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

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

Anderson DA, Walz ME, Weil E, Tonellato P, Smith MC. 2016. RNA-Seq of the Caribbean reef-building coral *Orbicella faveolata* (Scleractinia-Merulinidae) under bleaching and disease stress expands models of coral innate immunity. PeerJ 4:e1616 https://doi.org/10.7717/peerj.1616

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, Orbicella 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 Orbicella faveolata (Scleractinia-Meruliniidae) David A. Anderson^{1, 2}, Marcus Walz³, Ernesto Weil², Peter Tonellato^{3,4} & 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

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

61

62

63

64

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, Orbicella 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.

65 **Keywords:** RNA-seq, non-model organism, coral, Cnidaria, innate immunity, disease, 66 transcriptome, Wnt, Dicer

Introduction

68

69

70

71

72

73

74

75

76

77

78

79

80

81

82

83

84

85

86

87

88

89

90

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

92

93

94

95

96

97

98

99

100

101

102

103

104

105

106

107

108

109

110

111

112

113

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

114

115

116

117

118

119

120

121

122

123

124

126

127

128

129

130

131

132

133

134

135

136

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 nonclonal given their large distances of separation (>10 m), and the low levels of clonality (3.5%)

142

143

144

145

146

147

148

149

150

151

152

153

154

155

156

157

158

159

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, Velno 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

183

184

185

186

187

188

189

190

191

192

193

194

195

196

197

198

199

200

201

202

203

204

205

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.

207

208

209

210

211

212

213

214

215

216

217

218

219

220

221

222

223

224

225

226

227

228

Phylogenetic Analysis, Protein Domain Architecture and Immune Signaling Pathways

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 hmmcan 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 hostvirus 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

276

277

278

279

280

281

282

283

284

285

287

288

289

290

291

292

293

294

295

296

297

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 immunityrelated 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.

368

369

370

371

372

373

374

375

376

377

379

380

381

382

383

384

385

386

387

388

389

2013; Davy et al. 2006; Soffer et al. 2013; Wilson et al. 2005). 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

391

392

393

394

395

396

397

408

409

410

411

412

413

414

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

- 398 MCS received startup funds from the Department of Marine Science at the University of Puerto
- 399 Rico at Mayaguez to support this work. This material is based upon work supported by the
- 400 National Science Foundation (NSF) Graduate Research Fellowship under Grant No. DGE-
- 401 1143954 to DAA. Partial funding was provided by NSF grants IOS # 1017510 and OCE -
- 402 1105143 to EW.

403 References

- 404 53851 F. 2014. Endangered and Threatened Wildlife and Plants: Final Listing Determinations on Proposal
 405 To List 66 Reef-Building Coral Species and To Reclassify Elkhorn and Staghorn Corals. In:
 406 Commerce Do, editor. Federal Register: National Oceanic and Atmospheric Administration. p
 407 53851-54123.
 - Ainsworth T, Kvennefors E, Blackall L, Fine M, and Hoegh-Guldberg O. 2007. Disease and cell death in white syndrome of Acroporid corals on the Great Barrier Reef. *Marine Biology* 151:19-29.
 - Altizer S, Ostfeld RS, Johnson PT, Kutz S, and Harvell CD. 2013. Climate change and infectious diseases: From evidence to a predictive framework. *science* 341:514-519.
 - Anderson D, and Gilchrist S. 2008. Development of a novel method for coral RNA isolation and the expression of a programmed cell death gene in white plague-diseased Diploria strigosa (Dana, 1846). Proc 11th Int Coral Reef Symp Ft Lauderdale, Florida. p 211-215.
- Antonius A. 1973. New observations on coral destruction in reefs. Tenth Meeting of the Association of Island Marine Laboratories of the Caribbean: University of Puerto Rico (Mayaguez).
- Atad I, Zvuloni A, Loya Y, and Rosenberg E. 2012. Phage therapy of the white plague-like disease of Favia favus in the Red Sea. *Coral Reefs* 31:665-670.
- Augustin R, Fraune S, and Bosch TCG. 2010. How< i> Hydra</i> senses and destroys microbes. Seminars in immunology: Elsevier. p 54-58.

- Barr JJ, Auro R, Furlan M, Whiteson KL, Erb ML, Pogliano J, Stotland A, Wolkowicz R, Cutting AS, and Doran KS. 2013. Bacteriophage adhering to mucus provide a non-host-derived immunity.

 Proceedings of the National Academy of Sciences.
- Barshis DJ, Ladner JT, Oliver TA, Seneca FO, Traylor-Knowles N, and Palumbi SR. 2013. Genomic basis for coral resilience to climate change. *Proceedings of the National Academy of Sciences* 110:1387-1392.
 - Bayer T, Aranda M, Sunagawa S, Yum LK, DeSalvo MK, Lindquist E, Coffroth MA, Voolstra CR, and Medina M. 2012. Symbiodinium transcriptomes: genome insights into the dinoflagellate symbionts of reef-building corals. *PloS one* 7:e35269.
 - Boehm M, Hentschel U, Friedrich A, Fieseler L, Steffen R, Gamulin V, Mueller I, and Müller W. 2001. Molecular response of the sponge Suberites domuncula to bacterial infection. *Marine Biology* 139:1037-1045.
 - Boratyn GM, Camacho C, Cooper PS, Coulouris G, Fong A, Ma N, Madden TL, Matten WT, McGinnis SD, and Merezhuk Y. 2013. BLAST: a more efficient report with usability improvements. *Nucleic acids research* 41:W29-W33.
 - Bosch TCG. 2012. What hydra has to say about the role and origin of symbiotic interactions. *The Biological Bulletin* 223:78-84.
 - Bosch TCG, Anton-Erxleben F, Augustin R, Franzenburg S, and Fraune S. 2011. Hydra Go Bacterial. Beneficial Microorganisms in Multicellular Life Forms: Springer, 313-322.
 - Bourne DG, Garren M, Work TM, Rosenberg E, Smith GW, and Harvell CD. 2009. Microbial disease and the coral holobiont. *Trends in microbiology* 17:554-562.
 - Bruckner AW. 2012. Factors contributing to the regional decline of Montastraea annularis (complex).

 Proceedings of the 12th International Coral Reef Symposium, Cairns, Australia, July. p 9-13.
 - Bruckner AW, and Hill RL. 2009. Ten years of change to coral communities off Mona and Desecheo Islands, Puerto Rico, from disease and bleaching. *Diseases of Aquatic Organisms* 87:19.
 - Bruno JF, Selig ER, Casey KS, Page CA, Willis BL, Harvell CD, Sweatman H, and Melendy AM. 2007. Thermal stress and coral cover as drivers of coral disease outbreaks. *PLoS Biology* 5:e124.
 - Burge CA, Mouchka ME, Harvell CD, and Roberts S. 2013. Immune response of the Caribbean sea fan, Gorgonia ventalina, exposed to an Aplanochytrium parasite as revealed by transcriptome sequencing. *Frontiers in physiology* 4.
 - Chen Z, Gibson TB, Robinson F, Silvestro L, Pearson G, Xu B-e, Wright A, Vanderbilt C, and Cobb MH. 2001. MAP kinases. *Chemical reviews* 101:2449-2476.
 - Chevenet F, Brun C, Bañuls A-L, Jacq B, and Christen R. 2006. TreeDyn: towards dynamic graphics and annotations for analyses of trees. *BMC bioinformatics* 7:439.
 - Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, and Robles M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21:3674-3676.
 - Cróquer A, Bastidas C, Elliott A, and Sweet M. 2013. Bacterial assemblages shifts from healthy to yellow band disease states in the dominant reef coral Montastraea faveolata. *Environmental microbiology reports* 5:90-96.
 - Davy S, Burchett S, Dale A, Davies P, Davy J, Muncke C, Hoegh-Guldberg O, and Wilson W. 2006. Viruses: agents of coral disease? *Diseases of Aquatic Organisms* 69:101-110.
- Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard J-F, Guindon S, Lefort V, and Lescot M. 2008. Phylogeny. fr: robust phylogenetic analysis for the non-specialist. *Nucleic acids research* 36:W465-W469.
- Desalvo M, Voolstra C, Sunagawa S, Schwarz J, Stillman J, Coffroth MA, Szmant A, and Medina M. 2008.

 Differential gene expression during thermal stress and bleaching in the Caribbean coral Montastraea faveolata. *Molecular ecology* 17:3952-3971.

- 469 Ding S-W, and Voinnet O. 2007. Antiviral immunity directed by small RNAs. *Cell* 130:413-426.
- Ducklow HW, and Mitchell R. 1979. Observations on naturally and artificially diseased tropical corals: a scanning electron microscope study. *Microbial ecology* 5:215-223.
- Duncan AW, Rattis FM, DiMascio LN, Congdon KL, Pazianos G, Zhao C, Yoon K, Cook JM, Willert K, and Gaiano N. 2005. Integration of Notch and Wnt signaling in hematopoietic stem cell maintenance. *Nature immunology* 6:314-322.
 - Dunlap WC, Starcevic A, Baranasic D, Diminic J, Zucko J, Gacesa R, Van Oppen MJ, Hranueli D, Cullum J, and Long PF. 2013. KEGG orthology-based annotation of the predicted proteome of Acropora digitifera: ZoophyteBase-an open access and searchable database of a coral genome. *BMC genomics* 14:509.
 - Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research* 32:1792-1797.
 - Finn RD, Clements J, and Eddy SR. 2011. HMMER web server: interactive sequence similarity searching. *Nucleic acids research* 39:W29-W37.
 - Franzenburg S, Fraune S, Künzel S, Baines JF, Domazet-Lošo T, and Bosch TCG. 2012. MyD88-deficient Hydra reveal an ancient function of TLR signaling in sensing bacterial colonizers. *Proceedings of the National Academy of Sciences* 109:19374-19379.
 - Gleeson M, and Strong A. 1995. Applying MCSST to coral reef bleaching. *Advances in Space Research* 16:151-154.
 - Glynn P, Peters EC, and Muscatine L. 1985. Coral tissue microstructure and necrosis: relation to catastrophic coral mortality in Panama. *Dis Aquat Org* 1:29-37.
 - Glynn PW. 1983. Extensive 'bleaching' and death of reef corals on the Pacific coast of Panama. Environmental Conservation 10:149-154.
 - Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, and Zeng Q. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature biotechnology* 29:644-652.
 - Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, and Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic biology* 59:307-321.
 - Guo X, Zhang R, Wang J, Ding S-W, and Lu R. 2013. Homologous RIG-I—like helicase proteins direct RNAi-mediated antiviral immunity in C. elegans by distinct mechanisms. *Proceedings of the National Academy of Sciences* 110:16085-16090.
 - Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, and Lieber M. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature protocols* 8:1494-1512.
 - Hamada M, Shoguchi E, Shinzato C, Kawashima T, Miller DJ, and Satoh N. 2013. The complex NOD-like receptor repertoire of the coral Acropora digitifera includes novel domain combinations. *Molecular biology and evolution* 30:167-176.
 - Harvell C, Kim K, Burkholder J, Colwell R, Epstein PR, Grimes D, Hofmann E, Lipp E, Osterhaus A, and Overstreet RM. 1999. Emerging marine diseases--climate links and anthropogenic factors. *science* 285:1505-1510.
- Käsbauer T, Towb P, Alexandrova O, David CN, Dall'Armi E, Staudigl A, Stiening B, and Böttger A. 2007.
 The Notch signaling pathway in the cnidarian< i> Hydra</i> Developmental biology 303:376-390.
- Kim K, Harvell C, Kim P, Smith G, and Merkel S. 2000. Fungal disease resistance of Caribbean sea fan corals (Gorgonia spp.). *Marine Biology* 136:259-267.

- Kimes NE, Van Nostrand JD, Weil E, Zhou J, and Morris PJ. 2010. Microbial functional structure of Montastraea faveolata, an important Caribbean reef-building coral, differs between healthy and yellow-band diseased colonies. *Environmental microbiology* 12:541-556.
- Kortschak RD, Samuel G, Saint R, and Miller DJ. 2003. EST Analysis of the Cnidarian< i> Acropora millepora</i> Reveals Extensive Gene Loss and Rapid Sequence Divergence in the Model Invertebrates. *Current Biology* 13:2190-2195.
 - Kusserow A, Pang K, Sturm C, Hrouda M, Lentfer J, Schmidt HA, Technau U, von Haeseler A, Hobmayer B, and Martindale MQ. 2005. Unexpected complexity of the Wnt gene family in a sea anemone. *Nature* 433:156-160.
 - Kvennefors ECE, Leggat W, Hoegh-Guldberg O, Degnan BM, and Barnes AC. 2008. An ancient and variable mannose-binding lectin from the coral< i> Acropora millepora</i> binds both pathogens and symbionts. *Developmental & Comparative Immunology* 32:1582-1592.
 - Lehnert EM, Burriesci MS, and Pringle JR. 2012. Developing the anemone Aiptasia as a tractable model for cnidarian-dinoflagellate symbiosis: the transcriptome of aposymbiotic A. pallida. *BMC genomics* 13:271.
 - Libro S, Kaluziak ST, and Vollmer SV. 2013. RNA-seq Profiles of Immune Related Genes in the Staghorn Coral Acropora cervicornis Infected with White Band Disease. *PloS one* 8:e81821.
 - Liew YJ, Aranda M, Carr A, Baumgarten S, Zoccola D, Tambutté S, Allemand D, Micklem G, and Voolstra CR. 2014. Identification of MicroRNAs in the Coral Stylophora pistillata. *PloS one* 9:e91101.
 - MacKay CR, Wang JP, and Kurt-Jones EA. 2014. Dicer's role as an antiviral: still an enigma. *Current Opinion in Immunology* 26:49-55.
 - Marlow H, Roettinger E, Boekhout M, and Martindale MQ. 2012. Functional roles of Notch signaling in the cnidarian< i> Nematostella vectensis</i> *Developmental biology* 362:295-308.
 - Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal* 17:pp. 10-12.
 - Meyer E, Aglyamova G, Wang S, Buchanan-Carter J, Abrego D, Colbourne J, Willis B, and Matz M. 2009. Sequencing and de novo analysis of a coral larval transcriptome using 454 GSFlx. *BMC genomics* 10:219.
 - Miller DJ, Ball EE, and Technau U. 2005. Cnidarians and ancestral genetic complexity in the animal kingdom. *Trends in Genetics* 21:536-539.
 - Moran Y, Praher D, Fredman D, and Technau U. 2013. The evolution of microRNA pathway protein components in Cnidaria. *Molecular biology and evolution* 30:2541-2552.
 - Moriya Y, Itoh M, Okuda S, Yoshizawa AC, and Kanehisa M. 2007. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic acids research* 35:W182-W185.
 - Moya A, Huisman L, Ball E, Hayward D, Grasso L, Chua C, Woo H, GATTUSO JP, Forêt S, and Miller D. 2012. Whole transcriptome analysis of the Coral Acropora millepora reveals complex responses to CO2-driven acidification during the initiation of calcification. *Molecular ecology* 21:2440-2454.
 - Mukherjee K, Korithoski B, and Kolaczkowski B. 2014. Ancient Origins of Vertebrate-Specific Innate Antiviral Immunity. *Molecular biology and evolution* 31:140-153.
 - Mydlarz LD, Couch CS, Weil E, Smith G, and Harvell C. 2009. Immune defenses of healthy, bleached and diseased Montastraea faveolata during a natural bleaching event. *Diseases of Aquatic Organisms* 87:67-78.
 - Mydlarz LD, Holthouse SF, Peters EC, and Harvell CD. 2008. Cellular responses in sea fan corals: granular amoebocytes react to pathogen and climate stressors. *PloS one* 3:e1811.
- Ocampo ID, Zárate-Potes A, Pizarro V, Rojas CA, Vera NE, and Cadavid LF. 2015. The immunotranscriptome of the Caribbean reef-building coral Pseudodiploria strigosa.

 Immunogenetics:1-16.

- Palmer C, and Traylor-Knowles N. 2012. Towards an integrated network of coral immune mechanisms. *Proceedings of the Royal Society B: Biological Sciences* 279:4106-4114.
- Palmer CV, Modi CK, and Mydlarz LD. 2009. Coral fluorescent proteins as antioxidants. *PloS one* 4:e7298.
- Pati A, Heath LS, Kyrpides NC, and Ivanova N. 2011. ClaMS: a classifier for metagenomic sequences. Standards in genomic sciences 5:248.
 - Pinzón JH, Kamel B, Burge CA, Harvell CD, Medina M, Weil E, and Mydlarz LD. 2015. Whole transcriptome analysis reveals changes in expression of immune-related genes during and after bleaching in a reef-building coral. *Royal Society Open Science* 2:140214.
 - Polato NR, Vera JC, and Baums IB. 2011. Gene discovery in the threatened elkhorn coral: 454 sequencing of the Acropora palmata transcriptome. *PloS one* 6:e28634.
 - Pollock FJ, Morris PJ, Willis BL, and Bourne DG. 2011. The urgent need for robust coral disease diagnostics. *PLoS pathogens* 7:e1002183.
 - Pooyaei Mehr SFP, DeSalle R, Kao H-T, Narechania A, Han Z, Tchernov D, Pieribone V, and Gruber DF. 2013. Transcriptome deep-sequencing and clustering of expressed isoforms from Favia corals. *BMC genomics* 14:546.
 - Putnam NH, Srivastava M, Hellsten U, Dirks B, Chapman J, Salamov A, Terry A, Shapiro H, Lindquist E, and Kapitonov VV. 2007. Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *science* 317:86-94.
 - Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, and Lopez R. 2005. InterProScan: protein domains identifier. *Nucleic acids research* 33:W116-W120.
 - Radtke F, Fasnacht N, and MacDonald HR. 2010. Notch signaling in the immune system. *Immunity* 32:14-27.
 - Rosenberg E, Koren O, Reshef L, Efrony R, and Zilber-Rosenberg I. 2007. The role of microorganisms in coral health, disease and evolution. *Nature Reviews Microbiology* 5:355-362.
 - Sabourault C, Ganot P, Deleury E, Allemand D, and Furla P. 2009. Comprehensive EST analysis of the symbiotic sea anemone, Anemonia viridis. *BMC genomics* 10:333.
 - Schwarz JA, Brokstein PB, Voolstra C, Terry AY, Miller DJ, Szmant AM, Coffroth MA, and Medina M. 2008. Coral life history and symbiosis: functional genomic resources for two reef building Caribbean corals, Acropora palmata and Montastraea faveolata. *BMC genomics* 9:97.
 - Severance EG, and Karl SA. 2006. Contrasting population genetic structures of sympatric, mass-spawning Caribbean corals. *Marine Biology* 150:57-68.
 - Shinzato C, Inoue M, and Kusakabe M. 2014. A Snapshot of a Coral "Holobiont": A Transcriptome Assembly of the Scleractinian Coral, Porites, Captures a Wide Variety of Genes from Both the Host and Symbiotic Zooxanthellae. *PloS one* 9:e85182.
 - Shinzato C, Shoguchi E, Kawashima T, Hamada M, Hisata K, Tanaka M, Fujie M, Fujiwara M, Koyanagi R, and Ikuta T. 2011. Using the Acropora digitifera genome to understand coral responses to environmental change. *Nature* 476:320-323.
 - Soffer N, Brandt ME, Correa AM, Smith TB, and Thurber RV. 2013. Potential role of viruses in white plague coral disease. *The ISME journal*.
 - Soza-Ried J, Hotz-Wagenblatt A, Glatting KH, del Val C, Fellenberg K, Bode HR, Frank U, Hoheisel JD, and Frohme M. 2010. The transcriptome of the colonial marine hydroid Hydractinia echinata. *FEBS Journal* 277:197-209.
- Staal FJ, Luis TC, and Tiemessen MM. 2008. WNT signalling in the immune system: WNT is spreading its wings. *Nature Reviews Immunology* 8:581-593.
- Sun J, Chen Q, Lun JC, Xu J, and Qiu J-W. 2012. PcarnBase: Development of a transcriptomic database for the brain coral Platygyra carnosus. *Marine Biotechnology*:1-8.
- Sutherland KP, Porter JW, and Torres C. 2004. Disease and immunity in Caribbean and Indo-Pacific zooxanthellate corals. *Marine ecology progress series* 266:265-272.

- Talavera G, and Castresana J. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic biology* 56:564-577.
 - Tchernov D, Kvitt H, Haramaty L, Bibby TS, Gorbunov MY, Rosenfeld H, and Falkowski PG. 2011.

 Apoptosis and the selective survival of host animals following thermal bleaching in zooxanthellate corals. *Proceedings of the National Academy of Sciences* 108:9905-9909.
 - Ting JPY, Lovering RC, Alnemri ESPD, Bertin J, Boss JM, Davis B, Flavell RA, Girardin SE, Godzik A, and Harton JA. 2008. The NLR gene family: an official nomenclature. *Immunity* 28:285.
 - Vidal-Dupiol J, Ladrière O, Destoumieux-Garzón D, Sautière P-E, Meistertzheim A-L, Tambutté E, Tambutté S, Duval D, Fouré L, and Adjeroud M. 2011. Innate immune responses of a scleractinian coral to vibriosis. *Journal of Biological Chemistry* 286:22688-22698.
 - Vidal-Dupiol J, Zoccola D, Tambutté E, Grunau C, Cosseau C, Smith KM, Freitag M, Dheilly NM, Allemand D, and Tambutté S. 2013. Genes Related to Ion-Transport and Energy Production Are Upregulated in Response to CO2-Driven pH Decrease in Corals: New Insights from Transcriptome Analysis. *PloS one* 8:e58652.
 - Voolstra CR, Schnetzer J, Peshkin L, Randall CJ, Szmant AM, and Medina M. 2009. Effects of temperature on gene expression in embryos of the coral Montastraea faveolata. *BMC genomics* 10:627.
 - Weil E, Cróquer A, and Urreiztieta I. 2009a. Yellow band disease compromises the reproductive output of the Caribbean reef-building coral Montastraea faveolata (Anthozoa, Scleractinia). *Diseases of Aquatic Organisms* 87:45.
 - Weil E, Croquer A, Urreiztieta I, and Irizarry E. 2009b. Temporal variability and impact of coral diseases and bleaching in La Parguera, Puerto Rico from 2003–2007. *Carib J Sci* 45:1-26.
 - Weil E, and Rogers CS. 2011. Coral reef diseases in the Atlantic-Caribbean. *Coral Reefs: An Ecosystem in Transition*: Springer, 465-491.
 - Weil E, Smith G, and Gil-Agudelo DL. 2006. Status and progress in coral reef disease research. *Diseases of Aquatic Organisms* 69:1-7.
 - Weis VM. 2008. Cellular mechanisms of Cnidarian bleaching: stress causes the collapse of symbiosis. *Journal of Experimental Biology* 211:3059-3066.
 - Weiss Y, Forêt S, Hayward DC, Ainsworth T, King R, Ball EE, and Miller DJ. 2013. The acute transcriptional response of the coral Acropora millepora to immune challenge: expression of GiMAP/IAN genes links the innate immune responses of corals with those of mammals and plants. *BMC genomics* 14:400.
 - Wilson W, Dale A, Davy J, and Davy S. 2005. An enemy within? Observations of virus-like particles in reef corals. *Coral Reefs* 24:145-148.
 - Yoneyama M, and Fujita T. 2007. Function of RIG-I-like receptors in antiviral innate immunity. *Journal of Biological Chemistry* 282:15315-15318.
- 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].

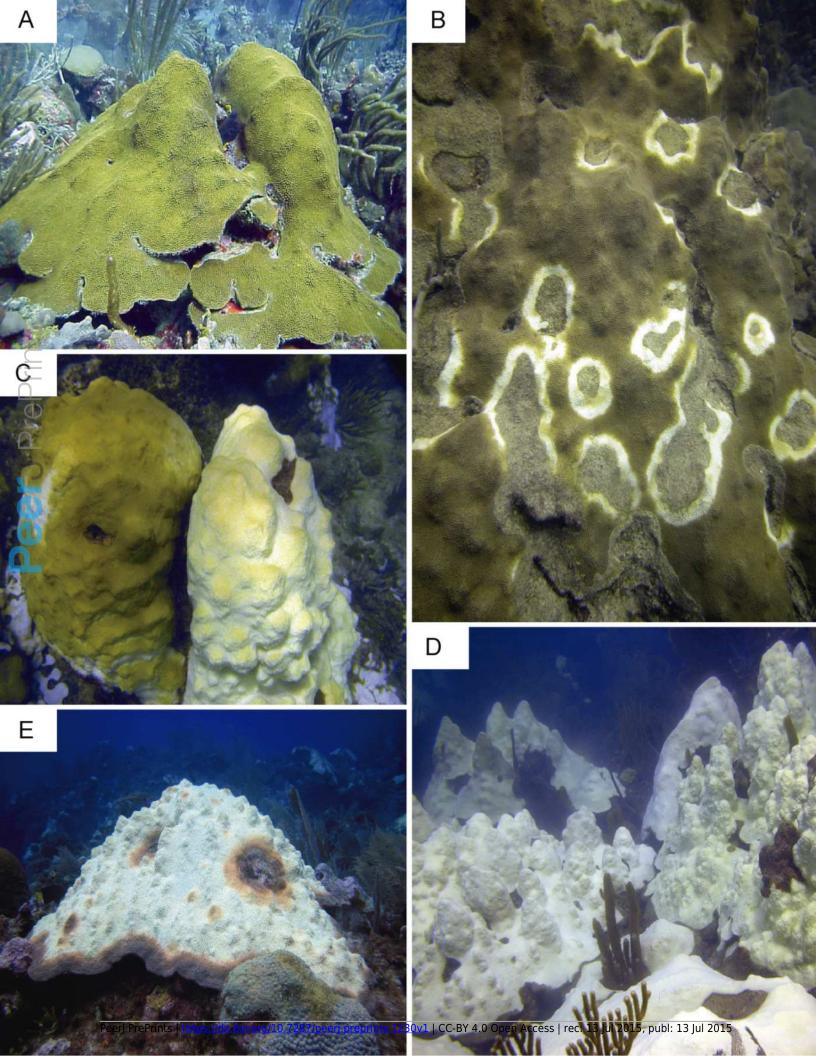


Figure 2(on next page)

Taxonomic hits.

Top hits to taxa in the Swissprot sequence database for the annotated *O. faveolata* reference transcriptome.

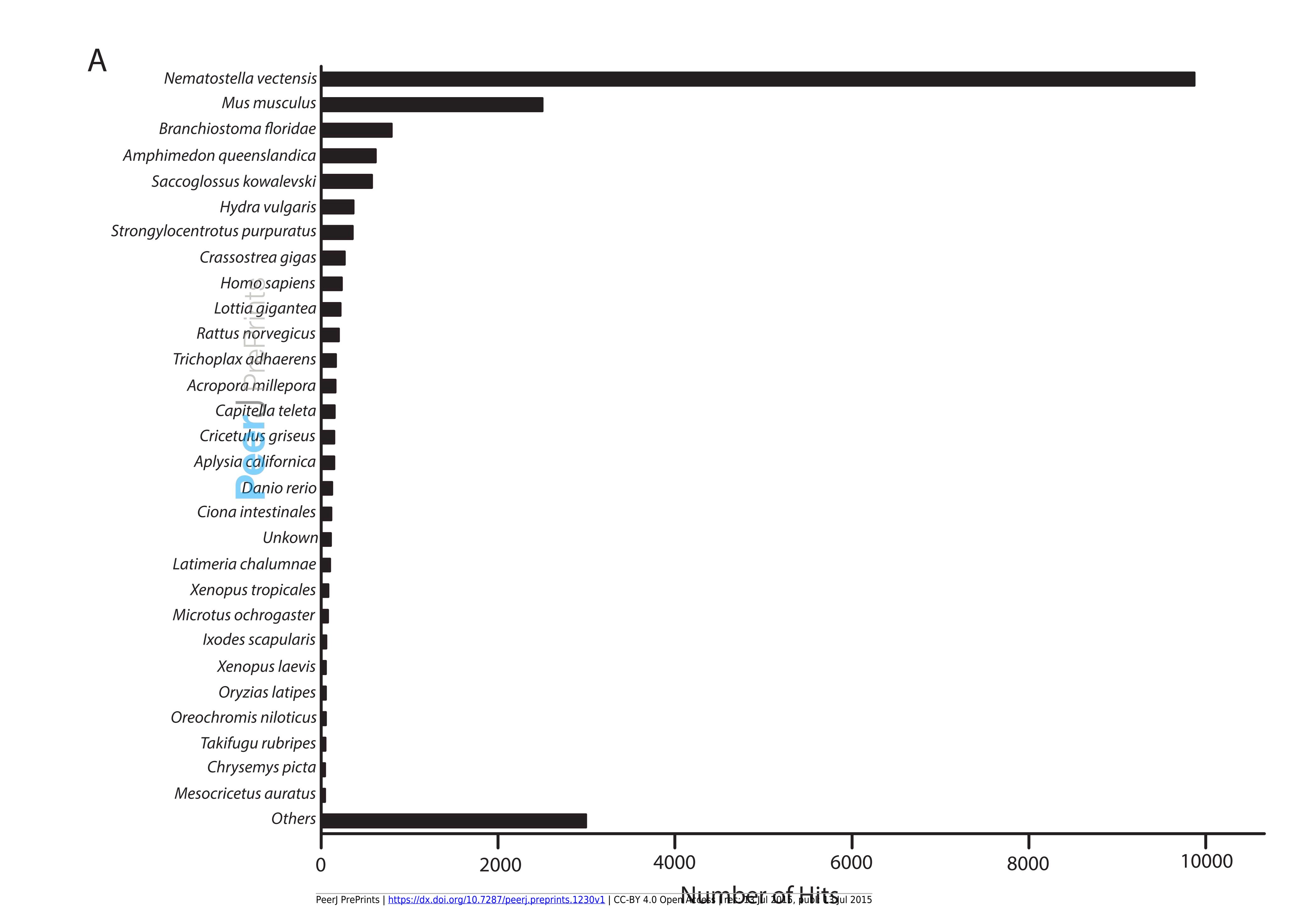
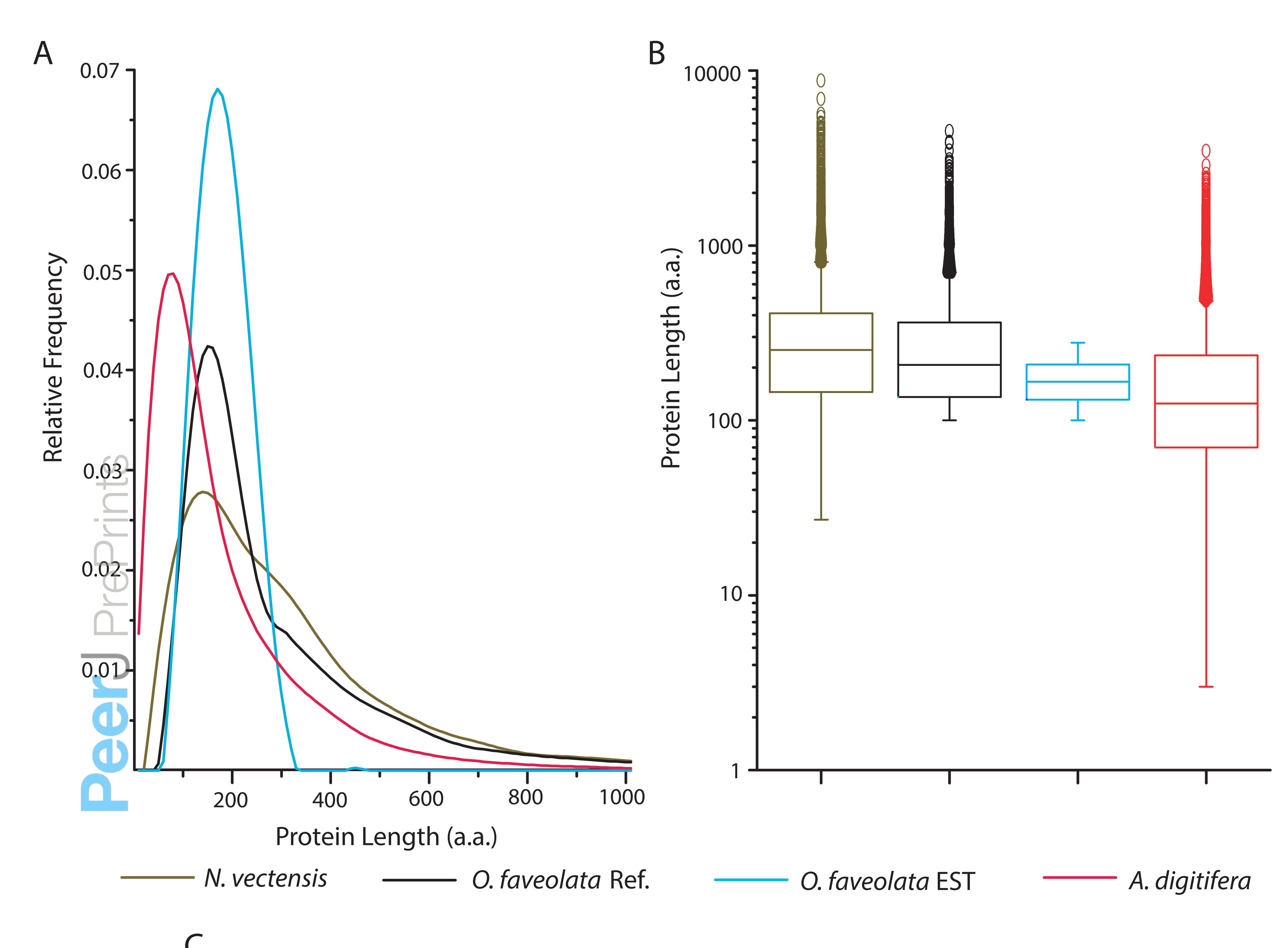


Figure 3(on next page)

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.

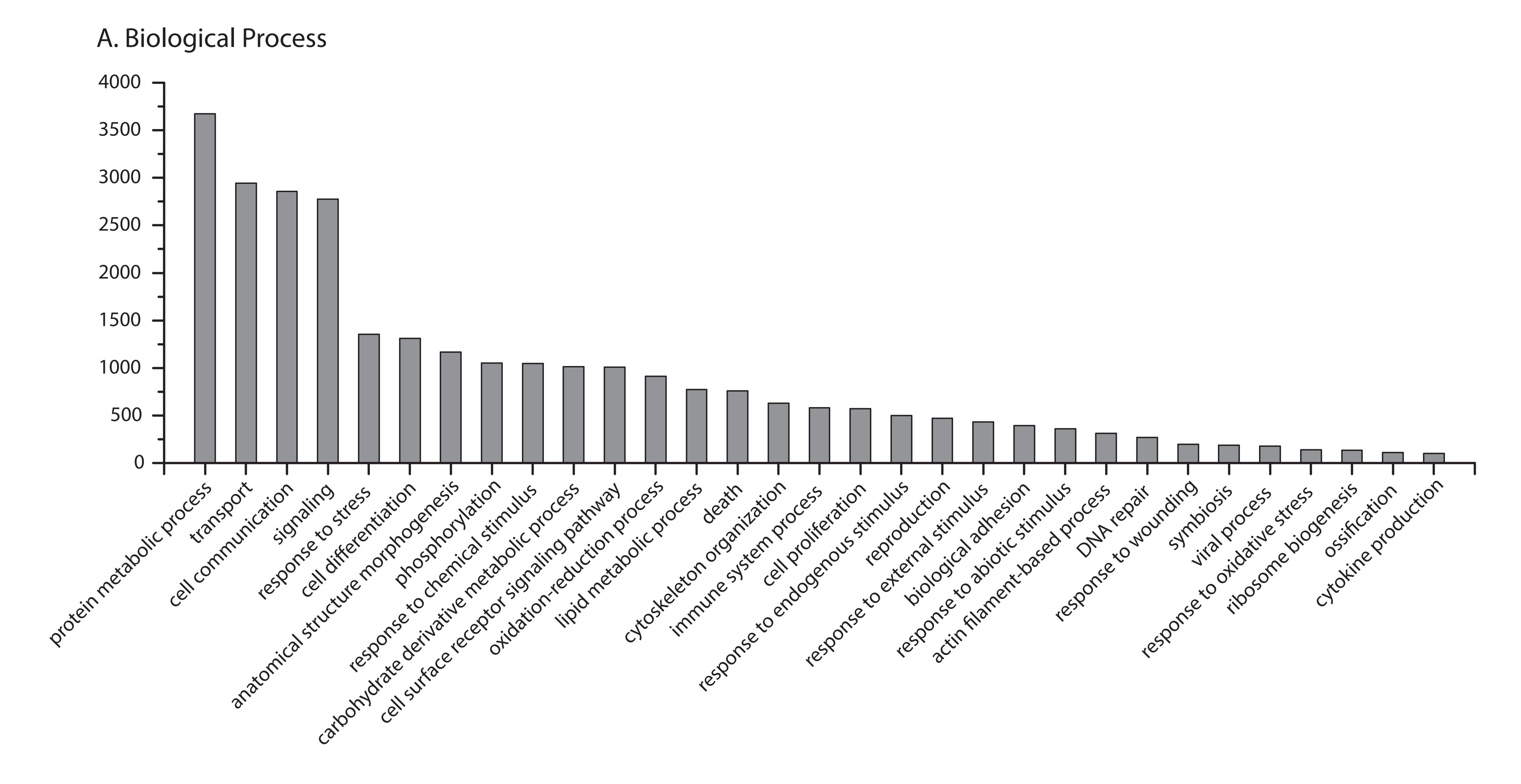


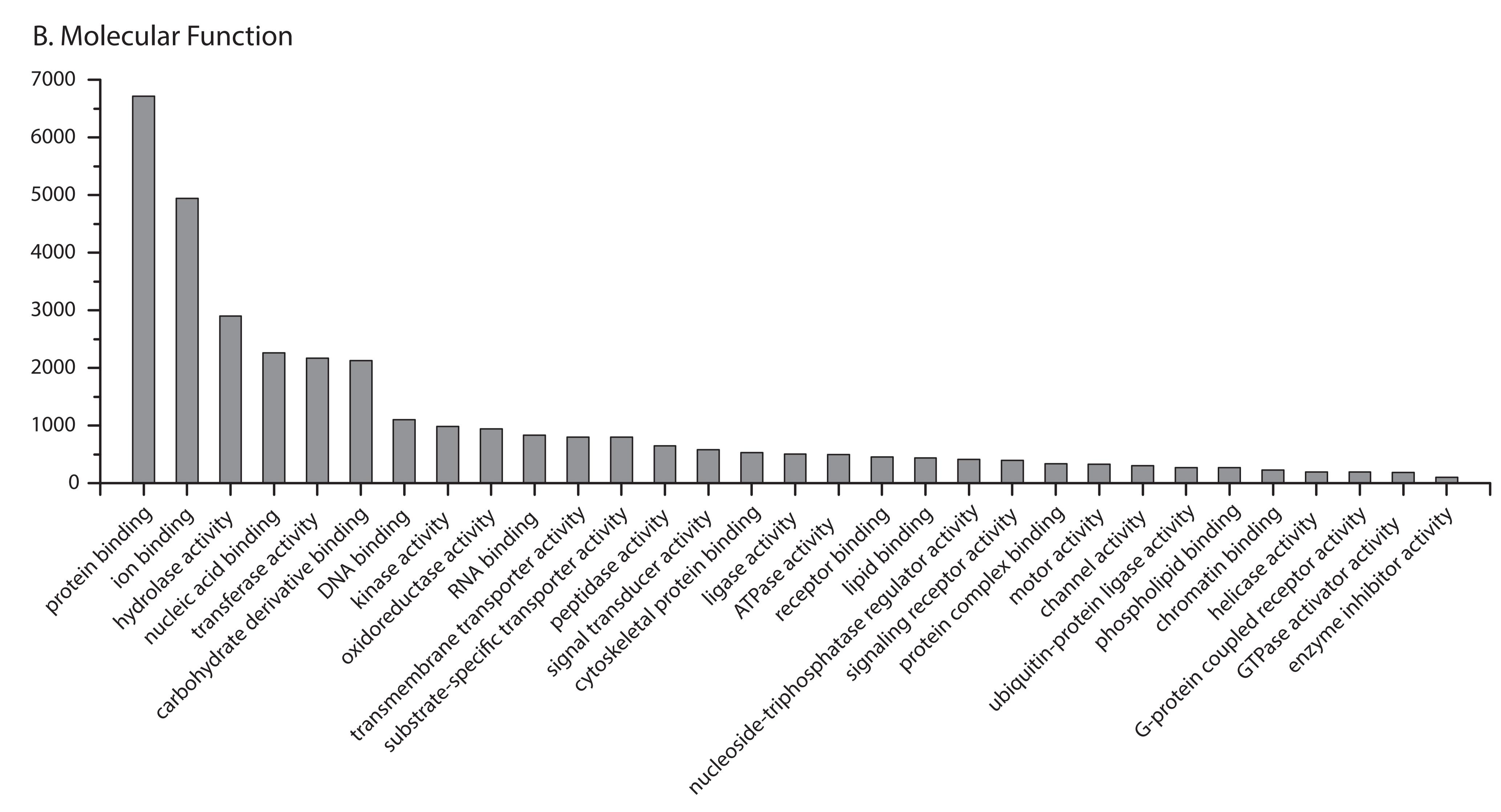
	Mean	Median	Maximum	Percentile	
Species	Length	Length	Length	95th	99th
N. vectensis	336	253	8745	882	1571
O. faveolata Ref.	297	207	4493	791	1279
O. faveolata EST	171	166	278	252	262
A. digitifera	183	125	3458	517	904

Figure 4(on next page)

GO terms represented in the coral reference transcriptome

Select gene ontology terms associated with at least 100 different sequences.





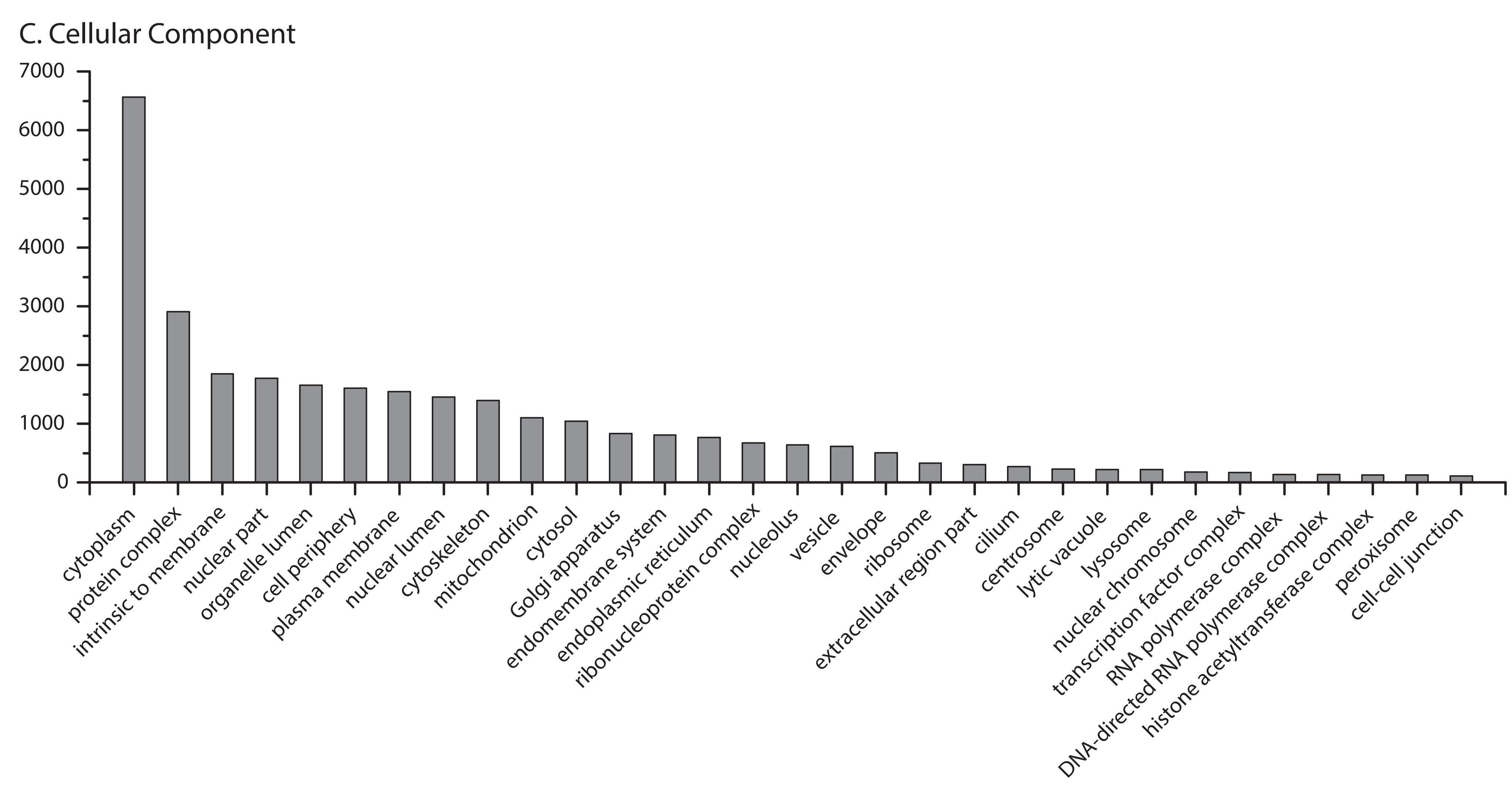


Figure 5(on next page)

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).

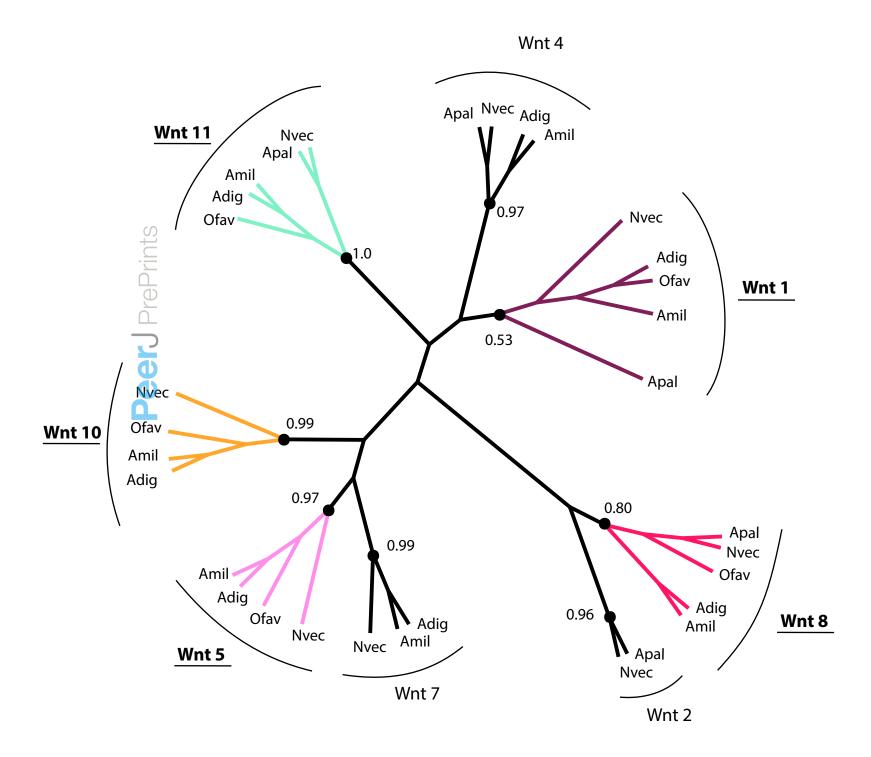
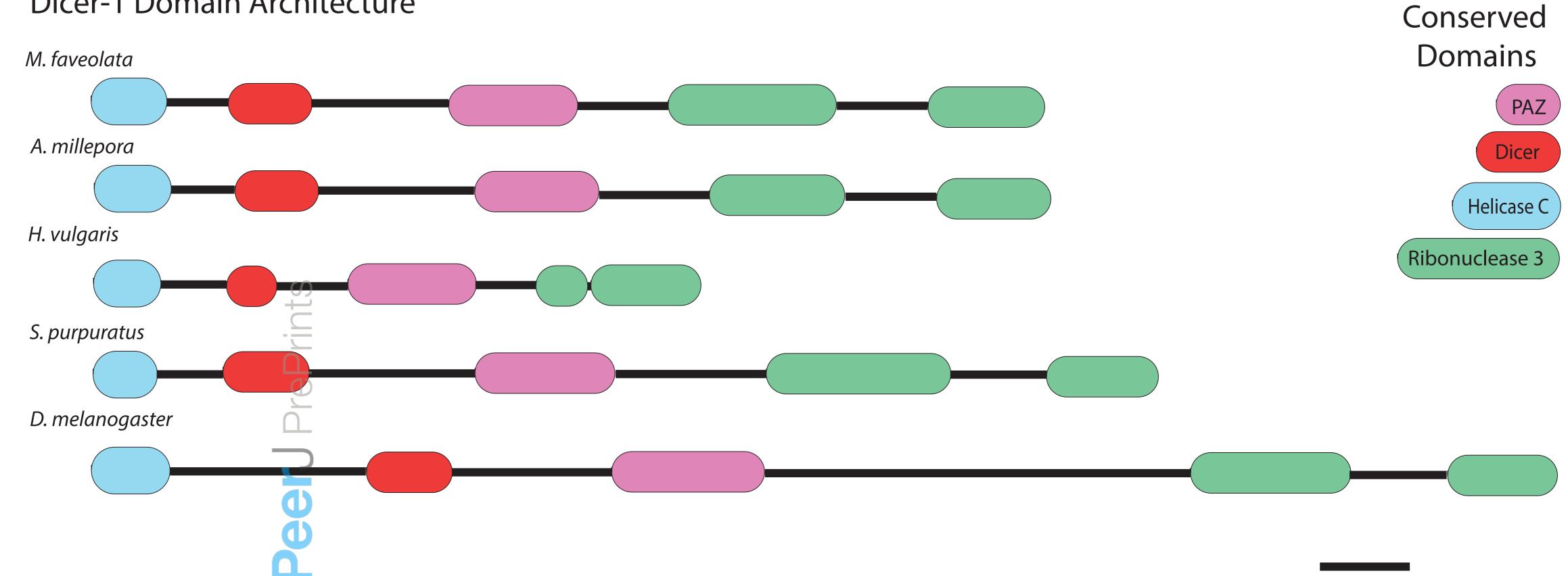


Figure 6(on next page)

Dicer-1-like domain architecture and pathway components

(A) Domain architecture of Dicer-1-like proteins. Scale bar representes 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.

Dicer-1 Domain Architecture



100 a.a. residues

B

Accession	Protein Name	O. faveolata Homologues
Q32KD4	Argonaute-1	comp43840, comp33938, comp46773, comp64615
Q9VCU9	Dicer	comp46340
Q7KNF1	Drosha	comp42908
U3MIH3	GW182/TNRC6	comp34495, comp51031
Q7K175	HEN1	comp39200
U3MH39	HYL1	comp38270
Q9V9V7	Pasha/DGCR8	comp23275
Q9V9K7	Serrate/Ars2	comp38150
Q9VLW8	R2D2	comp23275
Q6J5K9	Armitage	comp43557
G5EBQ3	RRF/RDRP/EGO	comp40697
O44406	ERI-1	comp39615

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.

	N. vectensis	A. millepora	M.faveolata		Presente	Presented Reference Transcriptome		
					% Relative to	% Relative to N.	%Relative to A.	
Metabolism	Hits	Hits	Present Hits	Prior Hits	Prior	vectensis	millepora	
Glycolysis & Gluconeogenesis	27	32	27	22	123	100	84	
Pentose Phosphate	17	20	18	12	150	106	90	
Citrate Cycle	23	23	20	18	111	87	87	
Biosynthesis of Amino Acids	54	52	49	34	144	91	94	
Valine, Leucine and Isoleucine Degradation	33	34	32	20	160	97	94	
Purine Metabolism	91	105	88	55	160	97	84	
Fatty Acid Metabolism	27	26	23	19	121	85	88	
Pyrimidine Metabolism	66	75	62	34	182	94	83	
Protein Complexes								
Spliceosome	96	106	99	74	134	103	93	
Ribosome	107	119	99	97	102	93	83	
Protein Export	20	21	20	20	100	100	95	
Oxidative Phosphorylation	79	87	85	74	115	108	98	
RNA Degradation	51	55	43	23	187	84	78	
Ubiquitin Pr <mark>ote</mark> olysis	81	95	84	49	171	104	88	
$\overline{\mathbf{Q}}$				Mean	140	96	89	
			Stand	dard Deviation	29	8	6	
Stress and Immunity								
MAPK	62	105	107	43	249	157	102	
Ras	50	80	86	39	221	151	108	
Wnt	47	63	57	29	197	108	90	
Notch	10	20	22	9	244	157	110	
Phagosome	45	57	59	35	169	95	104	
Peroxisome	53	54	49	28	175	75	91	
Toll-like Receptor	20	28	27	13	208	129	96	
Rig-like Receptor	18	21	22	10	220	122	105	
Bacterial Invasion	27	37	42	21	200	135	114	
Autophagy	15	11	11	3	367	73	100	
Apoptosis	19	28	26	14	186	130	93	
p53	24	33	29	14	207	181	88	
Nod-like Receptor	12	19	16	6	267	133	84	
NF-kB	17	30	26	10	260	153	87	
Tumor Necrosis Factor	23	42	35	18	194	146	83	
PI3K-Akt	63	107	93	67	139	126	87	
				Mean	219	130	96	
			Stand	dard Deviation	52	30	10	

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.

Accession	Protein Name	O. faveolata Homologues
094817	Ubiquitin-like protein ATG12	comp13112
Q9H1Y0	Autophagy protein 5	comp32308
Q14790	Casapse	comp42756, comp44852, comp43067, comp41842, comp39349
P02778	C-X-C motif chemokine 10	-
Q9NQC7	Ubiquitin carboxyl-terminal hydrolase CYLD	comp6468, comp45618
O00571	Rig-like receptor domain-containing proteins	comp42383, comp46396, comp45021, comp27027, comp31533, comp52755, comp12373, comp44096
Q3LXA3	Triokinase	comp29735
Q13158	FAS-associated death domain protein	-
P25963	NF-kappa-B inhibitor A/B	comp4914, comp6407
015111	Inhibitor of NF-kappa-B	comp44188
Q14653	Interferon regulatory factors	comp25035, comp40113, comp39249, comp50738, comp30904
P05161	Ubiquitin-like protein ISG15	comp41425, comp36790, comp36790
Q13233	MAPKs	comp45734, comp44708, comp32445, comp45894
043318	TGF-beta-activated kinase 1	comp110885, comp56867, comp14096
P45983	JNKs	comp10734, comp44864, comp54242, comp60240, comp38508
P19838	NFkBs	comp46270, comp44367
Q96G74	OTU domain-containing protein 5	comp52747, comp41206
Q13526	Rotamase Pin1	comp40298
Q13546	Serine/threonine-protein kinase RIP	-
Q96EQ8	E3 ubiquitin-protein ligase RNF125	-
Q9BRV8	Suppressor of IKBKE 1	comp43257
Q86WV6	Stimulator of interferon genes protein	comp46735, comp115636
Q9UHD2	Serine/threonine-protein kinase TBK1	comp66935
Q12933	TNF receptor-associated factors	comp46745, comp45812, comp43792, comp34933, comp60135, comp43035, comp46671
Q14258	E3 ubiquitin/ISG15 ligase TRIM25	-