SCIENTIFIC REPORTS

OPEN

Comparative analysis of the genomes of *Stylophora pistillata* and *Acropora digitifera* provides evidence for extensive differences between species of corals

Christian R. Voolstra[1], Yong Li[1], Yi Jin Liew[1], Sebastian Baumgarten[1,4], Didier Zoccola[2], Jean-François Flot[3], Sylvie Tambutté[2], Denis Allemand[2] & Manuel Aranda[1]

Stony corals form the foundation of coral reef ecosystems. Their phylogeny is characterized by a deep evolutionary divergence that separates corals into a robust and complex clade dating back to at least 245 mya. However, the genomic consequences and clade-specific evolution remain unexplored. In this study we have produced the genome of a robust coral, *Stylophora pistillata*, and compared it to the available genome of a complex coral, *Acropora digitifera*. We conducted a fine-scale gene-based analysis focusing on ortholog groups. Among the core set of conserved proteins, we found an emphasis on processes related to the cnidarian-dinoflagellate symbiosis. Genes associated with the algal symbiosis were also independently expanded in both species, but both corals diverged on the identity of ortholog groups expanded, and we found uneven expansions in genes associated with innate immunity and stress response. Our analyses demonstrate that coral genomes can be surprisingly disparate. Future analyses incorporating more genomic data should be able to determine whether the patterns elucidated here are not only characteristic of the differences between *S. pistillata* and *A. digitifera* but also representative of corals from the robust and complex clade at large.

Coral reefs are ecologically and economically highly important marine ecosystems, as they provide biodiversity hotspots for a large diversity of species and serve as a food source for millions of people[1,2]. Coral reefs are threatened by a combination of local (e.g., overfishing, eutrophication, pollution) and global (e.g., ocean warming and ocean acidification) factors[3–5]. Over the last decades, coral reef cover was significantly decimated and one-third of reef-building corals face elevated extinction risk from climate change and local impacts[6,7]. For this reason, it is important to understand the factors that contribute to ecosystem resilience.

At the heart of these ecosystems are the so-called coral holobionts, which provide the foundation species of reefs and consist of the coral animal host, its endosymbiotic photosynthetic algae, and a specific consortium of bacteria (among other organisms)[8,9]. While recent research highlights the contribution of all holobiont compartments to coral resilience[10–14], the majority of studies focus on the diversity of algal and bacterial symbionts associated with corals or on gene expression of the host under an array of stressors or across different environments[15–22]. Hence, although coral species display differing sensitivities to environmental stress[10], the genomic underpinnings of coral resilience are not clear.

A recent study by Bhattacharya, *et al.*[23] conducted a comparative analysis incorporating genomic and transcriptomic data from 20 coral species. Focusing on single-copy orthologs conserved across all analyzed corals (i.e., gene orthologs that have a clear one-to-one relationship across all coral species considered), the authors

[1]Red Sea Research Center, Division of Biological and Environmental Science and Engineering (BESE), King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia. [2]Centre Scientifique de Monaco, 8 quai Antoine Ier, 98000 Monaco, Monaco. [3]Université libre de Bruxelles, Avenue Franklin Roosevelt 50, 1050 Bruxelles, Belgium. [4]Present address: Biology of Host-Parasite Interactions Unit, Institut Pasteur, 25 rue du Dr Roux, 75015 Paris, France. Correspondence and requests for materials should be addressed to C.R.V. (email: christian.voolstra@kaust.edu.sa) or M.A. (email: manuel.aranda@kaust.edu.sa)

describe the presence of a variety of stress-related pathways (e.g., apoptotic pathways, reactive oxygen species scavenging pathways, etc.) that affect the ability of corals to respond to environmental stress. Importantly, the authors could show that corals harbor a highly adaptive gene inventory where important genes arose through horizontal gene transfer or went through rounds of evolutionary diversification. Similarly, the available genome of *Acropora digitifera*[24] highlights that the innate immunity repertoire of corals is presumably and notably more complex than those of the cnidarians *Nematostella vectensis*[25] and *Hydra magnipapillata*[26]. Seemingly so, the innate immunity repertoire of *A. digitifera* is also more complex than that of the symbiotic anthozoan sea anemone *Aiptasia*[27]. This has potential implications for our understanding of coral responses to environmental change, as a more complex innate immune system might reflect adaptations to better cope with environmental stress and increased flexibility to pathogen response[27,28]. Unfortunately, it is not straightforward to determine what a 'typical' coral genome looks like. This is because the phylogeny of scleractinian corals is characterized by a deep evolutionary split that separates corals into a robust and complex clade dating back to at least 245 mya[29,30]. Hence, several important questions, such as how well the available genome of *A. digitifera* indeed reflects general coral-specific traits or to what extent species from both coral clades diverged since their separation (giving rise to different adaptations), are currently unanswered due to the dearth of coral genomes. To this end, the Reef Future Genomics (ReFuGe) 2020 consortium has formed to sequence 9 hologenomes of coral species representing different stress susceptibilities in order to better understand conserved and lineage-specific traits, but a comprehensive analysis is pending[10]. In addition, Prada, *et al.*[31] sequenced the genomes of *Orbicella annularis*, *Orbicella faveolata*, and *Orbicella franksi*. However, at present, the accompanying genomic gene sets are not available.

In this study, we produced and analyzed the genome of *Stylophora pistillata*, a representative of the robust clade of corals, and compared it to the available genome of the complex coral *A. digitifera*. We were specifically interested in a comparison of (1) the set of orthologous genes, (2) genes that were independently expanded in either of the genomes or both, and (3) species-specific genes. These three classes of genes, we reasoned, provide complementary insight into the evolutionary history of both corals and may highlight important species-specific adaptive processes[32,33]. Further, such a comparative analysis may pinpoint genomic differences that arose from the different evolutionary trajectories that occurred in coral species from either clade and, as such, may represent clade-specific differences.

Results

Genome size and genic composition.
We assembled 400 Mb of the genome of the coral *S. pistillata* (Table 1, Fig. S1) with a scaffold N50 of 457 kb, representing ~92% of the 434 Mb genome as estimated via FACS (fluorescent activated cell sorting) (Fig. S2). 358 Mb were assembled into contigs, with a contig N50 of about 24 kb (Tables 1, S1). We identified 25,769 protein-coding genes encoded in the *S. pistillata* genome, of which 89% retrieved functional annotation from protein databases (Tables 1, S2). The genome size and the number of genes are comparable to the draft genome of *A. digitifera* that features a total scaffold length of about 419 Mb with a scaffold N50 of 191 kb and 23,523 protein-coding genes (Table 1). However, genome completeness as assessed by CEGMA[34] was considerably higher in *S. pistillata* with about 94.76% of the core eukaryotic genes present compared to 82.26% in *A. digitifera* (Table S3).

To obtain general insight into the genic composition of coral genomes, we performed a BLASTP search with the gene sets encoded in both genomes against the 'nr' protein database (see Methods). The vast majority of genes from both species had best matches to *Aiptasia* (48.36% for *S. pistillata* vs. 43.82% for *A. digitifera*) and *Nematostella* (23.54% for *S. pistillata* vs. 25.82% *A. digitifera*) (Fig. 1A). The remaining genes generally matched non-cnidarian proteins or had no matches (7.30% *S. pistillata* vs. 10.55% for *A. digitifera*), presumably representing lineage- or species-specific genes. Strikingly, when this analysis was extended to allow for inter-coral matching, pronounced differences became apparent between both coral species (Fig. 1B). In particular, we found that matches of *A. digitifera* genes to *S. pistillata* genes were highly disproportional (17,866 *A. digitifera* genes matched to 10,945 *S. pistillata* genes, $p < 10^{-300}$, Fisher's exact test), suggesting pervasive gene duplication in *A. digitifera*. In addition, *A. digitifera* exhibited significantly fewer matches to the anemones *Aiptasia* (1,942 genes) and *Nematostella* (1,011 genes) than *S. pistillata* (6,994 gene matches to *Aiptasia*, 3,437 gene matches to *Nematostella*) (Fisher's exact test, $p < 10^{-300}$ and $p < 10^{-283}$, respectively), pointing towards increased divergence of protein sequences in *A. digitifera*.

Conservation of protein-encoding genes.
We first compared the genomes at the protein level, considering 25,769 *Stylophora* and 23,523 *Acropora* protein-encoding genes. The proteins were classified into four categories according to their evolutionary relationships (Fig. 2), as deduced using InParanoid[35] (for details refer to Methods). The first category included *S. pistillata* proteins with one clearly identifiable counterpart in *A. digitifera* and *vice versa* (one-to-one orthologs). The function of these proteins is likely conserved and can be interrogated to infer core functions of coral genomes. An analysis based on 4,751 one-to-one orthologs across 20 coral species (including *S. pistillata* and *A. digitifera*) was conducted in a recent study[23], where the authors collated and queried these orthologs to elucidate four major issues in coral evolution, i.e. coral calcification, environmental sensing, symbiosis machinery, and the role of horizontal gene transfer (HGT). In our study, reciprocal best matches produced 6,302 protein pairs classified as one-to-one orthologs (24% of the *S. pistillata* and 27% of the *A. digitifera* proteins) (Dataset S1). The second category included proteins in which gene duplication has occurred in one or both species after divergence, resulting in "many-to-one" and "many-to-many" ortholog relationships, respectively. This group consisted of 2,747 *S. pistillata* and 2,900 *A. digitifera* proteins (11% of the *S. pistillata* and 12% of the *A. digitifera* proteins) and presumably harbor genes that expanded independently in both lineages (Dataset S2). We hypothesized that the presence of species-specific gene expansions likely reflects functions relevant to either species- or clade-specific evolution. The third category included 15,442 *S. pistillata* and 12,925 *A. digitifera* proteins (60% and 56% of the proteins, respectively) that have homologs in corals or other species, but

|	*Stylophora pistillata*	*Acropora digitifera*
Genome		
Genome file used	v1.0	v1.0 (Jul 2011)
Estimated genome size (Mb)	434	420
Total scaffold length (bp)	400,108,361	419,317,576
Scaffold N50 (bp)	457,453	191,489
Total contig length (bp)	358,078,850	364,965,673*
Contig N50 (bp)	24,388	10,700*
GC content, N excluded (%)	38.5	39
Genes		
Number of genes (longest transcript per locus)	25,769	23,523
Mean gene length (bp)	8,432	N/A
Gene model EST support (%)	82.1	78*
Exons		
Mean coding region length (bp)	2,086	1,707*
Number of exons per gene	7.9	7.0*
Mean length (bp)	266	230
Total length (Mb)	53.8	40.2
Introns		
Genes with introns (%)	96.6	N/A
Mean length (bp)	918.3	N/A
Total length (Mb)	162.1	N/A
Average length (bp)	6,333	N/A

Table 1. Assembly statistics for the genomes of *S. pistillata* and *A. digitifera*. *From Shinzato *et al.*[24]. Some statistics were not available, as the *A. digitifera* v1.0 gff3 file was not made public.

without easily discernable orthologous relationships between both corals. The high number of this group of proteins likely also reflects our conservative approach for ortholog identification (see Methods). Finally, the fourth group consisted of 1,278 *S. pistillata* and 1,396 *A. digitifera* proteins that have no detectable homologs in any other species. These species-specific proteins putatively belong to the class of taxonomically restricted genes (TRGs) and might be encoded by fast-evolving genes[32].

Core set of conserved proteins. The notion that one-to-one orthologs constitute a core of conserved functions that encode for basic biological processes was corroborated by a Gene Ontology (GO)-based analysis (Dataset S3). The 50 most common GO terms were associated with regulation of metabolism and cellular processes, organelle function, and notably, nitrogen-related metabolic processes, the regulation of which were previously discussed as central to coral holobiont functioning[36] and putatively important in coral bleaching[37]. In line with a functional conservation of these proteins, the average sequence identity of the one-to-one orthologs of *S. pistillata* and *A. digitifera*, which are presumably at least 245 mya apart[29], was 62% on the protein level (Dataset S4).

To test whether differences in average sequence similarity coincided with distinct functional classes in this group of highly conserved proteins, we divided the set of one-to-one orthologs into three categories, i.e. orthologs displaying ≥50%, between <50% and ≥30%, and those displaying <30% sequence similarity, and tested for Gene Ontology enrichment[33] (Dataset S4). Orthologs with ≥50% sequence conservation were enriched for genes associated with transcription, translation, and ribosomes. In comparison, orthologs displaying similarities between 50% and 30% were enriched for genes associated with endocytosis, immune system activation (NF-κB), and superoxide metabolic processes, which putatively play a role in the endosymbiosis. In contrast, orthologs with similarities <30% were enriched for processes playing a role in cell adhesion, cell junctions, and calcium ion binding, which may be associated with resistance to ocean acidification[38]. Thus, even within this group of highly conserved proteins differing levels of sequence similarity associate with enrichment of distinct functional classes, where most highly conserved proteins align with most basic housekeeping processes.

Gene expansions and reductions identify a set of common and species-specific processes related to cnidarian-dinoflagellate endosymbiosis. In the category of orthologs with "many-to-one" and "many-to-many" relationships, a functional enrichment analysis using GO information highlighted several biological processes that were enriched. Several of these processes were shared between both coral species, although the majority of enriched processes were species-specific (Fig. 3A, Dataset S5). The common processes included several immunity-related GO categories associated with the regulation of NF-κB and in particular interferon production, but also categories related to cell adhesion and bicarbonate transport. All of these processes have been implicated in the cnidarian-dinoflagellate endosymbiosis[27,39]. Immunity-related GO terms were also present in the enriched categories specific to *S. pistillata*, but other GO terms prevailed. For instance, several processes related to receptor-mediated endocytosis, amine metabolism, osmosis, apoptosis, and hyperoxia were
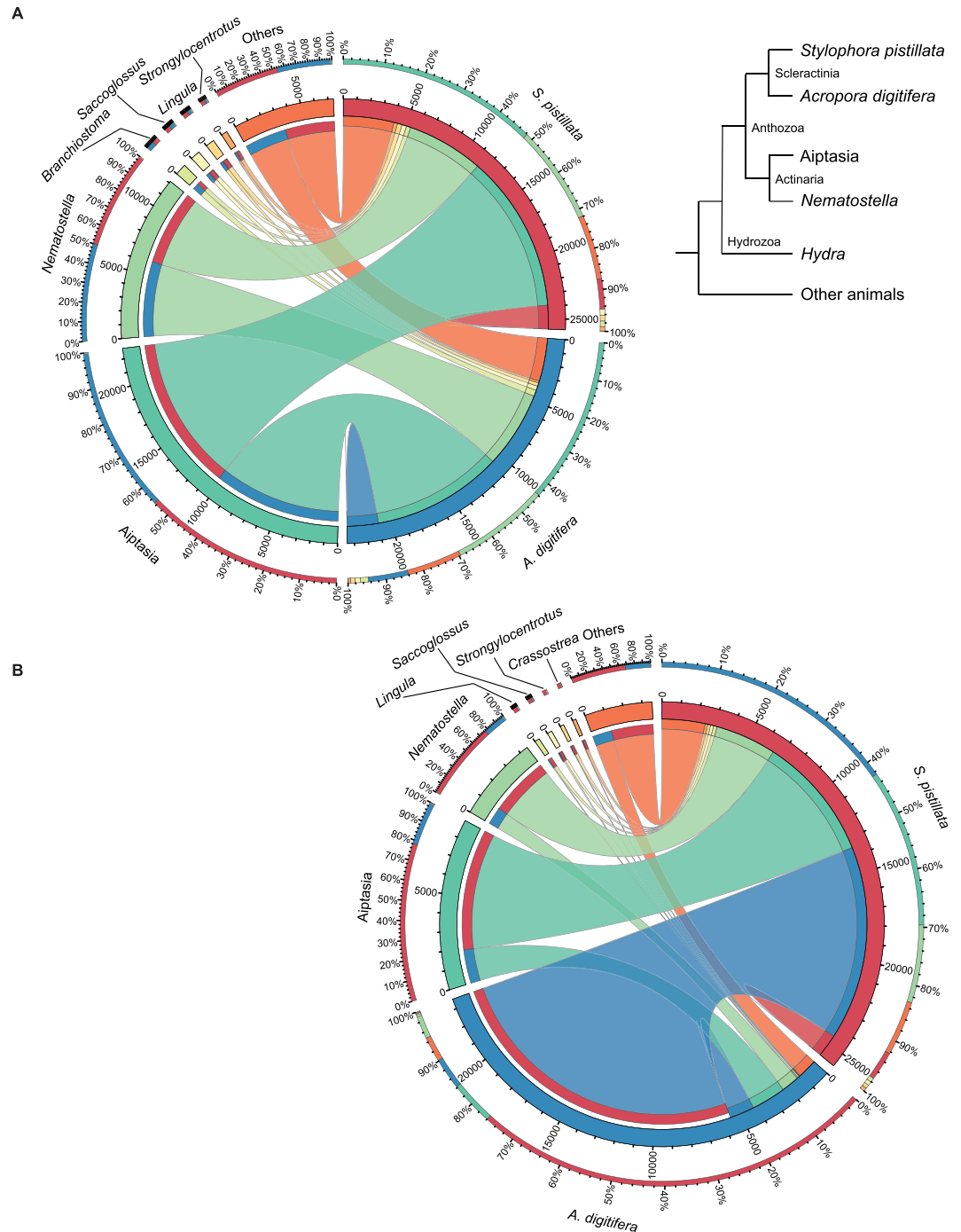
Figure 1. Comparison of genic composition of the genomes of *S. pistillata* and *A. digitifera*. Genes were assigned to closest genera based on BLASTP best hits, and distribution was visualized in chord diagrams (top 6 genera are shown, remaining hits are summed up in 'others'). The inlet shows a simplified phylogeny of the phylum Cnidaria to ease orientation and interpretation. (**A**) Both coral genomes appear similar in composition when queried against the combined NCBI 'nr' database and the *Aiptasia* gene models. The proportion of matching genes in *S. pistillata* and *A. digitifera* to other organisms is color-coded as follows: *S. pistillata/A. digitifera* genes with best matches to *Aiptasia* (dark green), to *Nematostella* (light green), to *Branchiostoma* (lancelet, greenish-yellow), to *Saccoglossus* (acorn worm, yellow), to *Lingula* (brachiopod, dark yellow), to *Strongylocentrotus* (sea urchin, orange), and to 'others' (dark orange). Genes without any matches are in red (*S. pistillata*) and blue (*A. digitifera*). (**B**) If coral genomic gene sets are included (i.e., genes from *S. pistillata* are allowed to match against *A. digitifera* and *vice versa*), many more *A. digitifera* genes match homologs in *S. pistillata* homologs than *vice versa*. As a consequence of the asymmetrical matching, the number of *A. digitifera* matches to other genera is vastly reduced. The proportion of matching genes between *S. pistillata* and *A. digitifera* to each other and other organisms is color-coded as follows: *S. pistillata* and *A. digitifera*

genes matching each other (blue); *S. pistillata*/*A. digitifera* genes with best matches to *Aiptasia* (dark green), to *Nematostella* (light green), to *Lingula* (brachiopod, dark yellow), to *Saccoglossus* (acorn worm, yellow), to *Strongylocentrotus* (sea urchin, dark yellow), to *Crassostrea* (oyster, orange); and to 'others' (dark orange); *S. pistillata* genes without any matches (red).

....

enriched. Again, these processes are conceivably related to the symbiotic relationship with zooxanthellae[27,39]. In *A. digitifera* we also found enrichment of processes related to innate immunity, in particular of genes associated with the Toll signaling pathway, interleukin production, and bacterial detection. Other enriched processes specific to *A. digitifera* were notably different however, such as miRNA metabolism, cytoskeletal organization, hydrogen peroxide metabolism, proteolysis, and pyroptosis.

Uneven expansions of proteins related to the immune system.

The group of proteins with gene duplications in one or both species revealed many uneven expansions (or reductions), as highlighted by the observation that in many cases a single protein in either species had multiple counterparts in the other species. Genes experiencing multiple rounds of duplications in either one or both species are arguably among the most interesting proteins to look at, as they may reveal information on processes independently selected in either or both species. For this reason, we further looked into ortholog groups with at least 3 proteins in either one or both corals (Dataset S6, Dataset S7).

We identified 76 ortholog groups with ≥3 genes in both *S. pistillata* and *A. digitifera* (Dataset S6). Both coral species expanded genes related to innate immunity receptors. For instance, we discovered 3 cases where a gene encoding for a NOD-like receptor family member (NLRC3) was independently expanded in both corals (Fig. 3B, Dataset S6). Further, we found 1 case where a gene encoding for a TLR (Toll-like receptor 1) and another case where a gene encoding for a TNFR (Tumor necrosis factor receptor superfamily member 19) showed independent duplication in both corals (Fig. 3B, Dataset S6). Importantly, the ortholog groups showed different degrees of expansions in both corals. In three of the above cases, we found more duplicated genes in *A. digitifera* than in *S. pistillata*. In contrast, in only one of the above cases we found more genes in *S. pistillata* than *A. digitifera* (Fig. 3B, Dataset S6). We found an extreme case of expansion for one of the NOD-like receptor family members, where *A. digitifera* harbored 52 proteins in comparison to *S. pistillata* that harbored only 5 proteins (Fig. 4A, Dataset S6). Overall, *A. digitifera* had a stronger tendency to show 'extreme' expansions (10 ortholog groups with more than 10 proteins) than *S. pistillata* (3 ortholog groups with more than 10 proteins) (Dataset S6). Besides innate immunity receptors, we also found 2 ortholog groups related to biomineralization, i.e. homologs of Carbonic Anhydrase (CA) and Bone Morphogenetic Protein 1 (BMP-1)[40], to be expanded in both species, but at very similar levels (3 vs. 4 proteins for CA and 3 proteins each for BMP-1 for *A. digitifera* and *S. pistillata*, respectively) (Fig. 3B, Dataset S6).

Orthologs that harbored gene duplications in only of the coral species were present at a similar number in *S. pistillata* and *A. digitifera* (Dataset S7). We identified 191 genes in *A. digitifera* that mapped to ortholog groups with ≥3 genes in *S. pistillata*. Among these, we found homologs of innate immunity related proteins, namely TRAF3 (TNF receptor-associated factor 3) (Fig. 4B) and TLRs (Toll-like receptor 2), to be present with 3 copies in *S. pistillata* mapping to 1 counterpart in *A. digitifera*. Also, a homolog of peroxidasin, important for oxidation-reduction, was present with 5 copies in *S. pistillata* that mapped to 1 corresponding counterpart in *A. digitifera* (Dataset S7). Looking at genes that were specifically expanded in *A. digitifera*, we identified 167 genes from *S. pistillata* that mapped to ortholog groups with ≥3 genes in *A. digitifera*. Most notably, a member of the NOD-like receptor family member (NLRC3) gave rise to 55 genes in *A. digitifera* with 1 counterpart in *S. pistillata* (Fig. 4C).

Discussion

The inference of evolutionary relationships within the Scleractinia is an ongoing subject of debate, complicated by the phenotypic plasticity in skeletal growth forms and unusual slow mitochondrial DNA sequence evolution[41,42]. Nevertheless, a major distinction into two clades ("complex" and "robust" corals) within the Scleractinia dating back to at least 245 mya[29] is corroborated by molecular analyses[41,43,44]. In line with this estimate, scleractinian corals first appeared in the fossil record about 245 mya[45]. However, the genomic consequences of this deep divergence remain unexplored. In this study we have assembled the genome of the robust coral *S. pistillata* and compared it to the available genome of the complex coral *A. digitifera* to gain a first look at coral species differences from both clades on a genomic scale (Fig. 1).

To thoroughly understand the extent of conservation at the protein level, we followed an ortholog-based approach where we assigned proteins into four categories according to their evolutionary relationships: (i) one-to-one orthologs, (ii) many-to-one and many-to-many orthologs, (iii) proteins without easily discernible orthologous relationships, and (iv) species-specific proteins without homologs in other species (Fig. 2). Importantly, our analysis complements the approach followed by Bhattacharya, *et al.*[23] in that the former study focused on the conserved set of single-copy orthologs to identify a core set of proteins and processes present in all corals that were subsequently compared to other lineages (of note, the set of single-copy orthologs was retrieved from the same genome assemblies of *A. digitifera* and *S. pistillata* as employed here). By contrast, in the current analysis, besides the analysis of single-copy orthologs, we were in particular focusing on the set of commonly and species-specific expanded genes to gain insight into the evolutionary history of both corals and to highlight potential species-specific adaptive processes. In other words, the analysis by Bhattacharya, *et al.*[23] centered on the commonalities between coral species, whereas the present study looked at differences between coral species. Given that analyses focusing on differences (i.e., presence vs. absence of genes) are particularly error-prone using transcriptome data, we restricted our analysis to the genomic gene sets.
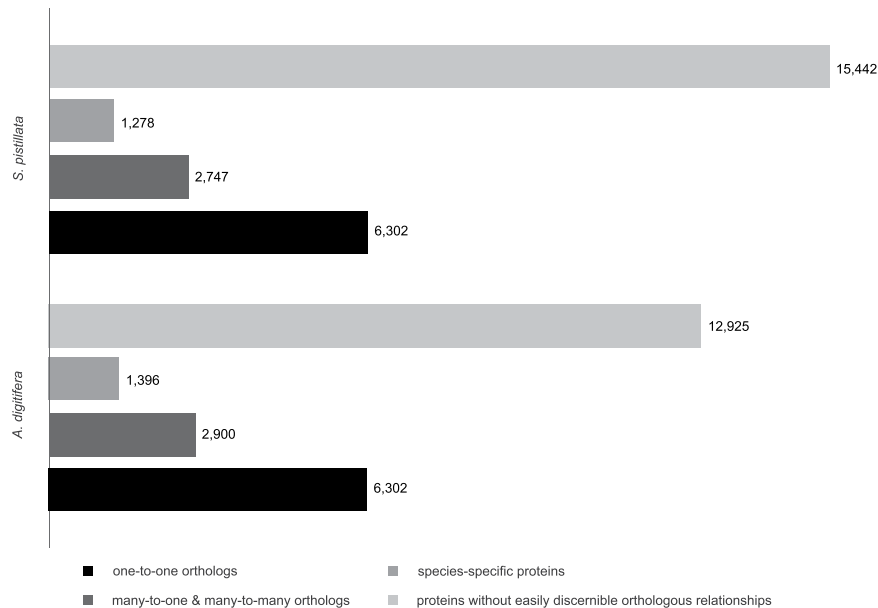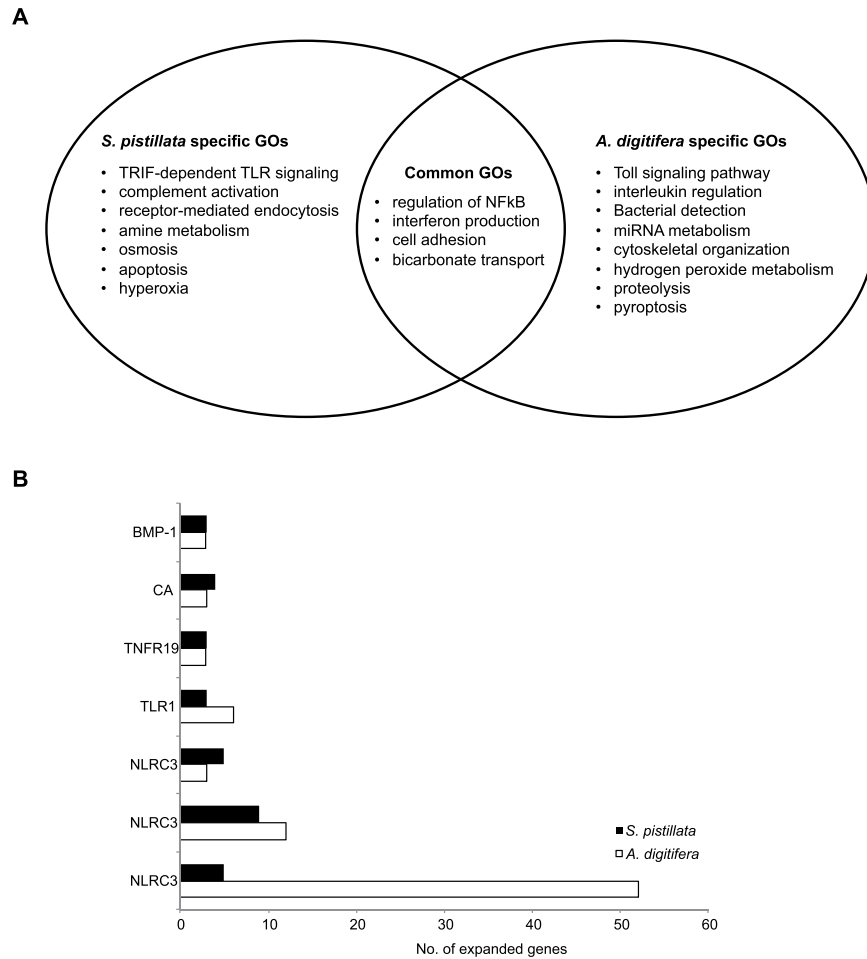
Figure 2. Classification of genomic protein sets of *S. pistillata* and *A. digitifera* according to evolutionary relationship. 25,769 *S. pistillata* proteins were compared to 23,523 *A. digitifera* proteins and assigned to four categories: (i) one-to-one orthologs (black), (ii) many-to-one and many-to-many orthologs (dark grey), (iii) species-specific proteins without homologs in other species (intermediate gray), and (iv) proteins without easily discernible orthologous relationships (light grey).

Notably, more than half of the proteins from both species (i.e., 15,442 *S. pistillata* and 12,925 *A. digitifera* proteins without easily discernable orthologous relationships plus the 1,278 *S. pistillata* and 1,396 *A. digitifera* species-specific proteins) could not be assigned clear orthologous relationships, putatively indicating a substantial divergence associated with the deep evolutionary split between both corals. This is further corroborated by the genic composition results, which shows that more than two thirds of proteins in both species match to homologs in other cnidarian species (*Aiptasia* and *N. vectensis*) (Fig. 1).

Of the remaining other half of proteins from the ortholog-based analysis, about two thirds of the proteins (6,302 protein pairs) displayed clear one-to-one relationships and can be considered core proteome members. This number closely resembles the number of one-to-one orthologs identified by Bhattacharya *et al.*[23] in a large meta-analysis of available coral genomes and transcriptomes (4,751 ortholog pairs). Within the set of conserved orthologs across corals, the authors characterized sets of proteins responsible for biomineralization, environmental sensing, and response to temperature, light, and pH. Our analysis of enriched GO terms largely supports the results of the study by Bhattacharya *et al.*[23] and further highlights the overarching emphasis on processes related to the cnidarian-dinoflagellate symbiosis in the coral host core set of conserved proteins. In line with functional conservation, the set of one-to-one orthologs revealed a comparatively high degree of sequence conservation of 62% at the protein level (Dataset S4). By comparison, the average sequence identity between *Anopheles* and *Drosophila*, which are separated by approximately the same time (~250 mya)[46], was estimated to be 56% in a previous study[33]. This indicates that despite the comparable divergence time in both pairs of species, coral proteins diverge at a lower rate than insect proteins, possibly because corals have much longer generation times[10], although a substantial portion of the genome can evolve at elevated rates[47].

A particular interesting category of orthologs is comprised of those with many-to-one and many-to-many relationships. Family expansion and reduction can be measured in different ways. The most basic measure is to annotate proteins to their domains and compare the normalized domain count between genomes[48]. Although this is a straightforward way to assess overall similarities and differences between genomes, it does not provide information on the relatedness of proteins with the same domains/domain compositions. A better resolution is provided by an analysis of the enriched functions of the many-to-one and many-to-many orthologs. About 10% of proteins from both genomes fall into this category. Although this group is less strictly defined than the group of one-to-one orthologs, they can still be assigned to a single ancestral gene, and hence, imply duplication within the species, i.e. after both species diverged. As such, analysis of these proteins allows inferences on adaptations, e.g. to different environments or life strategies. Following this reasoning, we interrogated the group of orthologs with many-to-one and many-to-many relationships in order to determine similarities and differences in evolutionary trajectories for the two coral species under investigation, as differentiation in function are suggested by increases and decreases in gene family sizes.

This analysis revealed several striking features. First, among the shared enriched processes for this category of orthologs, we found many processes directly related to the cnidarian-dinoflagellate symbiosis (Fig. 3A, Dataset S5). This partially resembles the results from the one-to-one ortholog analysis with the important difference that independent expansions of genes that map to common processes indicate that cnidarian-dinoflagellate

A

S. pistillata specific GOs

• TRIF-dependent TLR signaling
• complement activation
• receptor-mediated endocytosis
• amine metabolism
• osmosis
• apoptosis
• hyperoxia

Common GOs

• regulation of NFkB
• interferon production
• cell adhesion
• bicarbonate transport

A. digitifera specific GOs

• Toll signaling pathway
• interleukin regulation
• Bacterial detection
• miRNA metabolism
• cytoskeletal organization
• hydrogen peroxide metabolism
• proteolysis
• pyroptosis

B

Figure 3. Overview of processes and proteins associated with gene expansions in *S. pistillata* and *A. digitifera*. (**A**) Gene Ontology (GO) enrichment of Biological Process (BP) category ($p < 0.05$) of many-to-one and many-to-many orthologs for *S. pistillata* and *A. digitifera* (Dataset S5). (**B**) Ortholog groups with independently expanded genes in both coral species show different degrees of expansions in both corals and featured many immunity-related genes (Dataset S6). Notably, the NLRC3 examples belong to three distinct ortholog groups (Dataset S6). BMP-1 = Bone Morphogenetic Protein 1; CA = Carbonic Anhyrdase; TNFR19 = Tumor necrosis factor receptor superfamily member 19; TLR1 = Toll-like receptor 1; NLRC3 = NOD-like receptor C3.

symbioses are actively being shaped within coral species and that different hosts seem to converge on the same processes, indicating convergent evolution. Indeed, this seems to be the case for the endosymbionts, as shown by a recent genomic analysis between three *Symbiodinium* genomes[48], arguing that endosymbiosis affects evolution of hosts and symbionts alike. Future studies should explore this pattern further. For instance, genes being expressed during *Symbiodinium* exposure to larvae of *A. digitifera*[49] could be compared to expression patterns of symbiotic larvae of *S. pistillata*. This would also allow to further look into the differences between spawning and brooding corals. Since the latter release planula larvae with *Symbiodinium* symbionts, mechanisms of symbiont selection might underlie different evolutionary pressures, potentially explaining the here-observed differences. Second, within these processes, homologs of the same or similar genes are repeatedly being expanded across species, as highlighted by the example of three cases of expansion of NLRC3 (Fig. 3B, Dataset S6). This suggests that the same genes are potentially subjected to adaptation within and between coral species, arguing that convergent evolution might not only happen on the process level, but also at the protein level. Further evolutionary analysis incorporating more species might provide an avenue to identify genes important to coral adaptation. Last, even though we find expansions of the same genes between species, the extent of duplication is in some cases extremely uneven (Fig. 4A, Dataset S6). Again, incorporation of genomic data from additional species in the near future should help to answer whether this pattern is restricted to the species compared here or whether it is a general feature of coral (clade) evolution.

In particular, when considering ortholog groups that play a role in innate immunity, the emerging pattern is that both coral species independently expanded ortholog groups belonging to TLRs, TNFRs, and NLRs. These genes represent innate immune sensors and signaling proteins that recognize extracellular and intracellular microbial patterns or can activate downstream innate immunity signaling cascades. This is of relevance to the ability of a cnidarian host to distinguish among potential symbionts, pathogens, and particles of food[27]. Hence,
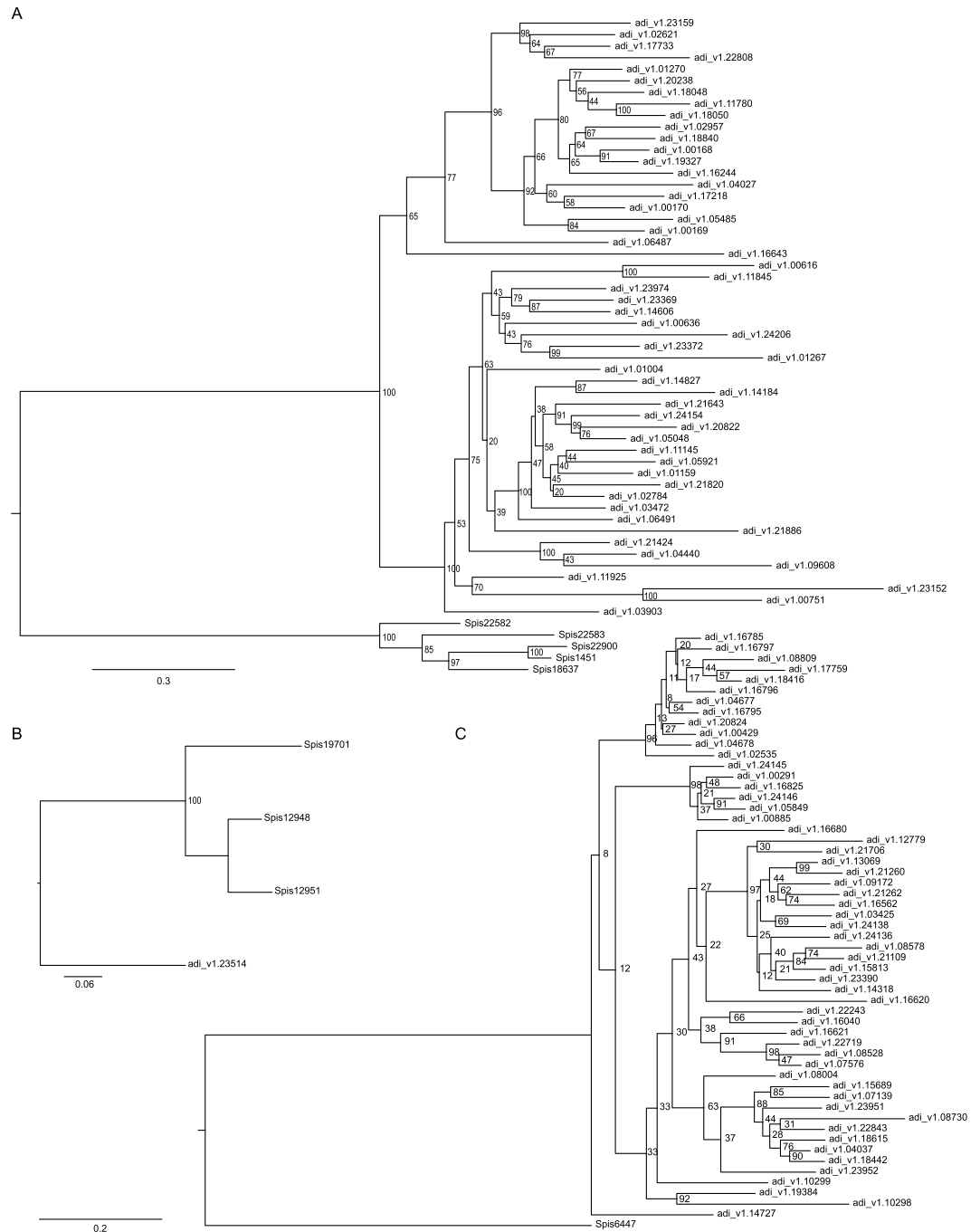
Figure 4. Gene expansion of orthologs in *S. pistillata* and *A. digitifera*. Maximum Likelihood phylogenetic trees were constructed using RAxML with 1,000 bootstraps. Bootstrap support is indicated at branches of phylogenetic trees. (**A**) Ortholog expansion displaying a many-to-many relationship for a NOD-like receptor family member, in which *A. digitifera* harbors 52 proteins in comparison to *S. pistillata* that harbors only 5 proteins. (**B**) Ortholog expansion displaying a many-to-one relationship of a TRAF (TNF receptor-associated factor 3) homolog with expansion in *S. pistillata* (3 genes) and one corresponding counterpart in *A. digitifera*. (**C**) Ortholog expansion displaying a many-to-one relationship of a member of the NOD-like receptor family member (NLRC3) with a particularly pronounced expansion in *A. digitifera* (55 genes) and only one corresponding counterpart in *S. pistillata*.

a more complex innate immunity network may partially reflect adaptations associated with the endosymbiosis with *Symbiodinium* and the ability to counter environmental insult by pathogens[24,27,39]. Our results from the analysis of independently expanded ortholog groups belonging to TLRs, TNFRs, and NLRs echoes the analysis of Shinzato *et al.*[24], where the authors found that the *A. digitifera* repertoire of Toll/TLR-related receptors was substantially more complex and diverse than that of *Nematostella*. Also, our findings are further in line with the extensive expansion of NLRs in *A. digitifera* as reported by Hamada *et al.*[50]. Last, our results are in line with

Baumgarten et al.[27] and Poole and Weis[51] who found that the TLR/ILR protein repertoires of the symbiotic sea anemone *Aiptasia* show close similarity to *N. vectensis* with apparently lineage-specific expansion in *A. digitifera*. From the admittedly limited analysis of two coral genomes, it appears, however, that *A. digitifera* shows a more pronounced tendency to duplicate genes in the ortholog groups that are expanded in both corals (many-to-many) (Dataset S6). This pattern was also apparent when considering innate immunity-related ortholog groups that are expanded in only one of both corals (many-to-one) (although expansions were similar when considering all ortholog groups) (Dataset S7). Hence, it will be interesting to see whether *A. digitifera* (and perhaps other *Acropora* species) represent indeed 'extreme' cases of expansion of innate immunity-related genes (even within corals) and whether this might even be a hallmark of coral species from the complex clade. With more coral genomes expected to becoming available soon[10], this will be a fascinating question to pursue. Complementary to this, expression-guided studies might help to further illuminate this point by provoking an immune response in a range of coral species and comparing overall expression patterns as well as identifying the genes and pathways that respond. This might also help to grasp why different coral species harbor such physiologically diverse tolerance levels to environmental stressors[10]. Further along this line of thought, the analysis of many-to-one ortholog groups also revealed that *A. digitifera* and *S. pistillata* seem to diverge on which innate immunity-related genes are expanded. In *A. digitifera* we find NLRs, whereas in *S. pistillata* we find TLRs and TRAF3 to be preferentially expanded. Thus, it will be interesting to determine whether these evolutionary differences might help to pinpoint groups of genes or individual proteins that determine differential specificity to algal symbionts or can be related to differences in physiology, such as thermal tolerance, stress resilience, symbiont transmission mode, and others[10]. In this context, it is important to note that the endosymbiosis with *Symbiodinium* evolved independently a number of times in the Triassic period[52]. In the case of zooxanthellate scleractinian corals, current estimates suggest that the symbiosis evolved about 210 mya[52]. Following this estimate would suggest that complex and robust corals acquired the ability to form symbioses independently. Further research on this represents an important undertaking, in particular with regard to our ability to better understand flexibility of forming symbioses with different *Symbiodinium* taxa[19,53].

Taken together, our analyses corroborate recent comparative genomic analyses that showcase how the proteomic information stored in coral genomes has provided the foundation for adapting to a symbiotic, sessile, and calcifying lifestyle of scleractinian corals[23,24,40]. In particular, our analyses of the core set of conserved proteins and the set of independently expanded ortholog groups in both species underscore the putative importance of the endosymbiotic relationship in determining evolutionary patterns. At the same time, our results demonstrate that coral genomes can be surprisingly disparate as highlighted by extremely uneven or independent expansions of some ortholog groups. It will be important to determine whether the patterns described here extend to differences between coral clades and, most importantly, if they are predictive and relevant to a coral's ability to respond to environmental change.

Methods

Organism and isolation of genomic DNA.   Colonies of *S. pistillata* collected at a depth of 5 m in front of the Marine Science Station, Gulf of Aqaba, Jordan[54], were transferred and maintained at the Centre Scientifique de Monaco in aquaria supplied with flowing seawater from the Mediterranean Sea (exchange rate: 2% h^{-1}) at a salinity of 38.2 ppt, pH 8.1 ± 0.1 under an irradiance of 300 μmol photons m^{-2} s^{-1} at 25 ± 0.5 °C. Corals were fed three times a week with frozen krill and live *Artemia salina* larvae. Based on nuclear ITS and mitochondrial COI, coral colonies were typed to be *S. pistillata* clade 4, which is found throughout the northwest Indian Ocean including the Red Sea, the Persian/Arabian Gulf and Kenya[55] (Fig. S1). DNA for sequencing libraries was extracted from *S. pistillata* nubbins using a nuclei isolation approach to minimize contamination with algal symbiont DNA. Briefly, cells from a *S. pistillata* nubbin of about 3 cm were harvested in 50 ml of 0.2 M EDTA solution using a water pick and refrigerated at 4 °C for 2 hours. Extracts were first passed through a 100 μm and subsequently through a 40 μm cell strainer (Falcon, Corning, Tewksbury MA, USA) to eliminate most of the zooxanthellae. Next, extracts were centrifuged at 2,000 g for 10 min at 4 °C. The supernatant was discarded and the resulting pellets were homogenized in lysis buffer (G2) of the Qiagen Genomic DNA isolation kit (Qiagen, Hilden, Germany). DNA was extracted following manufacturer's instructions using genomic-tip 100/G. DNA concentration was determined by optical density using an Epoch Microplate Spectrophotometer (BioTek, Winooski, VT, USA). A check for potential co-isolation of *Symbiodinium* DNA was assessed via PCR targeting the multicopy gene RuBisCO (Genbank accession number AY996050) and did not yield any amplification. The primers used were 5′-GGCGATGCCTCAGACAAGA-3′ (forward) and 5′-TGGGAGTGGTCTGCTTCATG-3′ (reverse), detailed in Moya, et al.[56].

Genome size estimation.   To assess genome size and validate the bioinformatically estimated genome size, we performed a physical measurement of nuclei DNA content using chicken red blood cells (CRBC) as a reference (DNA QC Particles kit, BD Biosciences, San Jose, CA, USA). Extraction and staining of nuclei were performed following the 'CyStain PI absolute T' kit (PARTEC #05-5023, Partec, Muenster, Germany) following the manufacturer's recommendation. Briefly, cells from *S. pistillata* from a nubbin of about 3 cm were harvested using a Water Pick in 50 ml of 0.2 M EDTA solution refrigerated at 4 °C and centrifuged at 2,000 g for 10 min at 4 °C. The cell pellet was resuspended in nuclei extraction buffer, incubated for 15 min at 22 °C, and subsequently filtered through a 40 μm cell strainer. Cell lysates were stained with propidium iodide for 60 min at 22 °C, protected from light. Fluorescence signals from nuclear suspensions of separate (i.e., *S. pistillata* or CRBC) and mixed nuclei (i.e., *S. pistillata* and CRBC) were measured on a LSRII Fortessa (BD Biosciences, San Jose, CA, USA) using a 561 nm laser and BP605/40 filter. Based on the known diploid DNA content of chicken erythrocytes of 2.33 pg per cell, coral genome size calculation was determined as follows: sample genome size [pg] = 1.165 x/y (x: fluorescence intensity of unknown sample; y: fluorescence intensity of CRBC). After calculating mean DNA content per copy of genetic information (1 C), genome size was determined by considering that 1 pg DNA equals 978 Mb. The measurements yielded an estimated *Stylophora pistillata* haploid genome size of 434 Mb (Fig. S2).

Genome sequencing and assembly.   Sequencing libraries were prepared using the Illumina TruSeq DNA kits for paired-end or mate-pair libraries respectively according to the manufacturer's instructions. A total of 5 paired-end and 8 mate-pair libraries were generated and sequenced on the Illumina HiSeq platform at the KAUST Bioscience Core Facility with exception of the library "miseq", which was sequenced on the Illumina MiSeq platform (Table S4). An additional mate-pair library (mp05) was generated and sequenced at GATC Biotech (Konstanz, Germany) (Table S4). All data were uploaded to NCBI and are available under Bioproject ID PRJNA281535 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA281535/).

All sequencing libraries (435x coverage) were trimmed using Trimmomatic version 0.32[57] to remove adaptors, primers, and low quality bases at the ends of sequence reads. Putative PCR duplicates were removed using FastUniq version 1.1[58] to compact the dataset for higher assembly performance. Three-pass digital normalization was performed on all paired-end libraries to reduce data redundancy with khmer[59] version 0.7.1 (k = 20 C = 20, then k = 20 C = 10). Four paired-end libraries (221x coverage) and four mate-pair libraries (96x coverage) were *de novo* assembled with ALLPATHS-LG release 48961[60] using parameter HAPLOIDIFY = True, and transcriptome data was used to scaffold the assembly with L_RNA_Scaffolder[61]. All Illumina sequencing libraries were used for scaffolding and gap filling using SSPACE version 1.2[62] and GapFiller version 1.11[63] iteratively for 3 rounds. The above-described procedure yielded an assembly of 358,078,850 bp total contig size and 400,108,361 bp total scaffold size with respective N50s of 24,388 bp and 457,453 bp. Basic genome statistics for contigs and scaffolds were generated using the perl script (http://korflab.ucdavis.edu/datasets/Assemblathon/Assemblathon2/Basic_metrics/assemblathon_stats.pl) used to validate assemblies in the "Assemblathon 2 Contest"[64]. The estimated genome size as per ALLPATHS-LG was reported at 433 Mb, and the assembled contig and scaffold lengths provide a genome coverage of ~83% and 92%, respectively. Further information regarding genome statistics are provided in Tables 1 and S1.

Identification and removal of contaminating sequences.   To identify and remove sequences likely to have originated from dinoflagellate, bacterial, or viral contaminants, a custom script was written employing the following strategy. BLASTN searches were conducted against six databases: the genomes of *S. minutum*[65] and *S. microadriaticum*[48]; complete bacterial genomes (ftp://ftp.ncbi.nih.gov/genomes/Bacteria/all.fna.tar.gz), draft bacterial genomes (ftp://ftp.ncbi.nih.gov/genomes/Bacteria_DRAFT/), and complete viral genomes (ftp://ftp.ncbi.nih.gov/genomes/Viruses/all.fna.tar.gz) databases from NCBI; and the viral database PhAnToMe (http://phantome.org). All databases were retrieved in July 2014. As the lengths of the query and hit sequences were up to hundreds of kilobases, a combination of cutoffs (total bit score >1,000, e-value $\leq 10^{-20}$) was used to identify scaffolds with significant sequence similarities to non-coral sequences representing potential contaminants. This procedure yielded 41 scaffolds that displayed significant similarity in over 50% of their non-N sequences, and thus were considered to have originated from bacterial contaminants and removed from the final assembly.

Annotation of repetitive elements.   *De novo* identification of species-specific repeat regions in the genome assembly of *S. pistillata* was performed using RepeatScout (version 1.0.5)[66] with an l-mer size of 16 bp. Using the default settings, 10,224 distinct repeat motifs were identified that occurred $\geq$ 10 times. Annotation of these repeats was performed as described previously[27] using three different methods: (i) RepeatMasker (version 4.0.2)[67] using RepBase version 19.07, (ii) TBLASTX against RepBase version 19.07, and (iii) BLASTX against a custom-made non-redundant database of proteins encoded by transposable elements (TEs; NCBI keywords: retrotransposon, transposase, reverse transcriptase, gypsy, copia). The best annotation among the three methods was chosen based on alignment coverage and score. The repeat motifs identified in this way and the set of known eukaryotic TEs from RepBase (May 2014 release) were then used to locate and annotate the repeat elements in the assembled genome using RepeatMasker (version 4.0.2). The repeat identification and annotation for the *A. digitifera* genome assembly was performed *sensu* Baumgarten, *et al.*[27] (Table S5).

Reference transcriptome sequencing and assembly.   Total RNA was extracted from *S. pistillata* nubbins subjected to different pH treatments. Briefly, nubbins were cultured in triplicates at pH 7.2, 7.6, 7.8, and 8.1 for 24 months prior to extraction. The RNA preparation was performed as described in Liew, *et al.*[68] and strand-specific sequencing libraries were generated using the NEBNext Ultra Directional RNA Library Prep Kit for Illumina (NEB, Ipswitch, MA, USA). A total of 12 libraries were generated and sequenced on 2 lanes of the Illumina HiSeq platform at the KAUST Bioscience Core Facility (KAUST, Thuwal, KSA), producing 924 million read pairs with 900x coverage. Sequence reads from all libraries were trimmed using Trimmomatic version 0.32 to remove adaptors, primers, and low quality read ends (base quality <30). Further, all reads shorter than 35 bp were removed. PhiX reads were removed using Bowtie2[69] and putative PCR duplicates were removed with PRINSEQ-lite version 0.20.3[70]. After that, all libraries were merged and error correction was carried out using ErrorCorrectReads.pl (ALLPATHS-LG). The resulting merged library was assembled *de novo* using Trinity[71] release 20140413 with strand-specific parameter (–SS_lib_type RF–min_kmer_cov 5–normalize_reads) yielding a total of 89,208 assembled transcripts. The reference transcriptome is available at http://spis.reefgenomics.org[72].

Gene model prediction.   All 89,208 transcripts from the reference transcriptome (see above) were mapped to the genome assembly and filtered by PASA release 20140417[71] to create a training set for AUGUSTUS version 3.0.2[73,74]. The training set was filtered using the following steps: (1) incomplete transcripts were removed, (2) transcripts with less than 3 exons were removed, (3) transcripts with ambiguous 5′ or 3′ untranslated regions (UTRs) were removed, (4) redundant transcripts were removed as indicated by BLASTP, (5) transcripts harboring repeat sequences were removed based on BLASTN against the repeat library generated by RepeatScout (see above). This yielded 2,844 transcripts and corresponding mapping information that were used to train AUGUSTUS. To improve prediction accuracy, a 'hints' file indicating the locations of matching transcripts was generated by

mapping all 89,208 transcripts from the reference transcriptome to the genome assembly using BLAT and the AUGUSTUS script blat2hints.pl. Using the 'hints' file, AUGUSTUS was used for *ab initio* prediction of gene models from the genome assembly and PASA was used subsequently for comparison and completion of the gene models.

Genome protein set completeness analysis. Completeness of the *S. pistillata* genome was assessed using the CEGMA (Core Eukaryotic Genes Mapping Approach) pipeline[75] that is based on a set of core eukaryotic genes (CEGs) from six model organisms (*Homo sapiens, Drosophila melanogaster, Arabidopsis thaliana, Caenorhabditis elegans, Saccharomyces pombe*, and *Saccharomyces cerevisiae*). CEGMA searches for existence of 248 highly conserved genes in the genomic protein set using an approach based on BLASTP and subsequent validation using Hidden Markov Models generated for the core gene set in order to estimate the completeness of a given genomic gene set.

Coral genomic protein set composition. BLASTP searches of gene models from *S. pistillata* and *A. digitifera* were performed against two different databases. Using *S. pistillata* to illustrate the procedure, two databases were created: a 'non-coral' database, and a 'with-coral' database. The former consisted of *Aiptasia* gene models[27] and the NCBI 'nr' database (Nov. 2015 release); the latter included gene models from *A. digitifera*, in addition to the sequences from the 'non-coral' database. Species names contained within annotations for the best hits (e-value $\leq 10^{-5}$) were parsed and fed into a python script that obtained the full taxonomic hierarchy for the respective organisms via an API hosted by Encyclopedia of Life (http://eol.org/api)[76]. Based on the resulting hierarchies, best hits were grouped based on their respective genus. Chord diagrams were drawn using Circos[77]. For visual clarity, only the six most frequent genera were shown and all others were collapsed into 'others'.

Protein set annotation. The final set of predicted proteins derived their annotations from UniProt (i.e., SwissProt and TrEMBL) or the NCBI 'nr' databases (Table S1, Dataset S8), similar to the pipeline described previously[27,68]. Briefly, genomic protein models were subjected to a BLASTP search against SwissProt and TrEMBL databases (June 2014 release). GO terms associated with SwissProt and TrEMBL hits were obtained from UniProt-GOA (July 2014 release)[78]. If the best-scoring hit of the BLASTP search did not yield any GO annotation, further hits (up to 20 hits, e-value $\leq 10^{-5}$) were considered and the best-scoring hit with available GO annotation was used. If none of the SwissProt hits had GO terms associated with them, the TrEMBL hits were processed similarly. Using this procedure, 21,446 genes (83.2% of the 25,769 gene models) were annotated and had at least one GO term associated with them (17,506 proteins had GO annotations via Swiss-Prot, while the remaining 3,940 were from TrEMBL) (Table S1). A majority of the annotations were based on strong alignments to existing sequences within the SwissProt and TrEMBL databases: 19,060 genes had e-values $\leq 10^{-10}$, 15,637 of these had e-values $\leq 10^{-20}$. Proteins that had no matches to either database were subjected to an additional search against the NCBI 'nr' database (e-value $\leq 10^{-5}$). An additional 1,466 proteins were annotated this way. A small fraction of proteins (2,857, 11.1%) had no hits to any of the three databases. A similar procedure was performed for the *A. digitifera* gene models to eliminate potential biases stemming from the use of different annotation pipelines and databases (Dataset S9).

Ortholog identification, category assignment, and GO enrichment analyses. Orthologs between *S. pistillata* (n = 25,679 genes) and the *A. digitifera* V1 gene set (n = 23,523 genes), originally obtained from http://marinegenomics.oist.jp/genomes/downloads?project_id=3 and available at http://spis.reefgenomics.org/download/), were identified using InParanoid v4.1[35] and subsequently assigned to four groups according to their evolutionary relationships: (i) one-to-one orthologs, (ii) many-to-one and many-to-many orthologs, (iii) proteins without easily discernible orthologous relationships, and (iv) species-specific proteins without homologs in other species. One-to-one orthologs were considered those orthologs with exactly one clearly identifiable counterpart in *S. pistillata* and *A. digitifera* in the InParanoid output. This ortholog group presumably is highly conserved and represents core functions of corals given their presence in both genomes. The second set of many-to-one and many-to-many orthologs consisted of proteins in which independent gene duplication has occurred in either only one (resulting in many-to-one gene mappings in the InParanoid output) or both coral species (resulting in many-to-many mappings in the InParanoid output) after divergence. This group of ortholog genes presumably can be used to identify species-wise evolution relevant for one species (in the case of many-to-one ortholog genes) or both corals (in the case of many-to-many ortholog genes). The third set of proteins did not produce gene mappings in the InParanoid output. Although these proteins have homologs in corals and other lineages, orthologous relationships between *S. pistillata* and *A. digitifera* were not easily discernible. Similarly, the fourth group of species-specific proteins retrieved no corresponding ortholog mappings in the Inparanoid analysis (i.e., there were only present in either *S. pistillata* or *A. digitifera*). Different from the proteins without easily discernible orthologous relationships however, these proteins had no detectable homologs in any other species as assessed from the absence of any annotation (see above).

Gene Ontology (GO) enrichment analyses of orthologs from all categories were conducted by testing annotations from genes belonging to a given group relative to annotations from all genomic genes of that species. For instance, in order to investigate enrichment of biological functions of the one-to-one orthologs, GO terms within the list of 6,302 *S. pistillata* genes were tested for enrichment relative to all *S. pistillata* genes with at least one annotated GO term (21,446 genes). Similarly, the 6,302 *A. digitifera* genes were tested against *A. digitifera* genes with at least one annotated GO term (18,544 genes). GO enrichment analyses were conducted using topGO (v2.24.0)[79] with the "weight01" settings. The threshold for significance was $p < 0.05$. The p values were not corrected for multiple testing as non-independent tests were carried out on each GO term.

Phylogenetic trees. Sequences from ortholog groups of interest were first aligned using MUSCLE[80], and aligned sequences were then trimmed with trimAl v1.4.1[81] with the "-automated1" flag. The alignments were subsequently constructed using RAxML v8.2.9[82] with 1,000 bootstraps (-m PROTGAMMAJTT -x 12345 -p 12345 -N 1000 -f a). Trees were viewed and exported to a graphical format using FigTree v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/).

Accession codes. The genome assembly, gene models, and protein models described in this study are available for download at http://spis.reefgenomics.org/download. A JBrowse genome browser is available at http://spis.reefgenomics.org/jbrowse. A BLAST server for the *Stylophora pistillata* genome is available at http://spis.reefgenomics.org/blast/. Raw sequence data reported are deposited at NCBI under the accession number PRJNA281535 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA281535/).

References

1. Porter, J. W. & Tougas, J. I. In *Encyclopedia of Biodiversity* (ed Simon Asher Levin) 73–95 (Elsevier, 2001).
2. Wilkinson, C. *Status of Coral Reefs of the World: 2008.* (2008).
3. Hughes, T. P. *et al.* Climate change, human impacts, and the resilience of coral reefs. *Science* **301**, 929–933 (2003).
4. Hoegh-Guldberg, O. *et al.* Coral reefs under rapidclimate change and ocean acidification. *Science* **318**, 1737–1742 (2007).
5. Maynard, J. *et al.* Projections of climate conditions that increase coral disease susceptibility and pathogen abundance and virulence. *Nature Climate Change* **5**, 688–694 (2015).
6. Carpenter, K. E. *et al.* One-third of reef-building corals face elevated extinction risk from climate change and local impacts. *Science* **321**, 560–563 (2008).
7. Hughes, T. P. *et al.* Global warming and recurrent mass bleaching of corals. *Nature* **543**, 373–377, https://doi.org/10.1038/nature21707 (2017).
8. Rohwer, F., Seguritan, V., Azam, F. & Knowlton, N. Diversity and distribution of coral-associated bacteria. *Marine Ecology Progress Series* **243**, 1–10 (2002).
9. Knowlton, N. & Rohwer, F. Multispecies microbial mutualisms on coral reefs: The host as a habitat. *The American Naturalist* **162**, 51–62 (2003).
10. Voolstra, C. *et al.* The ReFuGe 2020 Consortium—using "omics" approaches to explore the adaptability and resilience of coral holobionts to environmental change. *Frontiers in Marine Science* **2**, 1–8, https://doi.org/10.3389/fmars.2015.00068 (2015).
11. Daniels, C. *et al.* Metatranscriptome analysis of the reef-buidling coral *Orbicella faveolata* indicates holobiont response to coral disease. *Frontiers in Marine Science* **2**, https://doi.org/10.3389/fmars.2015.00062 (2015).
12. McFall-Ngai, M. *et al.* Animals in a bacterial world, a new imperative for the life sciences. *Proceedings of the National Academy of Sciences* **110**, 3229–3236 (2013).
13. Rosenberg, E., Koren, O., Reshef, L., Efrony, R. & Zilber-Rosenberg, I. The role of microorganisms in coral health, disease and evolution. *Nature Reviews: Microbiology* **5**, 355–362 (2007).
14. Theis, K. R. *et al.* Getting the hologenome concept right: an eco-evolutionary framework for hosts and their microbiomes. *mSystems* **1**, https://doi.org/10.1128/mSystems.00028-16 (2016).
15. Bay, R. A. & Palumbi, S. R. Rapid acclimation ability mediated by transcriptome changes in reef-building corals. *Genome Biology and Evolution* **7**, 1602–1612 (2015).
16. Seneca, F. O. & Palumbi, S. R. The role of transcriptome resilience in resistance of corals to bleaching. *Molecular Ecology* **24**, 1467–1484 (2015).
17. Barshis, D. J. *et al.* Genomic basis for coral resilience to climate change. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 1387–1392 (2013).
18. Kenkel, C. D., Meyer, E. & Matz, M. V. Gene expression under chronic heat stress in populations of the mustard hill coral (*Porites astreoides*) from different thermal environments. *Molecular Ecology* **22**, 4322–4334 (2013).
19. Hume, B. C. C. *et al.* Ancestral genetic diversity associated with the rapid spread of stress-tolerant coral symbionts in response to Holocene climate change. *Proceedings of the National Academy of Sciences* **113**, 4416–4421 (2016).
20. Ziegler, M. *et al.* Coral microbial community dynamics in response to anthropogenic impacts near a major city in the central Red Sea. *Marine Pollution Bulletin* **105**, 629–640 (2016).
21. Ziegler, M., Seneca, F. O., Yum, L. K., Palumbi, S. R. & Voolstra, C. R. Bacterial community dynamics are linked to patterns of coral heat tolerance. *Nature Communications* **8**, 14213, https://doi.org/10.1038/ncomms14213 (2017).
22. Ziegler, M. *et al.* Biogeography and molecular diversity of coral symbionts in the genus *Symbiodinium* around the Arabian Peninsula. *Journal of Biogeography* **44**, 674–686, https://doi.org/10.1111/jbi.12913 (2017).
23. Bhattacharya, D. *et al.* Comparative genomics explains the evolutionary success of reef-forming corals. *eLife* **5**, e13288, https://doi.org/10.7554/eLife.13288 (2016).
24. Shinzato, C. *et al.* Using the *Acropora digitifera* genome to understand coral responses to environmental change. *Nature* **476**, 320–323 (2011).
25. Putnam, N. H. *et al.* Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science* **317**, 86–94 (2007).
26. Chapman, J. A. *et al.* The dynamic genome of *Hydra*. *Nature* **464**, 592–596 (2010).
27. Baumgarten, S. *et al.* The genome of *Aiptasia*, a sea anemone model for coral symbiosis. *Proceedings of the National Academy of Sciences* **112**, 11893–11898 (2015).
28. Reed, K. C., Muller, E. M. & van Woesik, R. Coral immunology and resistance to disease. *Diseases of Aquatic Organisms* **90**, 85–92, https://doi.org/10.3354/dao02213 (2010).
29. Park, E. *et al.* Estimation of divergence times in cnidarian evolution based on mitochondrial protein-coding genes and the fossil record. *Molecular Phylogenetics and Evolution* **62**, 329–345 (2012).
30. Simpson, C., Kiessling, W., Mewis, H., Baron-Szabo, R. C. & Müller, J. Evolutionary diversification of reef corals: A comparison of the molecular and fossil records. *Evolution* **65**, 3274–3284 (2011).
31. Prada, C. *et al.* Empty niches after extinctions increase population sizes of modern corals. *Current Biology: CB* **26**, 3190–3194, https://doi.org/10.1016/j.cub.2016.09.039 (2016).
32. Khalturin, K., Hemmrich, G., Fraune, S., Augustin, R. & Bosch, T. C. G. More than just orphans: are taxonomically-restricted genes important in evolution? *Trends in Genetics* **25**, 404–413 (2009).
33. Zdobnov, E. M. *et al.* Genome and proteome analysis of *Anopheles gambiae* and *Drosophila melanogaster*. *Science* **298**, 149–159 (2002).
34. Parra, G., Bradnam, K. & Korf, I. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* **23**, 1061–1067 (2007).
35. Sonnhammer, E. L. L. & Östlund, G. InParanoid 8: orthology analysis between 273 proteomes, mostly eukaryotic. *Nucleic Acids Research*, https://doi.org/10.1093/nar/gku1203 (2014).

36. Rädecker, N., Pogoreutz, C., Voolstra, C. R., Wiedenmann, J. & Wild, C. Nitrogen cycling in corals: the key to understanding holobiont functioning? *Trends in Microbiology* **23**, 490–497 (2015).
37. Pogoreutz, C. *et al.* Sugar enrichment provides evidence for a role of nitrogen fixation in coral bleaching. *Global Change Biology* **23**, 3838–3848, https://doi.org/10.1111/gcb.13695 (2017).
38. Ganot, P. *et al.* Structural molecular components of septate junctions in cnidarians point to the origin of epithelial junctions in eukaryotes. *Molecular Biology and Evolution* **32**, 44–62 (2015).
39. Davy, S. K., Allemand, D. & Weis, V. M. Cell biology of cnidarian-dinoflagellate symbiosis. *Microbiology and Molecular Biology Reviews* **76**, 229–261, https://doi.org/10.1128/mmbr.05014-11 (2012).
40. Zoccola, D. *et al.* Bicarbonate transporters in corals point towards a key step in the evolution of cnidarian calcification. *Scientific Reports* **5**, 9983, https://doi.org/10.1038/srep09983 (2015).
41. Kitahara, M. V., Cairns, S. D., Stolarski, J., Blair, D. & Miller, D. J. A Comprehensive phylogenetic analysis of the Scleractinia (Cnidaria, Anthozoa) based on mitochondrial CO1 sequence data. *PLoS ONE* **5**, e11490, https://doi.org/10.1371/journal.pone.0011490 (2010).
42. Shearer, T. L., van Oppen, M. J., Romano, S. L. & Wörheide, G. Slow mitochondrial DNA sequence evolution in the Anthozoa (Cnidaria). *Molecular Ecology* **11**, 2475–2487 (2002).
43. Kitahara, M. V. *et al.* The "naked coral" hypothesis revisited – evidence for and against scleractinian monophyly. *PLoS ONE* **9**, e94774, https://doi.org/10.1371/journal.pone.0094774 (2014).
44. Fukami, H. *et al.* Mitochondrial and nuclear genes suggest that stony corals are monophyletic but most families of stony corals are not (order scleractinia, class anthozoa, phylum cnidaria). *PLoS ONE* **3**, e3222, https://doi.org/10.1371/journal.pone.0003222 (2008).
45. Wells, J. In *Treatise on Invertebrate Paleontology*. Part F. Coelenterata (ed R. C. Moore) 328–440 (Geological Society of America & University of Kansas Press, 1956).
46. Gaunt, M. W. & Miles, M. A. An insect molecular clock dates the origin of the insects and accords with palaeontological and biogeographic landmarks. *Molecular Biology and Evolution* **19**, 748–761 (2002).
47. Voolstra, C. R. *et al.* Rapid evolution of coral proteins responsible for interaction with the environment. *PLoS ONE* **6**, e20392, https://doi.org/10.1371/journal.pone.0020392 (2011).
48. Aranda, M. *et al.* Genomes of coral dinoflagellate symbionts highlight evolutionary adaptations conducive to a symbiotic lifestyle. *Scientific Reports* **6**, 39734, https://doi.org/10.1038/srep39734 (2016).
49. Mohamed, A. R. *et al.* The transcriptomic response of the coral *Acropora digitifera* to a competent *Symbiodinium* strain: the symbiosome as an arrested early phagosome. *Molecular Ecology* **25**, 3127–3141, https://doi.org/10.1111/mec.13659 (2016).
50. Hamada, M. *et al.* The complex NOD-like receptor repertoire of the coral *Acropora digitifera* includes novel domain combinations. *Molecular Biology and Evolution* **30**, 167–176 (2013).
51. Poole, A. Z. & Weis, V. M. TIR-domain-containing protein repertoire of nine anthozoan species reveals coral-specific expansions and uncharacterized proteins. *Developmental and Comparative Immunology* **46**, 480–488 (2014).
52. Stambler, N. In *Coral Reefs: An Ecosystem in* Transition (eds Zvy Dubinsky & Noga Stambler) 87–106 (Springer Netherlands, 2011).
53. Berkelmans, R. & van Oppen, M. J. H. The role of zooxanthellae in the thermal tolerance of corals: a 'nugget of hope' for coral reefs in an era of climate change. *Proceedings of the Royal Society B: Biological Sciences* **273**, 2305–2312, https://doi.org/10.1098/rspb.2006.3567 (2006).
54. Tambutté *et al.* A compartmental approach to the mechanism of calcification in hermatypic corals. *The Journal of Experimental Biology* **199**, 1029–1041 (1996).
55. Keshavmurthy, S. *et al.* DNA barcoding reveals the coral "laboratory-rat" *Stylophora pistillata* encompasses multiple identities. *Scientific Reports* **3**, 1520, https://doi.org/10.1038/srep01520 (2013).
56. Moya, A. *et al.* Cloning and use of a coral 36B4 gene to study the differential expression of coral genes between light and dark conditions. *Marine Biotechnology* **10**, 653–663, https://doi.org/10.1007/s10126-008-9101-1 (2008).
57. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
58. Xu, H. *et al.* FastUniq: A fast *de novo* duplicates removal tool for paired short reads. *PLoS ONE* **7**, e52249, https://doi.org/10.1371/journal.pone.0052249 (2012).
59. Crusoe, M. R. *et al.* The khmer software package: enabling efficient nucleotide sequence analysis. *F1000Research* **4**, 900, https://doi.org/10.12688/f1000research.6924.1 (2015).
60. Gnerre, S. *et al.* High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proceedings of the National Academy of Sciences* **108**, 1513–1518 (2011).
61. Xue, W. *et al.* L_RNA_scaffolder: scaffolding genomes with transcripts. *BMC Genomics* **14**, 604, https://doi.org/10.1186/1471-2164-14-604 (2013).
62. Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D. & Pirovano, W. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **27**, 578–579 (2011).
63. Boetzer, M. & Pirovano, W. Toward almost closed genomes with GapFiller. *Genome Biology* **13**, 1–9 (2012).
64. Bradnam, K. R. *et al.* Assemblathon 2: evaluating de novo methods of genome assembly in three vertebrate species. *GigaScience* **2**, 10–10 (2013).
65. Shoguchi, E. *et al.* Draft assembly of the *Symbiodinium minutum* nuclear genome reveals dinoflagellate gene structure. *Current Biology: CB* **23**, 1399–1408 (2013).
66. Price, A. L., Jones, N. C. & Pevzner, P. A. *De novo* identification of repeat families in large genomes. *Bioinformatics* **21**, 351–358 (2005).
67. Smit, A., Hubley, R. & Green, P. *RepeatMasker Open-4.0.*, http://www.repeatmasker.org (2013–2015).
68. Liew, Y. J. *et al.* Identification of microRNAs in the coral *Stylophora pistillata*. *PLoS ONE* **9**, https://doi.org/10.1371/journal.pone.0091101 (2014).
69. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**, 357–359 (2012).
70. Schmieder, R. & Edwards, R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **27**, 863–864 (2011).
71. Haas, B. J. *et al.* De novo transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity. *Nature Protocols* **8**, https://doi.org/10.1038/nprot.2013.084 (2013).
72. Liew, Y. J., Aranda, M. & Voolstra, C. R. Reefgenomics.Org - a repository for marine genomics data. *Database* **2016**, baw152, https://doi.org/10.1093/database/baw152 (2016).
73. Stanke, M. *et al.* AUGUSTUS: *ab initio* prediction of alternative transcripts. *Nucleic Acids Research* **34**, 435–439 (2006).
74. Stanke, M. & Waack, S. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* **19**(Suppl 2), 215–225 (2003).
75. Parra, G., Bradnam, K., Ning, Z., Keane, T. & Korf, I. Assessing the gene space in draft genomes. *Nucleic Acids Research* **37**, 289–297 (2009).
76. Parr, C. S. *et al.* The Encyclopedia of Life v2: Providing global access to knowledge about life on Earth. *Biodiversity Data Journal*, e1079, https://doi.org/10.3897/BDJ.2.e1079 (2014).
77. Krzywinski, M. *et al.* Circos: An information aesthetic for comparative genomics. *Genome Research* **19**, 1639–1645 (2009).
78. Dimmer, E. *et al.* The UniProt-GO Annotation database in 2011. *Nucleic Acids Research* **40**, 565–570 (2012).
79. Alexa, A. & Rahnenfuhrer, J. topGO: enrichment analysis for gene ontology. *R package version* 2.8 (2010).

80. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**, 1792–1797 (2004).
81. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
82. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).

Author Contributions

C.R.V., M.A., Y.L., Y.J.L., S.B. designed and conceived the study; D.Z., J.F., S.T., D.A., M.A. generated data; Y.J.L., S.B., Y.L., M.A., C.R.V. analyzed data; M.A., D.Z., J.F., S.T., D.A., C.R.V. contributed reagents/tools/materials; C.R.V. wrote the manuscript with contributions from M.A., Y.J.L., Y.L., and S.B.; all authors read and approved the final manuscript.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-017-17484-x.

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.