2000). In legume roots, antioxidant proteins such as SOD, CAT and thioredoxin play an important role during nodule formation and show increased expression in nodules necessary for lowering the reactive oxygen species levels (Lee *et al*., 2005). A reduced expression of the *thioredoxin* gene via RNA interference resulted in impaired nodule formation (Lee *et al*., 2005). Similar to root nodules, where the symbiotic relationships occurs between bacteria and plant, preserved antioxidant genes of *Symbiodinium* suggest that perhaps these antioxidants are important in the establishment of coral–algal symbiosis.

The discovery of calcium-dependent protein kinases within the shared *Symbiodinium* transcripts is consistent with the fact that these enzymes have previously been found in plants, ciliates and some protists (Harper and Harmon, 2005). In the Apicomplexa, Ca^{2+} acts as a secondary messenger

via a range of calcium-dependent protein kinases, and initiates a number of signaling processes that are important for communication between eukaryotic cells (Nagamune and Sibley, 2006). Calcium signaling and calcium-regulated protein kinases have also been recognized for their critical role in establishing plant–microbe symbioses, especially in the plant–rhizobium system (Harper and Harmon, 2005; Oldroyd *et al.*, 2009). Genes such as CCaMK that play a role in the establishment of rhizobial symbiosis between the legume plants and fungi (Mitra *et al.*, 2004) were found to be conserved in symbiotic dinoflagellates (Table 6) and represented by several isoforms organized into two evolutionarily distant groups (Figure 3). They have, however, also been discovered in nonsymbiotic photosynthetic dinoflagellates like *Karenia brevis*, playing a role in intracellular signaling pathways (Lidie *et al.*, 2005), as well as in the coral host (Shinzato *et al.*, 2011; Dunlap *et al.*, 2013) where they play a role in the motility of coral sperm (Morita *et al.*, 2009). Interestingly, CCaMKs have been highly conserved in plants that establish rhizobial symbiosis with nitrogen-fixing soil bacteria, but not in nonleguminous plants such as *Arabidopsis* and wheat that do not have this type of symbiotic interaction (Mitra *et al.*, 2004). In plants, mutants of the *DMI3* gene that encodes CCaMK are not capable of establishing symbiosis with mycorrhizal fungi (Mitra *et al.*, 2004). Consequently, in photosynthetic symbioses such as that between corals and *Symbiodinium*, the role of CCaMK may occur via Ca^{2+} signaling, allowing molecular recognition and the establishment of the coral–algal symbiosis.

Genes from the enriched KEGG pathway 'Phosphatidylinositol signaling system' were identified within the orthologous genes, shared among all four *Symbiodinium* clades, encoding proteins including phosphatidylinositol 4-phosphate 5-kinases, phosphatidylinositol 4-kinase gamma 5, phosphatidate cytidylyltransferase and myo-inositol oxygenase (Supplementary Table S4). An interesting characteristic of these genes is that they encompass multiple copies of the membrane occupation and recognition nexus (MORN) repeat. Proteins containing MORN repeats could be involved in cell division, as in the parasite *Toxoplasma gondii*, where they are located in the cell division apparatus and are proposed to play a role in the cytoskeleton interaction between the parasite and host (Gubbels et al., 2006; Takeshima *et al.*, 2000). In plants, the MORN repeat motif has been discovered in the *ARC3* gene that is important for the replication of chloroplasts (Shimada *et al.*, 2004). In addition, the symbiotic interaction and signaling between legumes and rhizobacteria is mediated via phosphatidylinositide-regulated endocytosis (Peleg-Grossman *et al.*, 2007), whereas glycosylphosphatidylinositol is important for the cell recognition and detection of microbes by the host (Davy *et al.*, 2012). Our results support the importance of the conserved 'Phosphatidylinositol signaling system' pathway for symbiotic dinoflagellates and suggest its possible involvement in symbiotic interactions. Conclusive evidence could come from i*n vitro* studies that would include mutants or silencing of particular genes from this pathway, followed by an evaluation of the capacity of *Symbiodinium* to establish symbiotic interaction with the host.

Within the sequences from the *Symbiodinium* cultures we identified almost 2% of sequences of bacterial and other origins (Supplementary Table S2). This indicates a possible gene intake via horizontal gene transfer or endosymbiosis that has been shown to be an important mechanism for acquiring innovative features during dinoflagellate evolution (Wisecaver and Hackett, 2011; Wisecaver *et al.*, 2013). However, the possibility of a sequence match due to conserved genes cannot be excluded.

Of transcripts aligned to the 1053 genes conserved in all four *Symbiodinium* clades, we revealed up to 17% of the transcripts aligning to five other eukaryotic genomes/transcriptomes (Supplementary Table S5). Approximately 80% of these transcripts matched only to our transcriptomes and the genome of *Symbiodinium minutum* (Shoguchi *et al.*, 2013) and could not be found in the nonsymbiotic dinoflagellate *Alexandrium minutum* nor in the unicellular red alga *Cyanidioschyzon merolae*. From the 1053 conserved SP genes, 432 were present only in *Symbiodinium*. The lack of alignment between these *Symbiodinium* transcripts and the sequences of the analyzed eukaryotes does not exclude the presence of the associated SP genes in other organisms, as many of them actually encode proteins involved in housekeeping and maintenance of the regular cellular functions, but does indicate sequence and probable functional divergence of these genes conserved within symbiotic lineages of dinoflagellates (Supplementary Tables S6 and S7). Consequently, these results present a potential pool of symbioses-related transcripts that should be evaluated under symbiotic conditions. Furthermore, future studies are needed to explore the genetic bases of differences between different *Symbiodinium* clades.

In this research, we used a conservative approach by targeting and evaluating functions and pathways unique to and preserved in all symbiotic dinoflagellate types. This was to avoid possible bias arising from the following circumstances: (1) *Symbiodinium* maintained in culture have a tendency to represent only a portion of the population from the *in hospite* environment (Santos *et al.*, 2001); (2) cultures used here were polyclonal, which inevitably introduces additional genetic variability; and (3) possibility of bacterial contamination despite antibiotic treatment (Shoguchi *et al.*, 2013). Evaluating the features of only conserved, shared

and well-described proteins of these symbiotic dinoflagellates is one important piece in a puzzle that includes many different genes and a number of pathways. Consequently, these results present an important foundation for recognizing similarity within coral endosymbiotic algae and their unique capacity among dinoflagellates to establish a symbiotic relationship with corals and other marine species.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

This work was supported by a University of Queensland Postdoctoral Fellowship for Women and a Discovery Early Career Researcher Award (DE120101412) to NNR, as well as the ARC Centre of Excellence for Coral Reef Studies (CE0561435) (www.coralcoe.org.au). We also acknowledge funding support from the Australian Research Council (Project LP0989200).

Data Deposition

This Transcriptome Shotgun Assembly project has been deposited at DDBJ/EMBL/GenBank under accession number GBGW00000000 (clade A), GBRZ00000000 (Clade B), GBSC00000000 (Clade C) and GBRR00000000 (Clade D).

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. (1990). Basic local alignment search tool. *J Mol Biol* **215**: 403–410.

Arnér ESJ, Holmgren A. (2000). Physiological functions of thioredoxin and thioredoxin reductase. *Eur J Biochem* **267**: 6102–6109.

Bachvaroff TR, Place AR. (2008). From stop to start: tandem gene arrangement, copy number and Trans-splicing sites in the dinoflagellate *Amphidinium carterae*. *PLoS One* **3**: e2929.

Baker AC. (2003). Flexibility and specificity in coral-algal symbiosis: diversity, ecology, and biogeography of *Symbiodinium*. *Annu Rev Ecol Evol Syst* **34**: 661–689.

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

Chen CA, Wang JT, Fang LS, Yang YW. (2005). Fluctuating algal symbiont communities in *Acropora palifera* (Scleractinia: Acroporidae) from Taiwan. *Mar Ecol Prog Ser* **295**: 113–121.

Davy SK, Allemand D, Weis VM. (2012). Cell biology of cnidarian-dinoflagellate symbiosis. *Microbiol Mol Biol Rev* **76**: 229–261.

Douglas AE. (2003). Coral bleaching–how and why? *Mar Pollut Bull* **46**: 385–392.

Dunlap WC, Starcevic A, Baranasic D, Diminic J, Zucko J, Gacesa R *et al.* (2013). KEGG orthology-based annotation of the predicted proteome of *Acropora digitifera*: ZoophyteBase-an open access and searchable database of a coral genome. *BMC Genom* **14**: 509.

Felsenstein J. (1989). Mathematics vs. Evolution: Mathematical Evolutionary Theory. *Science* **246**: 941–942.

Ghavam Mostafavi P, Fatemi SMR, Shahhosseiny MH, Hoegh-Guldberg O, Loh WKW. (2007). Predominance of clade D *Symbiodinium* in shallow-water reef-building corals off Kish and Larak Islands (Persian Gulf, Iran). *Mar Biol* **153**: 25–34.

Gubbels MJ, Vaishnava S, Boot N, Dubremetz JF, Striepen B. (2006). A MORN-repeat protein is a dynamic component of the *Toxoplasma gondii* cell division apparatus. *J Cell Sci* **119**: 2236–2245.

Guillard RR, Ryther JH. (1962). Studies of marine planktonic diatoms. I. *Cyclotella nana* Hustedt, and *Detonula confervacea* (cleve) Gran. *Can J Microbiol* **8**: 229–239.

Hackett JD, Anderson DM, Erdner DL, Bhattacharya D. (2004). Dinoflagellates: a remarkable evolutionary experiment. *Am J Bot* **91**: 1523–1534.

Harper JF, Harmon A. (2005). Plants, symbiosis and parasites: a calcium signalling connection. *Nat Rev Mol Cell Biol* **6**: 555–566.

Hoegh-Guldberg O, Mumby PJ, Hooten AJ, Steneck RS, Greenfield P, Gomez E *et al.* (2007). Coral reefs under rapid climate change and ocean acidification. *Science* **318**: 1737–1742.

Howells EJ, Beltran VH, Larsen NW, Bay LK, Willis BL, Van Oppen MJH. (2012). Coral thermal tolerance shaped by local adaptation of photosymbionts. *Nat Climate Change* **2**: 116–120.

Huang da W, Sherman BT, Lempicki RA. (2009a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* **37**: 1–13.

Huang da W, Sherman BT, Lempicki RA. (2009b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**: 44–57.

Huggett J, Dheda K, Bustin S, Zumla A. (2005). Real-time RT-PCR normalisation; strategies and considerations. *Genes Immun* **6**: 279–284.

Hughes TP, Baird AH, Bellwood DR, Card M, Connolly SR, Folke C *et al.* (2003). Climate change, human impacts, and the resilience of coral reefs. *Science* **301**: 929–933.

Jeong HJ, Yoo YD, Kang NS, Lim AS, Seong KA, Lee SY *et al.* (2012). Heterotrophic feeding as a newly identified survival strategy of the dinoflagellate *Symbiodinium*. *Proc Natl Acad Sci USA* **109**: 12604–12609.

Kim S, Bachvaroff TR, Handy SM, Delwiche CF. (2011). Dynamics of actin evolution in dinoflagellates. *Mol Biol Evol* **28**: 1469–1480.

Ladner JT, Barshis DJ, Palumbi SR. (2012). Protein evolution in two co-occurring types of *Symbiodinium*: An exploration into the genetic basis of thermal tolerance in *Symbiodinium* clade D. *BMC Evol Biol* **12**: 217.

LaJeunesse TC. (2002). Diversity and community structure of symbiotic dinoflagellates from Caribbean coral reefs. *Mar Biol* **141**: 387–400.

Lee MY, Shin KH, Kim YK, Suh JY, Gu YY, Kim MR *et al.* (2005). Induction of thioredoxin is required for nodule development to reduce reactive oxygen species levels in soybean roots. *Plant Physiol* **139**: 1881–1889.

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**: 1010–1021.
Leggat W, Yellowlees D, Medina M. (2011). Recent progress in *Symbiodinium* transcriptomics. *J Exp Marine Biol Ecol* **408**: 120–125.
Lesser MP. (2006). Oxidative stress in marine environments: biochemistry and physiological ecology. *Annu Rev Physiol* **68**: 253–278.
Lesser MP, Shick JM. (1989). Effects of irradiance and ultraviolet radiation on photoadaptation in the zooxanthellae of *Aiptasia pallida*: primary production, photoinhibition, and enzymic defenses against oxygen toxicity. *Mar Biol* **102**: 243–255.
Lesser MP, Stat M, Gates RD. (2013). The endosymbiotic dinoflagellates (*Symbiodinium* sp.) of corals are parasites and mutualists. *Coral Reefs* **32**: 603–611.
Levy O, Achituv Y, Yacobi YZ, Stambler N, Dubinsky Z. (2006). The impact of spectral composition and light periodicity on the activity of two antioxidant enzymes (SOD and CAT) in the coral *Favia favus*. *J Exp Mar Biol Ecol* **328**: 35–46.
Li L, Hastings JW. (1998). The structure and organization of the luciferase gene in the photosynthetic dinoflagellate *Gonyaulax polyedra*. *Plant Mol Biol* **36**: 275–284.
Lidie KB, Ryan JC, Barbier M, Van Dolah FM. (2005). Gene expression in Florida red tide dinoflagellate *Karenia brevis*: analysis of an expressed sequence tag library and development of DNA microarray. *Mar Biotechnol (NY)* **7**: 481–493.
Lin S. (2011). Genomic understanding of dinoflagellates. *Res Microbiol* **162**: 551–569.
McClanahan TR. (2004). The relationship between bleaching and mortality of common corals. *Mar Biol* **144**: 1239–1245.
McCord JM, Fridovich I. (1969). Superoxide dismutase. An enzymic function for erythrocuprein (hemocuprein). *J Biol Chem* **244**: 6049–6055.
McGinley MP, Aschaffenburg MD, Pettay DT, Smith RT, LaJeunesse TC, Warner ME. (2012). Transcriptional response of two core photosystem genes in *Symbiodinium* spp. exposed to thermal stress. *PLoS One* **7**: e50439.
Meyer E, Aglyamova GV, Wang S, Buchanan-Carter J, Abrego D, Colbourne JK *et al.* (2009). Sequencing and *de novo* analysis of a coral larval transcriptome using 454 GSFlx. *BMC Genom* **10**: 219.
Mitra RM, Gleason CA, Edwards A, Hadfield J, Downie JA, Oldroyd GED *et al.* (2004). A Ca2+/calmodulin-dependent protein kinase required for symbiotic nodule development: gene identification by transcript-based cloning. *Proc Natl Acad Sci USA* **101**: 4701–4705.
Morita M, Iguchi A, Takemura A. (2009). Roles of calmodulin and calcium/calmodulin-dependent protein kinase in flagellar motility regulation in the coral *Acropora digitifera*. *Mar Biotechnol* **11**: 118–123.
Moya A, Huisman L, Ball EE, Hayward DC, Grasso LC, Chua CM *et al.* (2012). Whole transcriptome analysis of the coral *Acropora millepora* reveals complex responses to CO(2)-driven acidification during the initiation of calcification. *Mol Ecol* **21**: 2440–2454.
Muscatine L, Pool RR, Trench RK. (1975). Symbiosis of algae and invertebrates: aspects of the symbiont surface and the host-symbiont interface. *Transac Am Microscop Soc* **94**: 450–469.
Nagamune K, Sibley LD. (2006). Comparative genomic and phylogenetic analyses of calcium ATPases and calcium-regulated proteins in the apicomplexa. *Mol Biol Evol* **23**: 1613–1627.
Ness RW, Siol M, Barrett SC. (2011). *De novo* sequence assembly and characterization of the floral transcriptome in cross- and self-fertilizing plants. *BMC Genom* **12**: 298.
Ogawa D, Bobeszko T, Ainsworth T, Leggat W. (2013). The combined effects of temperature and CO2 lead to altered gene expression in *Acropora aspera*. *Coral Reefs* **32**: 895–907.
Oldroyd GED, Harrison MJ, Paszkowski U. (2009). Reprogramming plant cells for endosymbiosis. *Science* **324**: 753–754.
Peleg-Grossman S, Volpin H, Levine A. (2007). Root hair curling and Rhizobium infection in *Medicago truncatula* are mediated by phosphatidylinositide-regulated endocytosis and reactive oxygen species. *J Exp Bot* **58**: 1637–1649.
Pochon X, Gates RD. (2010). A new *Symbiodinium* clade (Dinophyceae) from soritid foraminifera in Hawai'i. *Mol Phylogenet Evol* **56**: 492–497.
Pochon X, LaJeunesse TC, Pawlowski J. (2004). Biogeographic partitioning and host specialization among foraminiferan dinoflagellate symbionts (*Symbiodinium*; Dinophyta). *Mar Biol* **146**: 17–27.
Rae PM. (1976). Hydroxymethyluracil in eukaryote DNA: a natural feature of the pyrrophyta (dinoflagellates). *Science* **194**: 1062–1064.
Rizzo PJ. (1981). Comparative aspects of basic chromatin proteins in dinoflagellates. *BioSystems* **14**: 433–443.
Rohwer F, Seguritan V, Azam F, Knowlton N. (2002). Diversity and distribution of coral-associated bacteria. *Mar Ecol Prog Ser* **243**: 1–10.
Rosic NN, Hoegh-Guldberg O. (2010). A method for extracting a high-quality RNA from *Symbiodinium* sp. *J Applied Phycol* **22**: 139–146.
Rosic NN, Leggat W, Kaniewska P, Dove S, Hoegh-Guldberg O. (2013). New-old hemoglobin-like proteins of symbiotic dinoflagellates. *Ecol Evol* **3**: 822–834.
Rosic NN, Pernice M, Dove S, Dunn S, Hoegh-Guldberg O. (2011a). Gene expression profiles of cytosolic heat shock proteins Hsp70 and Hsp90 from symbiotic dinoflagellates in response to thermal stress: possible implications for coral bleaching. *Cell Stress Chaperones* **16**: 69–80.
Rosic NN, Pernice M, Rodriguez-Lanetty M, Hoegh-Guldberg O. (2011b). Validation of housekeeping genes for gene expression studies in *Symbiodinium* exposed to thermal and light stress. *Mar Biotechnol (NY)* **13**: 355–365.
Rowan R. (1998). Diversity and ecology of zooxanthellae on coral reefs. *J Phycol* **34**: 407–417.
Rowan R. (2004). Coral bleaching: thermal adaptation in reef coral symbionts. *Nature* **430**: 742.
Rowan R, Whitney SM, Fowler A, 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.
Royer Jr WE, Zhu H, Gorr TA, Flores JF, Knapp JE. (2005). Allosteric hemoglobin assembly: diversity and similarity. *J Biol Chem* **280**: 27477–27480.
Santos SR, Taylor DJ, Coffroth MA. (2001). Genetic comparisons of freshly isolated versus cultured

symbiotic dinoflagellates: Implications for extrapolating to the intact symbiosis. *J Phycol* **37**: 900–912.
Santos SR, Taylor DJ, Kinzie Iii RA, Hidaka M, Sakai K, Coffroth MA. (2002). Molecular phylogeny of symbiotic dinoflagellates inferred from partial chloroplast large subunit (23S)-rDNA sequences. *Mol Phylogenet Evol* **23**: 97–111.
Schulz MH, Zerbino DR, Vingron M, Birney E. (2012). Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* **28**: 1086–1092.
Shimada H, Koizumi M, Kuroki K, Mochizuki M, Fujimoto H, Ohta H *et al.* (2004). ARC3, a chloroplast division factor, is a chimera of prokaryotic FtsZ and part of eukaryotic phosphatidylinositol-4-phosphate 5-kinase. *Plant Cell Physiol* **45**: 960–967.
Shinzato C, Shoguchi E, Kawashima T, Hamada M, Hisata K, Tanaka M *et al.* (2011). Using the *Acropora digitifera* genome to understand coral responses to environmental change. *Nature* **476**: 320–323.
Shoguchi E, Shinzato C, Kawashima T, Gyoja F, Mungpakdee S, Koyanagi R *et al.* (2013). Draft assembly of the *Symbiodinium minutum* nuclear genome reveals dinoflagellate gene structure. *Curr Biol* **23**: 1399–1408.
Sorek M, Levy O. (2012). Influence of the quantity and quality of light on photosynthetic periodicity in coral endosymbiotic Algae. *PLoS One* **7**: e43264.
Sørensen JG, Kristensen TN, Loeschcke V. (2003). The evolutionary and ecological role of heat shock proteins. *Ecol Lett* **6**: 1025–1037.
Sturzenbaum SR, Kille P. (2001). Control genes in quantitative molecular biological techniques: the variability of invariance. *Comp Biochem Physiol B Biochem Mol Biol* **130**: 281–289.
Swulius MT, Waxham MN. (2008). Ca(2+)/calmodulin-dependent protein kinases. *Cell Mol Life Sci* **65**: 2637–2657.
Takeshima H, Komazaki S, Nishi M, Iino M, Kangawa K. (2000). Junctophilins: a novel family of junctional membrane complex proteins. *Mol Cell* **6**: 11–22.
Telford MJ. (2007). Phylogenomics. *Curr Biol* **17**: R945–R946.
Thornhill DJ, Lewis AM, Wham DC, Lajeunesse TC. (2013). Host-specialist lineages dominate the adaptive radiation of reef coral endosymbionts. *Evolution* **68**: 352–367.
Tonk L, Bongaerts P, Sampayo EM, Hoegh-Guldberg O. (2013). SymbioGBR: a web-based database of *Symbiodinium* associated with cnidarian hosts on the Great Barrier Reef. *BMC Ecol* **13**: 7.
Trench RK. (1979). The cell biology of plant-animal symbiosis. *Annu Rev Plant Physiol* **30**: 485–531.
Voolstra CR, Sunagawa S, Schwarz JA, Coffroth MA, Yellowlees D, Leggat W *et al.* (2009). Evolutionary analysis of orthologous cDNA sequences from cultured and symbiotic dinoflagellate symbionts of reef-building corals (Dinophyceae: *Symbiodinium*). *Comp Biochem Physiol Part D Genomics Proteomics* **4**: 67–74.
Weber MX, Medina M. (2012). The role of microalgal symbionts (*Symbiodinium*) in holobiont physiology. *Adv Bot Res* **64**: 119–140.
Weis VM. (2008). Cellular mechanisms of Cnidarian bleaching: stress causes the collapse of symbiosis. *J Exp Biol* **211**: 3059–3066.
Wisecaver JH, Brosnahan ML, Hackett JD. (2013). Horizontal gene transfer is a significant driver of gene innovation in dinoflagellates. *Genome Biol Evol* **5**: 2368–2381.
Wisecaver JH, Hackett JD. (2011). Dinoflagellate genome evolution. *Annu Rev Microbiol* **65**: 369–387.
Zerbino DR, Birney E. (2008). Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res* **18**: 821–829.
Zhang H, Hou Y, Miranda L, Campbell DA, Sturm NR, Gaasterland T *et al.* (2007). Spliced leader RNA trans-splicing in dinoflagellates. *Proc Natl Acad Sci USA* **104**: 4618–4623.
Zhang S, Sui Z, Chang L, Kang K, Ma J, Kong F *et al.* (2014). Transcriptome de novo assembly sequencing and analysis of the toxic dinoflagellate *Alexandrium catenella* using the Illumina platform. *Gene* **537**: 285–293.

Supplementary Information accompanies this paper on The ISME Journal website (http://www.nature.com/ismej)