Full-length transcriptome maps of reef-building coral illuminate the molecular basis of calcification, symbiosis, and circadian genes

Tingyu Han1, Xin Liao2, Yunchi Zhu1, Yunqing Liu1, Na Lu1, Yixin Li1, Zhuojun Guo1, J.-Y. Chen3, Chunpeng He1,* and Zuhong Lu1,*

1State Key Laboratory of Bioelectronics, School of Biological Science and Medical Engineering, Southeast University, Nanjing 210096, China
2Guangxi Key Lab of Mangrove Conservation and Utilization, Guangxi Mangrove Research Center, Beihai 536000, China.
3Nanjing Institute of Paleontology and Geology, Nanjing 210008, China

Authors for correspondence:
Chunpeng He
e-mail: cphe@seu.edu.cn
Zuhong Lu
e-mail: zhlu@seu.edu.cn
Abstract

**Background:** Reef-building corals are critical species for sustaining coral reefs and are highly threatened by global climate change. However, relevant transcriptomic data largely rely on short-read sequencing, which severely limits the understanding of coral molecular mechanisms and leaves many important biological questions unresolved.

**Results:** We sequenced the full-length transcriptomes of four common and frequently dominant reef-building corals, including two Robusta clade species, *Pocillopora damicornis* and *Pocillopora verrucosa*, and two Complexa clade species, *Acropora muricata* and *Montipora foliosa*, using the PacBio Sequel II platform. We obtained information on gene functions, structures and expression profiles. Among them, a comparative analysis of biomineralization-related genes provided insights into the molecular basis of coral skeletal density. The gene expression profiles of the symbiote Symbiodiniaceae were also isolated and annotated from the holobiont sequence data; these profiles showed more highly convergent traits related to gene structure and expression level than those of coral hosts. Interestingly, we observed that intracellular algal cells share some evolutionary convergence between intracellular symbiosis in corals and intracellular digestion in amphioxus. Finally, a phylogenetic analysis of key circadian clock genes among 27 evolutionarily representative species indicated that there are four key members in early metazoans, including *cry* genes; *Clock* or *Npas2*; *cyc* or *Arntl*; and *tim*, while *per*, as the fifth member, occurs in Bilateria.

**Conclusions:** Our work overcomes the incompleteness of short-read sequencing and illuminates the molecular basis of calcification, symbiosis, and circadian genes, thus providing a foundation for further work on the manipulation of skeleton production or symbiosis to promote the survival of these important organisms.
Keywords
Reef-building corals, Symbiodiniaceae, holobionts, PacBio Sequel II, full-length transcriptome, gene expression profile, biomineralization, symbiosis, circadian clock

Background
Coral reefs are among the most productive and biodiverse ecosystems on earth [1], and they provide survival habitats for approximately 30% of marine life [2-5]. Approximately 500 million people worldwide depend on these reefs, demonstrating their great ecological and economic value [6-10]. The scleractinian corals that mainly produce coral reefs constitute one of the most primitive metazoan branches [2]. They show the hydrozoan body plan typical of Phylum Cnidaria [9] and retain many ancient gene features of ancestral metazoans [11], providing an important genetic background for studying the evolutionary origin of metazoans and bilaterians [12, 13]. Currently, our understanding of the molecular biology of reef-building corals largely derives from omics sequencing, and these data have provided crucial information on coral calcification [14, 15], symbiosis [16-21], heat stress [22-24], acid stress [25, 26], cnidocytes [27, 28], collagen secretion [29], immunity [28, 30-32], budding [33, 34] and circadian clocks [35-37]. Despite these advances, many open questions remain.

The greatest contribution of scleractinian corals to marine biota and ecosystems is the deposition of reefs via biomineralization in calicoblasts [28], which form skeletons. The reef-building coral skeleton is built through the continuous deposition of aragonite [38], which is formed by a mineral fraction consisting of calcium carbonate and an organic matrix molecule fraction that includes carbohydrates, lipids and...
proteins [14]. Current reports indicate that the main components used for calcification include calcium ATPase, carbonic anhydrase (CA), bicarbonate transporter (i.e., solute carrier 4 [SLC4] and solute carrier 26 [SLC26]) and core skeleton organic matrix proteins (SOMPs) (e.g., acid-rich protein, uncharacterized skeletal organic matrix proteins [USOMPs], galaxins and alpha IV collagen) [14, 15, 38-49]. However, due to the lack of full-length sequences for these genes, in-depth biological studies such as gene function reconstruction or gene editing cannot be performed. Moreover, the expression levels of these genes are rarely reported, limiting precise studies of their regulation under environmental stress. In addition, the relationship between coral skeletal density and the expression levels of calcification genes has been rarely discussed.

Phagocytosis targeting primitive algae was a prerequisite for the development of food webs involving multicellular animals, as well as the origin of mitochondria and chloroplasts [50-53]. Additionally, paleontological [51], geochemical [50, 51] and molecular clock [50-52] evidence suggests that various forms of eukaryotic predation occurred during the Neoproterozoic Era (1000-541 million years ago), from the proliferation of marine algae to the origin of multicellular animal clades. At that time, increasing algal abundance created food webs with more efficient nutrient and energy transfer, driving shifts in ecosystems towards larger and increasingly complex organisms [50]. This effect is recorded in the appearance of sponges and the subsequent radiation of eumetazoans in the Ediacaran period [50]. Currently, most hosts of intracellular symbionts belong to groups with polyp body plans, such as hydra,
soft corals, anemones and reef-building corals, and Mollusca, such as giant clam [16-21, 27, 54-56]. Among these groups, reef-building corals have particularly widespread and extensive endosymbiosis (with Symbiodiniaceae). Beyond intracellular symbionts, another interesting form occurred in amphioxus (Branchiostoma belcheri), a primitive deuterostome invertebrate that degrades various small-scale algae by phagocytic intracellular digestion within phagosomes [57]. Thus, there are two major forms of interaction with algae, intracellular symbiosis and intracellular digestion, which differ in terms of how algae are acquired and whether intracellular algae are ‘reared’ or ‘destroyed’. Intracellular symbiosis is generally thought to depend on molecular recognition for the acquisition of algae [58], while intracellular digestion involves direct preying by cell cilia [57]. Symbiodiniaceae, similar to other dinoflagellates, must produce low doses of dinotoxins as a defence against predation or as metabolic byproducts; hence, host cells must carry out certain reactions to reduce algal toxicity in an intracellular symbiotic system, similar to intracellular digestion [59-61]. An incompatible dynamic equilibrium relationship with intracellular algae may lead to host cell death, and reducing immune rejection is a method for alleviating immune conflict and benefiting from these algae during intracellular symbioses or digestion [54]. Because both mechanisms involve interaction with intracellular algae and evolved from the early stage of algal phagocytosis, it is expected that there should be some overlap between them; however, the extent of this overlap is unclear. Such connections should be revealed by omics comparisons across host species.

Reef-building corals show obvious circadian rhythms; interestingly, the circadian
clock system and related genes have not been deeply investigated, although they play
important roles in sustaining healthy coral growth [35-37]. In a marine sponge,
*Amphimedon queenslandica*, which is a representative of the oldest extant animal
phyletic lineage, some molecular components of circadian clock are shared with
eumetazoans, reflecting a core clock present in the last common animal ancestor [62,
63]. At the basic molecular level, the operations of the circadian clock system and
biological rhythm behaviour are regulated through a conserved negative transcription-
translation feedback loop, which is controlled by five key gene families [64].
Cryptochromes (CRYs) are a class of flavoproteins that can detect blue light [65].
Generally, circadian locomotor output cycles kaput (CLOCK) and its homologous
protein Neuronal PAS domain-containing protein 2 (NPAS2), as well as Cycle (CYC) and
its homologous protein Brain and muscle ARNT-like (BMAL), act as positive regulatory
factors, while Period (PER) and Timeless (TIM) act as negative regulatory factors [66].
However, the specific connections among these families vary across clades. In
*Drosophila*, CRY regulates the circadian clock in a light-dependent manner [67],
whereas in mice, CRY1 and CRY2 act as light-independent inhibitors of the CYC-CLOCK
component [68]. However, in some invertebrates, such as monarch butterflies, CRYs
exhibit both *Drosophila*-like and mammal-like functions, providing evidence of an
ancestral clock gene regulation state [69, 70]. Three cry genes and fifteen other
homologous bilaterian circadian clock genes have been found in the *Acropora
digitifera* coral genome [35], and 24 circadian genes in the *Acropora millepora* coral
transcriptome have orthologues in bilaterian species, including 6 insect and 18
mammalian genes [36, 37]. The composition and related phylogenetic pattern of circadian genes among the actinozoan lineage remain unclear; more omics data is needed to elucidate them.

On a global scale, *Pocillopora damicornis* (*P. damicornis*), *Acropora muricata* (*A. muricata*) and *Montipora foliosa* (*M. foliosa*) are common dominant reef-building corals in the Indo-Pacific region, and *Pocillopora verrucosa* (*P. verrucosa*) is a typical reef-front stony coral that protects fringing reefs [71-74]. These four corals play pivotal roles in the Indo-Pacific coral reef ecological system. Recently, increasing amounts of sequencing data derived from these reef-building corals have been deposited in public databases (see Table 1 and Additional file 2: Table S1). However, all currently available sequences have been generated using short-read (50-300 bp) sequencing approaches, leaving incomplete information and sequence splicing errors [75-77]. Critically, although some high-quality sequences have been published for both corals themselves [23, 24, 30-32, 78-93] and their associated microorganisms [33, 94-113], the reliance on short reads has prevented the precise delimitation of the gene expression profiles of reef-building corals and their endosymbiotic Symbiodiniaceae. Recent advances in long-read sequencing technology (e.g., PacBio Sequel II) have made it possible to obtain large amounts of full-length transcript data from many organisms and tissues [114-116]. In principle, such data should permit the characterization of all expressed transcripts as complete, contiguous mRNA sequences from the transcription start site to the transcription end site; in turn, this enables the more accurate and efficient analysis of the full spectrum of gene expression profile...
information, including data on gene expression, alternative splicing, gene fusion, expression regulation, coding sequences (CDSs), and protein structure [117-120].

To address these questions, we applied a full-length transcriptomic isoform sequencing (Iso-Seq) strategy using the PacBio Sequel II sequencing platform and quantitative gene expression analysis using the Illumina HiSeq X Ten sequencing platform to produce transcriptome maps for the four aforementioned important reef-building corals and their endosymbiotic Symbiodiniaceae. Based on the gene expression profiles, we performed a deeper analysis of several important physiological features of reef-building corals, including biomineralization-related gene groups, circadian gene families, and coral and Symbiodiniaceae factors related to the endosymbiotic interaction. These results enhance our understanding of the molecular biology and ecology of reef-building corals, which may aid in the recovery of marine ecosystems, and provide insights into the evolution of the circadian circuitry and holobiosis.

Results

Full-length transcriptome sequencing and data processing

Based on the standard processes of the PacBio Sequel II sequencing platform, the full-length raw transcriptomic sequencing data (polymerase reads) of four coral holobionts obtained included 25.34 Gb in P. damicornis, 27.8 Gb in P. verrucosa, 22.32 Gb in A. muricata and 21.44 Gb in M. foliosa (Table 2; Additional file 3: Table S2). These data were filtered to ensure quality and reliability, including removing reads of less than 50 bp in length and adapter sequences (for details and statistics, see Table 2; Additional file 1: Fig. S1, 2; Additional file 3: Table S2). The gene expression profiles of corals and
their symbiotic Symbiodiniaceae (whose transcripts will be analysed separately in a later subsection) were separated based on an alignment to previously published sequences. In terms of the corals themselves, there were 20609 transcripts and 14167 unigenes (N50 = 2954 bp) in *P. damicornis*, 24174 transcripts and 12822 unigenes (N50 = 2313 bp) in *P. verrucosa*, 31242 transcripts and 13800 unigenes (N50 = 2126 bp) in *A. muricata*, and 25460 transcripts and 10905 unigenes (N50 = 1678 bp) in *M. foliosa* (Table 2; Additional file 3: Table S2). Unigene lengths were concentrated in the range of 1-3 kbp, with few unigenes of less than 1 kbp, verifying that the low-quality sequencing data had been filtered out (Additional file 1: Fig. S1b; Additional file 3: Table S2.8). A previous study has reported that there were 297221 assembled transcripts (N50 = 1831 bp) and 209337 unigenes (N50 = 1435 bp) in *P. damicornis* using a short-read sequencing approach [80]. However, it is obvious that PacBio sequencing can obtain fewer redundant sequences and higher quality sequences. Then, to evaluate the level of redundancy in the data, we examined the unigene-to-transcript ratio. The percentage of unigenes with a one-to-one unigene-to-transcript ratio was 73.71% (*P. damicornis*), 64.75% (*P. verrucosa*), 57.15% (*A. muricata*) and 57.26% (*M. foliosa*), and the percentage of unigenes with a one-to-two unigene-to-transcript ratio was 17.44% (*P. damicornis*), 17.52% (*P. verrucosa*), 18.41% (*A. muricata*) and 17.18% (*M. foliosa*) (detailed in Additional file 3: Table S2.9; Additional file 1: Fig. S1c). The sum of the two ratios was more than 74% in all samples, indicating that data redundancy was largely reduced. Thus, we obtained high-quality full-length transcriptome sequencing data for four coral holobionts that is suitable for subsequent analysis.
Gene function annotation and structure analysis

To obtain comprehensive gene function information, seven authoritative databases (NR [121], NT, Pfam [122], KOG [123], Swiss-Prot [124], KEGG [125], and GO [126]) were utilized to annotate the unigenes, and related statistics are shown in Fig. 1 (Additional file 4: Table S3; Additional file 1: Fig. S3-11). In the four corals, 93.88% (P. damicornis), 89.32% (P. verrucosa), 95.10% (A. muricata) and 79.64% (M. foliosa) of the unigenes were annotated in at least one database (Fig. 1a; Additional file 4: Table S3.1). The percentage of annotated unigenes in A. muricata, P. damicornis and P. verrucosa was approximately 90%, suggesting that most of the unigenes in these species are orthologues of genes with functional annotations available. In contrast, the percentage of unigenes annotated in M. foliosa was only 79.64% because sequencing data from Montipora are rare, and the adaptive specialized evolution of these species means that their genomes show especially large differences from those of other reef-building corals with available omics data [92, 127], limiting their in-depth annotation. It was also obvious that the percentage of annotated unigenes in A. muricata was higher than that in the other three corals (Fig. 1b) because A. digitifera [128] and A. millepora [129], with reported genome data, both belong to the same genus as A. muricata, providing a good reference genome background.

To explore whether the sequence similarity of corals to other species is consistent with their phylogenetic position, we counted the number of sequences annotated based on species information and found that the top five species with the most highly similar genes to those of the four investigated corals were A. digitifera, Exaiptasia pallida, Nematostella vectensis, B. belcheri and Stylophora pistillata (or A. millepora) (Fig. 2; Table 3; Additional file 4: Table S3.6). These species (other than the amphioxus
B. belcheri) are all cnidarians, and the similarities of the investigated corals to the reference species were all greater than 93%, indicating relatively high accuracy of the annotations. Interestingly, the NR results indicate that the gene expression profiles of all four investigated corals were close to that of B. belcheri (Fig. 2; Table 3; Additional file 4: Table S3.6), implying that the presence of intracellular algal cells in both reef-building corals and amphioxus [57] probably impacts the component pattern of the gene expression profiles of the host, leading to a convergence of adaptive evolution.

The prediction of CDSs is helpful for potential unigene analysis and provides a basis for subsequent protein analysis. The CDS lengths of the four corals were concentrated in the range of 300–1700 nt (Methods; Additional file 1: Fig. S12a), and a total of 97.91% (P. damicornis), 95.67% (P. verrucosa), 95.51% (A. muricata) and 89.04% (M. foliosa) of the unigenes were predicted to contain CDS regions (Additional file 5: Table S4.1). We then identified coral TFs among the predicted protein CDS results based on Pfam TF families (Methods) and found that the five largest TF families in the corals were ZBTB, zf-C2H2, homeobox, bHLH and TF_bZIP (Additional file 1: Fig. S12b; Additional file 5: Table S4.2). The results of simple sequence repeat (SSR) detection showed that most SSR sequences corresponded to poly(A) tails (Additional file 1: Fig. S12c; Additional file 5: Table S4.3), reconfirming the absence of genomic DNA contamination. Transcripts without coding potential were classified as lncRNAs. Generally, lncRNAs are shorter than mRNAs (Additional file 1: Fig. S12e). The number of predicted lncRNAs in M. foliosa was much greater than those in the other corals (Additional file 1: Fig. S12d). After CDS, TF, SSR and lncRNA analyses, we quantified the gene expression profiles in each coral by mapping the Illumina sequencing reads to the full-length transcriptome data. This mapping directly providing read count values that
could be converted to the expected number of fragments per kilobase of transcript sequence per million base pairs sequenced (FPKM) or transcripts per kilobase million (TPM) values for further ensuring diverse quantitative analysis under different conditions (Additional file 6: Table S5). Based on these data, the Pearson correlation coefficient (R) analysis can be used to reflect the gene expression profile similarity among the samples and thus verify the experimental reliability and the rationality of sample selection.

**Gene expression profile analysis**

We next compared the gene expression profiles of the four corals to better understand their ecological and evolutionary relationships. Consistent with the known phylogenetic relationships [130], the gene expression profiles of *P. damicornis* and *P. verrucosa* (both in the Robusta clade) were more similar to each other than to *A. muricata* and *M. foliosa* (both in the Complexa clade), and vice versa (Fig. 3a-d). The gene expression levels of each coral were considered to be consistent across all replicates (Additional file 1: Fig. S13). However, combined analysis indicated that the overall gene expression level of *M. foliosa* was lower than those of the other corals (Fig. 3c), showing that the structure of the gene expression profile of *M. foliosa* was significantly different from those of the other corals (Additional file 6: Table S5). These results indicated that beyond evolutionary aspects, *M. foliosa* exhibits prominent ecological adaptive features [92, 127]. In addition, a Pearson’s R$^2$ > 0.92 among the three biological replicates of each coral suggested that the sample selection was reasonable and the experimental data were reliable (Fig. 3d). Then, the differences in the coral gene expression profiles were explored. A total of 871 (*P. damicornis*), 1296 (*P. verrucosa*), 4369 (*A. muricata*) and 3911 (*M. foliosa*) differentially expressed genes
(DEGs) were identified based on the parameters \( \text{padj} < 0.001 \& |\text{log2(FoldChange)}| \geq 2 \) (Additional file 1: Fig. S14a; Additional file 7: Table S6); of these, 122 (\textit{P. damicornis}), 167 (\textit{P. verrucosa}), 2333 (\textit{A. muricata}) and 1956 (\textit{M. foliosa}) were particularly prominent, with a \(|\text{log2(FoldChange)}| \geq 10 \) (Additional file 1: Fig. S14b; Additional file 7: Table S6). These results showed that the genus \textit{Pocillopora}, of the Robusta clade, presented fewer differentially expressed genes than the genera of the \textit{Acropora} and \textit{Montipora} Complexa clades (Additional file 1: Fig. S14; Additional file 7: Table S6). Consequently, according to the gene expression profile results, biomineralization, symbiosis, and circadian gene families, which provide crucial information on coral population structure and on the evolution of coral gene repertoires, will be highlighted hereafter.

**Expression analysis of the biomineralization-related gene group**

Although the biomineralization mechanisms of reef-building corals are similar (Fig. 4a), they are displayed in a wide range of morphological differences. To gain insight into the molecular-level causes of this phenomenon, we investigated the expression patterns and features of biomineralization-related gene families and found that the expression levels of calcium ATPases and bicarbonate transporters, which are essential for \( \text{Ca}^{2+} \) and \( \text{HCO}_3^- \) transport, were higher in the Robusta genus \textit{Pocillopora} than in the Complexa genera \textit{Acropora} and \textit{Montipora} (Fig.4b-c; Table 4; Additional file 8: Table S7). Among the four investigated corals, \textit{M. foliosa} has the lowest apparent capacity to transport calcium ions to calcifying fluid because of the low expression of calcium ATPases (Fig.4b; Table 4; Additional file 8: Table S7). CA2 is universally expressed in all four corals at very high levels and accounts for a major proportion of the CA content,
implying its key role in coral biomineralization (Fig. 4d; Table 4; Additional file 8: Table S7).

The expression levels of coral biomineralization-related extracellular matrix (ECM) proteins, including acid-rich proteins, USOMPs, galaxins and collagen alpha-6(VI) chain proteins, were all obviously higher in the Complexa clade than in the Robusta clade (Fig. 4e-g; Table 4; Additional file 8: Table S7). Regarding key acid-rich ECM proteins, *A. muricata* showed the highest expression levels, diversity and balance, followed by *M. foliosa*, while skeletal aspartic acid-rich protein 1 (SAARP1) was the only protein in this group expressed in genus *Pocillopora*. Regarding the ECM accessory proteins USOMPs, *A. muricata* and *M. foliosa* presented extremely higher levels and diversity, and USOMP-6 was the dominant USOMP gene, whereas members of the genus *Pocillopora* mainly expressed USOMP-5 (Table 4; Additional file 8: Table S7). Galaxins were the only ECM components with higher expression in *M. foliosa* than in *A. muricata* (Table 4; Additional file 8: Table S7). This result indicates that there are significant differences in the types and expression levels of core calcification genes among different reef-building corals, which may affect their morphological formation and growth model.

**Genes expressed by Symbiodiniaceae (intracellular symbionts of coral) are mainly involved in energy and nutrient production**

The secretion of calcium carbonate skeletons to form marine reefs and the reliance on intracellular symbiotic algae (Symbiodiniaceae) for nutrient acquisition are two signature features of reef-building coral physiology [14-21, 128]. Although skeletal formation has been discussed extensively, the contributions of the genes expressed by the Symbiodiniaceae are rarely mentioned. Depending on identifying coral or
symbiont transcripts based on alignment to previously published sequences, we found that 235 (P. damicornis), 351 (P. verrucosa), 220 (A. muricata) and 561 (M. foliosa) Symbiodiniaceae unigenes were expressed in the investigated coral holobionts (Additional file 9: Table S8). Similar to the analysis process used for corals, Symbiodiniaceae gene expression profiling was performed using seven functional databases, NR, NT, Pfam, KOG, Swiss-Prot, KEGG, and GO, and published Symbiodiniaceae genome data (Additional file 1: Fig. S15-26; Additional file 9: Table S8.5-9). We found 832 transcripts that were expressed by the symbiotic Symbiodiniaceae of all four investigated corals (Additional file 1: Fig. S27), accounting for approximately 70.75% (P. damicornis), 70.87% (P. verrucosa), 88.79% (A. muricata) and 93.69% (M. foliosa) of the symbiote transcripts. Overall, gene expression profiles of the Symbiodiniaceae were more similar than those of coral hosts, suggesting considerable gene expression convergence. On the other hand, the gene expression patterns of the symbionts of P. damicornis and P. verrucosa were more similar to each other than to A. muricata and M. foliosa, and vice versa, coinciding with the evolutionary relationships of the corals (Additional file 1: Fig. S28).

To explore the functions of these expressed genes, we classified them according to the annotation results using the GO database (Fig. 5; Additional file 1: Fig. S29; Additional file 9: Table S8.11-12). In the biological process category, the identified gene functions were mainly involved in oxidation reduction and carbohydrate metabolism (Fig. 5). In the cellular component category, most of the functional genes were related to photosynthesis, such as light-harvesting complex or photosystem (Additional file 1: Fig. S29). In the molecular function category, the largest group of genes was associated with oxidoreductase or carbonate dehydratase activity (Additional file 1: Fig. S29).
These results suggest that the symbiotes of all four corals have similar expression profiles that feature genes related to energy and nutrient production.

Phylogenetic analysis of the key circadian clock gene regulation network

The circadian clock system is one of the most universal and fundamental characteristics of life across almost all living species [66]. Corals also exhibit circadian behaviours, and there have been studies describing the molecular mechanisms underlying the regulation of these behaviors [35-37]. However, whether this mechanism is applicable to other corals, other cnidarians and even other animals remains to be clarified. To explore the evolutionary origins of the key members of the circadian clock gene regulation network of reef-building corals, including the four corals studied herein, we selected 27 evolutionarily representative species based on their related gene and genome data available in the NCBI database to investigate the key circadian clock gene regulation network, namely, the cry1, cry2, Clock, Npas2, cyc, Arntl (also known as Bmal1), Arntl2 (also known as Bmal2), per1, per2, per3 and tim genes (Table 5; Additional file 10: Table S9), and performed phylogenetic analyses (Fig. 6; Additional file 1: Fig. S30-36). The results showed that Ciliata (Paramecium tetraurelia), Ctenophora (Pleurobrachia bachei), Scyphozoa (Aurelia aurita), Cubozoa (Chironex fleckeri) and Turbellaria (Schmidtea mediterranea) had no circadian clock gene members (Table 5). Phylogenetic tree analysis illustrated that the key circadian clock gene regulatory network evolved in early metazoans, whereas protozoans and the ctenophore clade did not have this kind of gene regulatory network; moreover, the jellyfish and Turbellaria clades showed the complete loss of this network during
In the Hydrozoa and Ascidiacea clades, all members of the circadian clock gene regulatory network except for the *tim* gene have been lost. The *tim* gene was the only gene of this group existing in all species with a circadian clock system, suggesting that it is the most conserved and stable circadian clock gene (Table 5; Fig. 6; Additional file 10: Table S9) [92, 130]. Consequently, the phylogenetic analysis suggested that in the early metazoan stage, there were four key gene members of the circadian clock gene regulatory network: *cry*, *Clock* or *Npas2*, *cyc*, *Arntl*(*Bmal1*), or *Arntl2*(*Bmal2*); and *tim*. The *cry1*, *Clock* or *Npas2*, *Arntl* (*Bmal1*) and *tim* genes were present in all investigated reef-building corals, and *cyc* was present in other corals, as shown in Table 5. Thus, the data suggest the existence of four key conserved members of the reef-building coral circadian clock gene regulation network. The phylogenetic analysis also indicated that the *per* gene is a novel member of the circadian clock gene regulation network in the Bilateria, as it is found in protostomes and deuterostomes but not in early metazoans (Table 5). Interestingly, the phylogenetic analysis results also showed that the two most conserved clades with key circadian clock gene regulation networks are sessile-living anthozoaans and free-living vertebrates. Together, these findings indicate that there were four key members of the circadian clock gene regulatory network in early metazoans, while *per*, as the fifth member, first occurred in Bilateria.

Discussion

Despite much research, our understanding of reef-building coral holobiont transcriptomes remains incomplete, and the short-read lengths intrinsic to the
prevailing technologies have limited access to complete genetic information. Here, we sequenced and analysed the full-length transcriptome of four common dominant reef-building coral holobionts with respect to gene functions, structures, and expression. The results revealed differences in the members actually involved in biomineralization processes in reef-building corals and provided new insights into further understanding the molecular mechanisms of coral density. In particular, we isolated and demarcated the gene expression profiles of the symbiote Symbiodiniaceae, which showed higher convergence than the coral hosts and suggested that there may be some association between marine animal intracellular symbiosis and the intracellular digestion of algal cells. Moreover, we confirmed that there were four key members of the circadian clock gene regulatory network in early metazoans and that per first occurred in Bilateria, providing a molecular basis for the study of the evolutionary origin of the animal circadian clock system.

Coral biomineralization and skeleton density

Coral biomineralization has been studied intensively in previous studies, but descriptions of its formation mechanism vary; here, we summarized this process, which consists of four components (Fig. 4a). (1) Calcium ion transport. In calicoblasts and paracells, calcium ions enter the cells through calcium channels and exit via calcium ATPases that exchange two calcium ions for four protons across the cell membrane [39, 40]. Ca\(^{2+}\) diffusion among cells with chemical gradients may also participate in calcium deposition [15]. (2) An HCO\(_3^-\) source. The source of HCO\(_3^-\) is metabolic and environmental CO\(_2\). Metabolic CO\(_2\) can be converted into HCO\(_3^-\) both
intracellularly and extracellularly, which is catalysed by CA under a favourable pH [41-43], and intracellular HCO$_3^-$ exits cells via bicarbonate transporters belonging to two membrane protein families (SLC4 and SLC26) [44]. In calcifying fluid, the HCO$_3^-$ concentration is higher than the Ca$^{2+}$ concentration, and the amount of calcium carbonate deposition is determined by the Ca$^{2+}$ concentration [43]. (3) Acid-rich proteins. Coral acid-rich proteins are key ECM proteins involved in biomineralization and can interact with amorphous calcium carbonate (ACC) directly, promoting crystal nucleation, determining growth axes and controlling crystal growth [45, 46]. (4) Other organic matrix proteins of the ECM. The bioprecipitation of aragonite crystals in corals also requires additional skeletal organic matrix proteins as binders [47], including USOMPs, galaxins and alpha IV collagen [38, 48, 49]. In addition, magnesium plays a crucial role in regulating the formation of different calcium carbonate phases that cooperate with organic matrix molecules to stabilize ACCs [131].

Here, the above core biomineralization-related gene expression level analyses among four reef-building corals indicate that the Robusta genus *Pocillopora* has a greater capacity to produce calcium carbonate than the Complexa genera *Acropora* and *Montipora*, but the opposite situation was observed for biomineralization-related ECM proteins, suggesting that the species of the Complexa clade can produce a larger volume of skeleton than those of the Robusta clade using the same amount of calcium carbonate, which is in accordance with the reported skeletal density data of these stony corals [34, 132]. Combined with the observation of growth patterns [132], it can be inferred that after the Tertiary Period, the Complexa clade corals developed to
occupy more space (as observed *A. muricate* and *M. foliosa*), whereas the Robusta clade corals tended to form more dense colonies (as observed in *P. verrucosa*).

**Evolutionary convergence triggered by intracellular algal cells**

Coincidently, all of the gene expression profiling analyses of the four investigated corals indicated evolutionary convergence between corals and amphioxus. According to the NR results, the best-hit genes between the four corals and *B. belcheri* all belonged to the top five groups identified among annotated species ([Fig. 2; Table 3](#)); ([Additional file 4: Table S3.6](#)). All of these genes except for some housekeeping genes were related to interactions with intracellular algae, namely, mechanisms for algal entry into host cells, for reducing algal toxicity and for reducing immune rejection ([Additional file 11: Table S10](#)).

Upon the entry of algae into host cells, the packaging of algae by assembling symbiosomes (in intracellular symbioses) or phagosomes (in intracellular digestion) requires the direct or indirect participation of proteins that play important roles in clathrin-mediated endocytosis and endosome formation (DENN domain-containing protein 1B, muskelin, fibrillin-2, exocyst complex component 2 and huntingtin) ([133-137](#)) and genes that are required to regulate the generation of functional cilia (nephrocystin-4, jouberin, intraflagellar transport protein 74, the fuzzy homologue protein, the centrosomal protein of 104 kDa and the cytosolic Fe-S cluster assembly factor NUBP2) ([138-144](#)). Nephrocystin-4 contains a proline-rich region and is involved in the organization of the subapical actin network in multiciliated epithelial cells by recruiting INT to the basal bodies of motile cilia, and it subsequently interacts with...
actin-modifying proteins [138]. Jouberin is involved in recruiting the Ras-related protein Rab-8A to the basal body of the primary cilium as a component of a complex localized in the transition zone of primary cilia, acting as a barrier that prevents the diffusion of transmembrane proteins between cilia and plasma membranes [139, 140]. Intraflagellar transport protein 74, together with intraflagellar transport protein 81, binds and transports tubulin within cilia and is required for ciliogenesis [141]. The fuzzy homologue protein is a planar cell polarity effector involved in cilium biogenesis and regulates protein and membrane transport to the cilium [142]. The centrosomal protein of 104 kDa is required for elongating cilium at the tip during cillum formation [143]. On the other hand, the cytosolic Fe-S cluster assembly factor NUBP2 negatively regulates cilium structure and formation [144]. The observed evolutionarily convergence of these factors between corals and amphioxus indicates that the cilia make an important contribution to algal cell entry of host cells in intracellular symbiosis as well as intracellular digestion, alongside complex molecular recognition mechanisms.

All four investigated symbiotic Symbiodiniaceae (Additional file 9: Table S8.6-9) expressed genes encoding polyketide synthase family members, including putative polyketide synthase 3, phthiocerol synthesis polyketide synthase type I PpsA, phthiocerol synthesis polyketide synthase type I PpsB and phthiocerol synthesis polyketide synthase type I PpsC, demonstrating that symbiotic Symbiodiniaceae can produce dinotoxins [145]. Among the genes commonly expressed by all four corals and amphioxus, several have prospective roles in reducing algal toxicity. Epoxide hydrolase
4 (epoxide hydratase 4) is a biotransformation enzyme that catalyses the hydrolysis of arene and aliphatic epoxides to produce water-soluble dihydrodiols [146], converting highly toxic chemical groups to less reactive and easily excluded groups. THAP domain-containing protein 4 is involved in the detoxification of reactive nitrogen species (RNS) and reactive oxygen species (ROS) [147-148]. S-formylglutathione hydrolase is a kind of serine hydrolase involved in the detoxification of formaldehyde [149]. Carbamoyl-phosphate synthase plays an important role in removing excess ammonia from the cell by catalysing the reaction of ammonia with carboxy phosphate to form carbamoyl phosphate [150]. Transmembrane protein 181 (TMEM181) is the rate-limiting factor in intoxication because it binds to cytolethal distending toxins (CDTs) secreted by entered intracellular algae [151]. The above functional analysis reveals specific genes that reduce algal toxicity to support eukaryotic intracellular symbiosis or digestion by opposing cell intoxication in corals and other hosts of algal symbiotes.

Intracellular algae act as large antigens in algal host cells. For example, the four reef-building corals investigated herein and published coral gene expression profiles all include gene repertoires for innate immunity (Additional file 4: Table S3.2-5) [54], including host innate immune responses to invading microbes. These genes are also involved in intracellular digestion (as in amphioxus) [57]. Acid sphingomyelinase-like phosphodiesterase 3b and phosphoinositide 3-kinase adapter protein 1 can help algal cells adapt to the intracellular environment by regulating Toll-like receptors, which play a key role in the innate immune system. Acid sphingomyelinase-like phosphodiesterase 3b is a lipid-modulating phosphodiesterase that acts as a negative
regulator of Toll-like receptors [152, 153]. Phosphoinositide 3-kinase adapter protein 1 can prevent excessive inflammatory cytokine production by linking Toll-like receptor signalling to phosphoinositide 3-kinase (PI3K) activation [154]. Taken together, our findings confirm that there are some evolutionary convergences between coral and amphioxus that have been triggered by interactions with intracellular algal cells.

**Evolutionary origins of the per gene**

The oscillations of both the transcript and protein levels of the per gene have a period of approximately 24 hours and play a central role in the molecular mechanism of the biological clock driving circadian rhythms. In *Drosophila*, after PER is produced from per mRNA, it dimerizes with Timeless (TIM), and the complex enters the nucleus and inhibits the TFs per and tim, which in turn lowers the levels of PER and TIM [53]. When TIM is not complexed with PER, another protein, double-time, or DBT, phosphorylates PER, targeting it for degradation [58]. In mammals, an analogous transcription-translation negative feedback loop is observed. One of the three PER proteins (PER1, PER2 and PER3) that is translated from the mammalian homologs of drosophila-per dimerizes via its PAS domain with one of two cryptochrome proteins (CRY1 and CRY2) to form a negative element of the clock. This dimer then interacts with the CLOCK (or NPAS2) and ARNTL (or ARNTL2) heterodimer, inhibiting its activity and thereby negatively regulating its own expression [133]. Despite an extensive search, we failed to detect orthologs of the per gene in our coral data. This also may be the case for sea anemones and *Hydra* (Table 5; Additional file 10: Table S9). On the other hand, amphioxus [134,135] and *C. gigas* [136] likely have per genes, and our results show that the biological clock is a common feature of a wide range of Metazoa, while per
gene originates from Bilateria. It is unclear, however, whether the ancient circadian clock system initially operated without the *per* gene, whether the Radiata system lost this locus via deletion or other mutation, or whether unknown proteins may have functions similar to the *per* gene. These hypotheses need to be further explored in future research.

Overall, the full-length transcriptome maps of reef-building coral may serve as references for expression analyses of coral under environmental stressors linked to global change, which alter the normal function of reef-building corals. As long-read sequencing continues to evolve in throughput, accuracy, accessibility and cost efficiency, full-length transcriptomes will be adopted by researchers and provide us with unprecedented views of serious biological problems. However, genome, small RNA, proteome and single-cell data are still lacking to further understand the biology of reef-building corals. Future efforts will be required to construct a coral genome database as a reference for analysing functional genes and their associated regulatory networks at the transcriptomic and proteomic levels and further refine them to specific cellular lineages.

**Conclusion**

The sequencing and analysis of the full-length transcriptome maps of dominant reef-building coral holobionts contributed to a more in-depth understanding of coral physiology. Related insights improve our understanding of the evolution of circadian rhythms and of holobiosis, and they provide a foundation for further work to protect or even manipulate coral skeleton production or symbiosis to promote the survival of these important organisms. These results provide support for remodelling reef-building corals through advanced gene toolkits under current global climate change.
and ecological deterioration.

Methods

Ethics
All coral samples were collected and processed in accordance with local laws for invertebrate protection.

Sample collection
The corals in the study were collected from the Xisha Islands in the South China Sea (latitude 15°40'-17°10' N, longitude 111°-113° E).

Coral culture system
The coral samples were cultured in our laboratory coral tank with conditions conforming to their habitat environment. All samples were raised in RedSea® tank (redsea575, Red Sea Aquatics Ltd., London, UK) at 26°C and 1.025 salinity (Red Sea Aquatics Ltd., London, UK). The physical conditions of the coral culture system were as follows: three coral lamps (AI®, Red Sea Aquatics Ltd., London, UK), a protein skimmer (regal250s, Reef Octopus), a water chiller (tk1000, TECO Ltd., Port Louis, Mauritius), two wave devices (VorTech™ MP40, EcoTech Marine Ltd. Bethlehem, PA, USA), and a calcium reactor (Calreact 200, Reef Octopus).

Total RNA extraction
The three biological replicates samples for each coral were isolated from three healthy branches in the same coral independent colony to ensure that enough high-quality RNA (> 15 µg) could be obtained for a PacBio cDNA library and three Illumina cDNA libraries. All the RNA extraction procedures followed the manufacturer's instructions. The total RNA was isolated with TRizol® LS Reagent (Thermo Fisher Scientific, 10296028,
Waltham, MA, USA) and treated with DNase I (Thermo Fisher Scientific, 18068015, Waltham, MA, USA). The high-quality mRNA was isolated with a FastTrack MAG Maxi mRNA Isolation Kit (Thermo Fisher Scientific, K1580-02, Waltham, MA, USA). The RNA extraction procedure was performed according to the following instructions: (1) grinded coral samples (kept the samples submerged in liquid nitrogen at all times); (2) when the samples were ground into small pieces, the TRIzol® LS reagent was added, the ratio of sample to reagent was about 1:3; (3) let samples stand and thaw naturally; (4) continue adding TRIzol® LS reagent until the samples were dissolved, and dispensed into 50 mL centrifuge tubes; (5) centrifuged at 4 °C and 3000 rpm for 5-15 min; (6) dispensed the supernatant into 50 mL centrifuge tubes; (7) added BCP (Molecular Research Center, BP 151, Cincinnati, OH, USA) to the above centrifuge tubes, the ratio of sample to reagent is about 5:1, shook well and then stood for 10 min; (8) centrifuged at 4 °C and 10500 rpm for 15 min; (9) took the supernatant, added an equal volume of Isopropanol (Amresco, 0918-500ML, Radnor, PA, USA) and mixed well, stood them overnight at -20°C; (10) centrifuged at 4 °C and 10500 rpm for 30 min, discarded the supernatant; (11) rinsed 2 times with 75% Ice Ethyl alcohol, Pure (Sigma-Aldrich, E7023-500ML, Taufkirchen, München, Germany). Finally, three samples of each coral were extracted in equal amounts (total >10 µg) and mixed for PacBio full-length transcriptome sequencing, the remainders (>1.5 µg per sample) were used for Illumina sequencing.

**Total RNA quality testing**

Before establishing the library, the quality of total RNA must be tested. RNA
degradation and contamination were monitored on 1% agarose gels electrophoresis; RNA purity (OD260/280 ratio) was checked using the NanoPhotometer® spectrophotometer (IMPLEN, CA, USA); RNA concentration was quantified using Qubit® RNA Assay Kit in Qubit® 2.0 Flurometer (Life Technologies, CA, USA); RNA integrity was assessed using the RNA Nano 6000 Assay Kit of the Agilent Bioanalyzer 2100 system (Agilent Technologies, CA, USA).

**Illumina cDNA library construction and sequencing**

A total amount of 1.5 µg RNA per sample was used as input material for the RNA sample preparations. Sequencing libraries were generated using NEBNext® Ultra™ RNA Library Prep Kit (E7530L) for Illumina® (NEB, Ipswich, MA, USA) following the manufacturer’s recommendations and index codes were added to attribute sequences to each sample. Briefly, mRNA was purified from total RNA using poly-T oligo-attached magnetic beads. Fragmentation was carried out using divalent cations under elevated temperature in NEBNext First Strand Synthesis Reaction Buffer (5X). First-strand cDNA was synthesized using random hexamer primer and M-MuLV Reverse Transcriptase (RNase H⁻). Second strand cDNA synthesis was subsequently performed using DNA Polymerase I and RNase H. Remaining overhangs were converted into blunt ends via exonuclease/polymerase activities. After adenylation of 3’ ends of DNA fragments, NEBNext Adaptor with hairpin loop structure was ligated to prepare for hybridization. In order to select cDNA fragments preferentially 250-300 bp in length, the library fragments were purified with AMPure XP system (Beckman Coulter, Beverly, USA). Then 3 µl USER Enzyme (NEB, USA) was used with size-selected, adaptor-ligated cDNA at 37 °C for 15 min followed by 5 min at 95 °C before PCR. Then PCR was performed
with Phusion High-Fidelity DNA polymerase, Universal PCR primers and Index (X) Primer. At last, PCR products were purified (AMPure XP system) and library quality was assessed on the Agilent Bioanalyzer 2100 system. The clustering of the index-coded samples was performed on a cBot Cluster Generation System using TruSeq PE Cluster Kit v3-cBot-HS (Illumia) according to the manufacturer’s instructions. After cluster generation, the library preparations were sequenced on an Illumina HiSeq X Ten platform and paired-end reads were generated.

**PacBio cDNA library construction and sequencing**

The Isoform sequencing (Iso-Seq) library was prepared according to the Isoform Sequencing protocol (Iso-Seq) using the Clontech SMARTer® PCR cDNA Synthesis Kit (Clontech Laboratories (now Takara Laboratories), 634926, Mountain View, CA, USA) and the BluePippin Size Selection System protocol as described by Pacific Biosciences (PN 100-092-800-03). Briefly, Oligo(dT) enriched mRNA was reversely transcribed to cDNA by SMARTer PCR cDNA Synthesis Kit; the synthesized cDNA was then amplified by polymerase chain reaction (PCR) using BluePippin Size Selection System protocol; the Iso-Seq library was constructed by full-length cDNA damage repair, terminal repair and attaching SMRT dumbbell adapters; the sequences of the unattached adapters at both ends of the cDNA were removed by exonuclease digestion; the cDNA obtained above was combined with primers and DNA polymerase to form a complete SMRT bell library. While the library was qualified, the PacBio Sequel II platform was used for sequencing based on the effective concentration and data output requirements of the library.

**Data Filtering and Processing**

The Illumina sequencing raw reads of fastq format were firstly processed through in-
house perl scripts. In this step, clean data were obtained by removing reads containing adapter, reads containing ploy-N and low-quality reads from raw data. At the same time, Q20, Q30, GC-content and sequence duplication levels of the clean data were calculated. All the downstream analyses were based on clean data with high quality.

The PacBio sequencing raw data were processed by SMRTlink v8.0 software. Circular consensus sequence (CCS) was generated from subread BAM files, parameters: min_length 50, min_passes 1, max_length 15,000. CCS.BAM files were output, which were then classified into full-length and non-full-length reads using lima, removing polyA using refine. Full-length fasta files produced were then fed into the cluster step, which does isoform-level hierarchical clustering (n*log(n)), followed by final Arrow polishing, hq_quiver_min_accuracy 0.99, bin_by_primer false, bin_size_kb 1, qv_trim_5p 100, qv_trim_3p 30.

Coral and Symbiodiniaceae sequences separation
Aligned consensus reads to coral or Symbiodiniaceae reference genomes respectivly using GMAP v2017-06-20 software [155]. The sequences mapped to Symbiodiniaceae reference genomes belonged to Symbiodiniaceae sequences, the sequences mapped to coral reference genomes belonged to coral sequences.

Correction and de-redundancy
The RNA-seq data sequenced by the Illumina HiSeq X Ten platform was used to correct additional nucleotide errors in polish consensus sequences obtained in the previous step with LoRDEC v0.7 software [156]. Using CD-HIT v4.6.8 software (parameters: -c 0.95 -T 6 -G 0 - aL 0.00 -aS 0.99) all redundancies were removed in corrected consensus reads to acquire final full-length transcripts and unigenes for subsequent bioinformatics analysis [157].
Gene functional annotation

Gene functions were annotated using the following databases: NT (NCBI non-redundant nucleotide sequences); NR (NCBI non-redundant protein sequences); Pfam (Protein family); KOG/COG database (Clusters of Orthologous Groups of proteins); Swiss-Prot (A manually annotated and reviewed protein sequence database); KEGG (Kyoto Encyclopedia of Genes and Genomes); GO (Gene Ontology). We used BLAST 2.7.1+ software [158] with the e-value ‘1e-5’ for NT database analysis, Diamond v0.8.36 BLASTX software [159] with the e-value ‘1e-5’ for NR, KOG, Swiss-Prot and KEGG databases analysis, and HMMER 3.1 package [160] for Pfam database analysis.

Gene structure analysis

ANGEL v2.4 software [161] was used to predict protein CDS (Coding sequence). We used the same species or closely related species confident protein sequences for ANGEL training and then ran the ANGEL prediction for given sequences. Usually, the TFs were identified based on the Pfam files of TF families in AnimalTFDB 3.0 database [162], however, corals were not included in this database, so we identified coral TFs based on the Pfam files of TF families using hmmsearch program in HMMER 3.1 package. SSR of the transcriptomes was identified using MISA v1 [163]. We used CNCI v2 [164], CPC2 v0.1 [165], PfamScan v1.6 [166] and PLEK v1.2 [167] four tools to predict the coding potential of transcripts. Transcripts predicted with coding potential by either/all of the three tools above were filtered out, and those lacking coding potential were our candidate set of IncRNAs.

Gene expression quantification

The full-length transcriptome obtained above was used as the reference background, and then the clean reads of each sample obtained by Illumina sequencing were
mapped to it by bowtie2 V2.3.4 software [168]. The alignment results were estimated by RSEM v1.3.0 software [169] to further obtain the read count values for each transcript, and then transferred read count to FPKM for analyzing the gene expression levels. Using Pearson correlation coefficient analyzed the relationship among the samples.

**Gene differential expression analysis**

The unigenes with the same annotation result in the NR database were merged to form a new read count expression matrix. Differential expression analysis of two groups was performed using the DESeq2 R package (1.30.1) [170]. DESeq2 provides statistical routines for determining differential expression in digital gene expression data using a model based on the negative binomial distribution. The resulting p-values were adjusted using Benjamini and Hochberg’s approach for controlling the false discovery rate and were named as padj. Coral genes with padj<0.001&|log2(FoldChange)|>=2 and 10 by DESeq2 were assigned as differentially expressed. Venn diagrams were drawn by VennDiagram R package (1.6.20) and GO classification bar charts were drawn by ggplot2 R package (3.3.5) [171].

**Phylogenetic Analysis**

To construct the phylogenetic tree, we selected core clock genes in our own 5 reef-building corals (including a new coral *Montipora capricomis* for verification), and 22 species belonging to 17 classes in the NCBI database (Table 5; Additional file 10: Table S9). The sequences alignment was performed using ClustalW by MEGA X software [172], the sites with two or more gaps in more than 50% species were deleted, repeated the above steps until the missing data were cleaned completely. The trees
were constructed using the Neighbor-Joining method [173] under the p-distances model with bootstrapping (2000). All positions with less than 50% site coverage were eliminated, i.e., fewer than 50% alignment gaps, missing data, and ambiguous bases were allowed at any position (partial deletion option). iTOL v6.3.2 [174] was used to visualize the tree.

Supplementary Information

Additional file 1: Fig. S1 Overview of PacBio Sequel II SMRT sequencing data processing. Fig. S2 Intermediate processes of PacBio Sequel II SMRT sequencing data processing. Fig. S3 GO classifications of coral genes in *P. damicornis*. Fig. S4 GO classifications of coral genes in *P. verrucosa*. Fig. S5 GO classifications of coral genes in *A. muricata*. Fig. S6 GO classifications of coral genes in *M. foliosa*. Fig. S7 KOG classifications of coral genes. Fig. S8 KEGG metabolic pathway classifications of coral genes in *P. damicornis*. Fig. S9 KEGG metabolic pathway classifications of coral genes in *P. verrucosa*. Fig. S10 KEGG metabolic pathway classifications of coral genes in *A. muricata*. Fig. S11 KEGG metabolic pathway classifications of coral genes in *M. foliosa*. Fig. S12 Summary of coral gene structural analysis. Fig. S13 FPKM box plot of each coral gene expression. Fig. S14 Venn plots of coral differentially expressed genes. Fig. S15 Summary of Symbiodiniaceae gene functional annotation. Fig. S16 NR database annotation of Symbiodiniaceae. Fig. S17 GO classifications of Symbiodiniaceae genes in *P. damicornis*. Fig. S18 GO classifications of Symbiodiniaceae genes in *P. verrucosa*. Fig. S19 GO classifications of Symbiodiniaceae genes in *A. muricata*. Fig. S20 GO classifications of Symbiodiniaceae genes in *M. foliosa*. Fig. S21 KOG classifications of Symbiodiniaceae genes. Fig. S22 KEGG metabolic pathway classifications of Symbiodiniaceae genes in *P. damicornis*. Fig. S23 KEGG metabolic pathway classifications of Symbiodiniaceae genes in *P. verrucosa*. Fig. S24 Summary of Symbiodiniaceae gene functional annotation.
classifications of Symbiodiniaceae genes in *P. verrucosa*. **Fig. S24** KEGG metabolic pathway classifications of Symbiodiniaceae genes in *A. muricata*. **Fig. S25** KEGG metabolic pathway classifications of Symbiodiniaceae genes in *M. foliosa*. **Fig. S26** Summary of Symbiodiniaceae gene structural analysis. **Fig. S27** Venn plots of statistics on the number of Symbiodiniaceae sequences expressed in four reef-building corals. **Fig. S28** Summary of the Symbiodiniaceae gene expression level analysis. **Fig. S29** Bar graph of coexpressed Symbiodiniaceae sequences based on GO molecular functions and cellular components categories. **Fig. S30** Conserved domains of *cry*1 or 2 genes among different species. **Fig. S31** Phylogenetic tree of *cry*1 or 2 genes based on the Neighbor-Joining method. **Fig. S32** Conserved domains of *Clock* or *Npas2* genes among different species. **Fig. S33** Phylogenetic tree of *Clock* or *Npas2* genes based on the Neighbor-Joining method. **Fig. S34** Conserved domains of *cyc* or *Arntl* genes among different species. **Fig. S35** Phylogenetic tree of *cyc* or *Arntl* genes based on the Neighbor-Joining method. **Fig. S36** *Per1* or 2 or 3 gene phylogenetic analysis.

**Additional file 2:** Table S1. Detailed sequence data of four reef-building corals in the NCBI database.

**Additional file 3:** Table S2. Data processing statistics of four reef-building corals.

**Additional file 4:** Table S3. Gene function annotation results statistics.

**Additional file 5:** Table S4. Gene structural analysis results statistics.

**Additional file 6:** Table S5. Gene expression level analysis results statistics.

**Additional file 7:** Table S6. Analysis of differential gene expression in corals.

**Additional file 8:** Table S7. Kernel biomineralization-related genes in four reef-building corals.

**Additional file 9:** Table S8. Symbiodiniaceae bioinformatics analysis.
Additional file 10: Table S9. Detailed kernel circadian clock genes.

Additional file 11: Table S10. Convergent expression of genes between corals and amphioxus.

Declarations

Ethics approval and consent to participate

All coral samples were collected and processed in accordance with local laws for invertebrate protection.

Consent for publication

Not applicable.

Availability of data and materials

The datasets (reef-building coral holobionts full-length and short-read transcriptome sequencing raw data) generated during the current study are available at the Sequence Read Archive (SRA) publicly available repository, [https://www.ncbi.nlm.nih.gov/sra/]. The accession numbers are SRR9613489 and SRR12904788-90 for Pocillopora damicornis, SRR12963486 and SRR12959238-40 for Pocillopora verrucosa, SRR12963485 and SRR12959195, SRR12959206, SRR12959217 for Acropora muricata, SRR12963484 and SRR12959182-4 for Montipora foliosa.

The datasets (Unigene, CDS, PEP and UTR sequences) analysed during the current study are available in https://doi.org/10.6084/m9.figshare.1940302.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the Guangxi Key Research and Development Program...
Authors' contributions

CH conceived the project. ZL, XL, and JC made suggestions for the project. TH performed the experiment, data analysis, writing and editing. All authors discussed the results and commented on the manuscript. The authors read and approved the final manuscript.

Acknowledgments

We thank M. Zhu from the Nanjing Institute of Geology and Paleontology, CAS for scientific guidance, M. Zhu, J. Lu, and Z. Gai from the Institute of Vertebrate Paleontology and Paleoanthropology, CAS for technical and scientific guidance, and Y. Loya from the Israel Academy of Sciences for scientific guidance.

References

4. Smith SV. Coral-reef area and the contributions of reefs to processes and


In: Dubinsky Z, Stambler N, editors. Coral Reefs: An Ecosystem in Transition. Springer,

coral dinoflagellate symbionts highlight evolutionary adaptations conducive to a

Comparison of 15 dinoflagellate genomes reveals extensive sequence and structural


Symbiodinium genomes reveal adaptive evolution of functions related to coral-

divergent Symbiodinium genomes reveal conservation of a gene cluster for sunscreen

Draft assembly of the Symbiodinium minutum nuclear genome reveals dinoflagellate


on 11 December 2021.


Ziegler M, Grupstra CGB, Barreto MM, Eaton M, BaOmar J, Zubier K, et al. Coral bacterial community structure responds to environmental change in a host-


113. Sweet M, Bythell J. White syndrome in Acropora muricata: nonspecific


1132. Li W, Jaroszewski L, Godzik A. Tolerating some redundancy significantly


162. Hu H, Miao YR, Jia LH, Yu QY, Zhang Q, Guo AY. AnimalTFDB 3.0: a
comprehensive resource for annotation and prediction of animal transcription factors.


170. Love MI, Huber W, Anders S. Moderated estimation of fold change and


Table 1 Sequence data of four reef-building corals in the NCBI database.

<table>
<thead>
<tr>
<th>NO.</th>
<th>BioProject</th>
<th>Data Volume, Gb</th>
<th>Data Volume, Mbytes</th>
<th>Library Strategy</th>
<th>Platform</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P. damicornis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>PRJNA227785</td>
<td>27</td>
<td>17496</td>
<td>RNA-Seq</td>
<td>Illumina</td>
<td>Genome Analyzer Ix</td>
</tr>
<tr>
<td>2</td>
<td>PRJNA299443</td>
<td>24</td>
<td>10069</td>
<td>RNA-Seq</td>
<td>Illumina</td>
<td>HiSeq 4000</td>
</tr>
<tr>
<td>3</td>
<td>PRJNA327142</td>
<td>10</td>
<td>3750</td>
<td>RNA-Seq</td>
<td>Illumina</td>
<td>HiSeq 4000</td>
</tr>
<tr>
<td>4</td>
<td>PRJNA433950</td>
<td>60</td>
<td>23545</td>
<td>RNA-Seq</td>
<td>Illumina</td>
<td>HiSeq X Ten</td>
</tr>
<tr>
<td>5</td>
<td>PRJNA435620</td>
<td>67</td>
<td>26334</td>
<td>RNA-Seq</td>
<td>Illumina</td>
<td>HiSeq X Ten</td>
</tr>
<tr>
<td>6</td>
<td>PRJNA306839</td>
<td>18</td>
<td>8574</td>
<td>RNA-Seq</td>
<td>Illumina</td>
<td>MiSeq</td>
</tr>
<tr>
<td>7</td>
<td>PRJEB26650</td>
<td>45</td>
<td>21254</td>
<td>RNA-Seq</td>
<td>Illumina</td>
<td>HiSeq 2500</td>
</tr>
<tr>
<td>8</td>
<td>PRJNAS44778</td>
<td>1348</td>
<td>0.41T</td>
<td>RNA-Seq</td>
<td>PacBio Illumina</td>
<td>Sequel II</td>
</tr>
<tr>
<td>9</td>
<td>PRJNAS45379</td>
<td>67</td>
<td>27278</td>
<td>RNA-Seq</td>
<td>Illumina</td>
<td>HiSeq 2000</td>
</tr>
<tr>
<td>10</td>
<td>PRJNA611041</td>
<td>258</td>
<td>90416</td>
<td>RNA-Seq</td>
<td>Illumina</td>
<td>HiSeq 1500</td>
</tr>
<tr>
<td>P. verrucosa</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>PRJNA552592</td>
<td>15</td>
<td>7782</td>
<td>RNA-Seq</td>
<td>Illumina</td>
<td>HiSeq 2500</td>
</tr>
<tr>
<td>2</td>
<td>PRJNA551401</td>
<td>108</td>
<td>47540</td>
<td>RNA-Seq</td>
<td>Illumina</td>
<td>HiSeq4000</td>
</tr>
<tr>
<td>3</td>
<td>PRJNAS44778</td>
<td>1348</td>
<td>0.41T</td>
<td>RNA-Seq</td>
<td>PacBio Illumina</td>
<td>Sequel II</td>
</tr>
<tr>
<td>A. muricata</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>PRJNAS44778</td>
<td>1348</td>
<td>0.41T</td>
<td>RNA-Seq</td>
<td>PacBio Illumina</td>
<td>Sequel II</td>
</tr>
<tr>
<td>2</td>
<td>PRJDB8519</td>
<td>1785</td>
<td>1.09T</td>
<td>WGS</td>
<td>Illumina</td>
<td>HiSeq 2500</td>
</tr>
<tr>
<td>3</td>
<td>PRJDB85633</td>
<td>110</td>
<td>63481</td>
<td>WGS</td>
<td>Illumina</td>
<td>HiSeq 2500</td>
</tr>
<tr>
<td>M. foliosa</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>PRJNAS44778</td>
<td>1348</td>
<td>0.41T</td>
<td>RNA-Seq</td>
<td>PacBio Illumina</td>
<td>Sequel II</td>
</tr>
</tbody>
</table>

Only datasets of ≥ 10 Gb coral sequences (i.e., excluding Symbiodiniaceae sequences) are included. The bolded lines indicate data generated in this paper.
Table 2 Statistics of PacBio Iso-seq data.

<table>
<thead>
<tr>
<th>Sample Name</th>
<th>P. damicornis</th>
<th>P. verrucosa</th>
<th>A. muricata</th>
<th>M. foliosa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polymerase reads (Gb)</td>
<td>25.34</td>
<td>27.8</td>
<td>22.32</td>
<td>21.44</td>
</tr>
<tr>
<td>Subreads (Gb)</td>
<td>24.43</td>
<td>26.52</td>
<td>21.36</td>
<td>20.24</td>
</tr>
<tr>
<td>CCS (Number)</td>
<td>334111</td>
<td>292565</td>
<td>341490</td>
<td>305153</td>
</tr>
<tr>
<td>FLNC (Number)</td>
<td>245504</td>
<td>249577</td>
<td>273822</td>
<td>238498</td>
</tr>
<tr>
<td>Polished consensus (Number)</td>
<td>20994</td>
<td>24860</td>
<td>31571</td>
<td>26455</td>
</tr>
<tr>
<td>Transcripts (Number)</td>
<td>20609</td>
<td>24174</td>
<td>31242</td>
<td>25460</td>
</tr>
<tr>
<td>Unigenes (Number)</td>
<td>14167</td>
<td>12822</td>
<td>13800</td>
<td>10905</td>
</tr>
<tr>
<td>Mean length (bp)</td>
<td>2668</td>
<td>2092</td>
<td>1918</td>
<td>1472</td>
</tr>
<tr>
<td>Minimum length (bp)</td>
<td>51</td>
<td>107</td>
<td>183</td>
<td>69</td>
</tr>
<tr>
<td>Maximum length (bp)</td>
<td>9888</td>
<td>7861</td>
<td>6991</td>
<td>5942</td>
</tr>
<tr>
<td>N50 (bp)</td>
<td>2954</td>
<td>2313</td>
<td>2126</td>
<td>1678</td>
</tr>
<tr>
<td>N90 (bp)</td>
<td>1793</td>
<td>1346</td>
<td>1230</td>
<td>905</td>
</tr>
</tbody>
</table>

CCS: circular consensus sequence.
FLNC: full-length non-chimera sequences.
N50 or 90: the length for which the collection of all subreads of that length or longer contains at least 50% or 90% of the total of the lengths of the subreads.
Table 3 The top five species with the most similar genes to the four investigated corals.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Species</th>
<th>Gene Number</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Acropora digitifera</td>
<td>8310</td>
<td>62.66%</td>
</tr>
<tr>
<td>2</td>
<td>Exaiptasia pallida</td>
<td>2313</td>
<td>17.44%</td>
</tr>
<tr>
<td>3</td>
<td>Nematostella vectensis</td>
<td>1632</td>
<td>12.31%</td>
</tr>
<tr>
<td>4</td>
<td>Branchiostoma belcheri</td>
<td>145</td>
<td>1.09%</td>
</tr>
<tr>
<td>5</td>
<td>Stylophora pistillata</td>
<td>80</td>
<td>0.60%</td>
</tr>
<tr>
<td>1</td>
<td>Acropora digitifera</td>
<td>7144</td>
<td>62.27%</td>
</tr>
<tr>
<td>2</td>
<td>Exaiptasia pallida</td>
<td>1928</td>
<td>16.80%</td>
</tr>
<tr>
<td>3</td>
<td>Nematostella vectensis</td>
<td>1475</td>
<td>12.86%</td>
</tr>
<tr>
<td>4</td>
<td>Branchiostoma belcheri</td>
<td>86</td>
<td>0.75%</td>
</tr>
<tr>
<td>5</td>
<td>Stylophora pistillata</td>
<td>54</td>
<td>0.47%</td>
</tr>
<tr>
<td>1</td>
<td>Acropora digitifera</td>
<td>9814</td>
<td>78.14%</td>
</tr>
<tr>
<td>2</td>
<td>Exaiptasia pallida</td>
<td>1098</td>
<td>8.74%</td>
</tr>
<tr>
<td>3</td>
<td>Nematostella vectensis</td>
<td>1034</td>
<td>8.23%</td>
</tr>
<tr>
<td>4</td>
<td>Acropora millepora</td>
<td>110</td>
<td>0.88%</td>
</tr>
<tr>
<td>5</td>
<td>Branchiostoma belcheri</td>
<td>33</td>
<td>0.26%</td>
</tr>
<tr>
<td>1</td>
<td>Acropora digitifera</td>
<td>5902</td>
<td>69.44%</td>
</tr>
<tr>
<td>2</td>
<td>Exaiptasia pallida</td>
<td>1034</td>
<td>12.16%</td>
</tr>
<tr>
<td>3</td>
<td>Nematostella vectensis</td>
<td>951</td>
<td>11.19%</td>
</tr>
<tr>
<td>4</td>
<td>Acropora millepora</td>
<td>78</td>
<td>0.92%</td>
</tr>
<tr>
<td>5</td>
<td>Branchiostoma belcheri</td>
<td>48</td>
<td>0.56%</td>
</tr>
<tr>
<td>Class</td>
<td>Protein Name</td>
<td>P. damicornis</td>
<td>P. verrucosa</td>
</tr>
<tr>
<td>--------------------------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>---------------</td>
<td>--------------</td>
</tr>
<tr>
<td>Calcium ATPase</td>
<td>Plasma membrane calcium-transporting ATPase 1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Calcium ATPase</td>
<td>Plasma membrane calcium-transporting ATPase 2</td>
<td>1574.217</td>
<td>685.81</td>
</tr>
<tr>
<td>Calcium ATPase</td>
<td>Plasma membrane calcium-transporting ATPase 3</td>
<td>1.783333</td>
<td>503.5267</td>
</tr>
<tr>
<td>Calcium ATPase</td>
<td>Plasma membrane calcium-transporting ATPase 4</td>
<td>1502</td>
<td>1900.66</td>
</tr>
<tr>
<td>Calcium ATPase</td>
<td>Calcium-transporting ATPase type 2C member</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Bicarbonate Transporter</td>
<td>solute carrier 4 (SLC4)</td>
<td>9643.527</td>
<td>8724.997</td>
</tr>
<tr>
<td>Bicarbonate Transporter</td>
<td>solute carrier 26 (SLC26)</td>
<td>2450.687</td>
<td>9511.6</td>
</tr>
<tr>
<td>Alpha Carbonic Anhydrase</td>
<td>carbonic anhydrase 1</td>
<td>741.3333</td>
<td>216</td>
</tr>
<tr>
<td>Alpha Carbonic Anhydrase</td>
<td>carbonic anhydrase 2</td>
<td>68885.42</td>
<td>34358</td>
</tr>
<tr>
<td>Alpha Carbonic Anhydrase</td>
<td>carbonic anhydrase 3</td>
<td>439</td>
<td>206.6667</td>
</tr>
<tr>
<td>Alpha Carbonic Anhydrase</td>
<td>carbonic anhydrase 12</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Acidic Proteins</td>
<td>Skeletal aspartic acid-rich protein 1 (SAARP1)</td>
<td>6040.233</td>
<td>20674.29</td>
</tr>
<tr>
<td>Acidic Proteins</td>
<td>Skeletal aspartic acid-rich protein 2 (SAARP2)</td>
<td>165.6667</td>
<td>297.3333</td>
</tr>
<tr>
<td>Acidic Proteins</td>
<td>Acidic skeletal organic matrix protein (Acidic SOMP)</td>
<td>1715.333</td>
<td>1351.663</td>
</tr>
<tr>
<td>Acidic Proteins</td>
<td>Secreted acidic protein 1 (SAP1)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Acidic Proteins</td>
<td>Secreted acidic protein 2 (SAP2)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Acidic Proteins</td>
<td>Aspartic and glutamic acid-rich protein</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Unique Uncharacterized Proteins</td>
<td>Uncharacterized skeletal organic matrix protein-1 (USOMP-1)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Unique Uncharacterized Proteins</td>
<td>Uncharacterized skeletal organic matrix protein-2 (USOMP-2)</td>
<td>227.6633</td>
<td>120.33</td>
</tr>
<tr>
<td>Unique Uncharacterized Proteins</td>
<td>Uncharacterized skeletal organic matrix protein-3 (USOMP-3)</td>
<td>2042.107</td>
<td>2688.417</td>
</tr>
<tr>
<td>Unique Uncharacterized Proteins</td>
<td>Uncharacterized skeletal organic matrix protein-4 (USOMP-4)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Unique Uncharacterized Proteins</td>
<td>Uncharacterized skeletal organic matrix protein-5 (USOMP-5)</td>
<td>7567.913</td>
<td>18213.03</td>
</tr>
<tr>
<td>Unique Uncharacterized Proteins</td>
<td>Uncharacterized skeletal organic matrix protein-6 (USOMP-6)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Unique Uncharacterized Proteins</td>
<td>Uncharacterized skeletal organic matrix protein-7 (USOMP-7)</td>
<td>948.47</td>
<td>199.6667</td>
</tr>
<tr>
<td>Unique Uncharacterized Proteins</td>
<td>Uncharacterized skeletal organic matrix protein-8 (USOMP-8)</td>
<td>168.6667</td>
<td>3082</td>
</tr>
<tr>
<td>Galaxin</td>
<td>Galaxin</td>
<td>6541</td>
<td>6595.007</td>
</tr>
<tr>
<td>Galaxin</td>
<td>Galaxin 2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Collagen alpha-6(VI)</td>
<td>Collagen alpha-6(VI) chain</td>
<td>226</td>
<td>222.6667</td>
</tr>
</tbody>
</table>
Table 5 Presence of key circadian clock genes in 27 species.

<table>
<thead>
<tr>
<th>Class</th>
<th>Represent Species</th>
<th>cry1</th>
<th>cry2</th>
<th>Clock</th>
<th>Npas2</th>
<th>cyc</th>
<th>Arntl1 (Bmal1)</th>
<th>Arntl2 (Bmal2)</th>
<th>per1</th>
<th>per2</th>
<th>per3</th>
<th>tim</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mammalia</td>
<td>Homo sapiens</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>-</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Mus musculus</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>-</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Actinopterygii</td>
<td>Danio rerio</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>-</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Asciidiacea</td>
<td>Ciona intestinalis</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Y</td>
</tr>
<tr>
<td>Cephalochordata</td>
<td>Branchiostoma florid</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>-</td>
<td>Y</td>
<td>-</td>
<td>Y</td>
<td>-</td>
<td>-</td>
<td>Y</td>
</tr>
<tr>
<td>Echinoidea</td>
<td>Stronglyocentrotus purpuratus</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>-</td>
<td>Y</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Y</td>
</tr>
<tr>
<td>Enteropneusta</td>
<td>Saccoglossus kowalevskii</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Y</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Y</td>
</tr>
<tr>
<td>Nematoda</td>
<td>Caenorhabditis elegans</td>
<td>-</td>
<td>-</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>-</td>
<td>-</td>
<td>Y</td>
</tr>
<tr>
<td>Insecta</td>
<td>Drosophila melanogaster</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>-</td>
<td>-</td>
<td>Y</td>
</tr>
<tr>
<td>Bivalvia</td>
<td>Crassostrea gigas</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>-</td>
<td>Y</td>
<td>Y</td>
<td>-</td>
<td>Y</td>
<td>Y</td>
<td>-</td>
<td>Y</td>
</tr>
<tr>
<td>Turbellaria</td>
<td>Schmidtea mediterranea</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Actinzechia</td>
<td>Pocillopora damicornis</td>
<td>Y</td>
<td>-</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Pocillopora verrucosa</td>
<td>Y</td>
<td>-</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Acropora muricata</td>
<td>Y</td>
<td>-</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Montipora foliosa</td>
<td>Y</td>
<td>-</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Montipora capricornis</td>
<td>Y</td>
<td>-</td>
<td>Y</td>
<td>Y</td>
<td>-</td>
<td>-</td>
<td>Y</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Acropora digitifera</td>
<td>Y</td>
<td>-</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Stylophora pistillata</td>
<td>Y</td>
<td>-</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Orbicella faveolata</td>
<td>Y</td>
<td>-</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Nemastostella vectensis</td>
<td>Y</td>
<td>-</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Exaiptasia diaphana</td>
<td>Y</td>
<td>-</td>
<td>Y</td>
<td>Y</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Hydrozoa</td>
<td>Hydra vulgaris</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Cubozoa</td>
<td>Chironex fleckeri</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Scyphozoa</td>
<td>Aurelia aurita</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Sponges</td>
<td>Amphimedon queenslandica</td>
<td>-</td>
<td>Y</td>
<td>Y</td>
<td>-</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Y</td>
</tr>
<tr>
<td>Ctenophore</td>
<td>Pleurobranchia bachei</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Ciliata</td>
<td>Paramecium bachei</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

'Y' means that the genome of the indicated species contains the gene, while '-' means that the genome of the indicated species does not contain the gene.
Fig. 1 Summary of coral gene functional annotation. 

a. Statistics of the annotation results of four reef-building corals, excluding Symbiodiniaceae sequences, in the NR, Swiss-Prot, KEGG, KOG, GO NT and Pfam databases. The horizontal axis represents the different functional databases, and the vertical axis represents the number of sequences annotated in different functional databases, at least one database and all databases.

b. Venn plots of the number of annotated sequences of four reef-building corals excluding Symbiodiniaceae sequences obtained using the NR, KEGG, KOG, GO and NT databases. The sum of the numbers in each large circle represents the number of transcripts annotated in one database, and the overlapping circles indicate the numbers of transcripts annotated to these databases simultaneously.
**Fig. 2** NR database annotation of corals. The top 20 species with the greatest number of top sequence hits to *P. damicornis* (a), *P. verrucosa* (b), *A. muricata* (c) and *M. foliosa* (d) are shown. The horizontal axis represents the species ID, and the vertical axis represents the number of coral unigenes annotated to different species. The pink columns represent species belonging to the Actinozoa. The blue columns represent species belonging to the genus *Branchiostoma*. 

- **a**: *Acropora digitifera*
- **b**: *Eupaxipa a pallida*
- **c**: *Hemicentrotus vesicularis*
- **d**: *Hymeniacidon pseudopterina*

---

1. *Acropora digitifera*
2. *Eupaxipa a pallida*
3. *Hemicentrotus vesicularis*
4. *Hymeniacidon pseudopterina*
Fig. 3 Summary of the coral gene expression level analysis. a. FPKM interval distribution. The horizontal axis shows the sample name, different colours represent FPKM intervals, and the vertical axis represents the number of genes in each interval. b. FPKM density distribution. The horizontal axis represents the log10 (FPKM+1) values, and the vertical axis represents the density of genes with different expression levels. c. FPKM box plot. The horizontal axis shows the sample name, and the vertical axis represents the log10 (FPKM+1) values. Each box plot shows five statistical values, including the maximum, upper quartile, median, lower quartile and minimum, from top to bottom. d. Pearson correlation among samples. The closer the value is to 1, the better the correlation. Pd: *P. damicornis*; Pv: *P. verrucosa*; Am: *A. muricata*; Mf: *M. foliosa*. 
Fig. 4 Mechanism of coral biomineralization and related gene expression levels. a. The mechanism of reef-building coral biomineralization. (1) Ca\(^{2+}\) transport by calcium ATPase or diffusion. (2) CO\(_2\) can be converted into HCO\(_3^-\) by CA and then exits the cells via bicarbonate transporters (i.e., SLC4 and SLC26). (3) Coral acid-rich proteins and Mg\(^{2+}\) can promote crystal nucleation, determine growth axes and control crystal growth. (4) The bioprecipitation of aragonite crystals in corals requires additional skeletal organic matrix proteins (USOMPs, galaxins and alpha IV collagen). The average expression levels of genes belonging to calcium ATPases (b), bicarbonate transporters (c), carbonic anhydrases (d), acid-rich proteins(e), USOMPs (f), and galaxins (g). The horizontal axis shows the sample name, and the vertical axis represents the average read count values of each set of genes. ATPase: calcium ATPase; SLC: solute carrier 4 (SLC4) and solute carrier 26 (SLC26); CA: carbonic anhydrase; Sym: Symbiodiniaceae; Mito: mitochondrion; and SOMPs: coral acid-rich proteins, alpha IV collagen, galaxins and uncharacterized skeletal organic matrix proteins (USOMPs). The solid lines represent definite paths, and the dashed lines represent possible paths. Pd: *P. damicornis*; Pv: *P. verrucosa*; Am: *A. muricata*; and Mf: *M. foliosa*. 
Fig. 5 Bar graph of coexpressed Symbiodiniaceae sequences based on GO biological processes categories. The vertical axis represents the GO terms, and the horizontal axis represents the number of transcripts annotated to the terms (including subterms of the terms).