Deciphering omics atlases to aid stony corals in response to global change

Chunpeng He

cphe@seu.edu.cn

Tingyu Han
Wanlong Huang
Bing Wang
Xin Liao
J.-Y. Chen
Zuhong Lu

Research Article

Keywords: Stony corals, genome assembly, chromosome, PacBio, Hi-C, full-length transcriptome, small RNA, scRNA-seq, SIDT1

Posted Date: March 11th, 2024

DOI: https://doi.org/10.21203/rs.3.rs-4037544/v1

Additional Declarations: The authors declare no competing interests.

Abstract

Coral reefs are rapidly declining due to global climate change[1]. Alongside environmental protection, gene engineering technologies are essential in protecting reef-building (stony) corals[2-7]. Precise multi-omics backgrounds are crucial for gene technique application[8-13] and understanding stony coral biology[2-5]. However, high heterozygosity[4,14] and exogenous sequence contamination[15-17] impede high-quality stony coral genome assembly. To address this, we developed a workflow for stony coral genome assembly, achieving four high-quality chromosome-level genomes (*Acropora muricata*, *Montipora foliosa*, *M. capricornis*, and *Pocillopora verrucosa*). A global synteny analysis of chromosomes and homeobox genes highlighted the highly conserved features of Scleractinia stony coral genomes, even after an evolutionary divergence 250 million years ago (mya). Single-cell RNA sequencing (scRNA-seq), combined with a pan-genomic analysis, aided our understanding of symbiosis and calcification, both issues of conservation concern. A dynamics simulation of the mechanism of systemic RNA interference defective protein 1 transmembrane family member 1 (SIDT1) tetramer offered a potential approach to observe gene impacts on stony corals. This work enhances understanding of the evolutionary, molecular, and cellular aspects of stony corals and provides a detailed dataset for applying gene engineering technologies in response to global climate challenges.

Introduction

Coral reefs face severe environmental threats: global warming, ocean acidification, Crown-of-thorns starfish (COTS), algae blooms, overfishing, and pollution. The protection of stony corals and their habitats is crucial for marine ecosystem health and diversity[1-5,7,12-14,17]. As the climate changes, without gene regulation technologies, stony corals may undergo bleaching and disappear[7,12-14]. A significant number of stony coral transcriptomes and genomes were previously sequenced[2-5,14], and the scRNA-seq remains in progress[5]. However, the unique nature of the coral holobiont[15,16] poses challenges in stony coral genome assembly[4,14]. Intracellular endosymbiotic Symbiodiniaceae and environmental symbiotic algae and microbes, accompany stony coral throughout their life cycle[16], and removing sequence contaminations from these is an indispensable step in genome assembly[4].

To address these challenges, we developed an open workflow for stony coral genome sequencing and pseudochromosome assembly based on high-throughput chromosome conformation capture (Hi-C) technology, using the most readily available adult coral branches (Methods, Supplementary Table 1). Sequencing data for the four studied stony corals is in Extended Data Table 1 and aligns with data storage details in Supplementary Table 1.

Genome assembly

Following the workflow, an initial genome survey was conducted to set a reference (Extended Data Fig. 1a, b, Methods). Sequencing data from the PacBio Sequel II platform were assembled into contigs using

Canu software (version 2.2) (Supplementary Table 1). Consequently, after correcting sequence errors, removing heterozygous contigs, and eliminating the sequence contaminations from microbes, endosymbiotic Symbiodiniaceae, and other symbiotic algae (Extended Data Fig. 1c, Methods), these contigs were assembled into scaffolds using 10X Genomics linked-reads. The scaffolds were anchored into n=14 pseudochromosomes[18], guided by Hi-C sequencing data and gap-closing techniques (Fig. 1a, Extended Data Fig. 2b, c, d, Methods, Supplementary Table 1). The Hi-C anchored rates for *A. muricata*, *M. foliosa*, *M. capricornis*, and *P. verrucosa* genomes are between 92.88% and 99.36%, and their genome size is between 354.87 Mb and 810.20 Mb, with corresponding contig N50 values between 0.83 Mb and 2.21 Mb (Fig. 1c, Extended Data Table 1).

Genome assessment

The genomes were assessed using the Benchmarking Universal Single-Copy Orthologs (BUSCO) software by searching against the lineage dataset of metazoa_odb10 (version 5.1.2) (Methods). In *A. muricata*, *M. foliosa*, *M. capricornis*, and *P. verrucose* genomes, out of 954 orthologous single-copy genes, the BUSCO score for complete genes is between 91.9% and 93.8%, indicating a relatively high level of assembly results (Fig. 1b, Extended Data Table 2). In the sequence consistency evaluation, the alignment rate of all short-read fragments to the genome is between 91.07% and 98.15% with a coverage rate between 94.42% and 99.22%, indicating consistency between the reads and the assembled genomes (Supplementary Table 1). The assessment of single nucleotide polymorphisms (SNPs) indicates that these genomes exhibit high single-base accuracy due to a lower homozygous SNP rate (Supplementary Table 1).

Genome annotation

Approximately 58.87%, 62.17%, 62.28%, and 48.46% of the *A. muricata*, *M. foliosa*, *M. capricornis*, and *P. verrucosa* genomes consist of repeat sequences, and the majority are transposable elements (TEs), notably DNA transposons (DNA), long interspersed elements (LINE), and long terminal repeats (LTR) (Fig. 1d, Extended Data Fig. 2, Extended Data Table 1, Methods, Supplementary Table 1). This suggests that TEs play a significant role in stony coral genomes, particularly within the complex clade (*A. muricata*, *M. foliosa*, and *M. capricornis*)[14,19], with a genus level distribution feature. Between 19,493 and 22,816 genes were predicted in these four genomes, resulting in functional gene rates between 99.6% and 99.8% (Extended Data Table 1, Methods, Supplementary Tables 1, 2). For small RNA annotation, 2,042, 3,610, 3,844, and 457 microRNA (miRNA) loci were identified in *A. muricata*, *M. foliosa*, *M. capricornis* (complex clade), and *P. verrucosa* (robust clade) genomes (Fig. 1d). Similarly, the number of identified miRNAs is higher in the complex clade than the robust. In all stony corals examined, there was a pronounced overexpression of miR-100 (Supplementary Table 5), and its target genes show lower expression levels in bulk transcriptomes (Supplementary Table 3) and single-cell transcriptomes (Supplementary Tables 9, 10). Therefore, miRNAs may play an essential role in stony corals' physiological regulation[11].

Chromosome macrosynteny analysis

The macrosynteny analysis illuminates the highly conserved nature of gene order within Scleractinia stony coral chromosomes. There is a pronounced alignment in the chromosome loci of coding genes, with few cross-chromosomal correspondences, in *A. muricata*, *M. foliosa*, *M. capricornis*, and *P. verrucosa* genomes (Fig. 2). The macrosynteny results cover two major branches of Scleractinia stony corals (complex and robust) and three genera (*Acropora*, *Montipora*, and *Pocillopora*), indicating evolutionary chromosomal conservation among Scleractinia stony corals and reflecting the core genome[18]. Moreover, while some Scleractinia stony corals have polyploid genomes[18], a karyotype with a diploid chromosome number of 2n=28 is likely an ancestral trait. Extending this comparative pattern to *Nematostella vectensis* highlights the conserved feature of chromosomal homology in their coevolution, suggesting that, within the class Hexacorallia, both Scleractinia and Actiniaria (sea anemones) underwent events of homologous chromosome fusion and breakage (Fig. 2, Extended Data Fig. 3). Furthermore, as Actiniaria does not form reefs, the conservation of chromosomal structure may have profoundly impacted the lifestyle of stony corals, especially their reef-building ability.

Homeobox gene microsynteny analysis

The homeobox gene family encodes transcription factors that play crucial roles in patterning body plans[20]. Homeobox genes are recognized as the gold standard in microsynteny analysis due to their conserved distribution on chromosomes[14]. Based on the macrosynteny analysis (Fig. 2) and gff3 files (Supplementary Table 1), we conducted a comprehensive annotation of orthologous homeobox genes for the four stony corals genomes and *N. vectensis* (Methods, Supplementary Table 6), showing that the genomic loci distribution of ANTP-class homeobox genes is highly conserved with remarkable consistency (Extended Data Fig. 3). These genes can be divided into seven across-chromosome corresponding groups: Hox, ParaHox, and five NKL (NK-like) groups. Within the Hox and NKL groups, there is evidence of tandemly duplicated genes (TDG)[14,19]. This synteny can be extended to the chromosome distribution in the *N. vectensis* genome, corresponding to the overall chromosome changes observed in the macrosynteny analysis. Insights from the microsynteny analysis of homeobox genes further elucidates the inherent conservation of the Scleractinia stony corals' body plan (Extended Data Fig. 3, 4) and reinforces our assembly results. Both global chromosome macrosynteny and orthologous group (OG) microsynteny analyses of the homeobox genes indicate that the primary cause for the divergence between the complex and robust clades of Scleractinia stony corals is due to specific evolutionary adaptations, not karyotype changes[19].

Phylogenetic and pan-genomic analysis

Based on a BUSCO assessment, 24 species of cnidarian (including 16 Scleractinia stony corals) and one poriferan (*Amphimedon queenslandica*) were selected for phylogenetic analysis (Extended Data Table 2, Methods)[21,22]. The phylogenetic relationship was reconstructed using 1,554 orthologous genes. Then, we conducted a pan-genomics analysis of Scleractinia stony corals (Extended Data Fig. 5a-c, Methods, Supplementary Table 7), involving a systematic analysis of core gene families (Set1), exclusive core gene

families (Set2), and core gene families shared with other cnidarians (Set3), and the expansion (Set4), contraction (Set5), and positively selected gene families (Set6) (Extended Data Fig. 5d, Supplementary Table 7).

The phylogenetic analysis mirrors fossil evidence[23], revealing that Anthozoa emerged within the phylum Cnidaria during the Precambrian and diverged from Medusozoa (Fig. 3). During the Cambrian Explosion, Hexacorallia emerged, alongside other cnidarian branches[24]. In the Early Ordovician, Hexacorallia experienced further biological diversification, highlighted by Scleractinia evolving separately from Actiniaria. Around the Permian-Triassic extinction event[25,26], approximately 200 mya, Scleractinia first exhibited reef-building capabilities[27] and diversified into its complex and robust clades[20]. The gene functions of Set6 and its relationship with Set2-Set4, also support this (Fig. 3, Extended Data Fig. 5d, Supplementary Table 7).

Scleractinia underwent five phases of phylogenetic diversification (Fig. 3), becoming a fundamental component of contemporary marine ecosystems[26,27]. The first, second and fourth phases aligned with major extinction events[25,26], offering geological insights into the group development. The fourth phase culminated in a marked increase in genus-level biodiversity, with near-complete representation of modern genera[27]. As noted earlier, the TE distribution in the genomes of various Scleractinia genera demonstrates notable similarities (Extended Data Fig. 2). We observed a discernible pattern in the evolutionary history of genomic TEs within these genera during the Cenozoic era (Extended Data Fig. 6a), aligning with taxonomic distinctions (Supplementary Table 7). The fifth diversification phase occurred during the Miocene epoch[26,28], marked by significant increases in species diversity within *Acropora* and *Montipora* genera (Fig. 3)[19]. The Pairwise Sequentially Markovian Coalescent (PSMC) analysis (Extended Data Fig. 6b) showed the effective population size of Scleractinia stony corals peaked between one and three mya and is currently in a 100,000-year recovery phase[28]. Scleractinia stony corals in the Indo-Pacific region show less robust recovery compared to the Atlantic region, perhaps due to the detrimental effects of COTS[17].

The pan-genomic analysis shows that Set5 is relatively limited compared to Actiniaria, with three gene families (Fig. 3, Extended Data Fig. 5d). These primarily pertain to the G-protein-coupled receptors (GPCRs) gene family, whose physiological functions are linked to growth inhibition at the individual (polyp) level (Supplementary Table 7). Set4 is more pronounced with gene families maintaining genomic stability and enhancing the group attachment and rapid growth abilities of polyps and their shared coenosarc tissues[29]. The specific physiological functions of gene families related to Set1-Set4 in symbiosis and reef-building is discussed further in conjunction with scRNA-seq data (Supplementary Table 8).

scRNA-seq

From an initial set of 312,044 sequenced cells, we selected 59,382 for downstream analysis based on general screening criteria, including 17,277 cells for *A. muricata*, 19,369 cells for *M. foliosa*, 11,243 cells for *M. capricornis*, and 11,493 cells for *P. verrucosa* (Methods, Supplementary Tables 8-10). Among the scRNA-seq results, the gene expression profiles of the cell clusters in *A. muricata* exhibit the highest sensitivity (Extended Data Fig. 7a, b, Extended Data Fig. 8a, Supplementary Tables 8, 9). We utilized these profiles to investigate the physiological functions of symbiosis and calcification, an issue of conservation interest and concern, combined with pan-genomics Set1-4.

These four Scleractinia stony corals have 12 types of cell clusters[5,30]: ectoderm-derived cnidocytes, neurons 1, neurons 2, gland cells, epidermal cells, calicoblastic cells (calicoblasts), endoderm-derived alga-hosting cells (symbiotic cells or endosymbiotic cells), digestive filaments, gastrodermal cells, retractor muscle cells, uncertainly germ layer-derived progenitor cells, and immune cells (Extended Data Fig. 7a, b, Extended Data Fig. 8c-h, Supplementary Table 8).

In *A. muricata*, these 12 types can be divided into 24 cell clusters by cell marker genes[5,30] and specific high-expression genes (Extended Data Fig. 7a, b, Extended Data Fig. 8a, Supplementary Table 8). Universally expressed genes at high levels across all cell clusters (Supplementary Table 9) is consistent with general transcriptome data (Supplementary Table 4) showing all cell clusters overwhelmingly expressing *CCAAT/enhancer-binding protein gamma* (*CEBPG*) (Extended Data Fig. 7c) and highly expressing *heat shock protein HSP 90-alpha* (*HSP90AA1*) and other genes (Supplementary Table 8). Except for *tubulin alpha-1A chain* (*TUBA1A*, Set3), the basic functions of these genes are environmental stimuli and stress response. Both *CEBPG* and *HSP90AA1* belong to Set3, reflecting the common characteristics of Scleractinia stony corals as cnidarians in reaction to environmental stimuli and stress.

Alga-hosting cells with three states

According to gene expression patterns and functions, the alga-hosting cells can be divided into three states: endosymbiotic homeostasis (C7), expelling Symbiodiniaceae (C20), and degrading Symbiodiniaceae (C22) (Fig. 4a, b, Supplementary Table 8). We observed that the functions and metabolic substrates of the specific high-expression genes in the three states closely align with the long-term observations of Wiedenmann *et al.* showing that Scleractinia stony corals can digest some symbionts, gaining access to an important nutrient pool[31].

Under the state of endosymbiotic homeostasis, alga-hosting cells (C7) highly express three genes: *Carbonic anhydrase 2* (*CA2*), *Transferrin* (*TF*), and *solute carrier family 26 member 6* (*SLC26A6*) (Fig. 4b, Supplementary Table 8). SLC26A6 is a membrane protein that transports bicarbonate (HCO^-_3) ions and other anions across cellular membranes[29]. Maintaining the intracellular bicarbonate ion concentration is crucial for providing an effective carbon source for photosynthesis in endosymbiotic Symbiodiniaceae[32]. CA2 catalyzes the reversible conversion between bicarbonate ions and carbon dioxide (CO_2), which is critical for maintaining dissolved inorganic carbon (DIC) and forming an intracellular CO_2 buffer pool necessary for supplying CO_2 directly to intracellular Symbiodiniaceae as a substrate for

photosynthesis[33]. CA2 also maintains intracellular pH, preventing cellular acidification and leading to the activation of cathepsins and the degradation of symbiotic algae[33,34]. Transferrin is a glycoprotein with crucial roles in iron transport and homeostasis, effectively reducing the cellular toxicity of intracellular algae[29]. These three genes belong to Set2, illustrating the shared characteristics of symbiotic cells in Scleractinia stony corals and the distinction between the symbiotic mechanism in Scleractinia stony corals and other cnidarians. Moreover, the gene expression pattern of C7 showcases the common features of molecular metabolic substrates in symbiotic cells in endosymbiotic homeostasis. This includes bicarbonate ions, iron ions, cholesterol, ammonium ions, saccharopine, glucose, aldehydes, glutathione (GSH), long-chain fatty acids, polyunsaturated fatty acids (PUFAs), and gamma-butyrobetaine (Fig. 4a, Supplementary Table 8). Symbiotic cells metabolize these substrates to provide essential nutrients (carbohydrates, amino acids, and lipids) to the body of Scleractinia stony coral, maintaining its survival and growth[32].

When expelling Symbiodiniaceae, alga-hosting cells (C20) specifically and overwhelmingly express the *sphingolipid delta(4)-desaturase DES1* (*DEGS1*, Set2, Set4), *glutaredoxin* (*GLRX*, set2), *sphingosine-1-phosphate lyase 1* (*SGPL1*, Set3), *NADP-specific glutamate dehydrogenase* (*NADP-GDH*, Set2), *aldehyde dehydrogenase*, *dimeric NADP-preferring* (*ALDH-NADP*, Set3) and other metabolic enzymes (Fig. 4b, Supplementary Table 8). DEGS1 is an enzyme responsible for introducing a double bond at the delta(4) position of the sphingolipid backbone[33], critical in increasing cell membrane fluidity[29] and expelling Symbiodiniaceae from alga-hosting cells[32]. GLRX Catalyzes the redox reactions involving glutathione (GSH) with other substrates, detoxifying reactive oxygen species (ROS), and protecting cells from oxidative damage[33]. SGPL1 irreversibly cleaves sphingosine-1-phosphate (S1P), forming hexadecenal and phosphoethanolamine[33]. Lower intracellular S1P levels in cnidarians indicate an unsuitable environment for symbiotic algae, leading to their expulsion or degradation[32]. NADP-GDH catalyzes the reversible oxidative deamination of ammonia and alpha-ketoglutarate to glutamate[29]. The high expression of this enzyme suggests a higher intracellular nitrogen environment[33]. NADP-GDH dynamically regulates intracellular nitrogen components preventing cellular ammonia poisoning[29]. ALDH-NADP catalyzes the oxidation of aldehyde compounds into their corresponding carboxylic acids, protecting cells from aldehyde toxicity and oxidative damage[33]. The specific high expression of related metabolic enzymes indicates that the C20 represents alga-hosting cells currently in a state of higher concentration of metabolic substrates (Fig. 4b, Supplementary Table 8). DEGS1, in Set2 and Set4, highlights the reliance of alga-hosting cells on DEGS1-associated mechanisms for the regulation of Symbiodiniaceae transmembrane transport. SGPL1, in Set3, exemplifies the common trait of cnidarians relying on intracellular S1P levels to regulate the symbiotic state between alga-hosting cells and Symbiodiniaceae[32]. Expelling Symbiodiniaceae demonstrates the special approach of Scleractinia stony corals, as sessile cnidarians, in dealing with an intracellular environment with higher concentrations of metabolic substrates in alga-hosting cells (Fig. 4b, Supplementary Table 8).

As Symbiodiniaceae degrade, alga-hosting cells (C22) highly express *cathepsin* gene families, *GLRX*, *acid phosphatase type 7* (*ACP7*, Set2), *bactericidal permeability-increasing protein* (*BPI*, Set3), and *cholesteryl ester hydrolase* (*LIPA*, Set2) (Fig. 4b, Supplementary Table 8). The variety and amount of these *cathepsin* expressions surpass immune cell types (C1, C15, and C21), suggesting that alga-hosting cells are in a state of intracellular digestion[31,33,34], targeting conditions more severe than those encountered by immune cells. Most related *cathepsin* members belong to Set3, including cysteine proteases of cathepsin L1, B, and Z (CTSL1, CTSB, and CTSZ) and aspartyl protease of cathepsin D (CTSD), indicating a common behavior of cnidarians digesting Symbiodiniaceae under certain conditions[32].

Calicoblasts and biomineralization

The biomineralization mechanism of calicoblastic cells and the role of alga-hosting cells in supporting calcification processes are well studied[5,31,32,35-39]. Like in *Stylophora pistillata*[35], calicoblastic cells express four coral acid-rich protein (CARP) genes rich in aspartic and glutamic acids: *acidic skeletal organic matrix protein* (*ASOMP*), *skeletal aspartic acid-rich protein* (*SAARP*), *SAARP1*, and *SAARP2* (Extended Data Fig. 7d, Supplementary Table 8). These four CARPs belong to Set2, indicating a common trait for the biomineralization of calicoblastic cells among Scleractinia stony corals[5,35]. In vitro, the four CARPs can bind calcium ions spontaneously in a stoichiometric manner and precipitate aragonite nanoparticles, within a seawater environment, at a pH range of 7.6 to 8.2, through an electrostatic interaction with protons on bicarbonate ions[35]. In vivo, these amorphous calcium carbonate (ACC) nanoparticles aggregate and form ordered aragonitic structures in calcifying space through crystal growth facilitated by particle attachment, intertwining with the skeletal organic matrix (SOM, specifically ECM) to form primary skeletons, which subsequently convert into sedimentary skeletons[36], accompanied by magnesium ion (Mg^{2+}) action[40] and incomplete microbial degradation (Extended Data Fig. 7e)[38]. Additionally, calicoblastic cells highly express other genes belong to the core gene families of Scleractinia stony corals, including *golgin subfamily A member 4* (*GOLGA4*, Set2, Set4), *CUB domain-containing protein 1* (*CDP1*, Set2), *coadhesin* (*CADN*, Set3, Set4), *matrix metalloproteinase-24* (*MMP24*, Set2), *electroneutral sodium bicarbonate exchanger 1* (*SLC4A8*, Set3), *Zinc metalloproteinase nas-14* (*NAS-14*, Set2), *myoferlin* (*MYOF*, Set3), *Mucin* (*Mucin*, Set2, Set4), *CUB domain-containing protein* (CDP, Set2), and *Ectin* (Set2) (Extended Data Fig. 7d). CDP1, CADN, MYOF, Mucin, CDP, and Ectin are related to ECM components, and MMP24 and NAS-14 are their regulatory factors (Supplementary Table 8)[29,33,37]. The ECM components, MMP24 and NAS-14 assist in calcification and play a role in innate immune barrier function[39]. SLC4A8 is a membrane protein that functions as an exchanger, transporting bicarbonate ions into or out of the cell under specific physiological conditions. In addition to *CEBPG* and *HSP90AA1*, *GOLGA4* is the highest expressing gene in the gene expression profile of calicoblastic cells (Supplementary Table 9). *GOLGA4* is mainly expressed in calicoblastic cells and epidermal cells (C16) (Extended Data Fig. 7d, Supplementary Table 8), which are the two main types of exocrine epithelial cells in Scleractinia stony corals[5]. *GOLGA4* is essential for the sorting and packaging of proteins, ensuring efficient transportation to their final destinations (Extended Data Fig. 7e)[41].

Calicoblastic cells require calcium ions and bicarbonate ions as substrates for their biomineralization process[7,32,35,36]. In surface seawater with a pH ranging from 7.6 to 8.2, the concentration of calcium ions (Ca^{2+}) is significantly higher (approximately 400-450 mg/L, equivalent to 10.0-11.3 mmol/L) than bicarbonate ions (1.4 to 2.0 mmol/L)[41]. Typically, bicarbonate ions are the dominant carbonate species, representing about 90% of dissolved inorganic carbon (DIC)[32,35,36]. Like *S. pistillata*[5], the calicoblastic cells in an adult colony expresses genes for carbonic anhydrase and calcium channels at a lower level (Extended Data Fig. 8b, Supplementary Table 8). Therefore, the concentration of environmental substrates and the gene expression pattern of calicoblastic cells suggest that, under normal biomineralization processes, calcium ions can be directly obtained from seawater, while bicarbonate ions require enrichment. As a result, the CO_2 buffer pool of the symbiotic cell system is crucial for the calcification process of Scleractinia stony corals (Extended Data Fig. 7e). The gene expression pattern (Extended Data Fig. 7d) reveals that the intracellular CO_2 buffer pool constructed by symbiotic cell (C7) through SLC (SLC26A6) and CA2 provides a carbon source for the photosynthesis of endosymbiotic Symbiodiniaceaes and allows the excess CO_2 to be transferred in the form of bicarbonate ions, via SLCs such as SLC26A and SLC4A8, to calicoblastic cells for biomineralization. This process deposits CO_2 from the natural environment, achieving biological carbon sequestration in the ocean[1,7,35,36].

SIDT1 and potential pre-miRNA transport

Directly observing how the absence of a specific gene affects stony corals using simple and feasible methods, like gene knockdown, is a scientific community goal[7]. The systemic RNA interference defective protein 1 (SID1) and their role in RNA interference (RNAi) in *Caenorhabditis elegans* provides some insights[8]. Therefore, both rely on material diffusion through intercellular spaces and direct cell-to-cell communication to maintain the survival of somatic cells[8,32]. In addition to cell signaling pathways, nematodes also use SID1 to transport double-stranded RNA (dsRNA) across cells mediating related cell communications[8]. The SID1 transmembrane family member 1 (SIDT1) is evolutionarily conserved in sequence, basic structure, and function across metazoans (Fig. 5a-c, Supplementary Table 8)[42]. Pan-genome analyses identified SIDT1 (Set3) as a core gene present in all stony corals, expressed in almost all cell clusters of *S. pistillata*[5], and the four corals investigated here (Supplementary Table 8). Given the high conservation of miR-100 and SIDT1 across the four studied corals, we systematically reconstructed the operational mechanism of stony coral SIDT1 based on the amu-miR-100 precursor (Fig. 5b, d-f, Supplementary Table 5, Methods).

Pulling molecular dynamics simulation (PMDS) analysis shows that the stony coral SIDT1 tetramer[42] has the capability to transport the miRNA precursor (pre-miRNA) across the cell membrane, enabling its entry into the cell (Supplementary Movie). Further potential of mean force simulation (PMFS) analysis indicates that the pre-miRNA can stably exist at a position 128 Å away from the center of the SIDT1 tetramer channel (RC=128 Å) before transport across the phospholipid bilayer (Fig. 5e). At this position, the G-50 base at the 3' end of the pre-miRNA forms a stable cation-π interaction with the R72 residue of

the orange SIDT1 protein, and the U-53 base of the pre-miRNA forms stable hydrogen bonds with the S111 and S113 residues of the protein. Subsequently, the pre-miRNA passes through the SIDT1 tetramer channel undergoing three phases: induction, channel-opening, and transmembrane (Fig. 5f).

In the induction phase, the orange and yellow SIDT1 protein N-terminal guides the pre-miRNA to the SIDT1 tetramer channel (Fig. 5f). The pre-miRNA gradually moves from 128 Å to 116 Å from the channel center, overcoming an energy barrier of 10.04 kcal/mol and stabilizes at 111 Å, releasing 4.44 kcal/mol of thermodynamic energy.

In the channel-opening phase, the yellow SIDT1 protein N-terminal shifts to open the tetramer channel (Fig. 5f). After overcoming a 13.30 kcal/mol energy barrier and releasing 7.94 kcal/mol of thermodynamic energy, it moves away from the other three proteins. Here, interactions between the SIDT1 tetramer and pre-miRNA remain relatively stable. As the tetramer channel opens, the pre-miRNA moves from 111 Å to stabilize at 91 Å from the channel center.

In the transmembrane stage, the pre-miRNA passes through the SIDT1 tetramer channel phospholipid bilayer (Fig. 5f). Simultaneously, the yellow SIDT1 protein N-terminal returns to its original state, closing the channel. The pre-miRNA overcomes an energy barrier of 7.56 kcal/mol and releases 20.62 kcal/mol of thermodynamic energy. The pre-miRNA moves from 91 Å to 40 Å and -57 Å from the channel center. At 40 Å, the four proteins in the SIDT1 tetramer directly interact with the pre-miRNA. The pre-miRNA stabilizes at the -57 Å position and is completely separated from and ceases to interact with the SIDT1 tetramer channel.

The PMFS analysis indicates that the transport process of the pre-miRNA through the SIDT1 tetramer channel requires overcoming an energy barrier of approximately 18.90 kcal/mol while releasing 2.10 kcal/mol of thermodynamic energy, indicating that the process is feasible both kinetically and thermodynamically. Additionally, throughout the transport process, the effective resistance coefficient of the pre-miRNA is 6.83×10^{7} s/cm, corresponding to an effective permeation rate of 1.46×10^{-8} cm/s. In summary, this work provides insights for the potential application of RNA interference (RNAi) technology in stony corals.

Discussion

During the single-cell sampling process, *M. foliosa*, *M. capricornis*, and *P. verrucosa* secreted large amounts of collagen, prolonging the duration of cell digestion (Methods), resulting in sequencing outcomes that are less sensitive than *A. muricata* and *S. pistillata*[5]. Due to the intertwined nature of the polyps and their shared coenosarc tissue[7], molecular morphological experiments, like in situ hybridization (ISH) conducted in *S. pistillata*[5] and sea anemone[7,30], were not feasible in these stony corals. The assay for transposase-accessible chromatin using sequencing (ATAC-seq) was not successful in the four studied stony corals.

Our open workflow supports continuously upgrading the PacBio sequencing platform but not step reduction. During stony coral genome assembly, we attempted simplified assembly methods, but these did not yield the desired results. Also unsuccessful was replacing the 10X Genomics linked-reads technology with Oxford Nanopore Technologies (ONT); however, as sequencing technologies advance, integrating HiFi and Ultra Long ONT technology into the workflow could facilitate T2T level chromosome assembly and gap-free results. We hope that the workflow presented in this paper, particularly the method for removing exogenous sequence contaminations from algae, can aid genome assembly in other stony coral genome projects.

Scleractinia stony corals are uniquely capable of achieving negative carbon emissions in the ocean, which is crucial to slow climate change. Additionally, the ocean's carbon pump is slowing more quickly than anticipated, indicating that a threat once thought distant has arrived[43]. Therefore, the ecological protection and restoration of coral reefs is important and urgent. With increased understanding of stony coral multi-omics, it will be beneficial to comprehensively utilize gene engineering technologies[12] and marine slow-release techniques to facilitate the ecological protection of coral reefs.

Declarations

Data availability

We uploaded the genome sequencing data of Illumina reads, PacBio reads, 10X Genomics linked-reads, and Hi-C reads, contaminant sequence dataset of endosymbiotic Symbiodiniaceae and symbiotic algae sequenced by Illumina HiSeq X Ten System, transcriptomes sequenced by Illumina HiSeq X Ten System, full-length transcriptomes sequenced by PacBio Sequel II sequencing platform, small RNA transcriptomes sequenced by Illumina HiSeq X Ten System, and BD Rhapsody matrix of scRNA-seq sequenced by Illumina NovaSeq 6000 System to the NCBI (BioProject: PRJNA544778) (Supplementary Table 1). This Whole Genome Shotgun (WGS) project has been deposited at DDBJ/ENA/GenBank under the accessions JAJTNX000000000 (A. muricata), JAJTOP000000000 (M. foliosa), JAJTOQ000000000 (M. capricornis), and JAJSRX000000000 (P. verrucosa), with the corresponding GenBank accessions of GCA_036669905.1 (NCBI reference genome), GCA_036669925.1 (NCBI reference genome), GCA_036669935.1 (NCBI reference genome), and GCA_036669915.1, respectively (Supplementary Table 1). The genome files of cds.fa, pep.fa, longest.filter.gff3.gz, TE_all_without_trf.gff, genome.fa.miRNA.flt.gff3, genome.fa.snRNA.flt.gff3, genome.rRNA.gff3, genome.TRF.gff, and genome.tRNA.gff3 have been uploaded to https://doi.org/10.6084/m9.figshare.19168286.v1 (Supplementary Table 1). The CSV file of BD Rhapsody matrix of scRNA-seq have been uploaded to https://doi.org/10.6084/m9.figshare.25202630.v1 (Supplementary Table 1).

Acknowledgements

We thank Prof. Min Zhu, Jing Lu, and Zhikun Gai (Institute of Vertebrate Paleontology and Paleoanthropology, Chinese Academy of Sciences) for the instructions on phylogenetic analysis; Prof.

Douglas H. Erwin (the Smithsonian National Museum of Natural History and the Santa Fe Institute), and Zongjun Yin (Nanjing Institute of Geology and Palaeontology, Chinese Academy of Sciences) for the instructions on stony coral and metazoan evolution; Na Lu, Yunqing Liu, and Zhuojun Guo (Southeast University) for bioinformatics analysis; Zheng Liu (Southeast University) for stony coral cultivation and aquarium maintenance; All the members of the National Center for Biotechnology Information (NCBI) for their support in data submission, analysis, and storage. This work was supported by the Special Funding of Southeast University, grant number 6907038067, 6907038068, and 6907038148, and the Big Data Computing Center of Southeast University.

Author contributions

C.H., Z. L. X. L. and J. C. conceived and supervised the project. C.H. and T. H. designed the experiments. C.H., T. H., W. H. and B. W. performed the experiments. T. H., W. H. and B. W. analyzed the data and drew the Fig.s. C.H., T. H. and W. H. interpreted the data and wrote the manuscript.

Competing interests

The authors declare no competing interests.

References

1. Hughes, T. P. et al. Global warming and recurrent mass bleaching of corals. *Nature* **543**, 373–377 (2017).
2. Cowen, L. J. & Putnam, H. M. Bioinformatics of corals: Investigating heterogeneous omics data from coral holobionts for insight into reef health and resilience. *Annu. Rev. Biomed. Data Sci.* **5**, 205–231 (2022).
3. Shinzato, C. et al. Using the *Acropora digitifera* genome to understand coral responses to environmental change. *Nature* **476**, 320–323 (2011).
4. Fuller, Z. L. et al. Population genetics of the coral *Acropora millepora*: Toward genomic prediction of bleaching. *Science* **369**, eaba4674 (2020).
5. Levy, S. et al. A stony coral cell atlas illuminates the molecular and cellular basis of coral symbiosis, calcification, and immunity. *Cell* **184**, 2973-2987.e18 (2021).
6. Putnam, N. H. et al. Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science* **317**, 86–94 (2007).
7. Foster, B. et al. A novel in vivo system to study coral biomineralization in the starlet sea anemone (*Nematostella vectensis*). *bioRxiv* 2023.10.04.560932 (2023).
8. Winston, W. M., Molodowitch, C. & Hunter, C. P. Systemic RNAi in C. elegans requires the putative transmembrane protein SID-1. *Science* **295**, 2456–2459 (2002).
9. Cao, J. et al. Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science* **357**, 661–667 (2017).

10. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
11. Li, Y. & Hui, J. H. L. Small RNAs in Cnidaria: A review. *Evol. Appl.* **16**, 354-364 (2022).
12. Mali, P. et al. RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).
13. Cleves, P. A., Strader, M. E., Bay, L. K., Pringle, J. R. & Matz, M. V. CRISPR/Cas9-mediated genome editing in a reef-building coral. *Proc. Natl. Acad. Sci. U S A.* **115**, 5235-5240 (2018).
14. Yu, Y. et al. Genome of elegance coral *Catalaphyllia jardinei* (Euphylliidae). *Front. Mar. Sci.* **9**, 991391 (2022).
15. Chen, B., et al. Latitudinal variation in the molecular diversity and community composition of Symbiodiniaceae in coral from the South China Sea. *Front. Microbiol.* **10**, 1278 (2019).
16. Thompson, J. R., Rivera, H. E., Closek, C. J. & Medina, M. Microbes in the coral holobiont: partners through evolution, development, and ecological interactions. *Front. Cell Infect. Microbiol.* **4**, 176 (2015).
17. Hall, M. R. et al. The crown-of-thorns starfish genome as a guide for biocontrol of this coral reef pest. *Nature* **544**, 231–234 (2017).
18. Kenyon, J. C. Models of reticulate evolution in the coral genus *Acropora* based on chromosome numbers: Parallels with plants. *Evolution* **51**, 756–767 (1997).
19. Ying, H. et al. Comparative genomics reveals the distinct evolutionary trajectories of the robust and complex coral lineages. *Genome Biol.* **19**, 175 (2018).
20. Steinworth, B. M., Martindale, M. Q. & Ryan, J. F. Gene loss may have shaped the cnidarian and bilaterian Hox and ParaHox complement. *Genome Biol. Evol.* **15**, evac172 (2023).
21. https://www.ncbi.nlm.nih.gov/datasets/genome/?taxon=6125
22. Liew, Y. J., Aranda, M. & Voolstra, C. R. Reefgenomics.org—a repository for marine genomics data. *Database* **2016**, baw152 (2016).
23. Chen, J. Y. et al. Precambrian animal life: probable developmental and adult cnidarian forms from Southwest China. *Dev. Biol.* **248**, 182-196 (2002).
24. Conway Morris, S. The Cambrian "explosion": slow-fuse or megatonnage? *Proc. Natl. Acad. Sci. U S A.* **97**, 4426-4429 (2000).
25. Raup, D. M. & Sepkoski, J. J. Jr. Mass extinctions in the marine fossil record. *Science* **215**, 1501–1503 (1982)
26. Alroy, J. Dynamics of origination and extinction in the marine fossil record. *Proc. Natl Acad. Sci. USA* **105** (Suppl. 1), 11536–11542 (2008)
27. Stanley, G. D. Jr & Fautin, D. G. Paleontology and evolution. The origins of modern corals. *Science* **291**, 1913–1914 (2001).
28. Barnosky, A. D. et al. Has the Earth's sixth mass extinction already arrived? *Nature* **471**, 51–57 (2011).

29. The UniProt Consortium. UniProt: the universal protein knowledgebase in 2023. *Nucleic Acids Res*. **51**, D523–D531 (2023).
30. Ranade, V. V. *Springer Handbook of Enzymes*, D. Schomburg & I. Schomburg (Eds.). (Springer Nature, 2009).
31. Wiedenmann, J. et al. Reef-building corals farm and feed on their photosynthetic symbionts. *Nature* **620**, 1018–1024 (2023).
32. Weis, V. M. Cell biology of coral symbiosis: Foundational study can inform solutions to the coral reef crisis. *Integr. Comp. Biol*. **59**, 845-855 (2019).
33. Sebe-Pedros, A. et al. Cnidarian cell type diversity and regulation revealed by whole-organism single-cell RNA-seq. *Cell* **173**, 1520–1534 (2018).
34. He, C. et al. Phagocytic intracellular digestion in amphioxus (*Branchiostoma*). *Proc. Biol. Sci*. **285**, 20180438 (2018).
35. Mass, T. et al. Cloning and characterization of four novel coral acid-rich proteins that precipitate carbonates in vitro. *Curr. Biol*. **23**, 1126–1131 (2013).
36. Von Euw, S. et al. Biological control of aragonite formation in stony corals. *Science* **356**, 933–938 (2017).
37. Drake, J. L. et al. Proteomic analysis of skeletal organic matrix from the stony coral *Stylophora pistillata. Proc. Natl. Acad. Sci. USA* **110**, 3788–3793 (2013).
38. Drake, J.L., Whitelegge, J.P. & Jacobs, D.K. First sequencing of ancient coral skeletal proteins. *Sci. Rep*. **10**, 19407 (2020).
39. Levy, S. & Mass, T. The skeleton and biomineralization mechanism as part of the innate immune system of stony corals. *Front. Immunol*. **25**, 850338 (2022).
40. Ho, D. T., & Schanze, J. J. Precipitation-induced reduction in surface ocean pCO2: Observations from the eastern tropical Pacific Ocean. *Geophys. Res. Lett*. **47**, e2020GL088252 (2020).
41. Stolarski, J. et al. A modern scleractinian coral with a two-component calcite-aragonite skeleton. *Proc. Natl. Acad. Sci. U S A*.**118**, e2013316117 (2021).
42. Pratt, A. J., Rambo, R. P., Lau, P. W., & MacRae, I. J. Preparation and characterization of the extracellular domain of human Sid-1. *PloS one* **7**, e33607 (2012).
43. Gunn, K.L. et al. Recent reduced abyssal overturning and ventilation in the Australian Antarctic Basin. *Nat. Clim. Chang*. **13**, 537–544 (2023).

Methods

Sample collection and culture

Coral colonies of *A. muricata*, *M. foliosa*, and *P. verrucosa* were collected from the Xisha Islands (15°40'–17°10' N, 111°–113° E) at water depths from 5 to 10 m. The colonies were transferred to laboratory

aquariums for acclimation and culture. *M. capricornis* colonies were obtained from Haiyi Aquatic Breeding Co., Ltd. All coral branches sampled were from the same colony.

DNA isolation and sequencing libraries

Genomic DNA was extracted from coral branches using the DNeasy Blood & Tissue Kit (Qiagen, Cat. No. / ID: 69504) following the manufacturer's instructions. The DNA quantity and quality were assessed using Qubit, Nanodrop, and Femto Pulse machines. Four distinct sequencing technologies were employed to acquire the genome sequences. These included paired-end sequencing libraries with a 350 bp insert size, 10X Genomics linked-read libraries with a 600 bp insert size, Hi-C libraries with a 350 bp insert size, and single-molecule real-time sequencing (SMRT) bell libraries with a 30 kb insert size. The libraries were constructed following protocols specific to each sequencing strategy and subsequently sequenced on the Illumina HiSeq X Ten platform (for paired-end, 10X Genomics, and Hi-C libraries) and the PacBio Sequel platform (for PacBio libraries).

Genome sequencing and assembly workflow

The workflow for stony coral genome sequencing and assembly consisted of nine steps: 1) genome survey, 2) V1 Genome—genome assembly was performed using Canu software (v. 2.2) with sequencing data from the PacBio Sequel platform, 3) V2 Genome—sequence errors were corrected using NextPolish (v. 1.3.1) software, based on sequencing data from the Illumina HiSeq X Ten System and the PacBio Sequel platform, 4) V3 Genome—heterozygous contigs were removed using Purge Haplotigs software (v. 1.0.2), 5) V4 Genome—exogenous sequence contaminations of metagenomic data were removed using BlobTools software (v. 1.0), with reference to the NCBI Nucleotide (Nt) database, 6) V5 Genome—exogenous sequence contaminations of endosymbiotic Symbiodiniaceae and symbiotic algae were removed using BLAST software (v. 2.2.26), guided by relevant sections of BioProject PRJNA544778 (Supplementary Table 1), 7) V6 Genome—contigs of V5 genome were linked into scaffolds using fragScaff software (v. 140324.1), guided by 10X Genomics linked-read sequencing data, 8) V7 Genome—scaffolds of V6 genome were anchored into a pseudochromosome-level genome using ALLHiC software (v. 0.9.13), with guidance from Hi-C sequencing data, and 9) V8 Genome—gaps in the V7 genome were closed using TGS-GapCloser software (v. 1.2.1), utilizing sequencing data from the PacBio Sequel platform.

Genome survey

The genome sizes were assessed by analyzing the K-mer spectrum derived from the paired-end Illumina sequencing data. The frequency of each K-mer (k = 17) was determined using Jellyfish (v. 2.1.3). K-mer analysis (K = 17) was employed to estimate the genome size and heterozygote frequency of the four studied stony corals. GC-content separation analysis was utilized to identify any exogenous sequence contamination.

Symbiotic algae isolation

The Symbiodiniaceae and other symbiotic algae were isolated and enriched from stony coral tissues using stepwise filtration and density gradient centrifugation in sterile filtered seawater (FSW). After 24 hours of antibiotic treatment (penicillin, streptomycin sulfate), the cells were cultured in 96-well plates and 24-well plates using modified L1 medium. Single clones of algae were obtained through single cell isolation and culture. The species were identified using internal transcribed spacer 2 (ITS2) sequences. All cultured Symbiodiniaceae were mixed to meet the sample quantity requirements for next-generation sequencing, which was used for subsequent decontamination of the stony coral genomes.

Building contaminant sequence dataset

Since stony corals from specific ecological niches have unique types of endosymbiotic Symbiodiniaceae and symbiotic algae[16], solely relying on the NCBI NT database for removing their sequence contaminations is not sufficient. To address this, we constructed a dataset by isolating the endosymbiotic Symbiodiniaceae and symbiotic algae from the four studied stony corals and sequencing their genomic DNA using the Illumina HiSeq X Ten System. This dataset comprised a subset that included a mix of Symbiodiniaceae from all four stony corals, nine subsets of symbiotic algae, and one subset of crustose coralline algae (Supplementary Table 1). Relevant removal procedures were subsequently carried out, guided by this dataset.

Illumina and PacBio sequencing

The raw sequencing data generated by the Illumina HiSeq X Ten platform, which included adapters, more than 10% of unknown nucleotides (N), and 50% of low-quality bases (Q-value <= 10), were trimmed to obtain high-quality reads. The subreads of the PacBio sequencing data were filtered using default parameters.

10X Genomics sequencing

The Chromium library was prepared following the protocols provided by 10X Genomics, using the Genome Reagent Kit v2 (10x Genomics, San Francisco, California, United States). Sample quantity and quality were assessed using Qubit, Nanodrop, and Femto Pulse machines. For each library, approximately 10 ng of high molecular weight (HMW) genomic DNA (with a mean fragment length greater than 65 kb) was utilized. In the microfluidic Genome Chip, a library of Genome Gel Beads was combined with HMW template gDNA, master mix, and partitioning oil to generate Gel Bead-In-Emulsions (GEMs) within the Chromium apparatus. Each Gel Bead was then functionalized with millions of copies of a 10x™ barcoded primer. The Genome Gel Bead dissolution inside the GEM released primers containing (i) an Illumina R1 sequence (Read 1 sequencing primer), (ii) a 16 bp 10x Barcode, and (iii) a 6 bp random primer sequence. The R1 sequence and the 10x™ barcode were added to the molecules during the GEM incubation, while P5 and P7 primers, R2 sequence, and Sample Index were added during library construction. The library was then subjected to 10 cycles of PCR amplification. In total, four 10X Genomics linked-read libraries were constructed and sequenced using an Illumina HiSeq X Ten platform.

Hi-C sequencing

To perform Hi-C sequencing on stony coral branch tissues, we adjusted the concentration of PBS buffers to 0.035 M in order to match the osmotic pressure of the seawater in which stony corals reside. One Dovetail Genomics Hi-C library was prepared using the whole body of an individual stony coral branch that had been treated with 75% alcohol. The procedure involved fixing the chromatin in the nucleus with a 2% formaldehyde solution in MS buffer (10 mM potassium phosphate, pH 7.0; 50 mM NaCl; 0.1M sucrose) at room temperature for 30 minutes under vacuum, followed by extraction. The fixed chromatin was then digested with 400 U of HindIII restriction enzyme (NEB) at 37 °C for 16 hours. DNA ends were labeled with biotin and incubated at 37 °C for 45 minutes, with the enzyme subsequently inactivated using a 20% SDS solution. DNA ligation was carried out by adding T4 DNA ligase (NEB) and incubating at 16 °C for 4-6 hours. After ligation, proteinase K was added to reverse cross-linking during an overnight incubation at 65 °C. The DNA fragments were purified and dissolved in 86 μL of water. Purified DNA was treated to remove biotin that was not internal to ligated fragments and then sheared to an average fragment size of approximately 350 bp. DNA fragments labeled with biotin were finally separated using Dynabeads® M-280 Streptavidin (Life Technologies). The Hi-C libraries were quality-controlled and sequenced on an Illumina HiSeq X Ten platform.

Total RNA isolation

The coral branch samples were ground manually into fine powder using a frozen mortar and pestle in liquid nitrogen. TRIzol® LS Reagent (Thermo Fisher Scientific, 10296028, Waltham, MA, USA) was added to the ground samples at a ratio of approximately 1:3 (sample to reagent), and the mixture was further homogenized. After centrifugation, the supernatant was collected and combined with BCP reagent (Molecular Research Center, BP 151, Cincinnati, OH, USA). The resulting solution was then centrifuged, and the supernatant was mixed with Isopropanol (Amresco, 0918-500ML, Radnor, PA, USA) and allowed to stand overnight at -20 °C. After another round of centrifugation, the supernatant was discarded, and the RNA pellet was washed with 75% Ice Ethyl alcohol, Pure (Sigma-Aldrich, E7023-500ML, Taufkirchen, München, Germany). The RNA samples were then ready for transcriptome sequencing.

RNA-seq

The PacBio Sequel platform was used for full-length transcriptome sequencing, while the Illumina HiSeq X Ten platform was employed for short-read transcriptome and small RNA (miRNA) sequencing.

BUSCO assessment

The BUSCO (v. 5) assessment in this study was conducted according to the methodology described in a previous study[44].

Gene annotation

The basic genome annotation workflow in this study consists of three pipelines: repetitive sequence annotation, gene annotation (structure and function annotations), and non-coding RNA (ncRNA) annotation (Supplementary Tables 1, 2). To enhance annotation quality, both full-length and short-read transcriptomic data were employed (Supplementary Tables 3-5). After ensuring that the full-length transcriptome data had reached saturation (Supplementary Table 3), we used full-length transcriptomic data, short-read transcriptomic data, and our reference genomes to conduct a quantitative analysis on the gene structures of transcriptomic isoforms, alternative splicing (AS) genes, fusion genes, and LncRNAs (Supplementary Tables 3, 4). Additionally, employing short-read transcriptomic data and our reference genomes, we performed a quantitative analysis of miRNAs (Supplementary Table 5).

Chromosome macrosynteny analysis

Focusing on the characteristics of orthologous coding genes at chromosome loci and protein sequences, we employed the General Feature Format version 3 (GFF3) files (longest.filter.gff3.gz and pep.fa) (Supplementary Table 1) to conduct an integrated visual chromosome synteny analysis across genomes, utilizing Multiple Collinear Scanning toolkits (mcscanx)[45] based on the protein sequences derived from the investigated genomic data[6, 49, 50].

Homeobox gene analysis

Homeodomain domain sequences of each Hox protein subclass were obtained from the homeobox gene database (HomeoDB2) (Supplementary Table 6)[46]. The homeodomain domain sequences of each Hox protein subfamily were aligned using MAFFT software (v. 7.158) with default parameters, forming a multiple sequence alignment for each subfamily. Based on the multiple sequence alignment of each Hox protein subfamily, a probabilistic model (profile HMM) for each subfamily was constructed using the hmmbuild software (v. 3.2.1) with default parameters, to query sequence databases and find homologous sequences. Pfam domains of all protein sequences in each species were predicted using the PfamScan (v. 1.6) software with default parameters and the Pfam database (v. 32). Protein sequences with a predicted Homeodomain (PF00046.29) domain were processed in a manner similar to the HMM training set using the hmmsearch software (v. 3.2.1). The resulting hits for each Hox subfamily were ranked according to E-value with a cut-off of 1e-10, and the protein sequence ranked first was identified as the corresponding protein of the respective Hox subfamily (Supplementary Table 6).

Comparative genomics analysis

Genomes and annotation files of 20 cnidarians and one *A. queenslandica* were downloaded from the NCBI database, excluding the four genomes assembled in this study. For genes with multiple alternative isoforms, only the longest transcript was chosen to represent the gene in each species. The proteins of the 25 species were aligned using an all-against-all search algorithm with an E-value threshold of 1e-7. The orthomcl markov clustering program was then utilized with the parameter "-inflation 1.5" to cluster gene families. Few single-copy genes were identified, limiting the species tree reconstruction. Orthologous genes from the 25 species were assigned to address this. Gene families with at least one gene per

species were collected based on the blastp results obtained from the gene family clustering analysis. For gene families with multiple copies, the member with the best blast match to the genes of *A. queenslandica* was selected as the orthologue for analysis, with the longest orthologue in *A. queenslandica* designated as the reference gene. In total, 1,554 orthologous genes were gathered for subsequent phylogenetic analysis. For specific parameters, see Supplementary Table 7.

To analyze the expansion (Set4) and contraction (Set5) of orthologous gene families (Supplementary Table 7), the CAFÉ tool (v. 4) was employed. This tool utilized a stochastic birth and death model with variable lambda values to study changes in gene families along each lineage in the phylogenetic tree. A probabilistic graphical model was introduced to calculate the probability of transitions in gene family size from parent to child nodes. The corresponding P-values were calculated for each lineage based on the conditional likelihood. A threshold of 0.05 was used to identify significantly expanded or contracted gene families.

To identify genes undergoing positive selection (Set6) along the Scleractinia stony coral lineage (foreground branch), a comparison was made with other species from the phylum Cnidaria (background branch). Multiple sequence alignment of 2,092 single-copy genes was conducted using MUSCLE (v. 3.8.31). The branch-site model in the codeml program, which is part of the PAML package (v. 4.7), was employed to detect signals of positive selection.

Pan-genomic analysis

A pan-genomic map was created for the selected stony corals, including core genes, dispensable genes, and private genes. Core genes are present in all individuals (Set1-Set3, Supplementary Table 7), dispensable genes are present in some individuals (at least two), and private genes are found in a single individual. The genes were clustered using OrthoMCL software (v. 1.4) to identify single-copy gene families, multi-copy gene families, and species-specific gene families.

Insertion times of TEs.

To understand the temporal dynamics of TE activities during the evolution of stony corals, divergences of TEs from the consensus sequences extracted from RepeatMasker (v. 4.1.5) results were adjusted for multiple substitutions using the Jukes-Cantor formula $K = -300/ 4\times \mathrm{Ln}(1 - D \times 4/300)$, where D represents the distance between the fragmented repeat and the consensus sequence. Insertion times of TEs were estimated using the equation $T = K/2r$, where T is the insertion time, and r is the nucleotide substitution rate for cnidaria, 4.0e-9 referred to in previous study[4].

PMSC analysis

The PMSC analysis in this study was conducted according to the methodology described in a previous study[4].

scRNA-seq

To prepare the single-cell suspension, a 0.035 M PBS with a pH of 8.0 was created. The samples were washed three times with this buffered solution. A digestion solution was prepared by diluting 1‰ Tyrisin (Sigma-Aldrich, T1426-500MG, St. Louis, MO, USA) in the 0.035 M PBS. The branches of *A. muricata*, *P. verrucosa*, *M. capricornis*, and *M. foliosa* underwent digestion for 15 minutes, 45 minutes, 2 hour, and 3 hours, respectively. Next, the solution was filtered using 40 μm FisherbrandTM Sterile Cell Strainers (Thermo Fisher Scientific, 22-363-547, Waltham, MA, USA) and centrifuged at 1,000 rpm for 5-6 minutes. The resulting supernatant was collected, and the digestion was stopped by adding BSA (Sangon Biotech, A600332-0005, Shanghai, China) to a final concentration of 0.4 mg/ml. Living cell counting and assessment were performed using an automated cell counter (TC20TM, Bio-Rad Laboratories, Inc., Hercules, CA, USA). After washing and resuspension, the qualified cells were prepared into a single-cell suspension at an appropriate concentration for the library construction of the BD Rhapsody™ Single-Cell Analysis System. Sequencing was then performed on the Illumina NovaSeq 6000 System.

Single-cell transcriptome analysis

The Umi-tools pipeline (https://umi-tools.readthedocs.io/en/latest/index.html) was used to analyze the raw data with default parameters. The filtered_gene_bc_matrices output from umi-tools was loaded into the Seurat package (v. 3.0). To remove dead and low-quality cells, genes expressed in more than three single cells were kept, and each cell was required to have at least 200 expressed genes. Additionally, DoubletFinder was used to identify and remove doublets. After filtering, the remaining cells were used for downstream analysis (Supplementary Table 9).

The scaled data were first normalized using 'LogNormalize'. The 'FindVariableFeatures' function with mean.cutoff (0.0125-5) and dispersion.cutoff (1, Inf) parameters was used to identify highly variable genes for clustering analysis. For principal component analysis (PCA), the scaled data were reduced to 50 approximate principal components (PCs) based on the highly variable genes. The number of PCs used for t-SNE analysis was determined using the Elbowplot function. Clusters were identified using the 'FindClusters' function with selected PCs and a resolution of 0.6. Cluster-enriched genes were identified using the 'FindAllMarkers' function with a 'wilcox' test, using parameters of 'min.pct = 0.25' and 'logfc.threshold = 0.25' to define enriched genes. The single-cell data structures and trajectories were visualized and explored using t-SNE (using the 'RunTSNE' function) and UMAP (using the 'RunUMAP' function with parameters 'metric = correlation', 'min.dist = 0.3', and 'umap.method = umap-learn').

The Seurat (v. 3.0) was utilized for data normalization, dimensionality reduction, clustering, and differential expression analysis. Specifically, the Seurat alignment method, canonical correlation analysis (CCA), was employed for integrated analysis of datasets. To perform clustering, highly variable genes were selected and principal components based on these genes were used to construct a graph, which was then segmented with a resolution of 0.6.

To determine the cell clusters of *A. muricata*, the marker genes of each cell type[5,30] and cluster-enriched genes were utilized (Supplementary Table 8). The Wilcox algorithm was employed to calculate statistical

significance. Cluster-enriched genes meeting the following criteria were considered signature genes: (1) adjusted p-value < 0.05 after Bonferroni correction using all features in the dataset, (2) log fold-change of the average expression > 2.0, and (3) pct.1 > 0.5 (pct.1 refers to the percentage of cells where the feature is detected in the first group) (Supplementary Table 9).

For the cell cluster determination of *M. foliosa*, *M. capricornis*, and *P. verrucosa*, supervised clustering was performed on the single-cell sequencing data based on marker genes (Supplementary Table 8)[5,30]. The R package scSorter v0.0.2 (category) was used for initial clustering[47], followed by further clustering correction and dimensionality reduction using the R package ssingleCellNet v0.1.0 (group) to generate visualization (Supplementary Table 10)[48].

SIDT1 analysis

The structures of the amu-miR-100 precursor and SIDT1 protein tetramer were constructed using AlphaFold2[10]. For details on the gene analysis, preparation of simulation systems, molecular dynamics simulations, pulling molecular dynamics simulations, potential of mean force simulations, and calculations of relative permeability coefficients, please refer to Supplementary Table 8.

References:

44. Waterhouse, R. M. et al. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol. Biol. Evol.* **35**, 543–548 (2018).

45. Wang, Y. et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **40**, e49 (2012).

46. Zhong, Y.-F. & Holland, P. HomeoDB2: functional expansion of a comparative homeobox gene database for evolutionary developmental biology. *Evol. Dev.* **13**, 567–568 (2011).

47. Guo, H. & Li, J. scSorter: assigning cells to known cell types according to marker genes. *Genome Biol.* **22**, 69 (2021).

48. Tan, Y. & Cahan, P. SingleCellNet: A computational tool to classify single cell RNA-seq data across platforms and across species. *Cell Syst.* **9**, 207-213.e2. (2019).

49. Chapman, J. A. et al. The dynamic genome of *Hydra*. *Nature* **464**, 592–596 (2010).

50. Li, Y. et al. Chromosome-level reference genome of the jellyfish *Rhopilema esculentum*. *Gigascience* **9**, giaa036 (2020).

Supplementary Tables

Supplementary Tables are not available with this version.

Figures

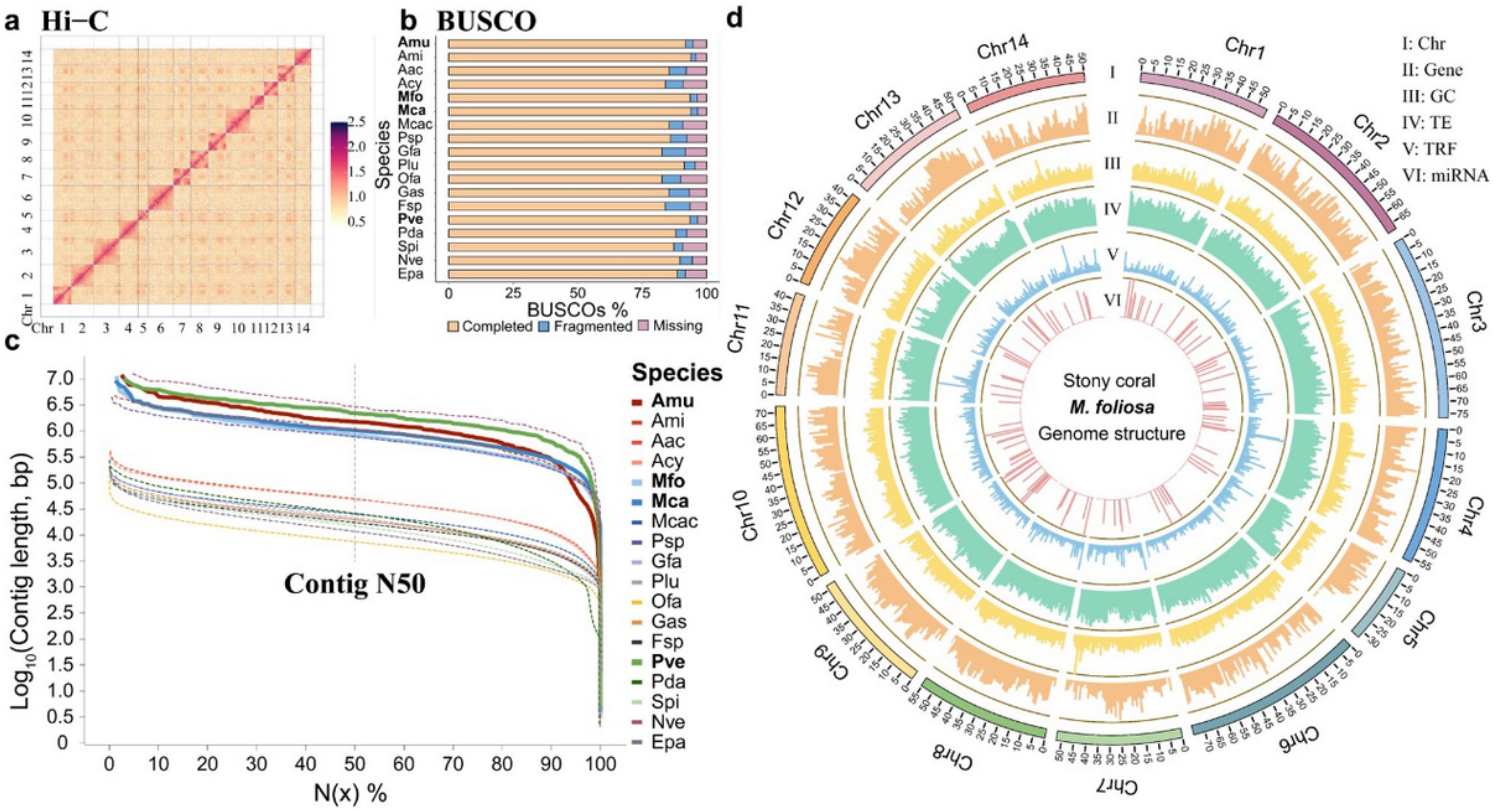

Figure 1

Genome assembly, assessment and annotation. **A.** *M. foliosa* genome Hi-C map. **b**. BUSCO comparison of stony coral genomes. Abbreviations listed in Extended Data Table 2. Bolded abbreviations denote genomes assembled in this study. **c**. Comparison of assembly contiguity with the x-axis as N(x) %, assembly percentage, and the y-axis is contig size. Bolded abbreviations and wider lines denote genomes assembled in this study. **d**. *M. foliosa* genome structure. Tracks from outer to inner (I-VI): chromosomes at 1Mb scales (Chr), gene density (Gene), GC density (GC), transposable element density (TE), tandem repeat sequence (TRF) density, and miRNA density.

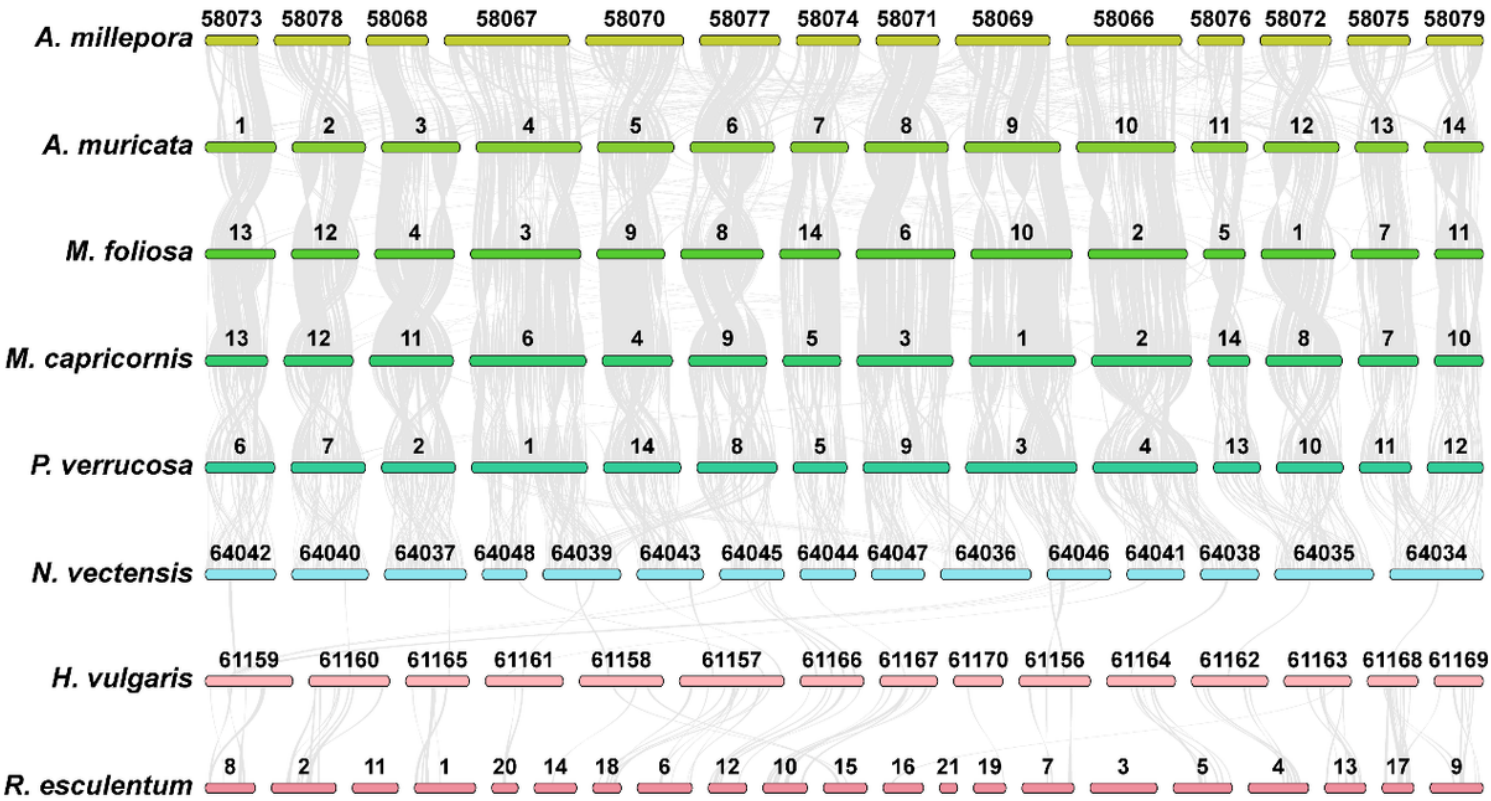

Figure 2

Chromosome macrosynteny analysis among stony corals and other cnidarians. The genomes in this study, along with the genome of *A. millepora* (assembled via linkage maps)[4], feature pseudochromosome configurations and exhibit contig N50 values that rise to the Mb scale (Extended Data Table 2), allowing for a global synteny analysis on orthologous genes using these five genomes in conjunction with the released chromosome-level genomes of other cnidarians, including sea anemone *Nematostella vectensis* (whose genome was sequenced using the Sanger sequencing method)[6], *Hydra vulgaris*[49], and true jellyfish *Rhopilema esculentum*[50] (Methods).

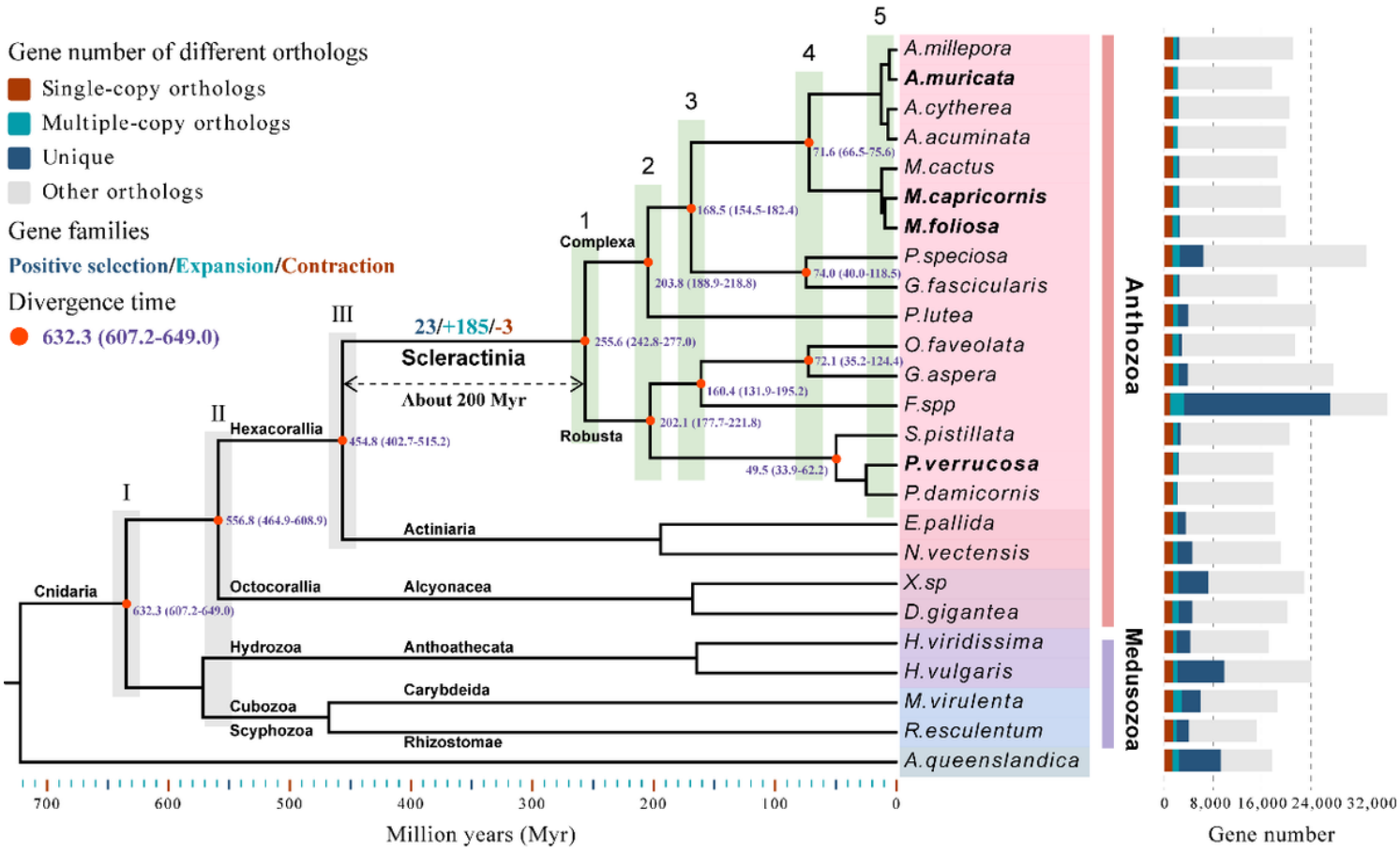

Figure 3

Phylogenetic analysis of Scleractinia stony corals. Timeline of major coral fossil records and developments from 650 mya to the present. I-III: Diverging times of Anthozoa, Hexacorallia, and Scleractinia. 1-5: Five distinct phases of phylogenetic diversity development in Scleractinia groups.

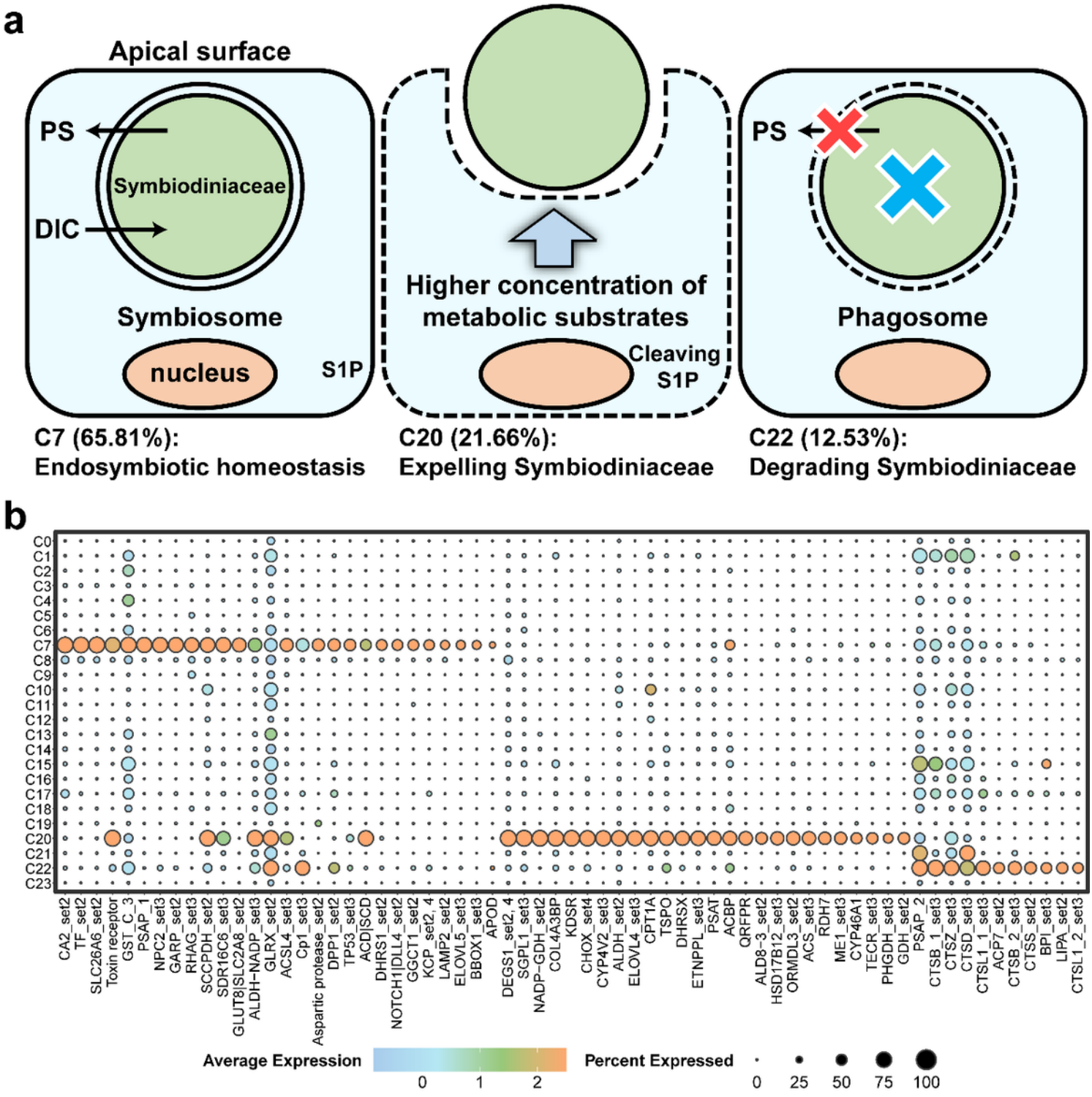

Figure 4

Alga-hosting cells and their three states. **a**. The physiological characteristics of alga-hosting cells in various states. **b**. Bubble plot depicting the gene expression patterns of functional genes in different states. Metabolic substrates during the Expelling Symbiodiniaceae state listed in Supplementary Table 8. PS: Photosynthesis. DIC: Dissolved inorganic carbon. C0-C23 are listed in Extended Data Fig. 7a.

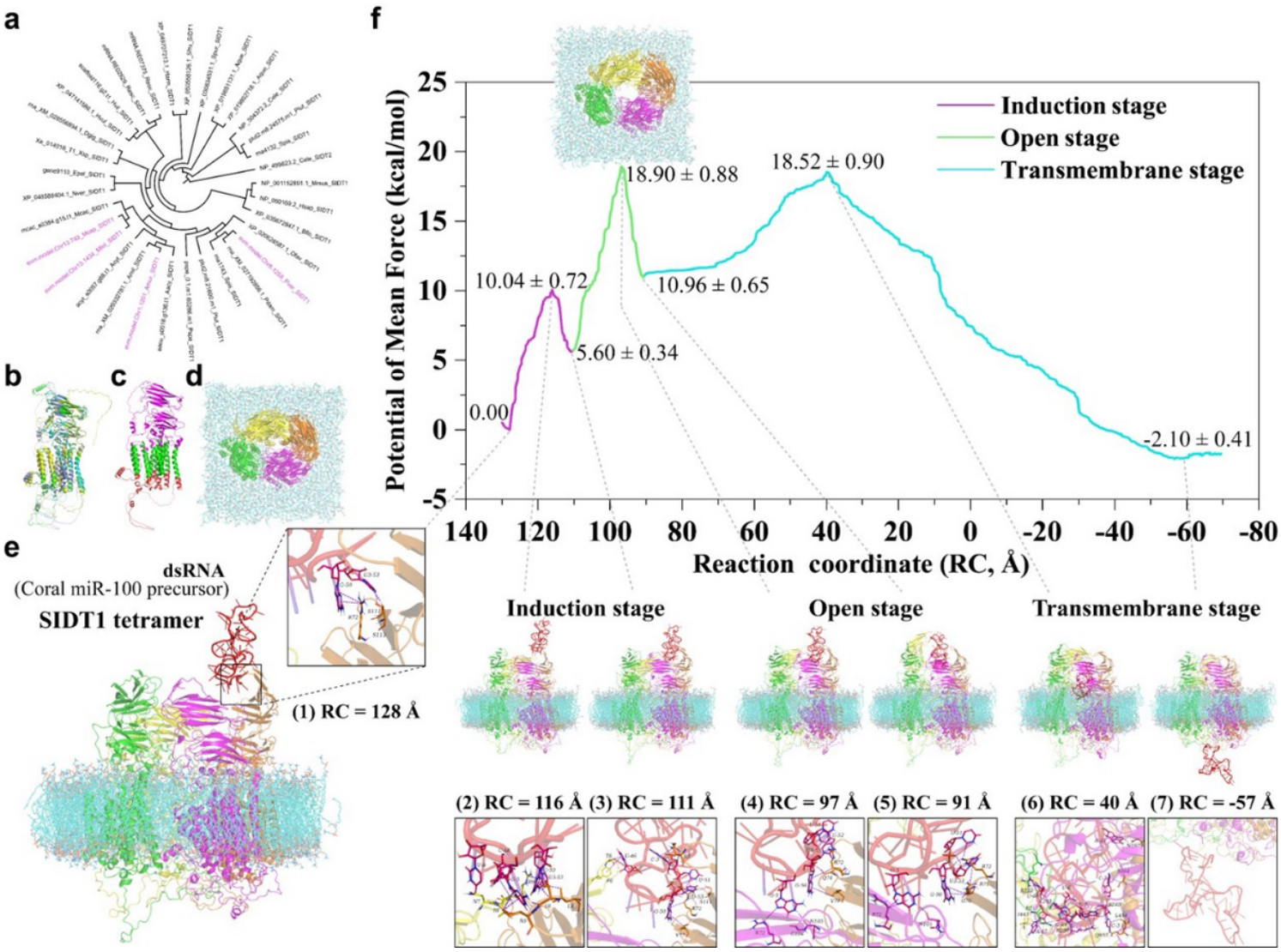

Figure 5

Scleractinia stony coral SIDT1 and its potential function in transporting miRNA precursors across cell membranes. **a**. Phylogenetic analysis of SIDT1 within metazoans visualized by the neighbor-joining tree. **b**. Results of the superposition of four predicted stony coral SIDT1 protein model structures compared with the Q9NXL6 (SIDT1_HUMAN) structure. *A. muricata* (white), *M. foliosa* (cyan), *M. capricornis* (green*), P. verrucosa* (slate), and Q9NXL6 (yellow).**c**. 3D structure of the *A. muricata* SIDT1 protein with each region differentiated: the extracellular (magenta), transmembrane (green), and intracellular (red). **d**. Cross-section analysis of the SIDT1 tetrameric structure showing that it remains in a closed state prior to contact with the pre-miRNA. **e**. Initial interaction between the pre-miRNA and the SIDT1 tetramer channel. **f**. The pre-miRNA passing the SIDT1 tetramer channel. Free energy was set to zero in the upper water layer. Free energy profile and system conformation change illustrate the translocation process of the miRNA precursor (red) from the upper water layer (Z=130 Å) of the phospholipid bilayer (cyan) to the lower water layer (Z=-70 Å) through the SIDT1 protein tetramer (yellow, green, magenta, and orange, respectively). This is the first dynamics simulation result regarding the SIDT1 tetramer transporting miRNA across cell membranes.

Supplementary Files

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

- 3ExtendedData.docx