Comparative Genomics Highlight the Importance of Lineage-Specific Gene Families in Evolutionary Divergence of the Coral Genus, Montipora

Yuki Yoshioka

Graduate School of Frontier Science, The University of Tokyo

Go Suzuki

Fisheries Technology Institute, Japan Fisheries Research and Education Agency

Yuna Zayasu

Marine Genomics Unit, Okinawa Institute of Science and Technology Graduate University

Hiroshi Yamashita

Fisheries Technology Institute, Japan Fisheries Research and Education Agency

Chuya Shinzato (✉ c.shinzato@aori.u-tokyo.ac.jp)

Atmosphere and Ocean Research Institute, The University of Tokyo

Research Article

Keywords: Corals, Montipora, Comparative genomics, Evolution, Rapidly evolving genes, Symbiosis, Transcriptome.

Posted Date: October 18th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-944849/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Scleractinian corals of the genus *Montipora* (Anthozoa, Cnidaria) possess some unusual biological traits, such as vertical transmission of algal symbionts; however, the genetic bases for those traits remain unknown. We performed extensive comparative genomic analyses among members of the family Acroporidae (*Montipora*, *Acropora* and *Astreopora*) to explore genomic novelties to explain unique biological traits of *Montipora* using improved genome assemblies and gene predictions for *M. cactus*, *M. efflorescens* and *Astreopora myriophthalma*.

Results: We obtained genomic data for the three species, of comparable high quality to other published coral genomes. Comparative genomic analyses revealed that the number of gene families restricted to *Montipora* are significantly higher than those of *Acropora* and *Astreopora*, but their functions are largely unknown. The number of gene families specifically expanded in *Montipora* was much lower than the number specifically expanded in *Acropora*. In addition, we found that evolutionary rates of the *Montipora*-specific gene families were significantly higher than other gene families shared with *Acropora* and/or *Astreopora*. Of 40 gene families under positive selection (Ka/Ks ratio > 1) in *Montipora*, 30 were specifically detected in *Montipora*-specific gene families. Comparative transcriptome analysis of early life stages of *Montipora*, which possesses maternally inherited symbionts, and *Acropora*, which lacks them, revealed that most gene families continuously expressed in *Montipora*, but not expressed in *Acropora* do not have orthologs in *Acropora*. Among the 30 *Montipora*-specific gene families under positive selection, 27 are expressed in early life stages.

Conclusions: Lineage-specific gene families were important to establish the genus *Montipora*, particularly genes expressed throughout early life stages, which under positive selection, gave rise to biological traits unique to *Montipora*. Our findings highlight evolutionarily acquired genomic bases that may support symbiosis in these stony corals and provide novel insights into mechanisms of coral-algal symbiosis, the physiological foundation of coral reefs.

Background

Coral reefs are the most biologically diverse shallow water marine ecosystems [1]. Reef-building corals and endosymbiotic algae of the family Symbiodiniaceae, photosynthetic products of which provide host corals with energy and nutrients, establish mutualistic relationships that are fundamental to coral reefs [2-4]. However, reef-building corals have declined in recent decades due to a variety of anthropogenic stresses, including ocean warming associated with climate change [5-7]. These stresses result in coral bleaching (the breakdown of the symbiosis between corals and their algal endosymbionts [8]), which ultimately leads to loss of habitat for numerous marine species and can precipitate the collapse of entire coral reef ecosystems [9].

The genus *Montipora* (family Acroporidae; Figure 1) is one of the most widespread reef-building corals in the Indo-Pacific [10]. Colony morphology in the genus varies from submassive to laminar, encrusting, and

branching colonies [10, 11]. *Montipora* has some unusual and interesting biological traits among acroporid corals, such as its mode of transmitting algal symbionts and higher stress tolerance. Symbiont transmission maintains symbioses across generations and strongly influences host evolution and adaptation to environments [12-14]. Two fundamental transmission modes predominate in nature (reviewed in [14]): horizontal transmission (symbionts acquired from the environment) and vertical transmission (symbionts acquired maternally). While most coral species (~71%), including *Acropora*, acquire symbionts from the ocean in each generation [15], all known *Montipora* species acquire algal symbionts vertically [16, 17] (Figure 1). Offspring of horizontal transmitters generally associate with a broad range of symbiont types and acquire optimal symbionts from new environments [18, 19]; however, there is no guarantee that optimal symbionts will be available. By contrast, offspring of vertical transmitters inherit symbionts that are suitable for their environments [20], but if they encounter an environment that differs significantly from that of their parents, or if the environment changes too drastically, the inherited symbionts may be disadvantageous. *Montipora* also exhibits low sensitivity to ocean acidification and thermal stressors compared to other coral species [21, 22]. These distinct differences between *Montipora* and its close congener, *Acropora*, may have occurred after their divergence (approx. 125 Mya [23]).

In the family Acroporidae, whole-genome data are becoming more readily available, now including 16 species of *Acropora* [23-26], 3 species of *Montipora* [23, 27, 28] and 1 *Astreopora* species [23], the latter being the basal genus of the Acroporidae [29] (Figure 1). Recently, Shinzato et al. [23] performed a large-scale genomic comparison of acroporids (using genomes of *Acropora*, *Montipora*, and *Astreopora*) and proposed that the evolutionary success of *Acropora* may have occurred by virtue of gene duplications. Although there have been some studies performing genome-wide analysis using *Montipora* genomes [27, 28], the genomic basis for their unique biological traits remains unknown. Exploiting abundant acroporid genomic resources, we performed comparative genomic analyses using improved genomic data of *Montipora* and *Astreopora*. We further identified genes with high evolutionary rates in *Montipora* that may be associated with adaptive evolution, and we specifically attempted to identify genes related to maintenance of maternally inherited symbionts by comparing gene expression during early life stages of *Montipora* and *Acropora*.

Results

Improvement of genome assemblies and gene predictions for Montipora and Astreopora

Assembly error, including retention of allelic contigs in haploid assemblies, is problematic for downstream analyses, mainly due to redundant genome sequences (alleles from the same genetic locus). We curated scaffold sequences of *M. cactus* and *M. efflorescens* by removing scaffold sequences with high or low coverage and those that may have originated from one of two allelic copies in heterozygous regions. Numbers of scaffold sequences were significantly reduced from the previous version, from 4,925 to 3,521 in *M. cactus* and from 5,162 to 3,589 in *M. efflorescens* (Table 1). For *Astreopora*, possible allelic scaffold sequences were removed from the genome assembly during the previous study [23]. The

previous version of gene models for *M. cactus*, *M. efflorescens*, and *Astreopora* were predicted using AUGUSTUS, based solely on a training set built for *Acropora* or for protein homology with gene models of other corals [23]. Thus, it was highly possible that lineage-specific genes were missed in the previous version. In this study, we performed gene prediction for *M. cactus*, *M. efflorescens*, and *Astreopora myriophthalma* using a combination of *ab initio* and RNA-seq evidence-based prediction. We predicted 29,158 protein-coding genes for *M. cactus*, 29,424 for *M. efflorescens* and 25,406 for *Astreopora myriophthalma* (Table 1). Benchmarking universal single-copy orthologs (BUSCO) completeness scores were 93.3% (of which 0.8% were duplicated) for *M. cactus*, 91.2% (of which 1.4% were duplicated) for *M. efflorescens* and 94.5% (of which 1.3% were duplicated) for *Astreopora myriophthalma*, which were considerably better scores than the previous version (Table 1). In comparison to other *Montipora* gene models, gene models reported by Shumaker et al. [28] may have contained a higher fraction of diploid copies (93.4% complete BUSCO, with 18.3% duplicated; Table1). Completeness of gene models reported by Helmkampf et al. [27] was lower than that reported by Shumaker et al. [28] (64.2%, with 0.5% duplicated; Table1). Thus, the gene models reported by Shumaker et al. [28] contained many duplicates, but those reported by Helmkampf et al. [27] lacked many genes. In contrast, BUSCO completeness scores of *M. cactus*, *M. efflorescens* and *Astreopora myriophthalma* reported in this study were comparable to published gene models of other coral species, including *A. millepora*, predicted using the NCBI annotation pipeline (Table 1). These improvements to the *Montipora* and *Astreopora* genomes enabled more accurate comparative genomics among acroporids.

Comparison of gene families within the Acroporidae

Identifying orthologous relationships between sequences is fundamental for comparative genomic analyses. To obtain orthologous relationships among acroporid genomes, we used three *Acropora* species (*A. digitifera*, *A. millepora*, and *A. tenuis*), for which BUSCO completeness scores are high (Table 1), two *Montipora* species (*M. cactus* and *M. efflorescens*), and *Astreopora myriophthalma*, representing the basal clade of the Acroporidae [29]. We obtained 12,769 gene families for *Montipora*, 11,007 for *Acropora* and 11,309 for *Astreopora* (Figure 2). We then categorized each gene family into seven groups, (1) common to all three genera (9,690 gene families), (2) common to *Montipora* and *Acropora* (743 gene families), (3) common to *Montipora* and *Astreopora* (665 gene families), (4) common to *Acropora* and *Astreopora* (257 gene families), (5) restricted to *Montipora* (1,670 gene families), (6) restricted to *Astreopora* (696 gene families) and (7) restricted to *Acropora* (316 gene families) (Figure 2). 75.8% (9,690/12,769) of the gene families in *Montipora*, 88% (9,690/11,007) in *Acropora*, and 85.7% (9,690/11,309) in *Astreopora* were shared among all three genera (Figure 2), indicating that a large number (~ 80 - 90%) of gene families are shared throughout the Acroporidae, and these are likely to be the core-gene families the Acroporidae.

The two major clades of reef-building corals possess different metabolic pathways [30]. From the six species, we compared 303 functional modules comprising ten categories in the Kyoto Encyclopedia of Genes and Genomes (KEGG) metabolic pathways and found that metabolic pathways were basically conserved in the three genera (Supplementary Table S1). An enzyme involved in cysteine biosynthesis

(KEGG module ID: M00338) and methionine degradation (KEGG module ID: M00035) was not detected among the six species (Supplementary Table S1), as reported in Shinzato et al. [23, 24]. Although one gene (KEGG entry ID: K04486) involved in the histidine biosynthetic pathway (KEGG module ID: M00026) was detected in acroporid corals used in this study, the remaining genes required to complete the pathway were not detected (Supplementary Table S1), as reported in Ying et al. [30]. Taken together, gene families involved in common features, such as amino acid synthesis, are widely conserved in the three genera.

While we identified 696 lineage-specific gene families in *Astreopora* and 316 in *Acropora*, we identified 1,670 gene families restricted to *Montipora* (2,307 genes in *M. cactus* and 2,303 in *M. efflorescens*) (Figure 2). The proportion of lineage-specific gene families in *Montipora* (13.07%) was significantly larger than those in *Acropora* (2.87%) and *Astreopora* (6.15%) (Pairwise proportion test: $p < 0.05$). In addition, although we performed gene annotation with BLAST searches against the SwissProt database (BLASTP, e-value cutoff: $1e-5$), the proportion of *Montipora*-specific gene families with SwissProt annotation was significantly lower than in *Acropora* and *Astreopora* (Pairwise proportion test: $p < 0.05$ for *Montipora* versus *Acropora*, $p < 0.05$ for *Montipora* versus *Astreopora*, and $p = 0.59$ for *Acropora* versus *Astreopora*; Figure 2). This indicates that functions of gene families restricted to *Montipora* are largely unknown.

Gene expansions in Montipora and comparisons among acroporids

Gene duplication has contributed to acquisition of new gene functions during evolution [31, 32]. To explore gene families that underwent expansions, we first compared gene numbers of 9,690 gene families common to the three genera and 743 gene families common to *Montipora* and *Acropora* (Figure 2). In these two groups, genes in families that underwent gene expansions in either *Montipora* or *Acropora* might have been duplicated after *Montipora* and *Acropora* diverged from their common ancestor. Three gene families, similar to dimethylsulfoniopropionate (DMSP) lyase (*Alma*; HOG0000829), Endonuclease-reverse transcriptase (*GP1*; HOG0000531), and Spondin (*Spon1*; HOG0001590), and three non-annotated gene families (NA; HOG0000965, HOG0001135, and HOG0001312), were significantly expanded in *Acropora* (Fisher's exact test: $p < 0.05$; Figure 3a and 3b). Recently, it was reported that DMSP lyase is the most expanded gene family in *Acropora* [28], and our result is consistent with a previous report, supporting the accuracy of this analysis. We found that three gene families, transient receptor potential protein (*TRPC*; HOG0002487), collagen alpha-1 (VII) chain (*COL7A1*; HOG0003259) and non-annotated gene family (NA; HOG0001797) are significantly expanded in *Montipora* compared with *Acropora* (Fisher's exact test: $p < 0.05$; Figure 3a and 3b).

Next, we compared gene numbers of 665 gene families common to *Montipora* and *Astreopora* (Figure 2), in which gene duplication may have occurred after divergence of *Montipora* or *Astreopora*. These genes may have been lost in *Acropora*. Two gene families (HOG0003949 and HOG0004557) lacking SwissProt annotation were significantly expanded in *Astreopora* (Fisher's exact test: $p < 0.05$; Figure 3c), whereas one other gene family, tetratricopeptide repeat protein 28 (*TTC28*; HOG0000387), which is involved in the

cell cycle in humans [33], was significantly expanded in *Montipora* compared with *Astreopora* (Fisher's exact test: $p < 0.05$; Figure 3c).

Estimation of evolutionary rate in each Montipora gene family group

The ratio of nonsynonymous (Ka) to synonymous substitutions (Ks) reflects the strength of selective pressure on protein sequences [34]. For example, when Ka is less than Ks ($Ka/Ks < 1$), selection has occurred to eliminate mutations of protein sequences (negative or purifying selection). In contrast, when Ka is larger than Ks ($Ka/Ks > 1$), selection has occurred to mutate the protein sequences (positive selection). In order to evaluate the strength of selective pressure acting on protein sequences in each *Montipora* gene family, we calculated pairwise Ka/Ks ratios between *Montipora* single-copy orthologous gene pairs (*M. cactus* versus *M. efflorescens*) for each of the four groups: 1) gene families common to the three Acroporidae genera, 2) gene families common to *Montipora* and *Acropora*, 3) gene families common to *Montipora* and *Astreopora*, and 4) gene families restricted to *Montipora* (Figure 4). When we compared Ka/Ks ratio between groups, gene families restricted to *Montipora* showed a highest Ka/Ks ratio (Wilcoxon rank sum test: $p < 0.05$; Figure 4), indicating that this gene family group has undergone a relaxation of negative selection, and that functional constraints on this gene family group are relaxed. This could explain why the deduced gene functions of gene families restricted to *Montipora* are largely unknown.

Positive selection specific to Montipora

To identify genes with fast evolutionary rates that may be associated with adaptive evolution in *Montipora*, we focused on gene families exhibiting $Ka/Ks > 1$. We found evidence of positive selection in 40 gene families (rapidly evolving gene families) (Table 2). Of those, 10 families are common to the three genera or common to *Montipora* and *Acropora*, while the remaining 30 families are restricted to *Montipora* (Table 2), suggesting that these 30 gene families arose specifically in that lineage and likely contribute to biological traits unique to *Montipora*. Although 28 of the 30 gene families restricted to *Montipora* were without annotation, their possible subcellular localization ranging from membrane to organelle was predicted by DeepLoc, a deep learning neural networks model (Table 2).

Gene expression unique to early life stages of Montipora

Presence of maternally inherited algal symbionts at an early life stage is the most obvious difference between vertical and horizontal transmitters (Figure 1). In order to identify gene families specifically involved in symbiosis in vertical transmitters, we compared the repertoire of expressed genes in early life stages of *Montipora* with those expressed in *Acropora*. In this analysis, a gene family was considered expressed if even only one gene in that family was expressed (Transcript per million (TPM) > 1). We confirmed that 11,930 and 10,838 gene families were expressed at early life stages of *Montipora* and *Acropora*, respectively (Figure 5a). Of these, 10,051 gene families (84% in *Montipora* and 93% in *Acropora*) were common to both at early life stages (Figure 5a), suggesting that these are essential for early development of acroporid corals; thus, we did not focus on these in the present study. We identified

1,879 gene families that were exclusively expressed in *Montipora* (Figure 5b). Among those, 60% (1,132 gene families) were expressed in planula larvae, metamorphosed larvae and recruit stages (Figure 5b), suggesting that these genes may be related to maintenance of algal symbionts in *Montipora*. Interestingly, 97% of these gene families ((753 + 344) / 1,132, Figure 5b) that were expressed throughout the three life stages were specific to *Montipora* or shared by *Astreopora* (Supplementary Table S2). In contrast, the remaining 3% of gene families ((22 + 13) / 1,132, Figure 5b) have orthologs in *Acropora*, but were not expressed in *Acropora*. Nonetheless, they were expressed throughout early life stages of *Montipora* (Supplementary Table S3). Within gene families containing gene duplications in the *Montipora* genomes above, two gene families (HOG0001797 and HOG0000387) were exclusively expressed in at least one early life stage in *Montipora*, and one of them (HOG0000387) was expressed throughout all three early life stages (Supplementary Table S2). Among the identified 30 rapidly evolving gene families restricted to *Montipora*, we detected gene expression of 90% of these families. Expression of nine families was detected in at least one early life stage of *Montipora*, and the remaining 18 gene families were continuously expressed throughout all three early life stages (Table 2).

Discussion

Improved genome information for genus Montipora and Astreopora

BUSCO completeness scores for improved gene models of *M. cactus*, *M. efflorescens*, and *Astreopora myriophthalma* were 93.3% (0.8% duplicates), 91.2% (1.4% duplicates), and 94.5% (1.3% duplicates), respectively (Table 1). They are considerably better than those of *M. capitata* (93.4% (18.3% duplicates) from Shumaker et al. [28] and 64.2% (0.5% duplicates) from Helmkamp et al. [27]; Table 1), and were comparable to those of other coral species (Table 1). These numbers indicate that we successfully obtained high-quality gene models from *Montipora* and *Astreopora* species. Numbers of genes in *M. cactus* and *M. efflorescens* genomes were not quite as large as that of *M. capitata* reported by Shumaker et al. [28]. Previously, it was reported that *M. capitata* has fewer exons and shorter coding regions per gene than other corals [27, 28]; however, this was not the case with *M. cactus* and *M. efflorescens* (Table 1). Fewer exons and shorter coding regions per gene could be an unusual feature of the *M. capitata* genome or could reflect the quality of the genome assembly. Indeed, the N50 size, one of the indices to evaluate the quality of genome assembly, was larger for both *M. cactus* and *M. efflorescens* genome assemblies than for *M. capitata* (Table 1).

Possible genomic evolutionary strategy unique to Montipora

Recent large-scale genome comparisons of acroporid genomes showed that 28 gene families were specifically expanded in *Acropora*, but none in *Montipora* [23], but we identified four expanded gene families in *Montipora* (Figure 3). Although the number of gene families in *Montipora* is not much different from those of *Acropora* and *Astreopora*, the proportion of lineage-specific gene families in *Montipora* was significantly larger than in *Acropora* and *Astreopora* (Figure 2). *Montipora* does not appear to have duplicated existing gene families, as has *Acropora*. Lineage-specific gene families contribute larger gene

numbers in *Montipora* genomes, and emergence of lineage-specific genes may have helped to establish unique biological traits of *Montipora* corals. In particular, *Montipora*-specific gene families under positive selection may be major contributors.

Three gene families, similar to *Trpc6*, *TTC28*, and *COL7A1*, and one gene family without annotation were significantly expanded in *Montipora* compared with *Acropora* or *Astreopora* (Figure 3). Known functions of transient receptor potential (TRP) proteins encoded by *Trpc* are diverse (reviewed in [35]). For example, TRP proteins respond to hypertonicity in yeasts [36, 37], detect and avoid noxious chemicals in nematodes [38], and discriminate warmth, heat, and cold in humans [35]. In each case, TRP proteins mediate sensory transduction in cells [35]. In corals, expression levels of *Trp*-like genes change when the concentration of CO₂ in seawater changes [39]. They also change diurnally [40, 41] or when exposed to symbiotic algae [42, 43]. The *Trpc6*-like gene family, specifically expanded in *Montipora*, may also be involved in sensory transduction during environmental transitions. The *TTC28*-like gene family has tetratricopeptide repeats (PF12176 and/or PF13424) and caspase HetF associated with Tprs (CHAT) domains (PF12770) (Supplementary Table S4). Canonical *TTC28* is composed of tetratricopeptide repeats and CHAT domains (Q96AY4: TTC28_HUMAN; [33]) and genes in the gene family (HOG0000387) are also composed of tetratricopeptide repeats and CHAT domains, indicating that this gene family may have been duplicated from canonical *TTC28*, which is conserved in all acroporids examined in this study (HOG0016559 in Supplementary Data S1). *TTC28* is required for the cell cycle in humans [33]. The expanded *TTC28*-like gene family may also be involved in cell cycle in *Montipora*. Collagen is expressed in gastrodermis at a specific developmental stage of cnidarian larvae [44-46] and the expanded collagen-like gene family may function in early development of *Montipora*.

In this study, we identified 40 genes under positive selection in *Montipora* (Table 2). Positive selection has often been detected in genes involved in immunity in vertebrates [47]. In corals, genes related to immunity, such as lectin and antimicrobial peptide, are also under positive selection [23, 48, 49]. In the 40 rapidly evolving gene families found in this study, with one exception, no genes appeared homologous to immune-related genes (Table 2). In addition, 28 of 30 rapidly evolving gene families restricted to *Montipora* have no annotated function (Table 2). Generally, genes with no homology to genes of other lineages are called orphan genes [50]. They may arise principally by two processes: gene duplication or *de novo* evolution from non-coding regions [50]. If a gene originates by duplication, the protein domains tend to be conserved in the new genes, since a functional protein domain cannot easily be changed by mutations [51], suggesting that the 28 rapidly evolving gene families originated by *de novo* evolution from non-coding regions. Orphan genes are expected to interact specifically with the environment as a consequence of lineage-specific adaptation [50]. Therefore, orphan genes may contribute to adaptive evolution in *Montipora*. In particular, 18 rapidly evolving gene families with expression throughout the three early life stages, planula larvae, metamorphosed larvae and recruits, may have important functions in symbiosis during early life stages of *Montipora*.

Conclusions

In this study, we highlighted possible genomic underpinnings of unusual biological traits of *Montipora* using high-quality genomic information of *Montipora* and *Astreopora*. We found that the driving force behind evolution of *Montipora* was lineage-specific gene families, rather than gene duplication, as among *Acropora* corals. The importance of rapidly evolving gene families in *Montipora* for their unique biological traits was particularly highlighted. Our dataset and findings offer novel insights into mechanisms of coral-algal symbiosis. Although genetic tools for manipulating corals have been established [52, 53], development of more efficient methods to deliver gene-knockdown or -knockout reagents into large numbers of zygotes will facilitate rapid screening for relevant phenotypes of candidate genes. In addition, coral cell lines which have the capacity to incorporate algal symbionts has been developed [54], allowing us to observe ongoing symbiosis at the single cell level. Together, these advances will facilitate a deeper understanding of cellular and molecular mechanisms of coral-algal symbiosis.

Methods

Sample preparation, RNA extraction, and RNA-Seq

Colonies of *Montipora cactus*, *M. efflorescens*, and *Astreopora myriophthalma* were collected in Sekisei Lagoon, Okinawa, Japan in May 2015, and were maintained in aquaria at the Research Center for Subtropical Fisheries, Seikai National Fisheries Research Institute, until spawning. Permits for coral collection were kindly provided by the Okinawa Prefectural Government for research use (Permits #29-74). We isolated total RNA from adult colonies of *M. cactus*, *M. efflorescens*, and *Astreopora myriophthalma* using an RNeasy Plant Mini Kit (QIAGEN). A TruSeq Stranded mRNA Library Kit (Illumina) was used for mRNA sequencing library preparation, and each library was sequenced from 100-bp paired-end libraries using a NovaSeq 6000 (Illumina). For *Montipora*, we also isolated total RNA from eggs, sperm, planula larvae (1- and 4-days post-fertilization) using TRIzol reagent (Thermo Fisher Scientific) according to the manufacturer's protocol. KAPA RNA HyperPrep Kit (Kapa Biosystems) and MGIEasy RNA Directional Library Prep Set (MGI) were used for total RNA and mRNA sequencing library preparation, and each library was sequenced on a NovaSeq 6000 in 150-bp paired-end and a DNBSEQ-G400RS (MGI) in 100-bp paired-end mode. This information is summarized in Supplementary Table S5.

Curating scaffold sequences of M. cactus and M. efflorescens and gene prediction

We downloaded scaffold sequences of *M. cactus* and *M. efflorescens* from the genome browser of the OIST Marine Genomics Unit (<https://marinegenomics.oist.jp>). We identified scaffold sequences with high or low coverage or those that may have originated from one of the two allelic copies of heterozygous regions, using Purge Haplotigs v1.1.1 [55] with default option and excluded these from subsequent analyses.

In addition to the above RNA samples, we used publicly available RNA-seq data from NCBI SRA for gene prediction (Supplementary Table S6). Low-quality reads (quality score < 20 and length < 20 bp) and sequence adaptors were trimmed using CUTADAPT v1.18 [56]. A total of 31 and 2 RNA-seq libraries were used for *Montipora* and *Astreopora* gene prediction, respectively. Repetitive elements in the scaffolds

were identified *de novo* with RepeatScout v1.0.6 [57] and RepeatMasker v4.1.0 (<http://www.repeatmasker.org>). Repetitive elements were filtered out by length (> 50 bp) and occurrence (more than 10 times for *Montipora*, more than 60 times for *Astreopora*). Gene prediction was first executed with the BRAKER pipeline v2.1.2 [58], with AUGUSTUS v3.3.3. RNA-seq reads were aligned to each genome sequence with HISAT v2.1.0 [59]. Then, the alignment information was used for BRAKER gene prediction with options “UTR=on”, “softmasking”, and “AUGUSTUS_ab_initio.” To improve gene prediction, we further executed genome-guided transcriptome assembly using StringTie [60] with option “-m 500.” Genome-based transcript structure was predicted with TransDecoder (<https://github.com/TransDecoder/TransDecoder/wiki>). During read alignment, we used soft-masked repeats for genome-guided transcriptome assembly and hard-masked repeats for BRAKER gene prediction. Finally, genes that were present in genome-guided assembly or *ab initio* prediction, but absent in predictions from the hint file were added to the prediction from the hint file using GffCompare [61], as summarized in Supplementary Figure S1. To evaluate the completeness of predicted genes, we used BUSCO v5.0 [62] with Metazoa OrthoDB10 dataset (2021-02-24, n=954).

Gene annotation, orthology inference within the Acroporidae

We used publicly available gene models for *A. digitifera* [23, 24], *A. tenuis* [23], and *A. millepora* [25] in addition to *Montipora* and *Astreopora* gene models. For *A. millepora*, we downloaded gene models from NCBI RefSeq (RefSeq assembly accession: GCF_004143615.1). We downloaded gene models of v2.0 for *A. digitifera* and v1.0 for *A. tenuis* from the genome browser of the OIST Marine Genomics Unit, respectively. We selected the longest transcript variants from each gene and translated them into protein sequences. All proteomes were annotated with BLASTP [63] (E-value cut off: 1e-5) against the SwissProt database (8 January 2021). Domains in proteomes were annotated using InterProScan v5.31-70.0 [64] with default settings. In addition, putative transposable elements in gene models were identified with TransposonPSI (<http://transposonpsi.sourceforge.net/>), Dfam scan (release 3.3; [65]), and Pfam keyword (“Reverse transcriptase” and “Integrase”). All proteins were also annotated with KEGG [66] in all eukaryote genes using GenoMaple v2.3.2 [67] with the GHOSTX search engine and the bi-directional best hit method. Module completion ratio (MCR) was calculated in each functional module defined by KEGG, also using GenoMaple v2.3.2. For clustering of orthologous genes (herein gene families) of the Acroporidae, we used OrthoFinder v2.4.0 [68] and *Porites australiensis* gene models (Shinzato et al., unpublished data) were also included as an outgroup for the Acroporidae. In this study, we used phylogenetic hierarchical orthogroups (HOG) as gene families. Gene families common to the three *Acropora* species were defined as *Acropora* gene families. Gene families shared by the two *Montipora* species were defined as *Montipora* gene families. Gene families containing transposon-like genes were excluded from subsequent analyses.

Transcriptomic comparisons between Montipora and Acropora

We used RNA-seq data of *M. efflorescens* (planula larvae), *A. tenuis* (blastula, gastrula, planula larvae and polyps) and *A. digitifera* (blastula, gastrula, planula larvae and polyps) (Supplementary Table S7). In

addition, publicly available RNA-seq data of *M. capitata* (planula larvae, metamorphosed larvae, and recruits) were also used in this study (Supplementary Table S7). Low-quality reads (quality score < 20 and length < 20 bp) and sequence adaptors were trimmed using CUTADAPT v1.18. Cleaned RNA-seq reads were mapped to each organism's gene models (For *M. capitata* RNA-seq data, we used *M. efflorescens* gene models as a reference) using SALMON v1.0.0 [69]. Expression levels were quantified using SALMON v1.0.0. Genes with TPM > 1 were considered expressed. Then expressed genes were classified into corresponding gene families based on the above gene family inference.

Estimation of the ratio of nonsynonymous to synonymous substitutions

Protein sequences of putative single-copy orthologs between *M. cactus* and *M. efflorescens* were aligned pairwise with MAFFT [70]. Aligned nucleotide codon sequences without alignment gaps were retrieved using the PAL2NAL script [71]. Genes with nucleotide alignment lengths longer than 120 bp were used for further analysis. We calculated pairwise nonsynonymous (Ka) and synonymous (Ks) substitution ratios of single-copy genes between *M. cactus* and *M. efflorescens* using KaKs_Calculator 2.0 [72] with option “-MA”. Following Villanueva-Canas et al. [73], we discarded gene families showing Ks < 0.01, as such low Ks values may result in inaccurate Ka/Ks estimates, and gene families showing Ks or Ka > 2 indicating saturation of substitutions. Genes exhibiting Ka/Ks ratios with $p < 0.05$ (Fisher's exact test) were used for further analysis. Subcellular localization of gene families showing Ka/Ks > 1 was predicted using the DeepLoc-1.0 online server [74].

Statistical analysis

Pairwise proportion tests were conducted to compare lineage-specific gene families (“number of lineage-specific gene families” / “number of gene families in lineage”) and gene annotation proportions of lineage-specific gene families (“number of genes with annotation” / “number of genes without annotation”). Fisher's exact test was conducted to identify expanded gene families in each group (“number of genes in one gene family in species A” / “number of genes in the rest of the gene family in species A” versus “number of genes in one gene family in species B” / “number of genes in the rest of the gene family in species B”). We considered a $p < 0.05$ as significantly expanded. The Wilcoxon rank sum test was conducted to compare median Ka/Ks values between gene family groups. All statistical tests were performed in R v4.0.3 [75].

Abbreviations

BUSCO: Benchmarking universal single-copy orthologs.

KEGG: Kyoto encyclopedia of genes and genomes.

DMSP: Dimethylsulfoniopropionate.

Ka: nonsynonymous substitutions.

Ks: synonymous substitutions.

TPM: Transcript per million.

TRP: Transient receptor potential protein.

TTC: Tetratricopeptide repeat protein.

CHAT: Caspase HetF associated with Tprs.

HOG: Phylogenetic hierarchical orthogroups.

MCR: Module completion ratio.

Declarations

□Ethics approval and consent to participate

Not applicable.

□Consent for publication

Not applicable.

□Availability of data and materials

Raw RNA-sequencing data have been deposited in the DDBJ/EMBL/GenBank databases under accession number DRA011820 (BioProject ID: PRJDB11460). A genome browser for *M. cactus*, *M. efflorescens* and *Astreopora myriophthalma* is available from the Marine Genomics Unit web site (<https://marinegenomics.oist.jp/gallery>). Sequence IDs of retained scaffolds were prepared as Supplementary Data S2. Gene models in GTF format for *M. cactus*, *M. efflorescens* and *Astreopora myriophthalma* are provided as supplementary Data S3-S5. For *M. capitata*, we downloaded genome assembly and gene models reported from Shumaker et al. [28] (URL: <http://cyanophora.rutgers.edu/montipora/>) and Helmkamp et al. [27] (Data set DOI: 10.15482/USDA.ADC/1503958).

□Competing interests

Chuya Shinzato is an editorial board member for *BMC Ecology and Evolution*.

□Funding

This study was supported by JSPS KAKENHI grants (20H03235 and 20K21860 for CS, 21H04742 for HY, 19K15902 for YZ and 20H03066 for GS) and Grant-in-Aid for JSPS Fellows to YY (20J21301).

□Author's contributions

C.S. and G.S. conceptualized, and C.S. supervised the project. G.S., C.S. and Y.Z. performed coral sampling. H.Y. helped with sampling of *Montipora* larvae. Y.Y. performed molecular biological experiments and bioinformatic analyses. Y.Y. and C.S. wrote main manuscript.

□Acknowledgements

Computations were partially performed on the NIG supercomputer at ROIS National Institute of Genetics. We thank Dr. Steven D. Aird for carefully reviewing our manuscript and for helpful comments.

References

1. Roberts CM, McClean CJ, Veron JEN, Hawkins JP, Allen GR, McAllister DE, Mittermeier CG, Schueler FW, Spalding M, Wells F, Vynne C, Werner TB: **Marine Biodiversity Hotspots and Conservation Priorities for Tropical Reefs**. Science 2002, **295**(5558):1280-1284.
2. Muscatine L, Porter JW: **Reef corals: mutualistic symbioses adapted to nutrient-poor environments**. Bioscience 1977, **27**(7):454-460.
3. Muscatine L, R. McCloskey L, E. Marian R: **Estimating the daily contribution of carbon from zooxanthellae to coral animal respiration 1**. Limnol Oceanogr 1981, **26**(4):601-611.
4. Falkowski PG, Dubinsky Z, Muscatine L, Porter JW: **Light and the Bioenergetics of a Symbiotic Coral**. Bioscience 1984, **34**(11):705-709.
5. Carpenter KE, Abrar M, Aeby G, Aronson RB, Banks S, Bruckner A, Chiriboga A, Cortés J, Delbeek JC, DeVantier L: **One-third of reef-building corals face elevated extinction risk from climate change and local impacts**. Science 2008, **321**(5888):560-563.
6. Hoegh-Guldberg O, Poloczanska ES, Skirving W, Dove S: **Coral reef ecosystems under climate change and ocean acidification**. Frontiers in Marine Science 2017, **4**:158.
7. Hughes TP, Anderson KD, Connolly SR, Heron SF, Kerry JT, Lough JM, Baird AH, Baum JK, Berumen ML, Bridge TC: **Spatial and temporal patterns of mass bleaching of corals in the Anthropocene**. Science 2018, **359**(6371):80-83.
8. Brown BE: **Coral bleaching: causes and consequences**. Coral Reefs 1997, **16**(S1):S129-S138.
9. Bellwood DR, Hughes TP, Folke C, Nystrom M: **Confronting the coral reef crisis**. Nature 2004, **429**(6994):827-833.
10. Veron JEN, Stafford-Smith M: *Corals of the world*: 1. publ. ed. Townsville: Australian Inst. of Marine Science; 2000.
11. Veron J: **Scleractinia of Eastern Australia, Part 5: Acroporidae**. Aust Inst Mar Sci Monogr Ser 1984, **6**:485.
12. Yamamura N: **Vertical transmission and evolution of mutualism from parasitism**. Theor Popul Biol 1993, **44**(1):95-109.
13. Yamamura N: **Evolution of mutualistic symbiosis: a differential equation model**. Researches on Population Ecology 1996, **38**(2):211-218.

14. Bright M, Bulgheresi S: **A complex journey: transmission of microbial symbionts.** Nature Reviews Microbiology 2010, **8**(3):218-230.
15. Baird AH, Guest JR, Willis BL: **Systematic and biogeographical patterns in the reproductive biology of scleractinian corals.** Annual Review of Ecology, Evolution, and Systematics 2009, **40**:551-571.
16. Richmond RH, Hunter CL: **Reproduction and recruitment of corals: comparisons among the Caribbean, the tropical Pacific, and the Red Sea.** Marine ecology progress series. Oldendorf 1990, **60**(1):185-203.
17. Fadlallah YH: **Sexual reproduction, development and larval biology in scleractinian corals.** Coral Reefs 1983, **2**(3):129-150.
18. Douglas AE: **Host benefit and the evolution of specialization in symbiosis.** Heredity 1998, **81**(6):599-603.
19. Rowan R, Knowlton N: **Intraspecific diversity and ecological zonation in coral-algal symbiosis.** Proceedings of the National Academy of Sciences 1995, **92**(7):2850-2853.
20. Byler KA, Carmi-Veal M, Fine M, Goulet TL: **Multiple symbiont acquisition strategies as an adaptive mechanism in the coral *Stylophora pistillata*.** PLoS One 2013, **8**(3):e59596.
21. Putnam HM, Davidson JM, Gates RD: **Ocean acidification influences host DNA methylation and phenotypic plasticity in environmentally susceptible corals.** Evolutionary applications 2016, **9**(9):1165-1178.
22. Gibbin EM, Putnam HM, Gates RD, Nitschke MR, Davy SK: **Species-specific differences in thermal tolerance may define susceptibility to intracellular acidosis in reef corals.** Mar Biol 2015, **162**(3):717-723.
23. Shinzato C, Khalturin K, Inoue J, Zayasu Y, Kanda M, Kawamitsu M, Yoshioka Y, Yamashita H, Suzuki G, Satoh N: **Eighteen coral genomes reveal the evolutionary origin of *Acropora* strategies to accommodate environmental changes.** Mol Biol Evol 2021, **38**(1):16-30.
24. Shinzato C, Shoguchi E, Kawashima T, Hamada M, Hisata K, Tanaka M, Fujie M, Fujiwara M, Koyanagi R, Ikuta T: **Using the *Acropora digitifera* genome to understand coral responses to environmental change.** Nature 2011, **476**(7360):320-323.
25. Ying H, Hayward DC, Cooke I, Wang W, Moya A, Siemering KR, Sprungala S, Ball EE, Forêt S, Miller DJ: **The whole-genome sequence of the coral *Acropora millepora*.** Genome biology and evolution 2019, **11**(5):1374-1379.
26. Fuller ZL, Mocellin VJ, Morris LA, Cantin N, Shepherd J, Sarre L, Peng J, Liao Y, Pickrell J, Andolfatto P: **Population genetics of the coral *Acropora millepora*: Toward genomic prediction of bleaching.** Science 2020, **369**(6501).
27. Helmkampf M, Bellinger MR, Geib SM, Sim SB, Takabayashi M: **Draft genome of the rice coral *Montipora capitata* obtained from linked-read sequencing.** Genome biology and evolution 2019, **11**(7):2045-2054.
28. Shumaker A, Putnam HM, Qiu H, Price DC, Zelzion E, Harel A, Wagner NE, Gates RD, Yoon HS, Bhattacharya D: **Genome analysis of the rice coral *Montipora capitata*.** Scientific reports 2019,

9(1):1-16.

29. Fukami H, Omori M, Hatta M: **Phylogenetic relationships in the coral family Acroporidae, reassessed by inference from mitochondrial genes.** Zool Sci 2000, **17**(5):689-696.
30. Ying H, Cooke I, Sprungala S, Wang W, Hayward DC, Tang Y, Huttley G, Ball EE, Forêt S, Miller DJ: **Comparative genomics reveals the distinct evolutionary trajectories of the robust and complex coral lineages.** Genome Biol 2018, **19**(1):1-24.
31. Ohno S: *Evolution by gene duplication.* George Allen & Unwin, London; 1970.
32. Conant GC, Wolfe KH: **Turning a hobby into a job: how duplicated genes find new functions.** Nature Reviews Genetics 2008, **9**(12):938-950.
33. Izumiyama T, Minoshima S, Yoshida T, Shimizu N: **A novel big protein TPRBK possessing 25 units of TPR motif is essential for the progress of mitosis and cytokinesis.** Gene 2012, **511**(2):202-217.
34. Hurst LD: **The Ka/Ks ratio: diagnosing the form of sequence evolution.** Trends in genetics: TIG 2002, **18**(9):486.
35. Clapham DE: **TRP channels as cellular sensors.** Nature 2003, **426**(6966):517-524.
36. Denis V, Cyert MS: **Internal Ca²⁺ release in yeast is triggered by hypertonic shock and mediated by a TRP channel homologue.** J Cell Biol 2002, **156**(1):29-34.
37. Zhou X, Batiza AF, Loukin SH, Palmer CP, Kung C, Saimi Y: **The transient receptor potential channel on the yeast vacuole is mechanosensitive.** Proceedings of the National Academy of Sciences 2003, **100**(12):7105-7110.
38. de Bono M, Tobin DM, Davis MW, Avery L, Bargmann CI: **Social feeding in *Caenorhabditis elegans* is induced by neurons that detect aversive stimuli.** Nature 2002, **419**(6910):899-903.
39. Moya A, Huisman L, Ball EE, Hayward DC, Grasso LC, Chua CM, Woo HN, Gattuso J, Foret S, Miller DJ: **Whole transcriptome analysis of the coral *Acropora millepora* reveals complex responses to CO₂-driven acidification during the initiation of calcification.** Mol Ecol 2012, **21**(10):2440-2454.
40. Bertucci A, Forêt S, Ball EE, Miller DJ: **Transcriptomic differences between day and night in *Acropora millepora* provide new insights into metabolite exchange and light-enhanced calcification in corals.** Molecular Ecology 2015, **24**(17):4489-4504.
41. Oldach MJ, Workentine M, Matz MV, Fan T, Vize PD: **Transcriptome dynamics over a lunar month in a broadcast spawning acroporid coral.** Mol Ecol 2017, **26**(9):2514-2526.
42. Mohamed AR, Cumbo V, Harii S, Shinzato C, Chan CX, Ragan MA, Bourne DG, Willis BL, Ball EE, Satoh N, Miller DJ: **The transcriptomic response of the coral *Acropora digitifera* to a competent *Symbiodinium* strain: the symbiosome as an arrested early phagosome.** Mol Ecol 2016, **25**(13):3127-3141.
43. Yoshioka Y, Yamashita H, Suzuki G, Zayasu Y, Tada I, Kanda M, Satoh N, Shoguchi E, Shinzato C: **Whole-Genome Transcriptome Analyses of Native Symbionts Reveal Host Coral Genomic Novelty for Establishing Coral-Algae Symbioses.** Genome biology and evolution 2021, **13**(1):evaa240.

44. SebPedrs A, Saudemont B, Chomsky E, Plessier F, Mailh M, Renno J, Loe-Mie Y, Lifshitz A, Mukamel Z, Schmutz S: **Cnidarian cell type diversity and regulation revealed by whole-organism single-cell RNA-Seq**. Cell 2018, **173**(6):1520-1534. e20.
45. Levy S, Elek A, Grau-Bov X, Menndez-Bravo S, Iglesias M, Tanay A, Mass T, SebPedrs A: **A stony coral cell atlas illuminates the molecular and cellular basis of coral symbiosis, calcification, and immunity**. Cell 2021, **184**(11):2973-2987. e18.
46. Hu M, Zheng X, Fan C, Zheng Y: **Lineage dynamics of the endosymbiotic cell type in the soft coral *Xenia***. Nature 2020, **582**(7813):534-538.
47. Shultz AJ, Sackton TB: **Immune genes are hotspots of shared positive selection across birds and mammals**. Elife 2019, **8**:e41815.
48. Iguchi A, Shinzato C, Foret S, Miller DJ: **Identification of fast-evolving genes in the scleractinian coral *Acropora* using comparative EST analysis**. PloS one 2011, **6**(6):e20140.
49. Hayes ML, Eytan RI, Hellberg ME: **High amino acid diversity and positive selection at a putative coral immunity gene (tachylectin-2)**. BMC evolutionary biology 2010, **10**(1):1-15.
50. Tautz D, Domazet-Los T: **The evolutionary origin of orphan genes**. Nature Reviews Genetics 2011, **12**(10):692-702.
51. Alb MM, Castresana J: **On homology searches by protein Blast and the characterization of the age of genes**. BMC evolutionary biology 2007, **7**(1):1-8.
52. Yasuoka Y, Shinzato C, Satoh N: **The mesoderm-forming gene brachyury regulates ectoderm-endoderm demarcation in the coral *Acropora digitifera***. Current Biology 2016, **26**(21):2885-2892.
53. Cleves PA, Strader ME, Bay LK, Pringle JR, Matz MV: **CRISPR/Cas9-mediated genome editing in a reef-building coral**. Proceedings of the National Academy of Sciences 2018, **115**(20):5235-5240.
54. Kawamura K, Sekida S, Nishitsuji K, Shoguchi E, Hisata K, Fujiwara S, Satoh N: **In vitro symbiosis of reef-building coral cells with photosynthetic dinoflagellates**. Frontiers in Marine Science 2021, :900.
55. Roach MJ, Schmidt SA, Borneman AR: **Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies**. BMC Bioinformatics 2018, **19**(1):460.
56. Martin M: **Cutadapt removes adapter sequences from high-throughput sequencing reads**. EMBnet.journal 2011, **17**(1):10.
57. Price AL, Jones NC, Pevzner PA: **De novo identification of repeat families in large genomes**. Bioinformatics 2005, **21**(suppl_1):i351-i358.
58. Brna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M: **BRAKER2: Automatic eukaryotic genome annotation with GeneMark-EP and AUGUSTUS supported by a protein database**. NAR Genomics and Bioinformatics 2021, **3**(1):lqaa108.
59. Kim D, Langmead B, Salzberg SL: **HISAT: a fast spliced aligner with low memory requirements**. Nature methods 2015, **12**(4):357.
60. Pertea M, Pertea GM, Antonescu CM, Chang T, Mendell JT, Salzberg SL: **StringTie enables improved reconstruction of a transcriptome from RNA-seq reads**. Nat Biotechnol 2015, **33**(3):290-295.

61. Pertea G, Pertea M: **GFF utilities: GffRead and GffCompare**. F1000Research 2020, **9**.
62. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM: **BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs**. Bioinformatics 2015, **31**(19):3210-3212.
63. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL: **BLAST plus : architecture and applications**. BMC Bioinformatics 2009, **10**:421.
64. Jones P, Binns D, Chang H, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong S, Lopez R, Hunter S: **InterProScan 5: genome-scale protein function classification**. Bioinformatics 2014, **30**(9):1236-1240.
65. Storer J, Hubley R, Rosen J, Wheeler TJ, Smit AF: **The Dfam community resource of transposable element families, sequence models, and genome annotations**. Mobile DNA 2021, **12**(1):1-14.
66. Kanehisa M, Goto S: **KEGG: Kyoto Encyclopedia of Genes and Genomes**. Nucleic Acids Res 2000, **28**(1):27-30.
67. Arai W, Taniguchi T, Goto S, Moriya Y, Uehara H, Takemoto K, Ogata H, Takami H: **MAPLE 2.3. 0: an improved system for evaluating the functionomes of genomes and metagenomes**. Biosci Biotechnol Biochem 2018, **82**(9):1515-1517.
68. Emms DM, Kelly S: **OrthoFinder: phylogenetic orthology inference for comparative genomics**. Genome Biol 2019, **20**(1):1-14.
69. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C: **Salmon provides fast and bias-aware quantification of transcript expression**. Nature methods 2017, **14**(4):417-419.
70. Katoh K, Standley DM: **MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability**. Mol Biol Evol 2013, **30**(4):772-780.
71. Suyama M, Torrents D, Bork P: **PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments**. Nucleic Acids Res 2006, **34**(suppl_2):W609-W612.
72. Wang D, Zhang Y, Zhang Z, Zhu J, Yu J: **KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies**. Genomics, proteomics & bioinformatics 2010, **8**(1):77-80.
73. Villanueva-Canas JL, Laurie S, Alba MM: **Improving genome-wide scans of positive selection by using protein isoforms of similar length**. Genome biology and evolution 2013, **5**(2):457-467.
74. Almagro Armenteros JJ, Sønderby CK, Sønderby SK, Nielsen H, Winther O: **DeepLoc: prediction of protein subcellular localization using deep learning**. Bioinformatics 2017, **33**(21):3387-3395.
75. R core team: **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. 2020, .
76. Conway JR, Lex A, Gehlenborg N: **UpSetR: an R package for the visualization of intersecting sets and their properties**. Bioinformatics 2017, .
77. Wickham H: **Elegant graphics for data analysis**. Media 2009, **35**(211):10.1007.
78. Allen M, Poggiali D, Whitaker K, Marshall TR, Kievit RA: **Raincloud plots: a multi-platform tool for robust data visualization**. Wellcome open research 2019, **4**.stylefix

Tables

Due to technical limitations, Table 1 and Table 2 are only available as a download in the Supplemental Files section.

Figures

Figure 1

Phylogenetic relationships within the Acroporidae and their morphology. (a) Schematic phylogenetic tree representing evolutionary relationships within the Acroporidae. (b and e) Colonies of *Montipora efflorescens* (b) and *M. cactus* (e). (c and d) Eggs of *M. efflorescens* with algal symbionts photographed under visible light (c) and blue light (d). (f and g) A planula larva of *M. efflorescens* with algal symbionts, photographed under visible light (f) and blue light (g). (h) A colony of *Acropora tenuis*. (i and j) A planula larva of *A. tenuis* without algal symbionts photographed under visible light (i) and blue light (j). (k) A

colony of *Astreopora myriophthalma*. Algal symbionts (brown dots) in eggs and planula larvae of *Montipora* (c and f). Green fluorescence was from fluorescent proteins of *Montipora* and red fluorescence was from chlorophyll in algal symbionts (d and g). Orange and cyan-green fluorescence were from fluorescent proteins of *Acropora* (j).

Figure 2

Gene family composition in acroporid genomes and the higher proportion of function-unknown genes in *Montipora*. Left horizontal bars indicate numbers of gene families in each genus. Vertical bars indicate numbers of gene families conserved among genera. Pie charts indicate the generic composition in a given number of gene families (vertical bars). Gene annotation was performed using BLAST searches against the SwissProt database (e-value cutoff: 1e-5), and numbers in pie charts indicate the number of assigned or unassigned gene annotations. Proportions of gene annotation were compared among gene families specific to each lineage and asterisks indicate statistical significance (Pairwise proportion test: $p < 0.05$). Upset plot was produced using the “UpSetR” package [76].

Figure 3

Gene family expansions in *Montipora*. (a) Comparison of numbers of genes in *Montipora* and *Acropora* in each gene family common to the three genera. (b) Comparison of numbers of genes in *Montipora* and *Acropora* in each gene family common to *Montipora* and *Acropora*. (c) Comparison of numbers of genes in *Montipora* and *Astreopora* in each gene family common to *Montipora* and *Astreopora*. The diagonal solid line indicates 1:1 numbers of genes in orthologous families. Possible gene names and gene family

IDs are shown for significantly expanded gene families (Fisher's exact test: $p < 0.05$). Scatter plot was produced using the "ggplot2" package [77].

Figure 4

Relaxed negative selection in *Montipora*-specific gene families. The Y-axis represents the distribution of the ratio of nonsynonymous (Ka) to synonymous amino acid substitutions (Ks). Orthologous gene pairs in two *Montipora* species (*M. cactus* and *M. efflorescens*) are used for calculation of pairwise Ka/Ks rate. Ka/Ks ratios were compared among gene families and significant differences were observed in all pairwise combinations (Wilcoxon rank sum test: $p < 0.05$). A raincloud plot was produced using the "raincloudplots" package [78].

(a)

(b)

Figure 5

Expression patterns of gene families during early life stages of *Montipora* and *Acropora*. (a) Numbers of gene families that are commonly or exclusively expressed in early life stages of *Montipora* and *Acropora*. Numbers of gene families that are exclusively expressed in each genus are shown in upper parentheses after genera names. The number of gene families that are commonly expressed in both *Montipora* and *Acropora* early life stages are shown in an orange box. Gene families that are exclusively expressed in

early life stages of *Montipora* were further classified according to whether they are expressed in one of the stages (green), or throughout all stages (red). (b) Gene families expressed in early *Montipora* life stages. For *Montipora*, RNA-seq data from planula larvae, metamorphosed larvae and recruits were used. For *Acropora*, RNA-seq data from blastulae, gastrulae, planula larvae and polyps were used. SRA accession numbers for the RNA-seq data are provided in Supplementary Table S7.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryMaterial.docx](#)
- [Table2.xlsx](#)
- [SupplementaryFigureS1.png](#)
- [Table1.xlsx](#)
- [SupplementaryTableS1.xlsx](#)
- [SupplementaryTableS2.xlsx](#)
- [SupplementaryTableS3.xlsx](#)
- [SupplementaryTableS4.xlsx](#)
- [SupplementaryTableS5.xlsx](#)
- [SupplementaryTableS6.xlsx](#)
- [SupplementaryTableS7.xlsx](#)
- [supplementaryData1.txt](#)
- [SupplementaryData2.xlsx](#)
- [SupplementaryData3.gz](#)
- [SupplementaryData4.gz](#)
- [SupplementaryData5.gz](#)