

RESEARCH

Open Access



# Comparative genomics highlight the importance of lineage-specific gene families in evolutionary divergence of the coral genus, *Montipora*

Yuki Yoshioka<sup>1,2</sup>, Go Suzuki<sup>3</sup>, Yuna Zayasu<sup>4</sup>, Hiroshi Yamashita<sup>3</sup> and Chuya Shinzato<sup>1\*</sup>

## 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 that might 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 gene families restricted to *Montipora* are significantly more numerous 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 transcriptomic 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.

**Keywords:** Corals, *Montipora*, Comparative genomics, Evolution, Rapidly evolving genes, Symbiosis, Transcriptome, Vertical transmission of algal symbionts

## 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

\*Correspondence: c.shinzato@aori.u-tokyo.ac.jp

<sup>1</sup> Atmosphere and Ocean Research Institute, The University of Tokyo, Kashiwa, Chiba, Japan

Full list of author information is available at the end of the article

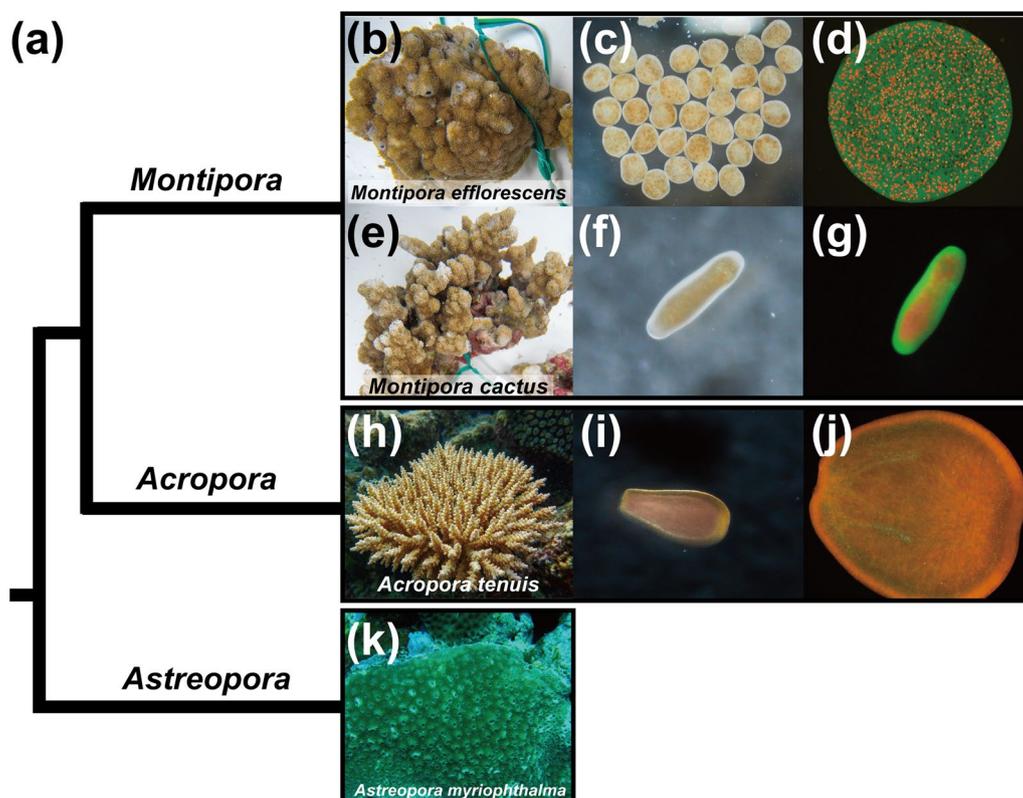


© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

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; Fig. 1) is one of the most widespread genera of 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 maternal transmission of symbionts and higher stress tolerance. Symbiont transmission maintains symbioses across generations and strongly influences host evolution and adaptation to environments [12–14]. Two

fundamental symbiont 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] (Fig. 1). Offspring of horizontal recipients 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 recipients inherit symbionts that are suitable for their physiology [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



**Fig. 1** Phylogenetic relationships in the Acroporidae and acroporid 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**)

distinct differences between *Montipora* and its close relative, *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 sister taxon of the remainder of the Acroporidae [29] (Fig. 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 resulted from gene duplications. Although some studies have performed 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 4925 to 3521 in *M. cactus* and from 5162 to 3589 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; Table 1). Completeness of gene models reported by Helmkamp et al. [27] was lower than that reported by Shumaker et al. [28] (64.2%, with 0.5% duplicated; Table 1). Thus, the gene models reported by Shumaker et al. [28] contained many duplicates, but those reported by Helmkamp 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*), two *Montipora* species (*M. cactus* and *M. efflorescens*), and *Astreopora myriophthalma*. We obtained 12,769 gene families for *Montipora*, 11,007 for *Acropora* and 11,309 for *Astreopora* (Fig. 2). We then categorized each gene family into seven groups, (1) shared by all three genera (9690 gene families), (2) shared by *Montipora* and *Acropora* (743 gene families), (3) shared by *Montipora* and *Astreopora* (665 gene families), (4) shared by *Acropora* and *Astreopora* (257 gene families), (5) restricted to *Montipora* (1670 gene families), (6) restricted to *Astreopora* (696 gene families) and (7) restricted to *Acropora* (316 gene families) (Fig. 2). 75.8% (9690/12,769) of the gene families in *Montipora*, 88% (9690/11,007) in *Acropora*, and 85.7% (9690/11,309) in *Astreopora* were shared among all three genera (Fig. 2), indicating that a large number (~88%) of gene families are shared throughout the Acroporidae, and these are likely to be the core-gene families among acroporids (Additional file 9: Data S1).

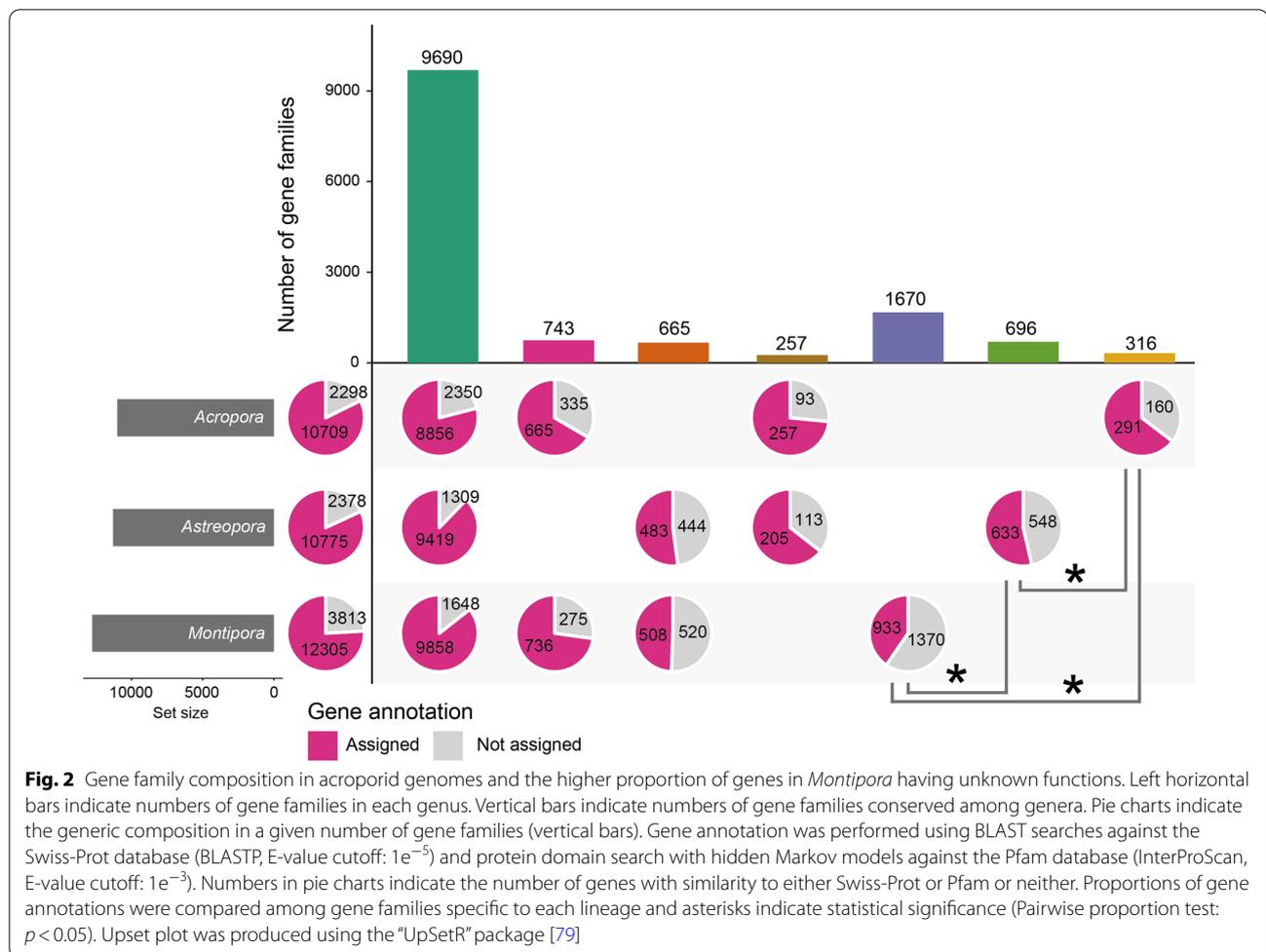
The two major clades of reef-building corals, known as Robusta and Complexa [30], possess different metabolic pathways [31]. From the six species, we compared 303 functional modules comprising ten categories in the

**Table 1** Statistics for gene prediction and assembly of *Montipora*, *Astreopora* and other coral genomes

Genus	<i>Montipora</i>		<i>Astreopora</i>		<i>Acropora</i>	
	<i>M. cactus</i>	<i>M. cactus</i>	<i>M. capitata</i>	<i>A. myriophthalma</i>	<i>A. tenuis</i>	<i>A. millepora</i> <sup>a</sup>
References	This study	Shinzato et al. [23]	Shumaker et al. [28]	This study	Shinzato et al. [23]	Shinzato et al. [23]
<b>Genome assembly</b>						
Assembly length	640,828,888	65,27,28,006	88,57,04,498	37,34,03,447	40,30,90,051	41,58,26,927
No. of scaffolds	3532	4925	3043	1149	1538	3869
N50	9,38,091	8,98,511	5,40,623	16,34,277	11,65,818	18,56,312
<b>Genes</b>						
No. of protein coding genes	29,158	21,983	63,227	25,406	22,905	22,327
Avg. length	12,905	15,343	6507 <sup>b</sup>	9550	10,298	10,445
<b>Transcripts</b>						
Avg. no. of CDS (exon) per gene	6.6 (7.2)	7.7 (7.9)	4.5 (NA)	6.0 (6.4)	7.2 (8.1)	6.6 (8.5)
Avg. CDS (exon) length per gene (bp)	1411 (2,033)	1871 (2,139)	1162 (NA)	1647 (2,132)	1562 (2,383)	1435 (2,415)
Avg. CDS (exon) length	215 (280)	244 (271)	259 (NA)	230 (279)	216 (294)	217 (285)
Total CDS (exon) length (Mbp)	41 (59)	41 (45)	73 (NA)	42 (54)	36 (55)	38 (57)
<b>BUSCO completeness (v5.0)</b>						
Complete	93.3%	87.9%	93.4% <sup>b</sup>	94.5%	92.9%	89.3%
Single-copy	92.5%	87.3%	75.1% <sup>b</sup>	93.2%	92.2%	88.6%
Duplicated	0.8%	0.6%	18.3% <sup>b</sup>	1.3%	0.7%	0.7%
Fragmented	4.1%	5.7%	4.1% <sup>b</sup>	3.1%	2.9%	4.7%
Missing	2.6%	6.4%	2.5% <sup>b</sup>	2.4%	4.2%	6.0%

<sup>a</sup> RefSeq assembly (GCF\_004143615.1) was used for the analysis

<sup>b</sup> CDS were used for the calculation



Kyoto Encyclopedia of Genes and Genomes (KEGG) metabolic pathways and found that metabolic pathways were basically conserved in the three genera (Additional file 1: 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 (Additional file 1: 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 (Additional file 1: Table S1), as reported in Ying et al. [31]. 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 1670 gene families restricted to *Montipora* (2307 genes in *M. cactus* and 2303 in *M. efflorescens*) (Fig. 2). The proportion of lineage-specific gene families in *Montipora*

(13.07%) was significantly larger than that in *Acropora* (2.87%) and *Astreopora* (6.15%) (Pairwise proportion test:  $p < 0.05$ ). Although we performed gene annotation with BLAST searches against the Swiss-Prot database (BLASTP, E-value cutoff:  $1e^{-5}$ ) and hidden Markov models against the Pfam database (InterProScan, E-value cutoff:  $1e^{-3}$ ), the proportion of *Montipora*-specific gene families with gene annotation was significantly lower than in *Acropora* and *Astreopora* (Pairwise proportion test:  $p < 0.05$  for all combinations; Fig. 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 [32, 33]. To explore gene families that underwent expansions, we first compared gene numbers of 9,690 gene families shared by the three genera and 743 gene families shared by *Montipora* and *Acropora* (Fig. 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 last common ancestor. Three gene families, homologous dimethylsulfonylproprionate (DMSP) lyase (*ALMA*; HOG0000829), Endonuclease-reverse transcriptase (*GPI*; HOG0000531), and Spondin (*SPONI*; HOG0001590), and three non-annotated gene families (NA; HOG0000965, HOG0001135, and HOG0001312), were significantly expanded in *Acropora* (Fisher's exact test:  $p < 0.05$ ; Fig. 3a and b). Recently, it was reported that DMSP lyase is the most expanded gene family in *Acropora* [28], and our result is consistent with that 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:  $p < 0.05$ ; Fig. 3a and 3b).

Next, we compared gene numbers of 665 gene families shared by *Montipora* and *Astreopora* (Fig. 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 Swiss-Prot annotation were significantly expanded in *Astreopora* (Fisher's exact test:  $p < 0.05$ ; Fig. 3c), whereas one other gene family, tetratricopeptide repeat protein 28 (*TTC28*; HOG0000387), which is involved in the cell cycle in humans [34], was significantly expanded in *Montipora* compared with *Astreopora* (Fisher's exact test:  $p < 0.05$ ; Fig. 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 [35]. For example, when Ka is less than Ks ( $Ka/Ks < 1$ ), selection has occurred to eliminate mutations of protein sequences (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 shared by the three Acroporidae genera, (2) gene families shared by *Montipora* and *Acropora*, (3) gene families shared by *Montipora* and *Astreopora*, and (4) gene families restricted to *Montipora* (Fig. 4). When we compared Ka/Ks ratio between groups, gene families restricted to *Montipora* showed the highest Ka/Ks ratio (Wilcoxon rank sum test:  $p < 0.05$ ; Fig. 4), indicating that this gene family group has undergone a relaxation of purifying 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 shared by the three genera or shared by *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*

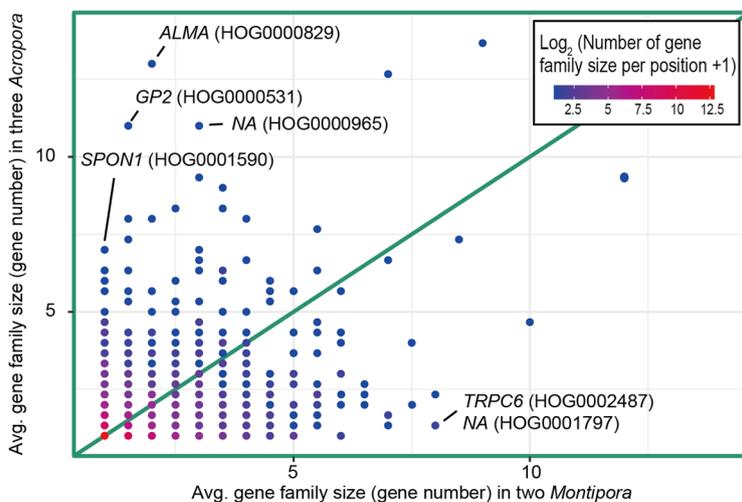
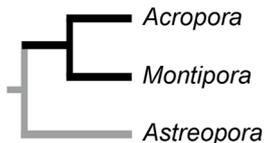
Presence of maternally inherited algal symbionts at an early life stage is the most obvious difference between vertical and horizontal transmitters (Fig. 1). In order to identify gene families specifically involved in symbiosis in vertical transmitters, we compared the repertoire of expressed genes in early life stages of *Montipora* with those expressed in *Acropora*. In this analysis, a gene family was considered expressed even if only one gene in that family was expressed (Transcripts per million (TPM)  $> 1$ ). We confirmed that 11,930 and 10,838 gene families were expressed during early life stages of *Montipora* and *Acropora*, respectively (Fig. 5a). Of these, 10,051 gene families (84% in *Montipora* and 93% in *Acropora*) were shared

(See figure on next page.)

**Fig. 3** Gene family expansions in *Montipora*. **a** Numbers of genes in *Montipora* and *Acropora* in each gene family shared by the three genera. **b** Numbers of genes in *Montipora* and *Acropora* in each gene family shared by *Montipora* and *Acropora*. **c** Numbers of genes in *Montipora* and *Astreopora* in each gene family shared by *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 [80]

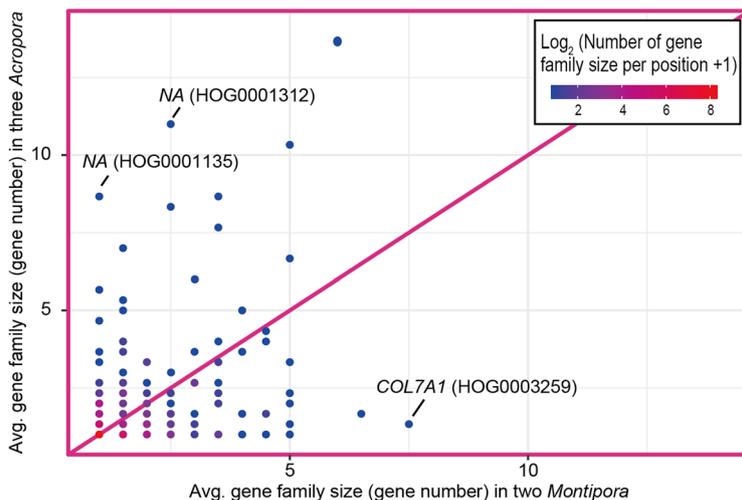
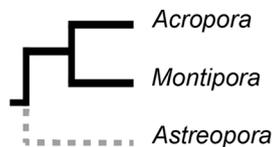
(a)

Expanded in *Montipora* or *Acropora*



(b)

Expanded in *Montipora* or *Acropora* but does not exist in *Astreopora*



(c)

Expanded in *Montipora* or *Astreopora* but does not exist in *Acropora*

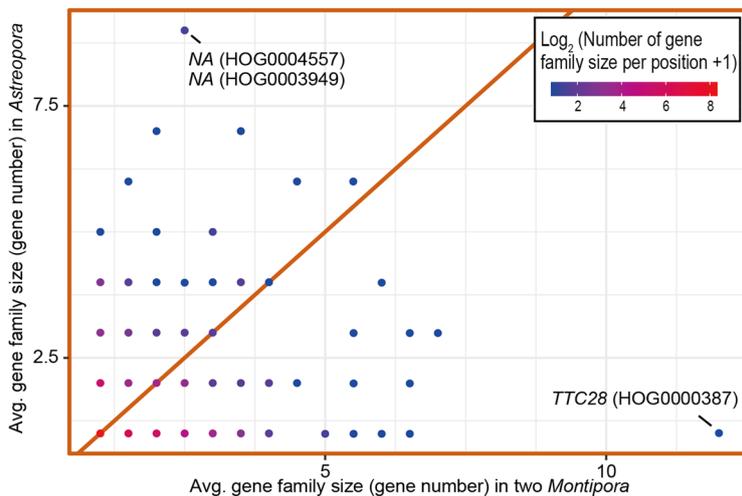
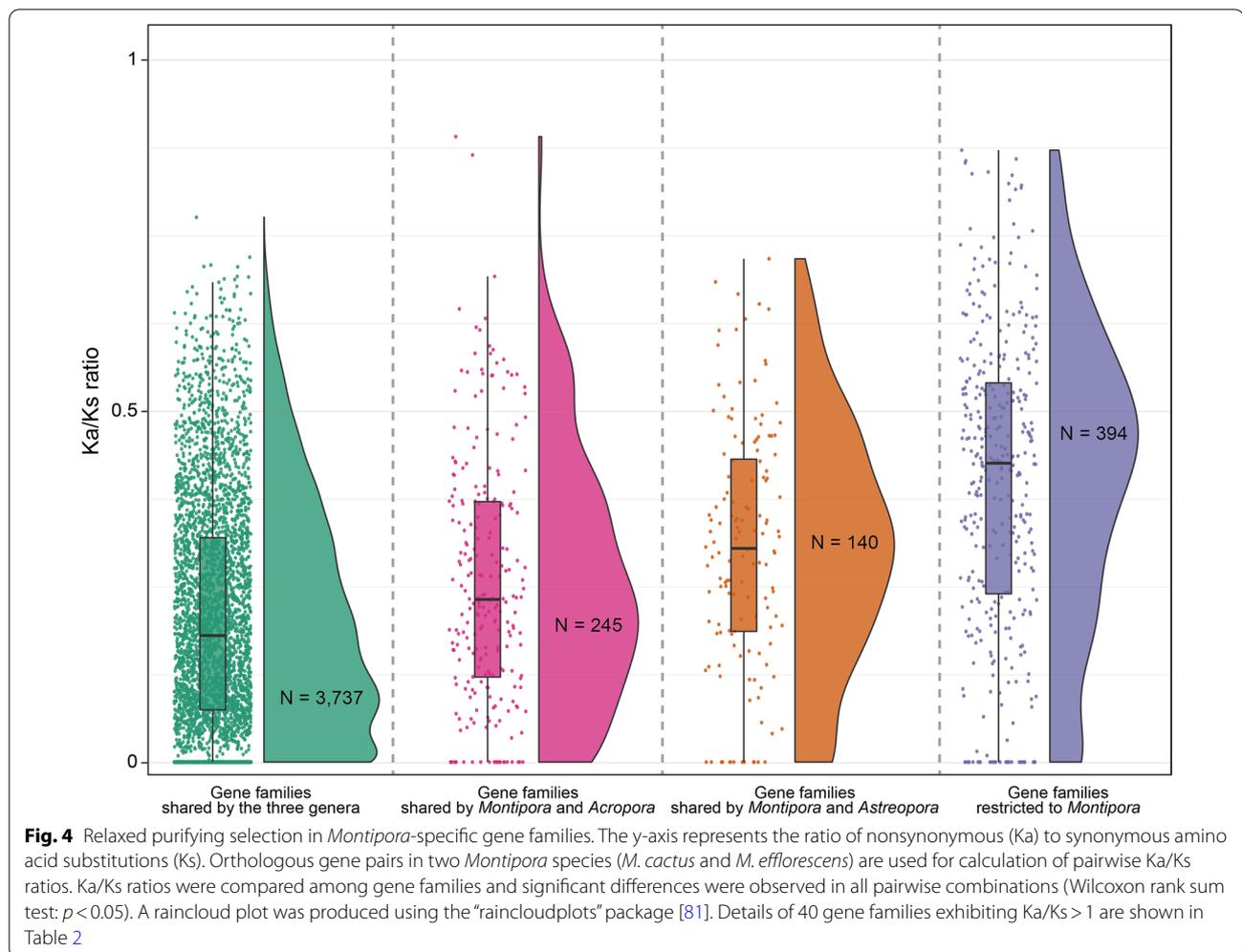


Fig. 3 (See legend on previous page.)



by both during early life stages (Fig. 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* (Fig. 5b). Among those, 60% (1132 gene families) were expressed in planula larvae, metamorphosed larvae, and recruit stages (Fig. 5a), suggesting that these genes may be related to maintenance of algal symbionts in *Montipora*. Interestingly, 97% of these gene families ((753 + 344) / 1132, Fig. 5b) that were expressed throughout the three life stages were specific to *Montipora* or shared by *Astreopora* (Additional file 2: Table S2). In contrast, the remaining 3% of gene families ((22 + 13) / 1,132, Fig. 5b) have orthologs in *Acropora*, but were not expressed in *Acropora*. Nonetheless, they were expressed throughout early life stages of *Montipora* (Additional file 3: 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 (Additional file 2: 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 the genera *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 Helmkampf et al.

**Table 2** Gene families under positive selection (Ka/Ks ratio > 1) in *Montipora*

Gene family group	Gene family ID	Expressed in early life stages of	Swiss-Prot annotation (BLASTP, e-value cutoff: $1e^{-5}$ )	Pfam domain (InterProScan, e-value cutoff: $1e^{-3}$ )	Subcellular localization	Ka	Ks	Ka/Ks ratio
Common to the three genera	HOG0011297	<i>Montipora</i> and <i>Acropora</i>	G patch domain and ankyrin repeat-containing protein 1	Ankyrin repeats (3 copies) (PF12796) G-patch domain (PF01585)	Nucleus	1.02946	0.868207	1.18573
	HOG0013033	<i>Montipora</i> and <i>Acropora</i>	NA	NA	Nucleus	0.047466	0.016942	2.80166
	HOG0013161	<i>Montipora</i> and <i>Acropora</i>	NA	NA	Extracellular	0.067747	0.011507	5.88723
	HOG0013504	<i>Montipora</i> and <i>Acropora</i>	NA	NA	Extracellular	0.250122	0.146258	1.71014
	HOG0016452	<i>Montipora</i> and <i>Acropora</i>	Nucleoredoxin-like protein 2	Thioredoxin-like (PF13905)	Cytoplasm	1.04061	0.866303	1.2012
Common to <i>Montipora</i> and <i>Acropora</i>	HOG0000399	-	NA	NA	Mitochondrion	1.04494	0.829832	1.25922
	HOG0007354	<i>Montipora</i> and <i>Acropora</i>	NA	NA	Cytoplasm	0.043171	0.01538	2.80698
	HOG0008211	<i>Montipora</i> and <i>Acropora</i>	E3 ubiquitin-protein ligase TRIM71	RING-type zinc-finger (PF13445) B-box zinc finger (PF00643)	Nucleus	1.02988	0.901833	1.14199
	HOG0008701	<i>Montipora</i> and <i>Acropora</i>	Collagen alpha-1(V) chain	Fibrillar collagen C-terminal domain (PF01410)	Extracellular	0.086894	0.031594	2.75035
	HOG0018764	<i>Montipora</i> and <i>Acropora</i>	Histamine H2 receptor	7 transmembrane receptor (rhodopsin family) (PF00001)	Cell membrane	1.03386	0.896216	1.15358
Restricted to <i>Montipora</i>	HOG0000410	<i>Montipora</i> <sup>a</sup>	NA	NA	Cytoplasm	1.0489	0.858182	1.22224
	HOG0006069	<i>Montipora</i> <sup>a</sup>	NA	NA	Cytoplasm	1.06207	0.789879	1.34459
	HOG0025744	<i>Montipora</i> <sup>a</sup>	NA	NA	Cytoplasm	1.04814	0.846526	1.23816
	HOG0025248	<i>Montipora</i> <sup>a</sup>	NA	NA	Golgi apparatus	1.06112	0.837214	1.26744
	HOG0025123	<i>Montipora</i> <sup>a</sup>	NA	NA	Mitochondrion	1.03568	0.871894	1.18785
	HOG0025167	<i>Montipora</i> <sup>a</sup>	NA	NA	Mitochondrion	1.05341	0.81252	1.29647
	HOG0025503	<i>Montipora</i> <sup>a</sup>	NA	NA	Mitochondrion	1.06094	0.794999	1.33451
	HOG0025176	<i>Montipora</i> <sup>a</sup>	NA	NA	Plastid	1.07662	0.736763	1.46128
	HOG0025663	<i>Montipora</i> <sup>a</sup>	NA	NA	Lysosome/ Vacuole	1.0435	0.883964	1.18048
	HOG0023383	<i>Montipora</i> <sup>b</sup>	NA	NA	Cytoplasm	1.0506	0.86614	1.21297
	HOG0024846	<i>Montipora</i> <sup>b</sup>	NA	NA	Cytoplasm	0.157342	0.054572	2.8832
	HOG0025044	<i>Montipora</i> <sup>b</sup>	NA	NA	Cytoplasm	1.04722	0.8635	1.21276
	HOG0025199	<i>Montipora</i> <sup>b</sup>	NA	NA	Cytoplasm	1.04978	0.833375	1.25967
	HOG0024900	<i>Montipora</i> <sup>b</sup>	NA	NA	Endoplasmic reticulum	0.072466	0.030348	2.38782
	HOG0025608	<i>Montipora</i> <sup>b</sup>	NA	NA	Endoplasmic reticulum	1.05623	0.836453	1.26275
	HOG0024879	<i>Montipora</i> <sup>b</sup>	NA	NA	Extracellular	1.04481	0.820572	1.27328
	HOG0024911	<i>Montipora</i> <sup>b</sup>	NA	NA	Extracellular	1.08591	0.764205	1.42097

**Table 2** (continued)

Gene family group	Gene family ID	Expressed in early life stages of	Swiss-Prot annotation (BLASTP, e-value cutoff: $1e^{-5}$ )	Pfam domain (InterProScan, e-value cutoff: $1e^{-3}$ )	Subcellular localization	Ka	Ks	Ka/Ks ratio
	HOG0024765	<i>Montipora</i> <sup>b</sup>	NA	NA	Mitochondrion	1.06916	0.745005	1.43511
	HOG0025496	<i>Montipora</i> <sup>b</sup>	NA	NA	Mitochondrion	1.03657	0.886546	1.16922
	HOG0025423	<i>Montipora</i> <sup>b</sup>	NA	NA	Mitochondrion	1.05678	0.799324	1.32209
	HOG0025535	<i>Montipora</i> <sup>b</sup>	Integrase/recombinase xerD homolog	NA	Mitochondrion	1.04223	0.880029	1.18431
	HOG0025606	<i>Montipora</i> <sup>b</sup>	NA	NA	Mitochondrion	1.05418	0.833955	1.26408
	HOG0024806	<i>Montipora</i> <sup>b</sup>	NA	Zinc knuckle (PF00098)	Nucleus	1.0438	0.876216	1.19126
	HOG0024834	<i>Montipora</i> <sup>b</sup>	NA	NA	Nucleus	1.04457	0.853951	1.22323
	HOG0024972	<i>Montipora</i> <sup>b</sup>	NA	NA	Nucleus	1.06879	0.737384	1.44943
	HOG0025048	<i>Montipora</i> <sup>b</sup>	NA	NA	Nucleus	1.0465	0.828206	1.26357
	HOG0025616	<i>Montipora</i> <sup>b</sup>	NA	NA	Nucleus	1.07293	0.838674	1.27932
	HOG0024996	-	NA	NA	Cell membrane	0.827813	0.407974	2.02908
	HOG0024975	-	NA	NA	Endoplasmic reticulum	1.02887	0.884795	1.16284
	HOG0025445	-	TNF receptor-associated factor 6	TRAF-type zinc finger (PF02176)	Nucleus	0.094242	0.019425	4.85163

<sup>a</sup> Expressed at least one early life stages

<sup>b</sup> Expressed at all three early life stages

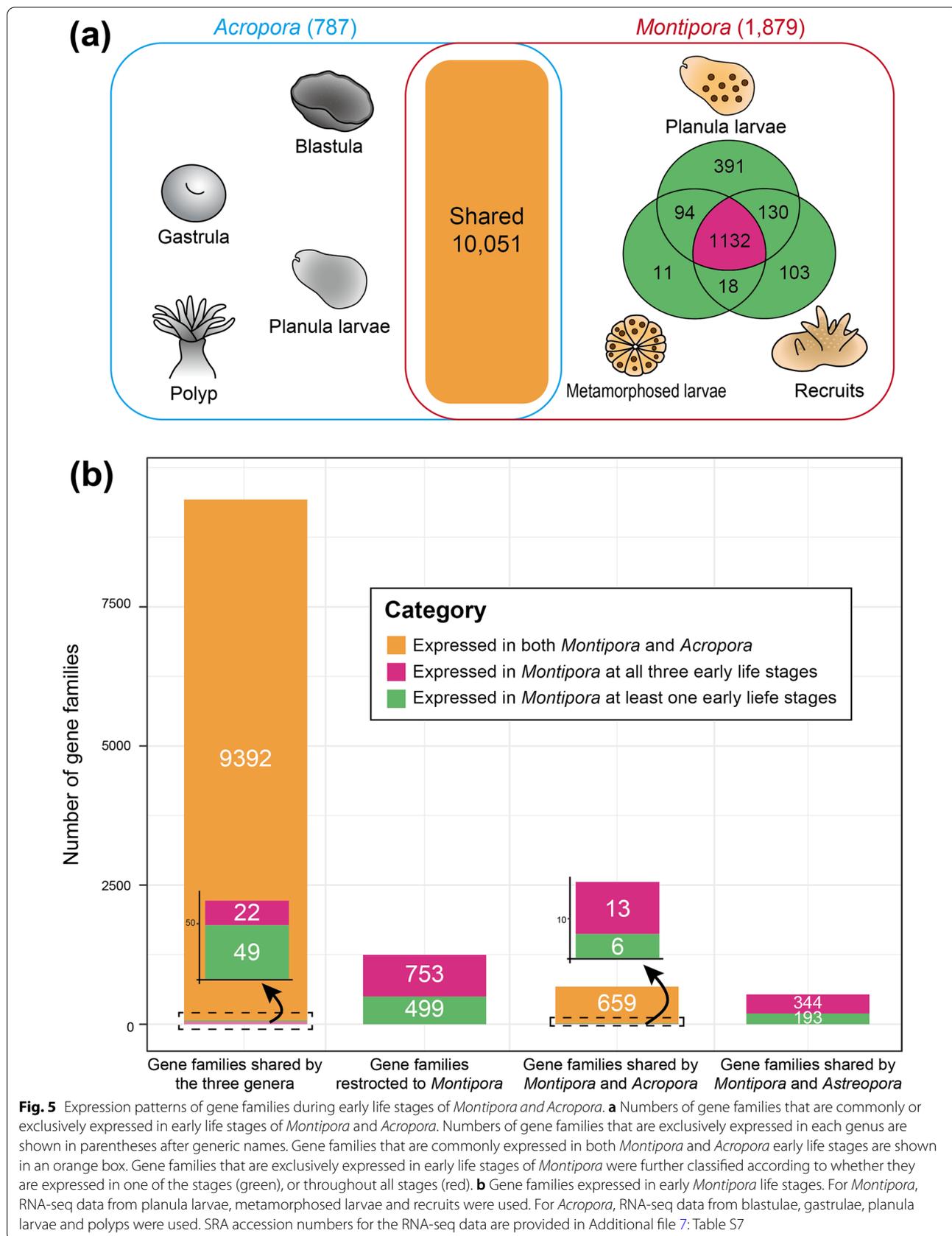
[27]; Table 1), and were comparable to those of other coral species (Table 1). These numbers indicate that we successfully obtained high-quality gene models from *Montipora* and *Astreopora* species. Numbers of genes in *M. cactus* and *M. efflorescens* genomes were not quite as large as those of *M. 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]. Nonetheless, we identified four expanded gene families

in *Montipora* (Fig. 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 (Fig. 2). *Montipora* does not appear to have duplicated existing gene families, as has *Acropora*. Lineage-specific gene families contribute to larger gene numbers in *Montipora* genomes, and emergence of lineage-specific genes may have helped to establish maternal transmission of symbionts in *Montipora* corals. In particular, *Montipora*-specific gene families under positive selection may be major contributors.

Three gene families, homologous to *TRPC6*, *TTC28*, and *COL7A1*, and one gene family without annotation were significantly expanded in *Montipora* compared with *Acropora* or *Astreopora* (Fig. 3). Known functions of transient receptor potential (TRP) proteins encoded by *TRPC* are diverse (reviewed in [36]). For example, TRP proteins respond to hypertonicity in yeasts [37, 38], detect and avoid noxious chemicals in nematodes [39], and discriminate warmth, heat, and cold in humans [36]. In each case, TRP proteins mediate sensory transduction in cells [36]. In corals, expression levels of *TRP*-like genes change when the concentration



of CO<sub>2</sub> in seawater changes [40]. They also change diurnally [41, 42] or when exposed to symbiotic algae [43, 44]. 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) (Additional file 4: Table S4). Canonical *TTC28* is composed of tetratricopeptide repeats and CHAT domains (Q96AY4: *TTC28\_HUMAN* [34]) 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 Additional file 9: Data S1). *TTC28* is required for the cell cycle in humans [34]. 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 [45–47] 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 [48]. In corals, genes related to immunity, such as lectins and antimicrobial peptides, are also under positive selection [23, 49, 50]. 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 [51]. They are thought to arise principally by two processes: gene duplication or de novo evolution from non-coding regions [51]. 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 [52], 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 [51]. 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 maternal transmission of symbionts in *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 maternal transmission of symbionts is inferred. Our dataset and findings offer novel insights into mechanisms of coral-algal symbiosis. Although genetic tools for manipulating corals have been established [53, 54], 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 [55], 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 *M. 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). Coral fragments (~3 cm diameter) from adult colonies of *M. cactus*, *M. efflorescens*, and *Astreopora myriophthalma* were snap frozen in liquid nitrogen and stored at -80°C until use. Fragments were then crushed in liquid nitrogen with an iron and hammer into powder. Total RNA was extracted from the powder 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*, eggs, sperm, planula larvae (1 and 4 d post-fertilization) were collected and preserved with TRIzol reagent (Thermo Fisher Scientific) at -80°C until use. Total RNA was extracted from preserved eggs, sperm, and planula larvae as in Yoshioka et al. [56]. KAPA RNA HyperPrep Kits (Kapa Biosystems) and MGIEasy RNA Directional Library Prep Sets (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 Additional file 5: Table S5.

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

We downloaded scaffold sequences of *M. cactus* and *M. efflorescens*, assembled from DNA sequences extracted from symbiotic algae-free coral sperm [23], 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 two allelic copies of heterozygous regions, using Purge Haplotigs v1.1.1 [57] with default settings. These were excluded from subsequent analyses.

In addition to the above RNA samples, we used publicly available RNA-seq data from NCBI SRA for gene prediction (Additional file 6: Table S6). Low-quality reads (quality score < 20 and length < 20 bp) and sequence adaptors were trimmed using CUTADAPT v1.18 [58]. 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 [59] 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 [60], with AUGUSTUS v3.3.3. RNA-seq reads were aligned to each genome sequence with HISAT v2.1.0 [61]. Then, the alignment information was used for BRAKER gene prediction with options “UTR=on”, “soft-masking”, and “AUGUSTUS\_ab\_initio.” To improve gene prediction, we further executed genome-guided transcriptome assembly using StringTie [62] 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 [63], as summarized in Additional file 8: Figure S1. To evaluate the completeness of predicted genes, we used BUSCO v5.0 [64] 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 three *Acropora* species for the following reasons. *A. digitifera* genome sequences were assembled with PacBio long-reads [23]. *A. millepora* genome assembly and gene models were curated by NCBI. *A. tenuis* had the second highest gene model completeness (BUSCO completeness scores) among *Acropora* genomes following *A. millepora* [23] and represents a distinct phylogenetic clade from *A. digitifera* and *A. millepora* [23]. We selected the longest transcript variants from each gene and translated them into protein sequences. All proteomes were annotated with BLASTP [65] (E-value cut off:  $1e^{-5}$ ) against the Swiss-Prot database (8 January 2021). In addition, domains in protein sequences were searched using hidden Markov models against the Pfam database with InterProScan v5.31–70.0 (E-value cutoff:  $1e^{-3}$ ) [66]. In addition, putative transposable elements in gene models were identified with TransposonPSI (<http://transposon.psi.sourceforge.net/>), Dfam scan (release 3.3; Storer et al. [67]), and Pfam keyword (“Reverse transcriptase” and “Integrase”). All protein sequences were also annotated with KEGG [68] in all eukaryote genes using GenoMaple v2.3.2 [69] 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 [70] and *Porites australiensis* gene models [71] were also included as an outgroup for the Acroporidae. In this study, we used phylogenetic hierarchical orthogroups (HOG) as gene families. Gene families shared by 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) (Additional file 7: Table S7). In addition, publicly available RNA-seq data of *M. capitata* (planula larvae, metamorphosed larvae, and recruits) were also used in this study (Additional file 7: 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 [72]. 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 [73]. Aligned nucleotide codon sequences without alignment gaps were retrieved using the PAL2NAL script [74]. 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 [75] with option “-MA”. Following Villanueva-Canas et al. [76], we discarded gene families showing  $K_s < 0.01$ , as such low  $K_s$  values may result in inaccurate Ka/Ks estimates, and gene families showing  $K_s$  or  $K_a > 2$  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  $K_a/K_s > 1$  was predicted using the DeepLoc-1.0 online server [77].

### 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 [78].

### 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.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12862-022-02023-8>.

**Additional file 1: Table S1.** Summary of comparison of KEGG metabolic pathways among acroporid genomes. MCR indicates the module completion ratio, calculated by GenoMaple.

**Additional file 2: Table S2.** IDs of gene families common to *Montipora* and *Astreopora* or restricted to *Montipora*, that are expressed during three life stages (planula larvae, metamorphosed larvae, and recruits).

**Additional file 3: Table S3.** IDs of gene families common to three genera (*Montipora*, *Acropora* and *Astreopora*) or common to *Montipora* and *Acropora*, that are expressed during three life stages (planula larvae, metamorphosed larvae, and recruits).

**Additional file 4: Table S4.** Summary of gene families significantly (Fisher’s exact test;  $p < 0.05$ ) expanded in the *Montipora* genome compared to those of *Acropora* or *Astreopora*.

**Additional file 5: Table S5.** Summary of samples used for RNA-seq.

**Additional file 6: Table S6.** Summary of publicly available RNA-seq samples that were included for gene prediction in *Montipora* and *Astreopora*.

**Additional file 7: Table S7.** Summary of RNA-seq samples used for comparative transcriptomic analyses.

**Additional file 8: Figure S1.** Summary of gene predictions for *M. cactus*, *M. efflorescens*, and *Astreopora myriophthalma*.

**Additional file 9: Data S1.** Result of orthologous clustering in Acroporidae genomes using OrthoFinder.

**Additional file 10: Data S2.** Summary of retained scaffolds after genome assembly curation for *Montipora cactus* and *M. efflorescens*.

**Additional file 11: Data S3.** Gene models for *M. cactus* in GTF format.

**Additional file 12: Data S4.** Gene models for *M. efflorescens* in GTF format.

**Additional file 13: Data S5.** Gene models for *Astreopora myriophthalma* in GTF format.

### 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.

### Author contributions

CS and GS conceptualized, and CS supervised the project. GS, CS and YZ performed coral sampling. HY helped with sampling of *Montipora* larvae. YY performed molecular biological experiments and bioinformatic analyses. YY and CS wrote main manuscript. All authors read and approved the final manuscript.

### 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).

### 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 Additional file 10: Data S2. Gene models in GTF format for *M. cactus*, *M. efflorescens* and *Astreopora myriophthalma* are provided as Additional file 11: Data S3, Additional file 12: Data S4, Additional file 13: Data 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 <https://doi.org/10.15482/USDA.ADC/1503958>).

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

Chuya Shinzato is an editorial board member for *BMC Ecology and Evolution*. All other authors declare that they have no competing interests.

### Author details

<sup>1</sup>Atmosphere and Ocean Research Institute, The University of Tokyo, Kashiwa, Chiba, Japan. <sup>2</sup>Graduate School of Frontier Science, The University of Tokyo, Kashiwa, Chiba, Japan. <sup>3</sup>Fisheries Technology Institute, Japan Fisheries Research and Education Agency, Ishigaki, Okinawa, Japan. <sup>4</sup>Marine Genomics Unit, Okinawa Institute of Science and Technology Graduate University, Onna, Okinawa, Japan.

Received: 28 September 2021 Accepted: 13 May 2022

Published online: 27 May 2022

## References

- 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):1280–4.
- Muscatine L, Porter JW. Reef corals: mutualistic symbioses adapted to nutrient-poor environments. *Bioscience*. 1977;27(7):454–60.
- Muscatine L, McCloskey LR, Marian RE. Estimating the daily contribution of carbon from zooxanthellae to coral animal respiration 1. *Limnol Oceanogr*. 1981;26(4):601–11.
- Falkowski PG, Dubinsky Z, Muscatine L, Porter JW. Light and the bioenergetics of a symbiotic coral. *Bioscience*. 1984;34(11):705–9.
- 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):560–3.
- Hoegh-Guldberg O, Poloczanska ES, Skirving W, Dove S. Coral reef ecosystems under climate change and ocean acidification. *Front Mar Sci*. 2017;4:158.
- 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):80–3.
- Brown BE. Coral bleaching: causes and consequences. *Coral Reefs*. 1997;16(S1):S129–38.
- Bellwood DR, Hughes TP, Folke C, Nystrom M. Confronting the coral reef crisis. *Nature*. 2004;429(6994):827–33.
- Veron JEN, Stafford-Smith M. Corals of the world. 1st ed. Townsville: Australian Inst. of Marine Science; 2000.
- Veron J. Scleractinia of Eastern Australia, Part 5: Acroporidae. *Aust Inst Mar Sci Monogr Ser*. 1984;6:485.
- Yamamura N. Vertical transmission and evolution of mutualism from parasitism. *Theor Popul Biol*. 1993;44(1):95–109.
- Yamamura N. Evolution of mutualistic symbiosis: a differential equation model. *Res Popul Ecol*. 1996;38(2):211–8.
- Bright M, Bulgheresi S. A complex journey: transmission of microbial symbionts. *Nat Rev Microbiol*. 2010;8(3):218–30.
- Baird AH, Guest JR, Willis BL. Systematic and biogeographical patterns in the reproductive biology of scleractinian corals. *Annu Rev Ecol Evol Syst*. 2009;40:551–71.
- Richmond RH, Hunter CL. Reproduction and recruitment of corals: comparisons among the Caribbean, the tropical Pacific, and the Red Sea. *Mar Ecol Prog Ser*. 1990;60(1):185–203.
- Fadlallah YH. Sexual reproduction, development and larval biology in scleractinian corals. *Coral Reefs*. 1983;2(3):129–50.
- Douglas AE. Host benefit and the evolution of specialization in symbiosis. *Heredity*. 1998;81(6):599–603.
- Rowan R, Knowlton N. Intraspecific diversity and ecological zonation in coral-algal symbiosis. *Proc Natl Acad Sci*. 1995;92(7):2850–3.
- 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.
- Putnam HM, Davidson JM, Gates RD. Ocean acidification influences host DNA methylation and phenotypic plasticity in environmentally susceptible corals. *Evol Appl*. 2016;9(9):1165–78.
- 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):717–23.
- Shinzato C, Khalturnin 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.
- 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):320–3.
- 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 Biol Evol*. 2019;11(5):1374–9.
- 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. <https://doi.org/10.1126/science.aba4674>.
- Helmkampf M, Bellinger MR, Geib SM, Sim SB, Takabayashi M. Draft genome of the rice coral *Montipora capitata* obtained from linked-read sequencing. *Genome Biol Evol*. 2019;11(7):2045–54.
- 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*. *Sci Rep*. 2019;9(1):1–16.
- Fukami H, Omori M, Hatta M. Phylogenetic relationships in the coral family Acroporidae, reassessed by inference from mitochondrial genes. *Zool Sci*. 2000;17(5):689–96.
- Romano SL, Palumbi SR. Evolution of scleractinian corals inferred from molecular systematics. *Science*. 1996;271(5249):640–2.
- Ying H, Cooke I, Sprungala S, Wang W, Hayward DC, Tang Y, Huttler 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.
- Ohno S. Evolution by gene duplication. London: George Allen & Unwin; 1970.
- Conant GC, Wolfe KH. Turning a hobby into a job: how duplicated genes find new functions. *Nat Rev Genet*. 2008;9(12):938–50.
- 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):202–17.
- Hurst LD. The Ka/Ks ratio: diagnosing the form of sequence evolution. *Trends Genet TIG*. 2002;18(9):486.
- Clapham DE. TRP channels as cellular sensors. *Nature*. 2003;426(6966):517–24.
- Denis V, Cyert MS. Internal Ca<sup>2+</sup> release in yeast is triggered by hypertonic shock and mediated by a TRP channel homologue. *J Cell Biol*. 2002;156(1):29–34.
- Zhou X, Batiza AF, Loukin SH, Palmer CP, Kung C, Saimi Y. The transient receptor potential channel on the yeast vacuole is mechanosensitive. *Proc Natl Acad Sci*. 2003;100(12):7105–10.
- 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.
- Moya A, Huisman L, Ball EE, Hayward DC, Grasso LC, Chua CM, Woo HN, Gattuso J, Forêt S, Miller DJ. Whole transcriptome analysis of the coral *Acropora millepora* reveals complex responses to CO<sub>2</sub>-driven acidification during the initiation of calcification. *Mol Ecol*. 2012;21(10):2440–54.
- 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. *Mol Ecol*. 2015;24(17):4489–504.

42. 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):2514–26.
43. 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):3127–41.
44. 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 novelties for establishing coral-algae symbioses. *Genome Biol Evol*. 2021;13(1):evaa240.
45. Seb e-Pedr os A, Saudemont B, Chomsky E, Plessier F, Mailh e 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.
46. Levy S, Elek A, Grau-Bov e X, Men endez-Bravo S, Iglesias M, Tanay A, Mass T, Seb e-Pedr os 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.
47. Hu M, Zheng X, Fan C, Zheng Y. Lineage dynamics of the endosymbiotic cell type in the soft coral *Xenia*. *Nature*. 2020;582(7813):534–8.
48. Shultz AJ, Sackton TB. Immune genes are hotspots of shared positive selection across birds and mammals. *Elife*. 2019;8: e41815.
49. 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.
50. Hayes ML, Eytan RI, Hellberg ME. High amino acid diversity and positive selection at a putative coral immunity gene (tachylectin-2). *BMC Evol Biol*. 2010;10(1):1–15.
51. Tautz D, Domazet-Lo so T. The evolutionary origin of orphan genes. *Nat Rev Genet*. 2011;12(10):692–702.
52. Alb a MM, Castresana J. On homology searches by protein Blast and the characterization of the age of genes. *BMC Evol Biol*. 2007;7(1):1–8.
53. Yasuoka Y, Shinzato C, Satoh N. The mesoderm-forming gene brachyury regulates ectoderm-endoderm demarcation in the coral *Acropora digitifera*. *Curr Biol*. 2016;26(21):2885–92.
54. Cleves PA, Strader ME, Bay LK, Pringle JR, Matz MV. CRISPR/Cas9-mediated genome editing in a reef-building coral. *Proc Natl Acad Sci*. 2018;115(20):5235–40.
55. 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. *Front Mar Sci*. 2021. <https://doi.org/10.3389/fmars.2021.706308>.
56. Yoshioka Y, Yamashita H, Suzuki G, Shinzato C. Larval transcriptomic responses of a stony coral, *Acropora tenuis*, during initial contact with the native symbiont, *Symbiodinium microadriaticum*. *Sci Rep*. 2022;12(1):1–11.
57. Roach MJ, Schmidt SA, Borneman AR. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinf*. 2018;19(1):460.
58. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetJ*. 2011;17(1):10.
59. Price AL, Jones NC, Pevzner PA. De novo identification of repeat families in large genomes. *Bioinformatics*. 2005;21(suppl1):i351–8.
60. Br una 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 Bioinf*. 2021;3(1):108.
61. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods*. 2015;12(4):357.
62. 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):290–5.
63. Pertea G, Pertea M. GFF utilities: GffRead and GffCompare. *F1000Research*. 2020. <https://doi.org/10.12688/f1000research.23297.1>.
64. Sim o FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 2015;31(19):3210–2.
65. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST plus: architecture and applications. *BMC Bioinformatics*. 2009;10:421.
66. 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):1236–40.
67. Storer J, Hubley R, Rosen J, Wheeler TJ, Smit AF. The Dfam community resource of transposable element families, sequence models, and genome annotations. *Mob DNA*. 2021;12(1):1–14.
68. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res*. 2000;28(1):27–30.
69. 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 function of genomes and metagenomes. *Biosci Biotechnol Biochem* 2018, 82(9):1515–1517.
70. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol*. 2019;20(1):1–14.
71. Shinzato C, Takeuchi T, Yoshioka Y, Tada I, Kanda M, Broussard C, Iguchi A, Kusakabe M, Marin F, Satoh N: Whole-genome sequencing highlights conservative genomic strategies of a stress-tolerant, long-lived scleractinian coral, *Porites australiensis* Vaughan, 1918. *Genome Biology and Evolution* 2021, <https://doi.org/10.1093/gbe/evab270>.
72. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods*. 2017;14(4):417–9.
73. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol Biol Evol*. 2013;30(4):772–80.
74. 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.
75. 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.
76. Villanueva-Canas JL, Laurie S, Alba MM. Improving genome-wide scans of positive selection by using protein isoforms of similar length. *Genome Biol Evol*. 2013;5(2):457–67.
77. Almagro Armenteros JJ, S nderby CK, S nderby SK, Nielsen H, Winther O. DeepLoc: prediction of protein subcellular localization using deep learning. *Bioinformatics*. 2017;33(21):3387–95.
78. R core team. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2020.
79. Conway JR, Lex A, Gehlenborg N. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics*. 2017.
80. Wickham H. Elegant graphics for data analysis. *Media*. 2009;35(211): 10.1007.
81. Allen M, Poggiali D, Whitaker K, Marshall TR, Kievit RA. Raincloud plots: a multi-platform tool for robust data visualization. *Wellcome Open Res*. 2019. <https://doi.org/10.12688/wellcomeopenres.15191.1>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

### Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

