A peer-reviewed version of this preprint was published in PeerJ on 15 February 2016.

View the peer-reviewed version (peerj.com/articles/1616), which is the preferred citable publication unless you specifically need to cite this preprint.

A transcriptome resource for the coral, *Orbicella faveolata* (Scleractinia-Meruliniidae) - an emerging model of coral innate immunity

David A Anderson, Marcus E Walz, Ernesto Weil, Peter Tonellato, Matthew C Smith

Research on climate change-driven coral disease outbreaks has greatly advanced our understanding of the coral innate immune system, and coral genomics has provided insight into the evolution of metazoan immune systems. However, only recently has next generation sequencing (NGS) technology been used to investigate coral immune responses to environmental stress and disease. In the present investigation, tissue of the threatened Caribbean reef-building coral, *O.* *faveolata*, was collected during a warm water thermal anomaly and coral disease outbreak in 2010 in Puerto Rico. Multiplex sequencing of messenger RNA that was extracted from tissue samples of diseased, bleached and apparently healthy coral colonies was conducted on the Illumina GAIIx platform, and total reads were pooled accordingly for de novo assembly into contigs by Trinity software. From 70,745,177 raw short-sequence reads, 32,463 contigs were assembled in silico and taxonomically identified as *O.* *faveolata* in origin. The resulting reference transcriptome was annotated with gene ontologies, and sequences were mapped to KEGG pathways. The reference transcriptome presented here expands upon the genetic data currently available for *O.* *faveolata*. By sampling coral colonies affected by infectious disease and thermal stress in situ, it was possible to enrich the data set for genes that are expressed under natural conditions as part of the coral innate immune response. The results of the present investigation advance our understanding of the coral immune system and expand upon bioinformatic data available for *O.* *faveolata* – a threatened coral species that is widely studied as a model for immune responses to climate-driven stress and disease.
Characterization of a transcriptome resource for the study of innate immunity in the Caribbean reef-building coral *Orcicella faveolata* (Scleractinia-Merulinidae)

David A. Anderson¹,², Marcus Walz³, Ernesto Weil², Peter Tonellato³,⁴ & Matthew C. Smith⁵

1. Department of Marine Sciences
   University of Puerto Rico at Mayagüez
   Mayagüez, PR 00681

2. Department of Pathology and Immunology
   Washington University School of Medicine
   St. Louis, Missouri

3. Joseph J. Zilber School of Public Health
   University of Wisconsin-Milwaukee
   Milwaukee, WI 53201

4. Center for Biomedical Informatics
   Harvard Medical School
   Harvard University
   Boston, MA 02115

5. School of Freshwater Sciences
   University of Wisconsin-Milwaukee
   Milwaukee, WI 53201
Abstract

Research on climate change-driven coral disease outbreaks has greatly advanced our understanding of the coral innate immune system, and coral genomics has provided insight into the evolution of metazoan immune systems. However, only recently has next generation sequencing (NGS) technology been used to investigate coral immune responses to environmental stress and disease. In the present investigation, tissue of the threatened Caribbean reef-building coral, *Orcicella faveolata*, was collected during a warm water thermal anomaly and coral disease outbreak in 2010 in Puerto Rico. Multiplex sequencing of messenger RNA that was extracted from tissue samples of diseased, bleached and apparently healthy coral colonies was conducted on the Illumina GAIIx platform, and total reads were pooled accordingly for *de novo* assembly into contigs by Trinity software. From 70,745,177 raw short-sequence reads, 32,463 contigs were assembled *in silico* and taxonomically identified as *O. faveolata* in origin. The resulting reference transcriptome was annotated with gene ontologies, and sequences were mapped to KEGG pathways. The reference transcriptome presented here expands upon the genetic data currently available for *O. faveolata*. By sampling coral colonies affected by infectious disease and thermal stress *in situ*, it was possible to enrich the data set for genes that are expressed under natural conditions as part of the coral innate immune response. The results of the present investigation advance our understanding of the coral immune system and expand upon bioinformatic data available for *O. faveolata* – a threatened coral species that is widely studied as a model for immune responses to climate-driven stress and disease.

Keywords: RNA-seq, non-model organism, coral, Cnidaria, innate immunity, disease, transcriptome, Wnt, Dicer
Introduction

Coral reefs of the Caribbean and Atlantic region have been experiencing a dramatic, climate change-driven decline in coral cover and reef biodiversity since as early as the 1970s as a result of massive coral bleaching events and disease outbreaks (Altizer et al. 2013; Antonius 1973; Bruno et al. 2007; Ducklow & Mitchell 1979; Glynn et al. 1985; Glynn 1983; Harvell et al. 1999; Weil & Rogers 2011; Weil et al. 2006). Investigations of molecular, cellular and humoral immunity in corals have led to the development of basic models of coral immune responses to disease and environmental stress (Palmer & Traylor-Knowles 2012). Climate change-driven stress can lead to disease outbreaks by altering coral-associated microbes, shifting communities from symbiont and commensal-dominated to pathogen-dominated (Bourne et al. 2009; Rosenberg et al. 2007). The basic model of the coral immune responses to disease involves the migration of pluripotent immunocytes, also known as amoebocytes, to physical wounds and disease lesions; the production of cytotoxic reactive oxygen species; the production of antioxidants to reduce self-harm; the accumulation of melanin as a barrier to pathogen invasion; and the production of antimicrobial compounds to regulate commensal microbiota (Mydlarz et al. 2008; Palmer et al. 2009; Tchernov et al. 2011; Vidal-Dupiol et al. 2011; Weis 2008). Failure to overcome an infection leads to the manifestation of lesions, which are associated with genetic and cellular signatures of stress and apoptosis, and the advancement of lesions and tissue mortality across coral colonies (Ainsworth et al. 2007; Anderson & Gilchrist 2008).

Despite an increased focus on coral immunity to better understand declines in coral communities across the globe, significant knowledge gaps remain with regard to physiological responses of coral to disease and environmental stress (Palmer & Traylor-Knowles 2012; Pollock et al. 2011). To that end, next generation sequencing (NGS) technologies promise to reveal the
genetic mechanisms that control the coral immune system on a whole-genome and whole-transcriptome scale. In addition, the versatility NGS allows for a wide variety of physiological responses to natural disease and climate stress events to be detected from samples collected and preserved in situ. To that end, several investigations have reported the use of NGS to detect differential expression of genes with putative roles in the regulation of the coral innate immune system in response to environmental stress and disease (Burge et al. 2013; Libro et al. 2013; Ocampo et al. 2015; Pinzón et al. 2015; Weiss et al. 2013).

Populations of *O. faveolata* have been severely impacted by coral bleaching and Caribbean yellow band disease (CYBD) across their geographic distribution in the Caribbean and Atlantic (Bruckner 2012; Bruckner & Hill 2009; Weil et al. 2009a; Weil et al. 2009b; Weil & Rogers 2011). Observed declines in this species have thus led to its recent classification as a threatened species under the federal endangered species act (53851 2014). In an effort to develop this species into a model for coral physiology under natural conditions, transcriptomics have been used to define changes in gene expression of this coral in response to environmental stress during larval development, the establishment of symbiosis, and the maintenance of homeostasis (Cróquer et al. 2013; Desalvo et al. 2008; Kimes et al. 2010; Pinzón et al. 2015; Schwarz et al. 2008; Voolstra et al. 2009). Most recently, Pinzón and colleagues (2015) used NGS to track temporal changes in gene expression of *O. faveolata* through a warm water thermal anomaly and bleaching event in 2010 in Puerto Rico. The present investigation supplements this work by using NGS to produce a reference transcriptome from tissue samples of *O. faveolata* infected with Caribbean yellow band disease (CYBD) that were collected during the same thermal anomaly, bleaching event and disease outbreak in Puerto Rico in 2010. The quality and completeness is assessed by comparison to current bioinformatics resources for other model
Cnidarians. By sequencing mRNA extracted from disease lesions and healthy areas of diseased colonies, enrichment for immunity-related genes was accomplished. Genes with putative roles in innate immune signaling pathways were assembled into gene sets. Finally, a set of these putative immunity genes and associated pathways were described in detail to support the further development of this model species of coral innate immune system.

Methods

Sample Collection

The concurrent thermal anomaly, coral bleaching event and Caribbean yellow band disease outbreak of 2010 in Puerto Rico provided a unique opportunity to sample colonies of *O. faveolata* affected by multiple environmental stressors that are known to induce innate immune responses (Mydlarz et al. 2009; Pinzón et al. 2015). All samples (25 cm²) were collected on a single dive at 10 m depth in October 2010 on Media Luna reef in La Parguera, Puerto Rico (17° 56.091 N, 67° 02.577 W). Samples were collected under a general collecting permit issued by the Department of Natural Resources of Puerto Rico to the Department of Marine Sciences at the University of Puerto Rico at Mayaguez. Reefs in this region experienced ten degree-heating weeks at the time of sample collection (Supp. Info. 1). Degree-heating weeks is a remote sensing metric that estimates accumulated thermal stress in corals during sea surface temperature anomalies (Gleeson & Strong 1995), and is reported by the National Oceanic and Atmospheric Administration. Five different health conditions were sampled from three separate colonies based on visual diagnoses made in the field: 1) bleached and asymptomatic tissue of a partially bleached colony; 2) asymptomatic tissue and lesion tissue from a CYBD-affected colony; and 3) tissue from a completely asymptomatic colony. It was assumed that the three colonies were non-clonal given their large distances of separation (>10 m), and the low levels of clonality (3.5%)
previously documented for the same species on the same reef (Severance & Karl 2006).

Photographic examples of each disease condition are presented in Figure 1. Within one hour of collection, samples were transported to the Department of Marine Sciences on Isla Magueyes, flash frozen in liquid nitrogen and stored at -80°C.

The Resource

RNA Extraction, mRNA Sequencing, and de novo Transcriptome Assembly

For every 1.00 cm² of sample area, 2.00 ml of Trizol (Life Technologies, CA, USA) was used to homogenize the coral tissue in 50 ml capped vials by vigorous shaking. A neutralization reaction occurs between the calcium carbonate coral skeleton and the acidic Trizol. Therefore, on removal of the homogenate to a new vial 1 to 5 µl of 6 M hydrochloric acid were added to each sample to minimize DNA contamination of the aqueous phase on the addition of chloroform. Total RNA was extracted from the acidified homogenate according to the manufacturer’s protocol. An additional chloroform extraction was performed, as required for samples with high levels of carbohydrate contaminants (i.e. coral mucus). The aqueous phase was removed and 700 µl were loaded onto RNeasy columns (Qiagen, Velmox Netherlands), and the purification of total RNA was performed according to the manufacturer’s protocols. A DNA digestion step was also performed using the same manufacturer’s on-column DNase treatment according to the manufacturer’s protocol. Following elution, RNA quality was verified by 2% denaturing agarose gel electrophoresis, and a NanoDrop spectrophotometer (Thermo Scientific, USA) was used to determine 260/280 nm ratios and total RNA concentrations. Total RNA was sent for mRNA multiplex sequencing on the Illumina GAIIx platform (Illumina Inc, USA) according to standard protocols of the genomic core facility at the Scripps Research Institute in Ft. Pierce, Florida, USA. Raw sequence reads from each sample are available at the NCBI Short Read Archive.
under Bioproject PRJNA236103. From 70,745,177 raw sequence reads of 72 nucleotide bases (bp) in length, adaptor sequences and low quality bases were trimmed and clipped using cutadapt (Martin 2011), which resulted in 59,114,519 reads with a mean length of 67 bp (standard deviation 5 bp). The Trinity software suite was chosen for de novo assembly of the raw sequencing reads (Grabherr et al. 2011; Haas et al. 2013), a tool that has been used for a variety of coral species (Ocampo et al. 2015; Pinzón et al. 2015; Pooyaei Mehr et al. 2013). Assembly of the trimmed and clipped reads yielded a metatranscriptome of 166,032 trinity transcripts (i.e. total number of assembled sequences) and 153,045 trinity components (i.e. total number of sequence contigs). Contigs were translated into a predicted proteome and compared to predicted proteomes of other model Cnidarians, N. vectensis and A. digitifera.

Taxonomic Identification of Metatranscriptome Sequences

To identify assembled transcripts originating from O. faveolata among the microbial sequences of the metatranscriptome, BLASTn was conducted against existing O. faveolata ESTs (n=29,334), the Nematostella vectensis complete genome in NCBI, and complete transcriptome of Acropora digitifera archived by the marine genomics unit of the Okinawa Institute of Science and Technology (Putnam et al. 2007). The same procedure was used to identify transcripts from the photosynthetic endosymbiont, Symbiodinium spp., using recently reported reference transcriptomes (Bayer et al. 2012). Sequences with no hits to aforementioned reference transcriptomes were identified by the classifier for metagenomic sequences (CLaMS) (Pati et al. 2011), using the complete genome of N. vectensis and whole transcriptome of Symbiodinium as training sets. Sequences that could not be identified by these methods or that were classified differently in the CLaMS analysis were removed from the assembled transcriptome. The O. faveolata-specific reference transcriptome is provided as a supplementary file (Supp. Info. 2).
The Framework

Transcriptome Annotation, Pathway Analysis, and Predicted Proteome

To assess the accuracy and quality of the assembled reference transcriptome, alignments of individual contigs to Sanger sequencing-generated *O. faveolata* expressed sequence tags (ESTs) in NCBI were conducted (Supp. Info. 3a) [16-18]. A complete genome has not been generated for *O. faveolata*, so the relative completeness of the *O. faveolata* reference transcriptome was assessed by comparisons to whole genome and transcriptome sequence data for the coral, *A. digitifera*, and the sea anemone, *N. vectensis* (Dunlap et al. 2013; Putnam et al. 2007). Transcriptome annotation was conducted using a combination of methods. First, gene ontologies were assigned to each contig by extracting GO annotations from best BLASTx alignments and EMBL InterProScan using the BLAST2GO platform (Boratyn et al. 2013; Conesa et al. 2005; Quevillon et al. 2005). Taxonomic identities of the top BLASTx hits to Swissprot protein sequences were used for a post-hoc analysis of the methods used to separate coral and *Symbiodinium* metatranscriptome sequences. With the lack of complete genomes for many coral models, pathway analysis is commonly used to assess the completeness of coral reference transcriptomes (Meyer et al. 2009; Polato et al. 2011). To accomplish this, *O. faveolata* sequences were annotated to KEGG orthologues (KO) using the KEGG Automatic Annotation Server (KAAS) (Moriya et al. 2007). To determine the completeness of this resource relative to other models, nucleotide collections in NCBI for the following species were annotated with KOs in parallel: *O. faveolata* ESTs (Supp. Info 3a, NCBI, n=29,334), transcriptome shotgun assemblies of *A. millepora* (Supp. Info 3b, NCBI TSA, n=101,726), and *N. vectensis* RNA sequences (Supp. Info 3c, NCBI RNA, n=25,131). Differences in the number of KEGG pathway components for each nucleotide collection were compared.
We have selected multiple gene families and associated signaling pathways to highlight the utility of the *O. faveolata* transcriptome and establish a framework for future research on the regulation of coral immune responses to disease. *O. faveolata* Wnt-like proteins are described and analyzed phylogenetically for the first time (Supp. Info. 4). Whole-length sequences were aligned by MUSCLE (Edgar 2004), conserved regions were curated using Gblocks (Talavera & Castresana 2007), maximum likelihood phylogenies were estimated using PhyML with 100 bootstraps and the WAG substitution model (Guindon et al. 2010), and trees were constructed using TreeDyn (Chevenet et al. 2006). This pipeline was executed using the Phylogeny.fr platform (Dereeper et al. 2008). Predicted Wnt protein sequences for *A. millepora* (Amil), *N. vectensis* (Nvec), and *Aiptasia pallida* (Apal) were downloaded from NCBI, and *A. digitifera* sequences were downloaded from Zoophytebase [30, 40] (Supp. Info. 5). Correct open reading frames were verified by hmmscan in the HMMER web server (Finn et al. 2011). The utility of the *O. faveolata* reference transcriptome to guide future immunological research was further demonstrated by using hmmscan to characterize the domain architecture of a Dicer-1-like predicted protein that spans 1245 amino acid residues. To support the hypothetical roles of Wnt and Dicer-like proteins in the coral innate immune system, the presence or absence of essential pathway components were identified by BLASTp alignments of predicted coral protein sequences against homologous proteins curated in KEGG Pathway map04310 and published literature (Ding & Voinnet 2007; Moran et al. 2013). Similarly, putative KOs are identified for the RIG-like receptor signaling pathway (KEGG map04622) to develop a model for coral host-virus interactions.

**Results and Discussion**
The reference transcriptome described here was assembled by first combining sequence reads that were produced separately for each RNA sample. Modifications of this general method are commonly used to maximize the diversity and coverage of expressed genes in the final reference transcriptome (Burge et al. 2013; Lehnert et al. 2012; Polato et al. 2011). In addition to providing a resource for future research on *O. faveolata* molecular physiology, the present investigation aimed to survey the repertoire of *O. faveolata* transcripts that have evolutionarily conserved sequences and putative functions in regulating metazoan innate immunity. From the reference transcriptome, gene sets were assembled based on the presence or absence of components of innate immune signaling pathways. Genes and pathways that have not been investigated in the context of coral immune responses to disease are described in depth to support the expansion of models of coral innate immunity.

**The Resource**

*De Novo Assembly, Taxonomic Identification, and Predicted Proteome*

Alignments of the assembled transcriptome against *N. vectensis* and *Symbiodinium* sequences resulted in the successful classification of 35,967 and 47,760 trinity transcripts as *O. faveolata* and *Symbiodinium*, respectively. However, 3504 transcripts were identified as common to both the coral and *Symbiodinium* transcriptomes after the CLaMS analysis. Therefore, those sequences were removed from both transcriptomes, resulting in 32,463 trinity transcripts classified as coral and 44,256 classified as *Symbiodinium*. In this report we only focus on sequences that make up the coral reference transcriptome (Supp. Info. 2).

The size of the coral assembly is similar to previous studies that have used the Illumina GAII platform, which report between 33,000 and 48,000 coral transcripts (Barshis et al. 2013; Libro et al. 2013). The respective GC content for the final coral transcriptome was 44%, which is
comparable to previous reports for corals (Sabourault et al. 2009; Soza-Ried et al. 2010; Vidal-Dupiol et al. 2013). Lastly, the frequency distribution of taxonomic classifiers assigned by BLASTx hits to Swissprot sequences reveals most frequent hits to the *N. vectensis* proteome, followed by other metazoans (Figure 2). Collectively, these results suggest that the methods chosen for assigning sequences to the coral host were sufficiently robust to exclude microbial sequences within the metatranscriptome.

To date, NGS investigations of corals use a diversity of methods to assess the quality of transcriptome assembly and sequence accuracy. The lack of widely applied quality metrics, such as those proposed for model organisms [32], is limited for corals due to the absence of whole reference transcriptomes or genomes for most coral species. In the present investigation, a comparative approach was adopted to assess the quality of the reference transcriptome. The N50 of the *O. faveolata* transcripts was 1736 bp, which is similar to a recent study that used Trinity to assemble a coral transcriptome (Pooyaei Mehr et al. 2013). This value is higher than many other transcriptome sequencing studies of Cnidarians (Barshis et al. 2013; Burge et al. 2013; Lehnert et al. 2012; Libro et al. 2013; Sun et al. 2012), but does not exceed maximums of the most recent reports (Moya et al. 2012; Pinzón et al. 2015; Shinzato et al. 2014). This demonstrates that the raw sequence reads and computational methods were robust in the assembly of the reference transcriptome.

Our analysis of the reference transcriptome and predicted proteome shows that they represent a significant expansion of the bioinformatic data available for *O. faveolata* on par with other Cnidarian models (Dunlap et al. 2013; Putnam et al. 2007; Shinzato et al. 2011) (Figure 3). BLASTp alignments of predicted *O. faveolata* and *A. digitifera* proteins to homologous proteins of the *N. vectensis* reference proteome showed similar levels of sequence conservation and
percent identity (Supp. Info. 6). The predicted proteome for *O. faveolata* is provided as a supplementary data file (Supp. Info. 7). To assess the sequence accuracy of high-throughput mRNA sequencing, transcript sequences were aligned to over 12,000 *O. faveolata* ESTs in NCBI, which were derived by Sanger sequencing of cloned mRNA. An average of 96% identity for aligned sequences greater than 300 bp suggests a high level of sequence accuracy (Supp. Info. 8).

The Framework

Annotation of the Reference Transcriptome

Pathway analysis revealed a significant coverage of KEGG metabolic, protein complex and immune signaling pathways compared with the *O. faveolata* ESTs present in NCBI (Table 1). The *O. faveolata* reference transcriptome has 1.6-fold greater coverage of metabolic and protein complex pathways, and a notable 2.3-fold enrichment of components of innate immunity-related pathways. This significantly expands the availability of annotated sequences to support future studies of the *O. faveolata* immune system at the genetic level. Coverage of metabolic and protein complex-related pathways is similar to the *N. vectensis* and *A. millepora* data sets. Comparison of the entire collection of KEGG pathways to which *O. faveolata* transcripts were annotated revealed a high relative level of completeness in comparison to the *N. vectensis* reference proteome (Supp. Info. 9, 10, 11). It is important to note that these completeness metrics are a “best estimate” in the absence of a complete reference genome or reference transcriptome derived from a variety of physiological conditions. Of the 32,463 coral sequences that make up the reference transcriptome, 20,861 were associated with one or more gene ontology terms by BLAST2GO (Supp. Info. 12). A summary of the most abundant gene ontology terms annotated to at least 100 different sequences is presented in Figure 4.
Molecular Physiology of Coral Innate Immunity

Models of coral immune responses to disease are generally characterized by inflammation that involves the production of ROS, production of antioxidants to reduce self-harm, the migration of phagocytic cells to sites of infection, the accumulation of melanin to prevent the spread of infection, and the production of antimicrobial compounds (Palmer & Traylor-Knowles 2012). In the arms race between invading pathogens and coral host resistance, a breakdown of host homeostasis leads to the activation of apoptosis and ultimately tissue mortality (Weis 2008). Components of the coral immune system can be organized into 4 levels within signaling pathway schemes: (1) pattern recognition receptors, (2) downstream signaling cascades, (3) inflammatory cytokine expression, and (4) effector mechanisms (Palmer & Traylor-Knowles 2012). Functional annotations of the O. faveolata reference transcriptome demonstrate full representation of these four basic levels of the coral innate immune system, as demonstrated by mapping KOs to KEGG pathways (Table 1) (Supp. Info. 11).

Some of the genes that code for components of developmental pathways, such as Wnt and Notch, have been studied in depth from an evolutionary and developmental perspective in Cnidarians (Käsbauer et al. 2007; Kusserow et al. 2005; Marlow et al. 2012; Miller et al. 2005; Radtke et al. 2010). It is also known that these genes have conserved roles in disease etiology across metazoan phyla (Duncan et al. 2005; Radtke et al. 2010; Staal et al. 2008). However, few studies have explored hypotheses about the role of developmental genes in regulating coral immune responses to disease. If one considers that pluripotent, phagocytic immune cells (i.e. amoebocytes) are at the front lines of wound healing and pathogen removal in corals, developmental pathways required for cell migration and differentiation are likely necessary to maintain host resistance to disease. Therefore, components of developmental signaling pathways
present in the *O. faveolata* reference transcriptome should be researched further. To that end, the
present investigation identifies five *O. faveolata* transcripts with predicted protein sequences that
are homologous to Wnt proteins. This family of proteins was first characterized for Cnidarians in
the model sea anemone, *N. vectensis* (Kusserow et al. 2005). Homologs are also present in the
transcriptomes of the corals *A. millepora* (Kortschak et al. 2003) and *A. digitifera* (Shinzato et al.
2011), and the sea anemone, *A. pallida* (Lehnert et al. 2012). In addition, 66 of the 75 transcripts
that putatively code for components of the KEGG Wnt pathway (map04310) were identified *O.
faveolata* reference transcriptome (Supp. Info. 11) (Figure 5). Twelve Wnt proteins were
originally identified in the *N. vectensis* genome. However, the extent to which Wnt proteins are
restricted to specific developmental life stages or environmental stimuli is not well understood.
Since the reference transcriptome for *O. faveolata* was derived from tissue from stressed
colonies, the set of *O. faveolata* Wnt proteins described here may only represent a subset of the
complete repertoire expressed under different environmental conditions and developmental life
stages.

Some of the innate immune signaling pathways that integrate the four levels of immunity
(pattern recognition receptors, downstream signaling cascades, inflammatory cytokine
expression, and effector mechanisms) that were enriched in the *O. faveolata* reference
transcriptome include: toll-like receptor (TLR), mitogen activated protein kinase (MAPK),
nuclear factor kappa beta (NF-κB), and apoptosis pathways. Although a wide variety of innate
immune signaling pathways have been cited as important in coral innate immunity, mechanistic
experiments to confirm the function of such pathways remains limited. MAPKs make up a
highly conserved family of protein kinases that regulate immune signaling pathways (Chen et al.
2001; Schwarz et al. 2008). In *O. faveolata*, differential expression under stress has been
reported for various members of MAPK signaling cascades (Schwarz et al. 2008; Voolstra et al. 2009), however, their specific function in corals has not been clearly defined. In sponges, it has been demonstrated mechanistically that MAPK signaling is activated by exposure of tissue to bacteria-derived lipopolysaccharide (Boehm et al. 2001). To regulate host-associated microbiota in *Hydra*, Toll-like receptor signaling has been shown to activate immune-related MAPK signaling cascades by MyD88-dependent phosphorylation of Jun-kinase (JNK), which in turn regulates the expression of pro-inflammatory transcription factor, NF-κB, and the pro-survival factor, Bcl-2 (Franzenburg et al. 2012).

Less cited immune signaling pathways in Cnidarians include Nod-like receptor (NLR), Rig-like receptor (RLR), and Dicer-1-like signaling, which all have conserved roles in intracellular pathogen recognition (MacKay et al. 2014; Ting et al. 2008; Yoneyama & Fujita 2007). Although genes with sequence homologies to components of NLR signaling pathways have been recently identified in sponges and Cnidarians, there is limited knowledge about their function in regulating coral innate immunity (Augustin et al. 2010; Bosch 2012; Hamada et al. 2013; Yuen et al. 2014). Preliminary models in *Hydra* suggest that NLRs are essential in regulating caspase-mediated cell death upon infection by bacterial pathogens (Bosch et al. 2011).

The *O. faveolata* reference transcriptome reported here expands the depth of coverage of components of NLR signaling that can be used to characterize the role of this pathway in bacterial-driven coral diseases in *O. faveolata* (Table 1).

To date, interactions between the coral host and associated microbes have focused on coral-fungal and coral-bacterial interactions (Kim et al. 2000; Kvennefors et al. 2008; Sutherland et al. 2004). However, recent investigations have demonstrated the important roles that viruses play in the coral immune system, from pathogenesis to symbiosis (Atad et al. 2012; Barr et al. 2013).
Therefore, future research should focus on the genetic mechanisms of coral-virus interactions. To that end, the present investigation reports a comprehensive set of components of the Dicer-1-like signaling pathway. Dicer-like proteins are endoribonucleases, and are essential members of the micro RNA (miRNA) and small interfering RNA (siRNA) pathways. These pathways have only recently gained attention in corals, and they may have conserved roles in antiviral immunity (Liew et al. 2014; MacKay et al. 2014; Moran et al. 2013). A full-length Dicer-1-like protein (4335 amino acids) and all of the essential components for siRNA and miRNA were identified in the O. faveolata predicted proteome. These components and the domain architecture of Dicer-1 are shown in Figure 6 (Supp. Info 12). Similar to Dicer-1, Rig-I-like receptors (RLR) have conserved roles in viral nucleic acid recognition. The present study reports components of RLR signaling pathways (Table 2) (Guo et al. 2013; Mukherjee et al. 2014). The data presented here can be used in future mechanistic investigations to elucidate the role of the O. faveolata innate immune system in regulating coral-virus interactions, and whether they are symbiotic, commensal or pathogenic.

Conclusions

By highlighting pathways and gene families with possible roles in coral innate immunity, the present study sets the stage for exploring new topics in coral innate immunity and further developing O. faveolata as a model for coral molecular physiology. Next generation sequencing of mRNA from thermally-stressed and diseased colonies collected during a natural stress event revealed an enrichment of genes with putative functions in regulating innate immunity. The quality of the reference transcriptome parallels those currently available for other Cnidarian models. The abundance of full-length sequences provides opportunities to study gene families...
with putative roles in the regulation of coral innate immunity in more depth. As NGS technologies become more accessible for the investigation of non-model organisms, and as open-access systems biology tools become integrated into current high-throughput computational pipelines, rapid advancements can be made in our understanding of the genetic factors that control coral physiology. The present investigation contributes to rapidly advancing body of work that has the ultimate goal of understanding the evolution of metazoan immune systems and applying this knowledge to coral reef conservation.

Acknowledgements

MCS received startup funds from the Department of Marine Science at the University of Puerto Rico at Mayaguez to support this work. This material is based upon work supported by the National Science Foundation (NSF) Graduate Research Fellowship under Grant No. DGE-1143954 to DAA. Partial funding was provided by NSF grants IOS # 1017510 and OCE - 1105143 to EW.

References


Bosch TCG. 2012. What hydra has to say about the role and origin of symbiotic interactions. The Biological Bulletin 223:78-84.


Yuen B, Bayes JM, and Degnan SM. 2014. The characterization of sponge NLRs provides insight into the origin and evolution of this innate immune gene family in animals. *Molecular biology and evolution* 31:106-120.
Figure 1 (on next page)

*O. faveolata* combined CYBD outbreak and BLE event in La Parguera, PR.

Asymptomatic colony, (B) multiple lesions of CYBD, (C) partial bleached colonies, (D) completely bleached colonies, (E) colony affected by BLE and CYBD. Photos by E. Weil, first published under creative commons license (CC BY) in Anderson et al. [95].
**Figure 2** (on next page)

Taxonomic hits.

Top hits to taxa in the Swissprot sequence database for the annotated *O. faveolata* reference transcriptome.
Figure 3 (on next page)

Comparative analysis of predicted proteomes.

(A) Frequency distribution of protein lengths in intervals of 50 amino acids (a.a.). (B) Box-Whisker plot of predicted protein lengths. (C) Summary statistics of predicted protein lengths.
C

<table>
<thead>
<tr>
<th>Species</th>
<th>Mean Length</th>
<th>Median Length</th>
<th>Maximum Length</th>
<th>Percentile 95th</th>
<th>Percentile 99th</th>
</tr>
</thead>
<tbody>
<tr>
<td>N. vectensis</td>
<td>336</td>
<td>253</td>
<td>8745</td>
<td>882</td>
<td>1571</td>
</tr>
<tr>
<td>O. faveolata Ref.</td>
<td>297</td>
<td>207</td>
<td>4493</td>
<td>791</td>
<td>1279</td>
</tr>
<tr>
<td>O. faveolata EST</td>
<td>171</td>
<td>166</td>
<td>278</td>
<td>252</td>
<td>262</td>
</tr>
<tr>
<td>A. digitifera</td>
<td>183</td>
<td>125</td>
<td>3458</td>
<td>517</td>
<td>904</td>
</tr>
</tbody>
</table>
Figure 4 (on next page)

GO terms represented in the coral reference transcriptome

Select gene ontology terms associated with at least 100 different sequences.
Figure 5 (on next page)

Radial cladogram of maximum likelihood analyses of *O. faveolata* Wnt-like protein sequences.

Internal nodes where Wnt-specific clades diverge are labeled with support values from 100 bootstrap trees. Species included in the analyses are *O. faveolata* (Ofav), *N. vectensis* (Nvec), *A. pallida* (Apal), *A. digitifera* (Adig), and *A. millepora* (Amil).
Figure 6 (on next page)

Dicer-1-like domain architecture and pathway components

(A) Domain architecture of Dicer-1-like proteins. Scale bar represents 100 amino acid (a.a.) residues. Domain architecture was determined by Hmmscan [53]. Sequence accession numbers and alignment results are presented in S.I. 7. (B) Pathway components of miRNA and siRNA required for antiviral immunity. BLASTp e-values < 1e-5 and validation by Hmmscan.
A

Dicer-1 Domain Architecture

M. faveolata
A. millepora
H. vulgaris
S. purpuratus
D. melanogaster

Conserved Domains
PAZ
Dicer
Helicase C
Ribonuclease 3

100 a.a. residues

B

<table>
<thead>
<tr>
<th>Accession</th>
<th>Protein Name</th>
<th>O. faveolata Homologues</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q32KD4</td>
<td>Argonaute-1</td>
<td>comp43840, comp33938, comp46773, comp64615</td>
</tr>
<tr>
<td>Q9VCU9</td>
<td>Dicer</td>
<td>comp46340</td>
</tr>
<tr>
<td>Q7KNF1</td>
<td>Drosha</td>
<td>comp42908</td>
</tr>
<tr>
<td>U3MIH3</td>
<td>GW182/TNRC6</td>
<td>comp34495, comp51031</td>
</tr>
<tr>
<td>Q7K175</td>
<td>HEN1</td>
<td>comp39200</td>
</tr>
<tr>
<td>U3MH39</td>
<td>HYL1</td>
<td>comp38270</td>
</tr>
<tr>
<td>Q9V9V7</td>
<td>Pasha/DGCR8</td>
<td>comp23275</td>
</tr>
<tr>
<td>Q9V9K7</td>
<td>Serrate/Ars2</td>
<td>comp38150</td>
</tr>
<tr>
<td>Q9VLW8</td>
<td>R2D2</td>
<td>comp23275</td>
</tr>
<tr>
<td>Q6J5K9</td>
<td>Armitage</td>
<td>comp43557</td>
</tr>
<tr>
<td>G5EBQ3</td>
<td>RRF/RDRP/EGO</td>
<td>comp40697</td>
</tr>
<tr>
<td>O44406</td>
<td>ERI-1</td>
<td>comp39615</td>
</tr>
</tbody>
</table>
Table 1 (on next page)

KAAS Pathway Analysis of *O. faveolata* reference transcriptome.

RNA sequences for *N. vectensis*, reference transcriptome of *A. millepora*, and *O. faveolata* ESTs present in NCBI prior to the present study were annotated by KAAS. The *O. faveolata* reference transcriptome of the present investigation was compared to the other datasets as percentage of components for each pathway relative to the other data sets.
<table>
<thead>
<tr>
<th>Metabolism</th>
<th>N. vectensis</th>
<th>A. millepora</th>
<th>M. faveolata</th>
<th>Presented Reference Transcriptome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glycolysis &amp; Gluconeogenesis</td>
<td>27</td>
<td>32</td>
<td>27</td>
<td>Prior hits</td>
</tr>
<tr>
<td>Pentose Phosphate</td>
<td>17</td>
<td>20</td>
<td>18</td>
<td>12</td>
</tr>
<tr>
<td>Citrate Cycle</td>
<td>23</td>
<td>23</td>
<td>20</td>
<td>18</td>
</tr>
<tr>
<td>Biosynthesis of Amino Acids</td>
<td>54</td>
<td>52</td>
<td>49</td>
<td>34</td>
</tr>
<tr>
<td>Valine, Leucine and Isoleucine Degradation</td>
<td>33</td>
<td>34</td>
<td>32</td>
<td>20</td>
</tr>
<tr>
<td>Purine Metabolism</td>
<td>91</td>
<td>105</td>
<td>88</td>
<td>55</td>
</tr>
<tr>
<td>Fatty Acid Metabolism</td>
<td>27</td>
<td>26</td>
<td>23</td>
<td>19</td>
</tr>
<tr>
<td>Pyrimidine Metabolism</td>
<td>66</td>
<td>75</td>
<td>62</td>
<td>34</td>
</tr>
<tr>
<td><strong>Protein Complexes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spliceosome</td>
<td>96</td>
<td>106</td>
<td>99</td>
<td>74</td>
</tr>
<tr>
<td>Ribosome</td>
<td>107</td>
<td>119</td>
<td>99</td>
<td>97</td>
</tr>
<tr>
<td>Protein Export</td>
<td>20</td>
<td>21</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>Oxidative Phosphorylation</td>
<td>79</td>
<td>87</td>
<td>85</td>
<td>74</td>
</tr>
<tr>
<td>RNA Degradation</td>
<td>51</td>
<td>55</td>
<td>43</td>
<td>23</td>
</tr>
<tr>
<td>Ubiquitin Proteolysis</td>
<td>81</td>
<td>95</td>
<td>84</td>
<td>49</td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td>140</td>
<td>96</td>
<td>89</td>
<td></td>
</tr>
<tr>
<td><strong>Standard Deviation</strong></td>
<td>29</td>
<td>8</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td><strong>Stress and Immunity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAPK</td>
<td>62</td>
<td>105</td>
<td>107</td>
<td>43</td>
</tr>
<tr>
<td>Ras</td>
<td>50</td>
<td>80</td>
<td>86</td>
<td>39</td>
</tr>
<tr>
<td>Wnt</td>
<td>47</td>
<td>63</td>
<td>57</td>
<td>29</td>
</tr>
<tr>
<td>Notch</td>
<td>10</td>
<td>20</td>
<td>22</td>
<td>9</td>
</tr>
<tr>
<td>Phagosome</td>
<td>45</td>
<td>57</td>
<td>59</td>
<td>35</td>
</tr>
<tr>
<td>Peroxisome</td>
<td>53</td>
<td>54</td>
<td>49</td>
<td>28</td>
</tr>
<tr>
<td>Toll-like Receptor</td>
<td>20</td>
<td>28</td>
<td>27</td>
<td>13</td>
</tr>
<tr>
<td>Rig-like Receptor</td>
<td>18</td>
<td>21</td>
<td>22</td>
<td>10</td>
</tr>
<tr>
<td>Bacterial Invasion</td>
<td>27</td>
<td>37</td>
<td>42</td>
<td>21</td>
</tr>
<tr>
<td>Autophagy</td>
<td>15</td>
<td>11</td>
<td>11</td>
<td>3</td>
</tr>
<tr>
<td>Apoptosis</td>
<td>19</td>
<td>28</td>
<td>26</td>
<td>14</td>
</tr>
<tr>
<td>p53</td>
<td>24</td>
<td>33</td>
<td>29</td>
<td>14</td>
</tr>
<tr>
<td>Nod-like Receptor</td>
<td>12</td>
<td>19</td>
<td>16</td>
<td>6</td>
</tr>
<tr>
<td>NF-kB</td>
<td>17</td>
<td>30</td>
<td>26</td>
<td>10</td>
</tr>
<tr>
<td>Tumor Necrosis Factor</td>
<td>23</td>
<td>42</td>
<td>35</td>
<td>18</td>
</tr>
<tr>
<td>PI3K-Akt</td>
<td>63</td>
<td>107</td>
<td>93</td>
<td>67</td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td>219</td>
<td>130</td>
<td>96</td>
<td></td>
</tr>
<tr>
<td><strong>Standard Deviation</strong></td>
<td>52</td>
<td>30</td>
<td>10</td>
<td></td>
</tr>
</tbody>
</table>
Table 2 (on next page)

Components of host-virus interaction pathways in the *O. faveolata* reference transcriptome.

Alignment results of human Rig-like receptor pathway components (Uniprot Accession Number) to *O. faveolata* transcript sequences by BLASTp. E-values < 1e-5 and predicted protein domain validation by Hmmscan.
<table>
<thead>
<tr>
<th>Accession</th>
<th>Protein Name</th>
<th>O. faveolata Homologues</th>
</tr>
</thead>
<tbody>
<tr>
<td>O94817</td>
<td>Ubiquitin-like protein ATG12</td>
<td>comp13112</td>
</tr>
<tr>
<td>Q9H1Y0</td>
<td>Autophagy protein 5</td>
<td>comp32308</td>
</tr>
<tr>
<td>Q14790</td>
<td>Casapse</td>
<td>comp42756, comp44852, comp43067, comp41842, comp39349</td>
</tr>
<tr>
<td>P02778</td>
<td>C-X-C motif chemokine 10</td>
<td>-</td>
</tr>
<tr>
<td>Q9NQC7</td>
<td>Ubiquitin carboxyl-terminal hydrolase CYLD</td>
<td>comp6468, comp45618</td>
</tr>
<tr>
<td>O00571</td>
<td>Rig-like receptor domain-containing proteins</td>
<td>comp42383, comp46396, comp45021, comp27027, comp31533, comp52755, comp12373, comp44096 comp29735</td>
</tr>
<tr>
<td>Q3LXA3</td>
<td>Triokinase</td>
<td>-</td>
</tr>
<tr>
<td>Q13158</td>
<td>FAS-associated death domain protein</td>
<td>-</td>
</tr>
<tr>
<td>P25963</td>
<td>NF-kappa-B inhibitor A/B</td>
<td>comp4914, comp6407</td>
</tr>
<tr>
<td>Q15111</td>
<td>Inhibitor of NF-kappa-B</td>
<td>comp44188</td>
</tr>
<tr>
<td>Q14653</td>
<td>Interferon regulatory factors</td>
<td>comp25035, comp40113, comp39249, comp50738, comp30904</td>
</tr>
<tr>
<td>P05161</td>
<td>Ubiquitin-like protein ISG15</td>
<td>comp41425, comp36790, comp36790</td>
</tr>
<tr>
<td>Q13233</td>
<td>MAPKs</td>
<td>comp45734, comp44708, comp32445, comp45894</td>
</tr>
<tr>
<td>O43318</td>
<td>TGF-beta-activated kinase 1</td>
<td>comp110885, comp56867, comp14096</td>
</tr>
<tr>
<td>P45983</td>
<td>JNKs</td>
<td>comp10734, comp44864, comp54242, comp60240, comp38508</td>
</tr>
<tr>
<td>P19838</td>
<td>NFkBs</td>
<td>comp46270, comp44367</td>
</tr>
<tr>
<td>Q96G74</td>
<td>OTU domain-containing protein 5</td>
<td>comp52747, comp41206</td>
</tr>
<tr>
<td>Q13526</td>
<td>Rotamase Pin1</td>
<td>comp40298</td>
</tr>
<tr>
<td>Q13546</td>
<td>Serine/threonine-protein kinase RIP</td>
<td>-</td>
</tr>
<tr>
<td>Q96EQ8</td>
<td>E3 ubiquitin-protein ligase RNF125</td>
<td>-</td>
</tr>
<tr>
<td>Q9BRV8</td>
<td>Suppressor of IKBKE 1</td>
<td>comp43257</td>
</tr>
<tr>
<td>Q86WV6</td>
<td>Stimulator of interferon genes protein</td>
<td>comp46735, comp115636</td>
</tr>
<tr>
<td>Q9UHD2</td>
<td>Serine/threonine-protein kinase TBK1</td>
<td>comp66935</td>
</tr>
<tr>
<td>Q12933</td>
<td>TNF receptor-associated factors</td>
<td>comp46745, comp45812, comp43792, comp34933, comp60135, comp43035, comp46671</td>
</tr>
<tr>
<td>Q14258</td>
<td>E3 ubiquitin/ISG15 ligase TRIM25</td>
<td>-</td>
</tr>
</tbody>
</table>