Gene Expression Variation Resolves Species and Individual Strains among Coral-Associated Dinoflagellates within the Genus *Symbiodinium*

John E. Parkinson¹,*, Sebastian Baumgarten², Craig T. Michell², Iliana B. Baums¹, Todd C. LaJeunesse¹, and Christian R. Voolstra²,*

¹Department of Biology, Pennsylvania State University
²Red Sea Research Center, Division of Biological and Environmental Science and Engineering, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia

*Corresponding author: E-mail: jparkinson@psu.edu; christian.voolstra@kaust.edu.sa.

Accepted: February 1, 2016

Data deposition: This project has been deposited at the NCBI Sequence Read Archive database under accession numbers PRJNA274852, PRJNA274854, PRJNA274855, and PRJNA274856, and the RNAseq transcriptomes at reefgenomics.org.

Abstract

Reef-building corals depend on symbiotic mutualisms with photosynthetic dinoflagellates in the genus *Symbiodinium*. This large microalgal group comprises many highly divergent lineages (“Clades A–I”) and hundreds of undescribed species. Given their ecological importance, efforts have turned to genomic approaches to characterize the functional ecology of *Symbiodinium*. To date, investigators have only compared gene expression between representatives from separate clades—the equivalent of contrasting genera or families in other dinoflagellate groups—making it impossible to distinguish between clade-level and species-level functional differences. Here, we examined the transcriptomes of four species within one *Symbiodinium* clade (Clade B) at ~20,000 orthologous genes, as well as multiple isoclonal cell lines within species (i.e., cultured strains). These species span two major adaptive radiations within Clade B, each encompassing both host-specialized and ecologically cryptic taxa. Species-specific expression differences were consistently enriched for photosynthesis-related genes, likely reflecting selection pressures driving niche diversification. Transcriptional variation among strains involved fatty acid metabolism and biosynthesis pathways. Such differences among individuals are potentially a major source of physiological variation, contributing to the functional diversity of coral holobionts composed of unique host–symbiont genotype pairings. Our findings expand the genomic resources available for this important symbiont group and emphasize the power of comparative transcriptomics as a method for studying speciation processes and interindividual variation in nonmodel organisms.

Key words: dinoflagellates, phylogenetics, RNAseq, symbiosis, transcriptome, zooxanthellae.

Introduction

The concept that adaptation and speciation are driven largely by natural selection on variant individuals of a population is central to evolutionary biology. Much like other types of genetic diversity, gene expression variation is extensive, highly heritable, and subject to selection (Ranz and Machado 2006; Voolstra et al. 2007; Wittkopp et al. 2008). The role of differential gene expression in ecological speciation has received renewed interest in the genomics era because the molecular biology of nonmodel organisms with unique evolutionary histories can now be studied in great detail at relatively low cost (Wolf et al. 2010). For example, among two recently diverged species of cordgrass, only one is successful at invading environments perturbed by climate change, and it exhibits unique expression patterns at growth- and stress-related genes (Chelaifa et al. 2010). A similar study in daisies illustrated that a comparative transcriptomic framework can be used to identify selective processes affecting ecological speciation (Chapman et al. 2013). Additionally, transcription-based assays of microbial metagenomes have revealed unique niche diversification (e.g., specialization for certain substrates, metabolic pathways, or environments) that is otherwise hidden due to functional redundancy in the genomes of
many bacteria (Gifford et al. 2013). Thus, comparative genomics can also provide a means to recognize important functional variation in organisms that are difficult to probe phenotypically, such as corals and their symbionts.

Coral reef ecosystems support tremendous marine biodiversity and ecological goods and services (Moberg and Folke 1999). Coral productivity and growth depend on a mutualism with endosymbiotic dinoflagellates known as *Symbiodinium* (Muscatine and Porter 1977; Muscatine 1990; Yellowlees et al. 2008). This microalgal “genus” is incredibly diverse, encompassing at least nine major lineages that show ribosomal divergence equivalent to that found among different genera, families, or even orders of other dinoflagellates (Rowan and Powers 1992). Likewise, *Symbiodinium* exhibit many unique ecologies, ranging from “host-specialized” taxa commonly found as symbiotic partners of corals (Parkinson, Coffroth, et al. 2015), to “ecologically cryptic” taxa with alternate non-symbiotic lifestyles (LaJeunesse et al. 2015), to completely “free-living” taxa that thrive independently in the water column (Jeong et al. 2014). Unlike their mostly obligate coral hosts, many *Symbiodinium* can survive ex hospite and are maintained in culture. In the natural environment, stressful conditions can cause the association between corals and host-specialized symbionts to break down in a process called coral bleaching, which can lead to colony mortality (Fitt et al. 2001).

Climate change is predicted to drive more frequent and intense bleaching events (Hoegh-Guldberg 1999), prompting a major research focus on how climate-related stressors might affect the two major evolutionary radiations within Clade B: The coral–dinoflagellate symbioses in the future. Accordingly, the last decade has generated many studies describing coral host transcription in various contexts (Meyer and Weis 2012), but comparable studies in *Symbiodinium* are still in their early stages (Leggat et al. 2007; Leggat, Yellowlees, et al. 2011; Lin 2011).

With the incorporation of next-generation sequencing technology, genomic resources for *Symbiodinium* have expanded greatly despite their status as a nonmodel organism. The first draft genome was released in 2013 (Shoguchi et al. 2013), with the complete chloroplast genome following shortly thereafter (Barbrook et al. 2014). Multiple mRNA transcriptomes are available (Bayer et al. 2012; Ladner et al. 2012; Baumgarten et al. 2013; Rosic et al. 2014; Xiang et al. 2015), representing the four major clades known to associate with scleractinian corals (Clades A, B, C, and D). Recent efforts have expanded in important new directions, such as the description of *Symbiodinium* microRNAs (Baumgarten et al. 2013), the comparison of orthologous genes among clades (Voolstra et al. 2009; Ladner et al. 2012; Barshis et al. 2014; Rosic et al. 2014), the completion of another draft genome (Lin et al. 2015), and the development of the *Aiptasia–Symbiodinium* system for in-depth cellular and physiological research (Weis et al. 2008; Sunagawa et al. 2009; Lehner et al. 2012, 2014; Xiang et al. 2013; Baumgarten et al. 2015).

Dinoflagellate genomes are unique among eukaryotes for multiple reasons (Leggat, Yellowlees, et al. 2011). Of particular note, dinoflagellates including *Symbiodinium* modulate nuclear-encoded protein levels predominantly by posttranscriptional processes (Morse et al. 1989; Leggat, Seneca, et al. 2011). It is now understood that dinoflagellates also exhibit some measure of transcriptional regulation, albeit changes in expression profiles are minimal when exposed to different environmental conditions (Erdner and Anderson 2006; Moustafa et al. 2010). For example, the number and magnitude of expression changes among *Symbiodinium* exposed to thermal stress are relatively small compared with their animal hosts (Leggat, Seneca, et al. 2011). Barshis et al. (2014) found that two *Symbiodinium* spp. in Clades C and D did not alter gene expression when exposed to temperature stress in hospite, even though the host response involved the modulation of hundreds of genes (Barshis et al. 2013, 2014). Interestingly, a large number of transcriptional differences were maintained, or fixed, for the two species from different clades regardless of temperature treatment (Barshis et al. 2014). This suggests that fixed expression differences are likely to be evident in strains cultured ex hospite under identical controlled environmental conditions. Differences in these “stable-state” expression profiles among lineages may strongly reflect evolutionary divergence, some of which may be adaptive. These expression patterns may also correspond to functional differences among distantly related species. If lineage-specific expression extends to the subcladal level—that is, between species within clades or among individual strains within species—it will be critical to recognize this source of variation when interpreting *Symbiodinium* genomic data and account for it in future experimental designs.

By comparing different isoclonal cell lines (strains), it is possible to reveal intraspecific variation in genomic features that underlie ecological and physiological phenotypes. For example, unique genes distinguish strains of nitrogen-fixing rhizobial bacteria with different symbiotic efficiencies and host specificities (Galardini et al. 2011; Österman et al. 2015). At the level of transcription, toxic and nontoxic strains of the dinoflagellate *Alexandrium minutum* maintain fixed expression differences at shared genes (Yang et al. 2010). We may expect similar patterns among *Symbiodinium* strains, but this idea has never been tested. *Symbiodinium* belonging to Clade B are ideal candidates for further genomic characterization because several ecologically distinct species within this group were recently described (LaJeunesse et al. 2012; Parkinson, Coffroth, et al. 2015), a draft genome exists for the member species *Symbiodinium minutum* (Shoguchi et al. 2013), and multiple genetically distinct cultures are available for several species.

Currently, the extent of variation among species within a single *Symbiodinium* clade and among individual strains within a single species is mostly unknown (Parkinson and Baums 2014). To address this knowledge gap, we analyzed stable-state gene expression among four species representing the two major evolutionary radiations within Clade B: The
Pleistocene (B1) radiation and the Pliocene (B19) radiation (sensu LaJeunesse 2005). For each radiation, two species with different ecologies were studied: Either host-specialized taxa or ecologically cryptic taxa. Although these latter species were originally cultured from coral tissues, they have never been detected as the numerically dominant symbionts in cnidarian mutualisms, and therefore were probably commensals or free-living contaminants isolated from the mucus or gastrovascular cavity (Parkinson, Coffroth, et al. 2015). Where available, we incorporated biological replication in the form of distinct isoclonal cell cultures. The genomic resources developed here should assist in the design and interpretation of future comparative transcriptional analyses among Symbiodinium strains, species, and clades, as well as broaden our understanding of speciation among microeukaryotes.

Materials and Methods

Culturing

Isoclonal cultures (strains) of Clade B Symbiodinium were maintained at the Pennsylvania State University. They were originally acquired from the Robert K. Trench and Buffalo Undersea Reef Research collections. This study included one strain of Symbiodinium aenigmaticum (mac04-487), four strains of S. minutum (mac703, Mf1.05b, rt002, and rt351), one strain of Symbiodinium pseudominutum (rt146), and four strains of Symbiodinium psygmophilum (HIAnp, Mf10.14b.02, PurPFlex, and rt141), for the analysis of ten individual transcriptomes. Most strains are available from the Provasoli-Guillard National Center for Marine Algae and Microbiota at Bigelow Laboratory for Ocean Sciences, East Boothbay, Maine (LaJeunesse et al. 2012; Parkinson, Coffroth, et al. 2015) or by request. Within S. minutum and S. psygmophilum, strains were confirmed to represent distinct genotypes based on repeat length variation at the microsatellite locus Sym15 (Pettay and LaJeunesse 2007) and haplotype differences in the psbA noncoding region (LaJeunesse and Thornhill 2011).

Single cells were originally isolated from host tissues by Schoenberg and Trench (1980) using modified methods of McLaughlin and Zarl (1959) or by Mary Alice Coffroth following the methods of Santos et al. (2001). To establish initial crude cultures, several drops of a heavy suspension of symbiont cells were transferred into nutrient-enriched filtered seawater (Provasoli 1968) and then spread onto semisolid agar (0.8%) containing the same seawater. Vegetative cells from viable colonies on agar were then transferred to liquid medium ASP-8A (Ahles 1967). To generate isoclonal lines, viable colonies on agar were then transferred to liquid medium containing the same seawater. Vegetative cells from the second week of growth postsynchronization, all target cultures were transferred to 50 ml tubes and centrifuged at 3000 RCF (relative centrifugal force). The media was decanted and the pellets were flash frozen in liquid nitrogen. Pellets were ground with a prechilled mortar and pestle and transferred into 1.5 ml tubes. Nucleic acids were extracted with TriReagent (Thermo Fisher Scientific, Waltham, MA) and RNA was isolated and cleaned with the RNeasy Kit (Qiagen, Venlo, The Netherlands) according to the manufacturer’s protocols.

Total RNA isolations were shipped on dry ice to the KAUST Red Sea Research Center, where they were quality checked using a Bioanalyzer 2000 (Agilent, Santa Clara, CA) or by library preparation. For Illumina 2 × 100 bp paired-end sequencing, 180 bp libraries were generated from oligo(dT)-enriched mRNA using the NEBNext Ultra Directional RNA Library Prep Kit for Illumina (New England Biolabs, Ipswich, MA) according to the manufacturer’s protocols. Each read pair ideally yielded a partially overlapping 180 bp contiguous sequence, allowing for additional quality control. mRNA sequencing libraries for each of the ten samples were multiplexed in equimolar concentrations and run on one lane on the Illumina HiSeq 2000 platform, producing a total of 142 million paired-end reads. All raw RNAseq data are available in the NCBI Sequence Read Archive database under accession numbers http://www.ncbi.nlm.nih.gov/bioproject/PRJNA274856/ (S. aenigmaticum), http://www.ncbi.nlm.nih.gov/bioproject/PRJNA274852/ (S. minutum), http://www.ncbi.nlm.nih.gov/bioproject/PRJNA274855/ (S. pseudominutum), and http://www.ncbi.nlm.nih.gov/bioproject/PRJNA274854/ (S. psygmophilum).

Transcriptome Assemblies and Annotation

Adapters and low quality nucleotides (<20 Phred score in ASCII 33 format) were removed from raw reads with Trimmomatic (Bolger et al. 2014). Reads were error corrected with the error correction module of the AllPATHS-LG pipeline (Gnerre et al. 2011; Ribeiro et al. 2012). Quality-controlled reads combined from all samples on a per-species basis were assembled using the Trinity package (Grabherr et al. 2011) with minimum k-mer coverage of 2 and minimum contig length of 250 bp to generate one reference assembly per species (four total). For mapping purposes, we further reduced each reference assembly to include only the longest transcript for a related set of splice variants of a gene. For each sample, reads were mapped back to the reduced assembly for the appropriate species with Bowtie2 (Langmead and
Orthologous genes were identified via reciprocal BLASTp of ORFs pairwise for each species within the program InParanoid (Remm et al. 2001), retaining hits with bitscores >300 (supplementary table S1, Supplementary Material online). Multiparanoid (Alexeyenko et al. 2006) was then used to identify orthologs that occurred exactly four times (once in each species). Additional domain-based functional annotations were assigned using the Pfam database v27 (Finn et al. 2014) and are provided in supplementary table S2, Supplementary Material online.

Most current software designed to analyze differential expression for RNAseq data assumes raw read counts among samples mapped to one common reference transcriptome, and therefore only accepts integer values as input. To compare expression at orthologous genes across species, it was necessary to normalize read counts by transcript length using FPKM to account for species-specific sequence length differences. This normalization produced many decimal expression values that were still informative. For this reason, we scaled FPKM by a common factor such that the lowest expressed gene’s value equaled 1 and then rounded values to the nearest integer. Thus, a scaled FPKM of 50 means the gene is expressed 50 times higher than the lowest expressed gene retained in the data set. This way, all orthologs could be compared in the scaled FPKM space.

Scaled FPKM data were then used as input for the R package edgeR (Robinson et al. 2010), which accommodates data sets with unequal replication when performing comparisons among treatments (in this case, species). No additional normalization procedures were carried out within the program. Significant differential expression was determined by pairwise comparisons among species with a false discovery rate–adjusted P value (FDR) of <0.1. To additionally test by lineage, all S. minutum and S. pseudominutum samples were grouped as “Pleistocene” and all S. aenigmaticum and S. psymophilum samples were grouped as “Plioene.” To test by ecology, all S. minutum and S. psymophilum samples were grouped as host-specialized, and all S. aenigmaticum and S. pseudominutum samples were grouped as ecologically cryptic.

Multidimensional scaling (MDS) plots were generated in edgeR using the plotMDS function. The distances between pairs of RNA samples correspond to the leading log2-fold-changes, which is the average (root-mean-square) of the largest absolute log2-fold-changes (Robinson et al. 2010). In all three comparisons of the MDS plots (non-DEG [differentially expressed gene] only, non-DEG + DEG, and DEG only), similar clustering was observed among the four replicates of the two species with replicates. It was therefore reasonable to assume that (hypothetical) replicates of the other two species would show a similar distribution. Because variation between replicates was consistent, we assumed that the distances between any of the species (which contributes to the number of DEGs identified) was not affected by the number of replicates within a species. This was incorporated into the between-species
Gene Expression between and within Symbiodinium Species

Differential Expression within Species

For both S. minutum and S. psygmophilum, isoclonal cultures of four individual strains each were available, providing two opportunities to test for differential expression within a species. Each species was analyzed separately in edgeR (Robinson et al. 2010). Raw read counts were normalized with the geometric mean method.

Unlike hybridization-based techniques, the level of technical variation from sequencing is predictable and can therefore be distinguished from biological variation (Chen et al. 2014). In this case, the technical coefficient of variation describes the measurement error derived from the uncertainty with which the abundance of every gene is estimated from the sequencing platform, which decreases with increasing total counts for each gene in an RNA sample. In contrast, the biological coefficient of variation (BCV) describes the variation of the unknown, true abundance of each gene among replicates of RNA samples that will remain, even if sequencing depth could be increased indefinitely. Thus, the BCV represents the most important and main source of variation in expression studies using a high-throughput deep-sequencing approach (McCarthy et al. 2012). In RNA expression studies, the BCV is usually determined from the biological replicates of RNA samples so the total variation of gene abundances can be calculated by considering the following equation: Total CV^2 = Technical CV^2 + Biological CV^2 (McCarthy et al. 2012).

Due to the lack of replicates for the comparison of the four genotypes of S. minutum and S. psygmophilum, we set the BCV to a fixed value a priori under the assumption that a majority of genes were not differentially expressed, which we considered appropriate for Symbiodinium of the same species under stable-state conditions. Although a value of ~1% is suggested for technical replicates and a value of ~10% is suggested for unique samples from separate but genetically identical model organisms, a value of ~40% is appropriate for independent biological samples (Chen et al. 2014) and was chosen for our expression analysis among the four genotypes of both Symbiodinium species.

Significance of DEGs was determined by pairwise comparisons among individuals based on the negative binomial distribution with FDR < 0.1. A pairwise Euclidian distance matrix for all strain comparisons within and between S. minutum and S. psygmophilum was computed based on scaled FPKM values using PRIMER v6 software (Clarke and Gorley 2006).

Visualization and Functional Analyses

DEGs between and within species were visualized as heatmaps by plotting scaled FPKM expression values with Gene-E (Gould 2015). Lists were tested for GO term functional enrichment with the R/Bioconductor package topGO (Alexa and Rahnenfuhrer 2010), using the default “weight01” Alexa algorithm with the recommended cutoff of P < 0.05.

Results

Transcriptome Assemblies

We targeted isoclonal strains from four Clade B Symbiodinium species: S. aenigmaticum (n=1 strain: mac04-487); S. minutum (n=4 strains: mac703, Mf1.05b, rt002, and rt351); S. pseudominutum (n=1 strain: rt146); and S. psygmophilum (n=4 strains: HIAp, Mf10.14b.02, PurPFlex, and rt141). We reared the ten cultures under identical conditions in one incubator to assess stable-state conditions in the absence of environmental variability. Symbiodinium minutum and S. pseudominutum belong to the Pleistocene (B1) radiation. The former is a host-specialized mutualist because it commonly associates with the anemone Aiptasia sp. The latter is ecologically cryptic—although it has been isolated from the background symbiont population of four cnidarians, it has never been identified as a dominant symbiont. Symbiodinium psygmophilum and S. aenigmaticum belong to the Pliocene (B19) radiation; the former is host-specialized, the latter is ecologically cryptic. Note the uneven distribution of strains within species. This limitation was based on which cultures were available in the collection and meant that certain contrasts (e.g., 4 S. minutum strains vs. 4 S. psygmophilum strains) potentially had more power to detect differential expression than others (e.g., 1 S. aenigmaticum strain vs. 1 S. pseudominutum strain). However, for these data it was unlikely that the number of replicates affected power given that the same variance distribution was used for all comparisons (see Materials and Methods).

From these cultures, we generated ten high-quality short-read RNAsq libraries (table 1). Across species, sequencing statistics were quite similar. Total reads per sample averaged 14.3 million, while on average 88.5% of reads per sample passed quality control. For each of the four species, we generated a single reference assembly from either a combination of all strains of a given species (in the cases of S. minutum and S. psygmophilum) or from the single representative strain (in the cases of S. aenigmaticum and S. pseudominutum). The number of assembled genes per transcriptome averaged 48,700, the number of predicted ORFs averaged 41,300, the contig N50 statistic averaged 1,515 bp, mean transcript length per transcriptome averaged 1,078 bp, and annotation success averaged 46.5%. These values are in agreement with previously published Clade B Symbiodinium transcriptomes (Bayer et al. 2012; Baumgarten et al. 2013; Shoguchi et al. 2013).

After uploading reference assemblies into IPA software, we identified 19 relevant canonical pathways with significant gene set representation in all species, including fatty acid.


Downloaded from https://academic.oup.com/gbe/article-abstract/8/3/665/2574104 by guest on 19 March 2019
beta-oxidation, nitric oxide signaling, oxidative stress response, cell cycle control, RNA processes, and protein ubiquitination (supplementary fig. S1, Supplementary Material online). We compared the ratio of genes observed to total associated genes per pathway among transcriptomes; each pathway was evenly represented in each Clade B *Symbiodinium* species. The four species were also roughly equivalent in terms of their proportions of microsatellite repeat motifs (supplementary fig. S2, Supplementary Material online).

**Between-Species Expression Differences**

In order to compare gene expression between *Symbiodinium* species, we identified orthologs via a reciprocal BLAST approach on predicted ORFs. All species shared a total of 19,359 orthologs after filtering out paralogs and low quality matches (average pairwise ortholog count: 27,784; supplementary table S1, Supplementary Material online). We subsequently conducted a GO term enrichment analysis on DEGs in order to assess which pathways were differentially represented (supplementary table S2, Supplementary Material online). The *S. psymgophilum*–*S. pseudominutum* contrast was enriched for processes including photosynthesis, response to cold, and transmembrane transport. The *S. pseudominutum–S. minutum* contrast was enriched for photosynthesis and heat acclimation. The *S. psymgophilum*–*S. aenigmaticum* contrast was enriched for photosynthesis and mitosis. The *

---

**Table 1**

Sequencing Statistics for the Ten Strains (A) and Transcriptome Assembly Statistics for the Four Species (B) of Clade B *Symbiodinium*

**A**

<table>
<thead>
<tr>
<th>Species</th>
<th>Strain</th>
<th>Radiation</th>
<th>Ecology</th>
<th>Total Read Count (M)</th>
<th>Remaining After QC (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Symbiodinium minutum</em></td>
<td>mac703</td>
<td>Pleistocene (B1)</td>
<td>Host-specialized</td>
<td>10.9</td>
<td>89.03</td>
</tr>
<tr>
<td><em>S. minutum</em></td>
<td>Mf1.05b</td>
<td>Pleistocene (B1)</td>
<td>Host-specialized</td>
<td>19.3</td>
<td>88.40</td>
</tr>
<tr>
<td><em>S. minutum</em></td>
<td>rt002</td>
<td>Pleistocene (B1)</td>
<td>Host-specialized</td>
<td>12.4</td>
<td>88.00</td>
</tr>
<tr>
<td><em>S. minutum</em></td>
<td>rt351</td>
<td>Pleistocene (B1)</td>
<td>Host-specialized</td>
<td>8.7</td>
<td>88.68</td>
</tr>
<tr>
<td><em>Symbiodinium psymgophilum</em></td>
<td>HIAp</td>
<td>Pliocene (B19)</td>
<td>Host-specialized</td>
<td>13.4</td>
<td>88.14</td>
</tr>
<tr>
<td><em>S. psymgophilum</em></td>
<td>Mf10.14b.02</td>
<td>Pliocene (B19)</td>
<td>Host-specialized</td>
<td>11.1</td>
<td>88.63</td>
</tr>
<tr>
<td><em>S. psymgophilum</em></td>
<td>PurPflex</td>
<td>Pliocene (B19)</td>
<td>Host-specialized</td>
<td>11.7</td>
<td>88.55</td>
</tr>
<tr>
<td><em>S. psymgophilum</em></td>
<td>rt141</td>
<td>Pliocene (B19)</td>
<td>Host-specialized</td>
<td>19.5</td>
<td>88.47</td>
</tr>
<tr>
<td><em>Symbiodinium pseudominutum</em></td>
<td>rt146</td>
<td>Pleistocene (B1)</td>
<td>Ecologically cryptic</td>
<td>23.7</td>
<td>88.83</td>
</tr>
<tr>
<td><em>Symbiodinium aenigmaticum</em></td>
<td>mac04-487</td>
<td>Pliocene (B19)</td>
<td>Ecologically cryptic</td>
<td>11.9</td>
<td>88.10</td>
</tr>
</tbody>
</table>

**B**

<table>
<thead>
<tr>
<th>Species</th>
<th>Assembly Length (Mbp)</th>
<th>Gene Count</th>
<th>Predicted ORF Count</th>
<th>Genes Annotated (%)</th>
<th>Longest Gene Length (bp)</th>
<th>Mean Gene Length (bp)</th>
<th>N50 (bp)</th>
<th>GC Content (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>S. minutum</em></td>
<td>57.2</td>
<td>51,199</td>
<td>42,929</td>
<td>47.3</td>
<td>37,483</td>
<td>1,118</td>
<td>1,579</td>
<td>51.33</td>
</tr>
<tr>
<td><em>S. psymgophilum</em></td>
<td>57.2</td>
<td>50,745</td>
<td>42,740</td>
<td>47.7</td>
<td>31,367</td>
<td>1,128</td>
<td>1,618</td>
<td>51.37</td>
</tr>
<tr>
<td><em>S. pseudominutum</em></td>
<td>51.3</td>
<td>47,411</td>
<td>40,716</td>
<td>46</td>
<td>31,393</td>
<td>1,081</td>
<td>1,508</td>
<td>51.51</td>
</tr>
<tr>
<td><em>S. aenigmaticum</em></td>
<td>44.6</td>
<td>45,343</td>
<td>38,923</td>
<td>44.9</td>
<td>24,202</td>
<td>984</td>
<td>1,355</td>
<td>51.39</td>
</tr>
</tbody>
</table>

Note.—The *S. minutum* and *S. psymgophilum* assemblies in (B) are composited from the reads of all respective strains listed in (A).
**Within-Species Expression Differences**

To understand the extent of gene expression differences among individuals within species, we first quantified expression variability by estimating the BCV and subsequently performed pairwise comparisons of the four distinct strains of *S. minutum* or *S. psygmophilum* using a fixed BCV that was more conservative (lower) than the original BCV estimate (Chen et al. 2014). We identified many pairwise expression differences among strains, ranging from 61 to 404 DEGs for *S. minutum* (fig. 4a) and 82 to 293 DEGs for *S. psygmophilum* (fig. 4b).

To further illustrate differences between pairs of strains for *S. minutum* and *S. psygmophilum*, we calculated an expression–strain distance matrix based on Euclidean distances between all pairwise strain comparisons using scaled FPKM values to assess variance (fig. 4c). Technical variation would be expected to be equally distributed across all samples. The distance matrix revealed a nonrandom distribution of variation in gene expression between pairs of strains for a given species. Also, the distance matrix showed that pairwise distances between strains from different species exceeded any within-species variation, and that both species exhibited distinct variance distributions among member strains.

Finally, we generated heatmaps to show the subset of all annotated genes differentially expressed in at least one individual among the four *S. minutum* strains (fig. 5a) and among the four *S. psygmophilum* strains (fig. 5b). In *S. minutum*, DEGs between strains were most highly enriched for the processes of malonyl-CoA biosynthesis, protein polymerization, long-chain fatty acid biosynthesis and metabolism, microtubule and nuclear envelope organization, GTP catabolism, and mitosis regulation (*supplementary table S2, Supplementary Material online*). In *S. psygmophilum*, DEGs between strains were most highly enriched for the processes of DNA replication and biosynthesis, sulfate assimilation and hydrogen sulfide biosynthesis, and microtubule organization (*supplementary table S2, Supplementary Material online*). A full list of annotations, expression values, and DEG list memberships can be found in *supplementary table S3, Supplementary Material online*.

**Discussion**

Fixed differences in gene expression ultimately influence the phenotypic variation available for selection to act upon. We anticipated that a comparative analysis of *Symbiodinium* spp. transcription would improve our understanding of adaptation and speciation among microeukaryotes. Indeed, we found that despite an overall similarity in gene content and expression among Clade B species with distinct ecologies, all cultures exhibited lineage-specific expression differences diagnostic for each species. Overrepresentation of photosynthesis-related gene expression variation among species likely reflects adaptation to unique light regimes over evolutionary time. Extensive disparity in the expression of fatty acid metabolism genes among strains within species may translate into differences in membrane composition, thermal tolerance, energy reserves, and growth rates. These differences may play a

---

**Fig. 1.**—(A) Phylogenetic relationships and ecologies for the Clade B *Symbiodinium* species used in this experiment. The maximum parsimony tree was generated based on microsatellite Sym15 flanker region data from Parkinson, Coffroth, et al. (2015) and the methodologies described therein. (B) The numbers of differentially expressed genes (DEGs) between Clade B *Symbiodinium* species. Counts are placed on the lines connecting the two species being contrasted. Line thickness is scaled by the number of DEGs. Also depicted are the numbers (**n**) of cultured strains (clonal cell lines) included per species. Counts below the diagram show the total number of genes differentially expressed in at least one species and the total number of comparable orthologs across all species.
**Fig. 2.**—Expression heatmaps of annotated DEGs among species. Colors are scaled to the minimum (purple) and maximum (orange) expression value per gene. Any gene that is differentially expressed in at least one species is included.
crucial role in coral–dinoflagellate symbiosis ecology and evolution. By examining the stable-state transcriptomes of cultures reared independently of their hosts under identical environmental conditions, we infer that these differences stem from genotypic rather than environmental factors. Our efforts reinforce the utility of comparative transcriptomics for studying speciation and functional variation in dinoflagellates and other nonmodel organisms (Chelaifa et al. 2010; Chapman et al. 2013; Gifford et al. 2013).

Partitioning the Variation in Gene Expression

When comparing multiple species, expression differences can be attributed to 1) technical variation, 2) within-species variation, and 3) among-species variation, with the proportion of variable genes expected to increase from 1) to 2) to 3) (Whitehead and Crawford 2006). Our results matched this general trend. Technical variation was inferred to be low based on the agreement between our data and transcriptome statistics from other studies that included the same S. minutum strain, the high mapping success achieved between our Mf1.05b reads and the draft genome derived from the same strain (73%), and the nonrandom distribution of expression differences among species (fig. 2c). The percentage of orthologous genes differentially expressed within species (1.54% for S. minutum and 1.18% for S. psygmophilum; fig. 4) was roughly half that found between species (2.33%; fig. 2b). Overall, DEGs make up a small proportion of the entire transcriptome, as has been found before for Symbiodinium and other dinoflagellates (Baumgarten et al. 2013; Barshis et al. 2014; Xiang et al. 2015).

Between-Species Variation

By comparing closely related species within a clade, we greatly expanded our comparative power to determine what genetic changes underlie speciation among Symbiodinium. We were able to identify at least four times as many orthologs shared between Clade B species as has been possible using similar methods to compare species across separate clades (Ladner et al. 2012; Barshis et al. 2014; Rosic et al. 2014). Genetic divergence between clades is massive (Rowan and Powers 1992), and thus comparisons among species within clades reveal finer-scale differences likely to be important in physiological and ecological processes. Overall, stable-state gene expression was similar among Clade B Symbiodinium. Of the nearly 20,000 orthologs shared by S. aenigmaticum, S. minutum, S. pseudominutum, and S. psygmophilum, only 452 (2.3%) were differentially expressed between species. Thus a substantial portion of the transcriptome maintains relatively constant expression levels across members of Clade B. This result mirrors similar studies in other systems such as flowering plants where only a small proportion of interspecific orthologs were differentially expressed (Chapman et al. 2013).

The species comparison with the greatest number of DEGs was S. minutum versus S. psygmophilum (fig. 1b), which fit expectations for several reasons. First, our replication scheme (four strains per species) may have enhanced our ability to detect fixed differences between these species’ transcriptomes (though this is unlikely; see Materials and Methods). Second, both species associate with different hosts and likely diverged in part due to coevolutionary constraints imposed by those hosts, whereas ecologically cryptic species may not have faced the same constraints. Third, they are from phylogenetically divergent lineages. Finally, S. minutum is warm water adapted, while S. psygmophilum is cold water adapted (Thornhill et al. 2008), likely contributing to expression differences.

Interestingly, the contrasts with the second- and third-most abundant DEG counts both involved S. aenigmaticum (fig. 1b), a very distinct species from the Pliocene radiation and...
one that appears to have undergone rapid evolution (fig. 1a; LaJeunesse 2005; Parkinson, Coffroth, et al. 2015). The three species pair comparisons with the least number of DEGs all involved S. pseudominutum (fig. 1b). In fact, this species was roughly equidistant from all other species based on DEG number and MDS position (fig. 3c). Its position might be explained on the one hand by its close evolutionary history with S. minutum, and on the other by its cryptic ecology shared with S. aenigmaticum. Based on these results, fixed differences in gene expression may not always correspond to phylogenetic similarity.

Multidimensional scaling offered a complementary analysis for visualizing the similarities in expression among all strains without a priori knowledge of species membership (fig. 3). By restricting the data set to only non-DEGs, almost all replicates from all species (8 of 10) clustered together (fig. 3a), matching the expectation that at stable-state these Clade B Symbiodinium generally maintain similar expression profiles. When both non-DEGs and DEGs were included in the analysis, each species was mostly resolved, showing that non-DEGs contributed little to either species-level signal or noise (fig. 3b). As expected, when only the DEGs were considered, all species resolved well (fig. 3c). Note however that the distant positioning of S. aenigmaticum in all three MDS plots indicates that a large proportion of expression variation for this species is unique.
**FIG. 5.**—Expression heatmaps of annotated DEGs among individual strains (clonal cell lines) within (A) *Symbiodinium minutum* and (B) *Symbiodinium psygmophilum*. Colors are scaled to the minimum (purple) and maximum (orange) expression value per gene. Any gene that is differentially expressed in at least one strain is included.
In addition to pairwise comparisons, we also contrasted groups of replicate species by lineage (2 species from the Pleistocene radiation vs. 2 Pliocene radiation species) or by ecology (2 host-specialized species vs. 2 ecologically cryptic species). The Pleistocene—Pliocene contrast was equivalent to the S. minutum—S. psygmophilum comparison in terms of identity of DEGs, meaning that the species contrast either captured all the differences between major lineages, or that adding just one more strain to each group did not affect expression variation sufficiently to alter our detection of DEGs, even though the strain belonged to a different species. Similarly, the ‘host-specialized’—‘ecologically cryptic’ contrast only recovered four unique genes that had not been identified in any of the species contrasts. The probable identity of only one of these DEGs was determined (a general mRNA splicing factor). These results indicate that differential expression of a particular set of genes does not necessarily explain shared ecological attributes of phylogenetically distinct species.

Photosynthesis Gene Expression Differences between Species

Expression differences among closely related species were consistently enriched for photosynthesis genes (supplementary table S2, Supplementary Material online). Here, overrepresentation of plastid genes cannot be attributed to light intensity differences because all cultures were reared under identical light conditions. In fact, although we might expect these genes to be regulated by light intensity in Symbiodinium as they are in other photosynthetic organisms (Escoubas et al. 1995; Pfannschmidt 2003), only minor (or no) changes in photosynthesis-related gene expression have been detected in cultures exposed to varying light levels (McGinley et al. 2013; Xiang et al. 2015). Thus, we conclude that different species evolved unique expression levels among photosynthesis-related genes. These differences may relate to inherent variation in the circadian rhythm among species (Van Dolah et al. 2007; Sorek and Levy 2012) or, more likely, to functional variation in photosynthesis biochemistry. For example, during heat stress, thermally sensitive Symbiodinium taxa suffer physiological disruption of PSII photochemistry (Warner et al. 1999; Robison and Warner 2006) and associated downregulation of core photosynthesis genes (McGinley et al. 2012), whereas thermally tolerant species do not. The maintenance of distinct expression patterns at key genes may underlie the capacity for certain Symbiodinium species to occupy distinct niches, as has been demonstrated for three diatom species in the genus Pseudonitzschia (Di Dato et al. 2015).

Evolutionary Significance of Gene Expression Variation

In biogeographic surveys of marine mutualisms, depth and latitude (correlates of light availability) are often primary factors explaining the distribution of Symbiodinium diversity (Rowan and Knowlton 1995; Laleunese et al. 2004, 2014; Frade et al. 2008; Finney et al. 2010; Sanders and Palumbi 2011). Thus, light availability represents a main axis of niche differentiation for this group. Symbiodinium possess a diverse array of light-harvesting proteins (Boldt et al. 2012), which may be both the cause and consequence of ecological specialization. Many such genes have been transferred to the nuclear genome (Bachvaroff et al. 2004), while others are encoded on plastid minicircles (Zhang et al. 1999; Moore et al. 2003; Barbrook et al. 2014). Minicircles are subject to different transcriptional mechanisms than nuclear encoded genes (Dang and Green 2010), which may also facilitate specialization to different light regimes. Given that a majority of expression variation between divergent species is expected to accumulate neutrally over time (Khaitovich et al. 2005), it is intriguing that expression differences between Symbiodinium species are consistently enriched for photosynthesis genes (Baumgarten et al. 2013; Barshis et al. 2014; Rosic et al. 2014; this study). This evidence suggests that species-specific differences in gene expression are functionally important and influenced by natural selection tied to niche diversification.

Within-Species Variation

Within each of the two species with four isoloclonal cultures, we detected hundreds of DEGs: 659 unique genes among S. minutum strains (fig. 4a) and 506 unique genes among S. psygmophilum strains (fig. 4b). Interestingly, only four annotated genes differentially expressed among S. minutum overlapped with those among S. psygmophilum, and enriched categories only overlapped for housekeeping genes which regulate biochemical processes like nucleic acid synthesis and microtubule organization (supplementary table S2, Supplementary Material online). Thus, transcriptional variation among strains differs from species to species (fig. 5). Furthermore, nonrandom gene expression differences among strains of a given species exist even under identical rearing conditions (fig. 4c), emphasizing that a degree of expression variation among Symbiodinium strains is genetically determined and potentially subject to natural selection. Thus, the extent of variation among isoloclonal strains may be much greater than previously assumed. Although inter-individual differences are known to play a significant role in symbiosis ecology and evolution in terrestrial systems (Shuster et al. 2006; Whitham et al. 2006; Hughes et al. 2008), such evidence has been lacking for coral–dinoflagellate associations (Parkinson and Baums 2014). Although ~500 of the ~40,000 genes represents a small fraction of the transcriptome, such differences may be important, especially because overall differential expression of genes within a Symbiodinium species responding to stress seems low (Barshis et al. 2014; but see Baumgarten et al. 2013).

For example, putative “symbiosis genes” have been identified by comparing symbiotic versus aposymbiotic cnidarian
hosts (Meyer and Weis 2012). The expression levels of similar genes in the symbiont may also play a role in maintaining functional associations. Two such genes varied among S. minutum strains: An ABC transporter (up to 4.2-fold) and a glutathione reductase (up to 9.5-fold). There were also clear differences in the expression of long chain fatty acid CoA ligase (up to 12.2-fold), long chain acyl-CoA synthetase (up to 8.8-fold), and six acetyl-CoA carboxylases (up to 12.5-fold), indicating that certain strains regulate fatty acid metabolism differently. These genes may be related to cell membrane composition, which in turn can affect thermal sensitivity (Tchernov et al. 2004; Diaz-Almeyda et al. 2011). They may also relate to energy storage and nutrient availability, perhaps contributing to different growth rates observed among some of these strains ex hospite (Parkinson and Baums 2014). Under environmental change, these functional differences may impact stress tolerance among genotypic host–symbiont combinations in a population (Parkinson and Baums 2014; Parkinson, Banaszak, et al. 2015), partly explaining why some coral colonies of a given species bleach while others do not, even when sharing the same symbiont species (Goulet et al. 2008; LaJeunesse et al. 2010). Similar fine-scale variation has been observed among maize strains with distinct flavonoid content (Casati and Walbot 2003) and among dinoflagellate strains with distinct toxicities (Yang et al. 2010).

Conclusions

Comparisons among deeply sequenced transcriptomes can reveal the extent and function of molecular variation that is critical to speciation in nonmodel organisms. Such work provides important baselines against which experimentally manipulated samples might be compared and more accurately interpreted. Our data reveal the extent of expression variation that occurs among strains of Symbiodinium and emphasizes how natural selection on existing populations may play a critical role in the response of coral–dinoflagellate symbioses to climate change. The genomic resources described here should improve functional investigations into marine symbiosis biology, particularly as model systems continue to be developed (Baumgarten et al. 2015). Future studies should examine the same strains exposed to different stressors (thermal, osmotic, and/or light), in order to characterize the relationships between physiological and gene expression phenotypes. Each strain should also be brought into an experimental host (e.g., the model Aiptasia [= Exaipatia]) and observed in symbiosis, which would provide insight into how changes in gene expression work to maintain stable cnidarian–dinoflagellate mutualisms. Our findings underscore that important transcriptional differences exist at different taxonomic ranks among dinoflagellates, from clades to species to strains. Future Symbiodinium genomics experiments should be designed such that clade-level questions incorporate different species to serve as a representative sampling of the clade under study, while species-level questions should incorporate distinct strains to serve as a representative sampling of the species under study. Such designs will improve our understanding of Symbiodinium genetic, functional, and phylogenetic diversity.

Supplementary Material

Supplementary tables S1–S3 and figures S1 and S2 are available at Genome Biology and Evolution online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

Mary Alice Coffroth provided cultures, and Maren Ziegler assisted with PRIMER software. This work was supported by the National Science Foundation (DEG-0750756 to J.E.P., OCE-0928764 to T.C.L. and I.B.B.), the Pennsylvania State University, and baseline research funds to CRV by King Abdullah University of Science and Technology (KAUST).

Literature Cited

Casati P, Walbot V. 2003. Gene expression profiling in response to ultra-
 violet radiation in maize genotypes with varying flavonoid content.

 Chapman MA, Hiscock SJ, Filatov DA. 2013. Genomic divergence during

 Chelaifa H, Mahe F, Ainouche M. 2010. Transcriptome divergence be-
 tween the hexaploid salt-marsh sister species Spartina maritima and

 Chen Y, Lun AT, Smyth GK. 2014. Differential expression analysis of com-
 plex RNA-seq experiments using edgeR. In: Nettleton D, Datta S, ed-
 itors. Statistical analysis of next generation sequencing data. Springer

 PRIMER-E. 192pp.

 Dang YK, Green BR. 2010. Long transcripts from dinoflagellate chloroplast

 stability of photosynthetic membranes and fatty acid composition at

 Di Dato V, et al. 2015. Transcriptome sequencing of three Pseudo-nitzschia
 species reveals comparable gene sets and the presence of Nitric Oxide
 Synthase genes in diatoms. Sci Rep. 5:12329


 Erdner DL, Anderson DM. 2006. Global transcriptional profiling of the
 toxic dinoflagellate Alexandrium fundyense using massively parallel
 signature sequencing. BMC Genomics 7:88

 regulation of cab gene: transcription is signaled by the redox state of


 Finney JC, et al. 2010. The relative significance of host-habitat, depth, and
 geography on the ecology, endemism, and speciation of coral endo-

 pretation of thermal tolerance limits and thermal thresholds in tropical

 Variation in symbiont distribution between closely related coral species

 Galardini M, et al. 2011. Exploring the symbiotic pangenome of the nitro-
gen-fixing bacterium Sinorhizobium meliloti. BMC Genomics 12:235


 Gnerre S, et al. 2011. High-quality draft assemblies of mammalian gen-
 omes from massively parallel sequence data. Proc Natl Acad Sci


 Goulet TL, Laduneuex TC, Fabriceus KE. 2008. Symbiont specificity and
 bleaching susceptibility among soft corals in the 1998 Great Barrier

 Grabherr MG, et al. 2011. Full-length transcriptome assembly from RNA-

 Haas BJ, et al. 2013. De novo transcript sequence reconstruction from
 RNA-seq using the Trinity platform for reference generation and anal-

 Hoegh-Guldberg O. 1999. Climate change, coral bleaching and the future


 Jeong HJ, et al. 2014. Genetics and morphology characterize the dinofla-
gellate Symbiodinium voratum, n. sp., (Dinophyceae) as the sole rep-
 resentative of Symbiodinium clade E. J Eukaryot Microbiol. 61:75–94.


 genome microsatellite search and investigation. Bioinformatics
 23:1683–1685.

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

 Lafontaine TC. 2005. “Species” radiations of symbiotic dinoflagellates in
 the Atlantic and Indo-Pacific since the Mocien-Pliocene transition.

 Symbiodinium necroappetens sp. nov. (Dinophyceae): an opportunist
 ‘zooxanthella’ found in bleached and diseased tissues of Caribbean

 Lafontaine TC, Parkinson JE, Reimer JD. 2012. A genetics-based descrip-
tion of Symbiodinium minutum sp. nov. and S. psammophillum sp. nov.
 (Dinophyceae), two dinoflagellates symbiotic with cnidaria. J Phycol.
 48:1380–1391.

 Lafontaine TC, Thornhill DJ. 2011. Improved resolution of reef-coral en-
dosymbiont (Symbiodinium) species diversity, ecology, and evolution

 Lafontaine TC, et al. 2004. Closely related Symbiodinium spp. differ in
 relative dominance in coral reef host communities across environmen-

 Lafontaine TC, et al. 2010. Host-symbiont recombination versus natural
 selection in the response of coral-dinoflagellate symbioses to environ-

 Lafontaine TC, et al. 2014. Ecologically differentiated stress tolerant en-
dosymbionts in the dinoflagellate genus Symbiodinium Clade D are

 Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie

 EST library from the dinoflagellate (Symbiodinium sp.) symbiont of

 and their algal symbiont to thermal stress. PLoS One 6:e26687

 Leggat W, Yellowlees D, Medina M. 2011. Recent progress in

 Lehnert EM, Burriesci MS, Pringle JR. 2012. Developing the anemone
 Aiptasia as a tractable model for cnidian-dinoflagellate symbiosis: the
 transcriptome of aposymbiotic A. pallida. BMC Genomics


 162:551–569.

 Lin SJ, et al. 2015. The Symbiodinium kawagutii genome illuminates din-

 McCarthy DJ, Chen Y, Smyth GK. 2012. Differential expression analysis of
 multifactor RNA-Seq experiments with respect to biological variation.

 McGinley MP, Suggett DJ, Warner ME. 2013. Transcript patterns of chlo-
 roplast-encoded genes in cultured Symbiodinium spp. (Dinophyceae):
testing the influence of a light shift and diel periodicity. J Phycol.

 McGinley MP, et al. 2012. Trancriptional response of two core photosys-
tem genes in Symbiodinium spp. exposed to thermal stress. PLoS One
 7:e50439

 Symbiodinium species in the Sargasso Sea. Proc Natl Acad Sci
 U S A. 111:15245–15250.

 Symbiodinium species in the Sargasso Sea. Proc Natl Acad Sci
 U S A. 111:15245–15250.


Thornhill DJ, Kemp DW, Bruns BU, Fitt WK, Schmidt GW. 2008. Correspondence between cold tolerance and temperate biogeography in a Western Atlantic Symbiodinium (Dinophyta) lineage. J Phycol. 44:1126–1135.


Associate editor: Eric Bapteste