EVOLUTIONARY BIOLOGY

The coral *Acropora loripes* genome reveals an alternative pathway for cysteine biosynthesis in animals

Octavio R. Salazar[1,2]*, Arun Prasanna N.[1,2], Guoxin Cui[1,2], Line K. Bay[3,4], Madeleine J. H. van Oppen[3,5], Nicole S. Webster[3,6,7], Manuel Aranda[1,2]*

The metabolic capabilities of animals have been derived from well-studied model organisms and are generally considered to be well understood. In animals, cysteine is an important amino acid thought to be exclusively synthesized through the transsulfuration pathway. Corals of the genus *Acropora* have lost cystathionine β-synthase, a key enzyme of the transsulfuration pathway, and it was proposed that *Acropora* relies on the symbiosis with dinoflagellates of the family Symbiodiniaceae for the acquisition of cysteine. Here, we identify the existence of an alternative pathway for cysteine biosynthesis in animals through the analysis of the genome of the coral *Acropora loripes*. We demonstrate that these coral proteins are functional and synthesize cysteine in vivo, exhibiting previously unrecognized metabolic capabilities of animals. This pathway is also present in most animals but absent in mammals, arthropods, and nematodes, precisely the groups where most of the animal model organisms belong to, highlighting the risks of generalizing findings from model organisms.

INTRODUCTION

Cysteine is one of two sulfur-containing amino acids, together with methionine, and is required for proper development, sulfur metabolism, and production of antioxidants. In animals, cysteine biosynthesis is considered to be exclusively carried out through the transsulfuration pathway (*1*), which relies on cystathionine β-synthase (CBS) and pyridoxal 5′-phosphate (PLP) (vitamin B₆) as a cofactor to catalyze the condensation of homocysteine and serine to produce cystathionine. Cystathionine is subsequently converted into L-cysteine by cystathionine γ-lyase. In humans, mutations in *CBS* lead to homocystinuria (*2*), a disease caused by the accumulation of homocysteine, resulting in disorders such as mental retardation, skeletal abnormalities, optic lens dislocation, and cardiovascular disease (*3*). CBS is not present in corals of the genus *Acropora* (*4*, *5*), and species within this genus might instead rely on their symbiosis with dinoflagellates of the family Symbiodiniaceae for the acquisition of cysteine (*4*). Cysteine is required for the production of glutathione, an important and crucial antioxidant for the response to reactive oxygen species generated by the symbiont under heat stress. It was, therefore, proposed that the reliance on symbiont-derived cysteine may contribute to the increased susceptibility of *Acropora* species to thermal bleaching compared to many other coral genera (*6*, *7*).

Plants and bacteria use a different pathway for cysteine biosynthesis in which cysteine is synthesized through the sulfate assimilation pathway, where L-serine is converted into *O*-acetyl-L-serine by a serine *O*-acetyltransferase and then converted into L-cysteine by an *O*-acetylserine sulfhydrylase (CysK/CysM) (*8*, *9*). Analysis of cysteine biosynthesis pathways in Kyoto Encyclopedia of Genes and Genomes (KEGG) (*10*, *11*) shows that metazoans lack the required enzymes for cysteine biosynthesis through the sulfate assimilation pathway. However, a recently discovered pathway for cysteine biosynthesis, not yet included in enzymatic reaction databases such as MetaCyc (*12*) or KEGG, has been identified in fungi (*13*), in which L-serine is converted into *O*-succinyl-L-serine by an L-serine-*O*-succinyltransferase (Cys2) and then into L-cysteine by *O*-succinylserine sulfhydrylase (Cys1a). Mutations in either *Cys2* or *Cys1a* lead to cysteine auxotrophy in *Schizosaccharomyces pombe* (*14–16*).

Here, we explored the genome of the coral *Acropora loripes* for alternative cysteine biosynthesis pathways in animals. We show that the loss of the *CBS* gene in *Acropora* corals was likely due to the activity of transposable elements. We identified the presence of an alternative cysteine biosynthesis pathway through the action of homologs of the proteins Cys2 and Cys1a, previously only described in fungi. Moreover, this alternative pathway was found throughout the animal kingdom except for vertebrates, arthropods, and nematodes and was also found in other eukaryotes except for higher plants. We further show that the Cys1a and Cys2 proteins of *A. loripes* can complement the *Cys1a*Δ and *Cys2*Δ mutants in *S. pombe*, demonstrating that the coral proteins can synthesize cysteine through this alternative pathway and revealing a previously unrecognized pathway for cysteine biosynthesis in animals.

RESULTS

CBS loss in the genus *Acropora*

To analyze the loss of CBS in *Acropora*, we first sequenced the genome of *A. loripes*, resulting in an assembly composing of 335 scaffolds and a BUSCO (Benchmarking Universal Single-Copy Orthologs) (*17*) score against the Metazoa dataset v.10 of: complete, 96.1%; fragmented, 1.7%; and missing, 2.2% (Table 1, figs. S1 to S4, tables S1 to S4, and data S1), being one of the most complete and least fragmented

[1]Marine Science Program, Biological and Environmental Sciences and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Kingdom of Saudi Arabia. [2]Red Sea Research Center, King Abdullah University of Science and Technology, Thuwal, Kingdom of Saudi Arabia. [3]Australian Institute of Marine Science, Townsville, Queensland, Australia. [4]AIMS@JCU, Division of Research and Innovation, James Cook University, Townsville, Australia. [5]School of BioSciences, University of Melbourne, Parkville, Victoria 3010, Australia. [6]Australian Centre for Ecogenomics, University of Queensland, St Lucia, Australia. [7]Australian Antarctic Division, Department of Agriculture, Water and the Environment, Kingston, Australia.

*Corresponding author. Email: octavio.salazarmoya@kaust.edu.sa (O.R.S.); manuel.aranda@kaust.edu.sa (M.A.)

Table 1. Genome assembly statistics of *A. loripes* compared to other sequenced *Acropora* species used in this study.

Parameter	*Acropora loripes* (this study)	*Acropora digitifera* (4)	*Acropora tenuis* (62)	*Acropora millepora* (64)
Assembly size of scaffolds (Mb)	401.87	447.49	486.81	386.60
Assembly size of contigs (Mb)	401.86	379.28	486.81	349.00
Number of scaffolds/contigs	335/398	2,421/54,033	614/614	3,869/20,440
Scaffold N50	2.82 Mb	483.55 kb	2.83 Mb	494.53 kb
Contig N50	2.35 Mb	11.00 kb	2.83 Mb	36.67 kb
Repetitive elements (%)	43.13	30.99*	43.39*	34.55
GC content (%)	38.99	39.04	39.07	38.85
Gap content (Ns) (%)	0.002	15.24	0	9.72
Number of predicted genes	27,152	33,865	30,327	26,615
Completeness by BUSCO (Metazoa) C:F:M (%)[†]	96.1:1.7:2.2	77.5:10.2:12.3	88.6:4.8:6.6	88.0:6.5:5.5

*Values generated by de novo masking genomes from other studies. proteins per genome against the Metazoa v.10 dataset. †Percentages of complete, fragmented, and missing BUSCO genes on predicted

coral genomes to date. The *CBS* locus was examined in different Anthozoa species, revealing conserved synteny across species that is traceable across the evolutionary tree (Fig. 1). To rule out synteny misinterpretation due to biases in gene prediction, we analyzed syntenic genes surrounding the *CBS* locus in all stony coral genomes using BLAST (*18*) to ensure that missing genes were truly missing in the vicinity of that locus of the corresponding genome. A duplication of *CBS* seems to have occurred in robust corals, as two copies are found in tandem duplication in *Stylophora pistillata* and *Pocillopora damicornis*, albeit the latter showing a disruption of the extra copy (Fig. 1). A high degree of synteny can be observed within stony corals, where *CBS* is preceded by a thiamine transporter (*SLC19A3*) and two facing copies of G protein–coupled receptor class E (*GPCR-E*) (Fig. 1). A region of ~17 kb, devoid of any genes, is found in all *Acropora* species at the location where *CBS* would be expected to occur. In this region, scars of two transposable elements were found: A *Dictyostelium* intermediate repeat sequence (*DIRS*) type (*19*) and a *Penelope*-like element (*PLE*) (*20*). These remnants suggest that the transposition of these elements might have caused the loss of *CBS* in *Acropora*. Nevertheless, other uncharacterized repetitive elements were also found in this region and cannot be ruled out as the cause of *CBS* loss. A retrotransposon Tf2-8 polyprotein (*TF2-8*) was found close to the *CBS* gene in *Porites lutea*, a member of complex corals together with *Acropora*. The appearance of *TF2-8* at the *CBS* locus is accompanied by an inversion of the *CBS* gene and a loss in synteny (Fig. 1). This synteny break is only found in complex corals and indicates a locus instability in their common ancestor. TF2s are long terminal repeat retrotransposons that form virus-like structures from a single transcript and primarily move through recombination with preexisting copies of other *TF2* genes (*21*). It is possible that the insertion and further movement of this retrotransposon led to the loss of *CBS* in the genus *Acropora*.

Cys1a is a previously unrecognized cysteine synthase in animals

Next, we analyzed the encoded protein domains in the genome of *A. loripes* for proteins involved in cysteine biosynthesis (Fig. 2)

and identified the existence of a gene (*alor_g9760*), from now on referred to as *Acys1a* for *Acropora Cys1a*, encoding for a protein of 362 amino acids with a cysteine synthase 1 domain (PTHR10314:SF211). This gene also encodes a pyridoxal-5′ phosphate–dependent enzyme domain (IPR001926) and a pyridoxal-phosphate attachment site (IPR001216), both of which are common to all cysteine synthases (Fig. 3). *Acys1a* contains introns, was identified in all previously analyzed anthozoans, and shows low percentages of identity against cysteine synthases from the canonical pathways (Fig. 3), indicating that this gene did not arise from contamination and implying a conserved function rather than a remnant of a lost pathway. Acys1a protein sequence exhibits 58% of identity against *S. pombe* Cys1a, being the highest among cysteine synthases. Furthermore, they share the same domain architecture (Fig. 3).

To classify Acys1a, we performed a phylogenetic analysis with similar sequences, including protein sequences similar to *S. pombe* Cys1a, as well as sequences from other cysteine biosynthesis pathways (data S3). A phylogenetic tree was reconstructed, classifying cysteine synthases into six main groups (Fig. 4): the CBS group, from the transsulfuration pathway; CysK and CysM groups, from the sulfate assimilation pathway; CysO, from the O-phosphoserine pathway (*22–26*); Cys1a, from the succinylserine pathway; and the unknown group, greatly resembling Cys1a proteins. Acys1a fell within the Cys1a group and formed a subgroup composed of sequences from species across all metazoan lineages except for vertebrates, arthropods, and nematodes. Search for homologous proteins in these three metazoan groups in UniProt (*27*) revealed that Cys1a is absent in them. Only one homologous protein was identified in the nematode *Necator americanus*, but closer inspection revealed that the protein belonged to a very small scaffold with two genes, both devoid of introns and showing 100% identity to proteins from *Paraburkholderia silvatlantica*, indicating that these proteins might stem from bacterial contamination in the genome assembly. Cys1a proteins were also present in nonmetazoan eukaryotes, except for higher plants (Fig. 4). Furthermore, they were identified in bacteria and archaea. A group of proteins greatly resembling Cys1a formed an independent clade, which we called Cys1a-like. Proteins of the group Cys1a-like have the
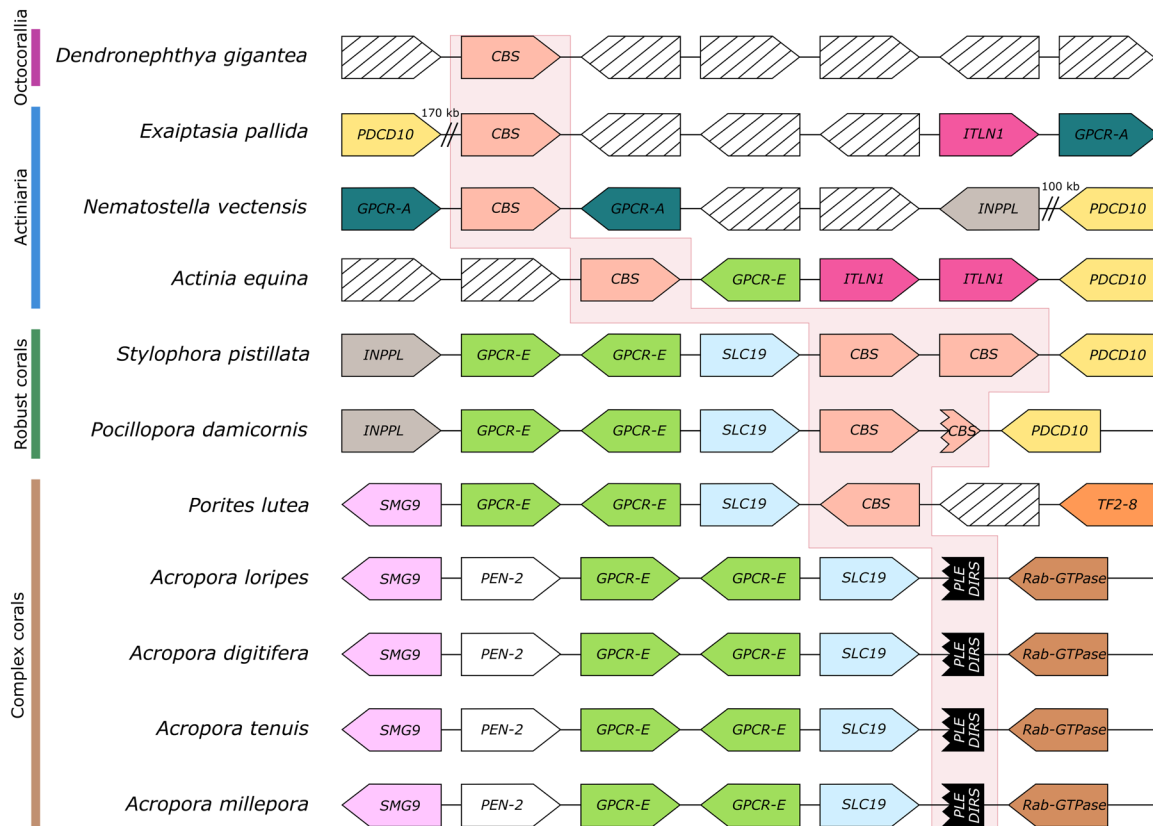
Fig. 1. *CBS* gene loss in the genus *Acropora*. Gene synteny at the *CBS* gene locus across anthozoans. Shape tip depicts gene orientation. The red background highlights the *CBS* locus. Homologous genes are depicted with the same color. *CBS* in red, programmed cell death protein 10 (*PDCD10*) in yellow, intelectin-1 (*ITLN1*) in magenta, phosphatidylinositol 3,4,5-trisphosphate 5-phosphatase 2 (*INPPL*) in gray, G protein–coupled receptor class A (*GPCR-A*) in teal, G protein–coupled receptor class E (*GPCR-E*) in green, thiamine transporter (*SLC19*) in light blue, protein SMG9 (*SMG9*) in lilac, transposon Tf2-8 polyprotein (*TF28*) in orange, gamma-secretase subunit pen-2 (*PEN-2*) in white, Penelope-like repeat (*PLE*) and *Dictyostelium* intermediate repeat sequence (*DIRS*) in black, Rab-GTPase (guanosine triphosphatase) (*Rab-GTPase*) in brown, and genes without conservation across species in black stripes.

cysteine synthase–related domain (PTHR10314:SF35) instead of the cysteine synthase 1 domain (PTHR10314:SF211) found in Cys1a (Fig. 3). Proteins from the coral symbionts of the family Symbiodiniaceae fall within this clade, but these organisms also have the *CBS* gene.

Cys2 is also present throughout metazoans

Cys1a requires *O*-succinyl-L-serine for the biosynthesis of cysteine, which, in the fungi *S. pombe*, is provided by the L-serine *O*-succinyltransferase (SST) Cys2 through the transfer of a succinyl group from succinyl-CoA (coenzyme A) to L-serine (Fig. 2) (*13*). We screened the *A. loripes* genome for homologous proteins of Cys2 from *S. pombe* and identified a gene (*alor_g16104*), subsequently referred to as *Acys2*, encoding for a putative Cys2 protein of 465 amino acids. Cys2 belongs to the MetX family of enzymes, previously thought to only be L-homoserine *O*-acetyltransferases (HATs). These enzymes are crucial for the biosynthesis of methionine (*28*) but were recently shown to also function as L-homoserine *O*-succinyltransferases (HSTs), L-serine *O*-acetyltransferases (SATs), and SSTs (*13*, *29*). Cys2 enzymes have been shown to be primarily SSTs and have almost 300 times greater efficiency with L-serine than with L-homoserine (*13*); furthermore, Cys2 mutants of *S. pombe* become cysteine auxotrophic (*14*, *15*). However, because of the recent discovery of the different activities of the

MetX proteins, it has been estimated that more than 60% of the MetX proteins are incorrectly annotated (*13*), rendering the classification of MetX proteins through the current annotation from homologous proteins ineffective. MetX proteins have been divided into six groups according to their primary enzymatic activities (*13*): MetX-G1, composed of HATs; MetX-G2, with HSTs; MetX-G3a, with HSTs and SSTs; MetX-G3b, with HSTs; MetX-G3c, with SSTs; and MetX-G4, proteins without any detected enzymatic activity. *S. pombe* Cys2 is classified as a MetX-G3c.

To classify the Acys2 protein into a MetX group and avoid possible previous misannotations, we retrieved sequences of experimentally validated MetX proteins (*13*) and sequences from all eukaryotic lineages with similarities to Acys2 and *S. pombe* Cys2 (data S3). We then reconstructed a phylogenetic tree with MetA proteins as the outgroup, a family of proteins with similar activities to MetX but not genetically related. The tree was resolved into six groups in line with their enzymatic activities (Fig. 5A). Acys2 fell within the MetX-G3c group and, similarly to Cys1a, formed a subgroup with other metazoans. Furthermore, no MetX-G3c sequences from vertebrates, arthropods, or nematodes were identified. A number of key residues have been previously identified in MetX-G3 proteins required for their SST activity (*13*, *30*): Gly-Leu-Ser-Ala/Pro and the similar

Cysteine biosynthesis pathways and their presence in *Acropora loripes*
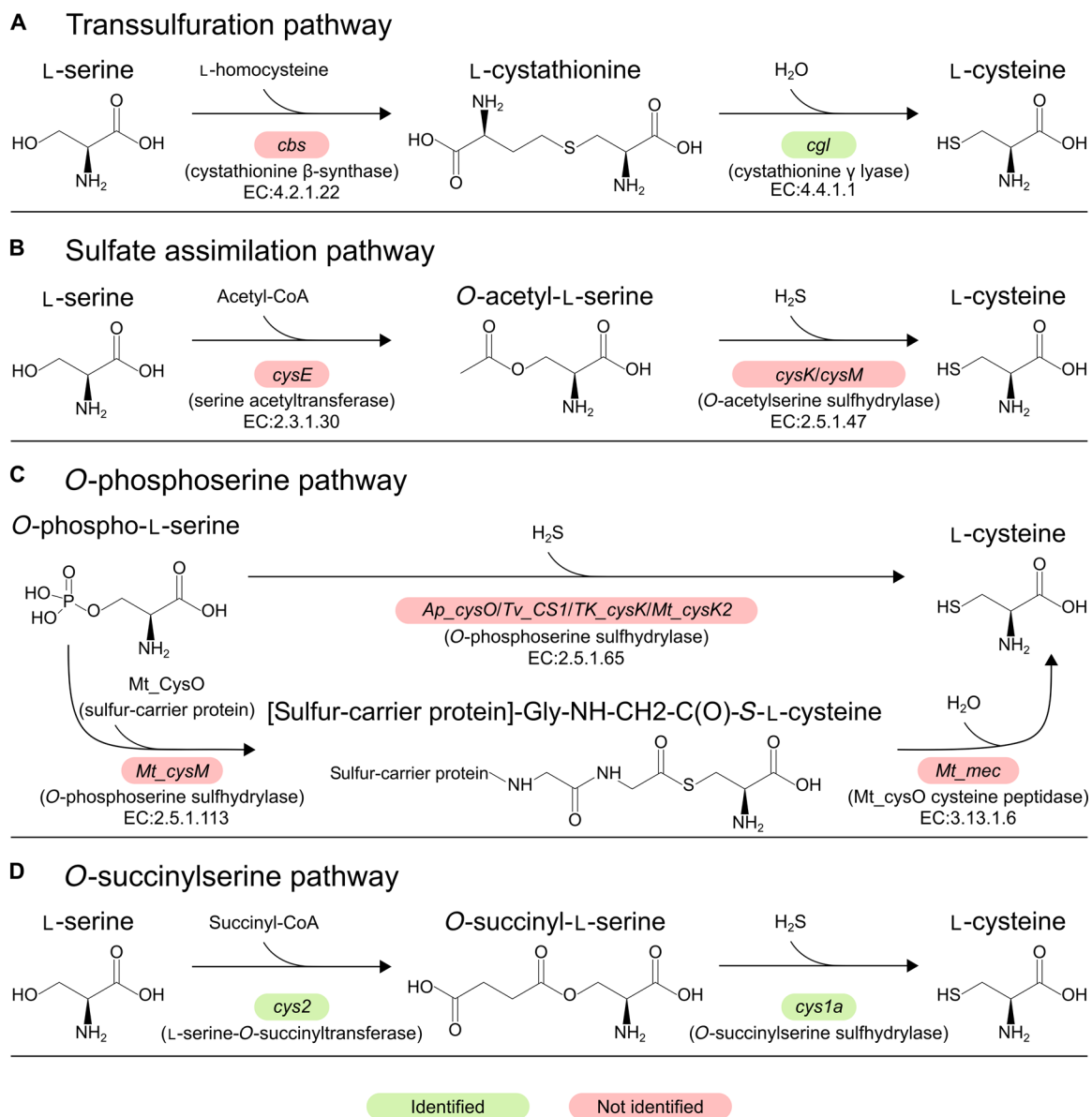
Fig. 2. Cysteine biosynthesis pathways and the identification of their genes in *A. loripes*. (**A**) Transsulfuration pathway. (**B**) Sulfate assimilation pathway. (**C**) *O*-phosphoserine pathway. (**D**) *O*-succinylserine pathway. Genes are denoted by colored boxes: Red not identified and green identified in the genome of *A. loripes*. EC (Enzyme Commission) numbers are included for the available classified reactions.

Gly-Ile-Ser-Ala at positions 52 to 55 based on *Xanthomonas campestris* MetX-3Ga; Ser-Leu/Met at positions 149 to 150; Arg[183] based on *Frateuria aurantia* MetX-3Ga; and Asp[360] based on *X. campestris* MetX-3Ga. Furthermore, it is considered that having Ser-Tyr instead of Ser-Leu/Met at positions 149 to 150 results in L-homoserine specificity instead of L-serine (*13*), as observed in MetX-3Gb enzymes. Acys2 contains all these important residues for SST activity and was thus classified as a MetX-3Gc enzyme (Fig. 5B). Moreover, putative Cys2 proteins from animals and other eukaryotic lineages also contain these residues (Fig. 5B).

A. loripes Cys1a and Cys2 proteins rescue cysteine biosynthesis in vivo

To confirm that the animal homologs of Cys1a and Cys2 can indeed catalyze the biosynthesis of cysteine, we transformed codon-optimized versions of the *A. loripes* genes *Acys1a* and *Acys2* into the cysteine auxotrophic mutants of *Cys1aΔ* and *Cys2Δ* in *S. pombe* (*31*), respectively. Similarly to *A. loripes*, *S. pombe* does not have the CBS enzyme from the transsulfuration pathway (*32*) and relies solely on the *O*-succinylserine pathway for cysteine biosynthesis (*13*). Complementation of the *S. pombe* mutants with *A. loripes* genes *Acys1a*
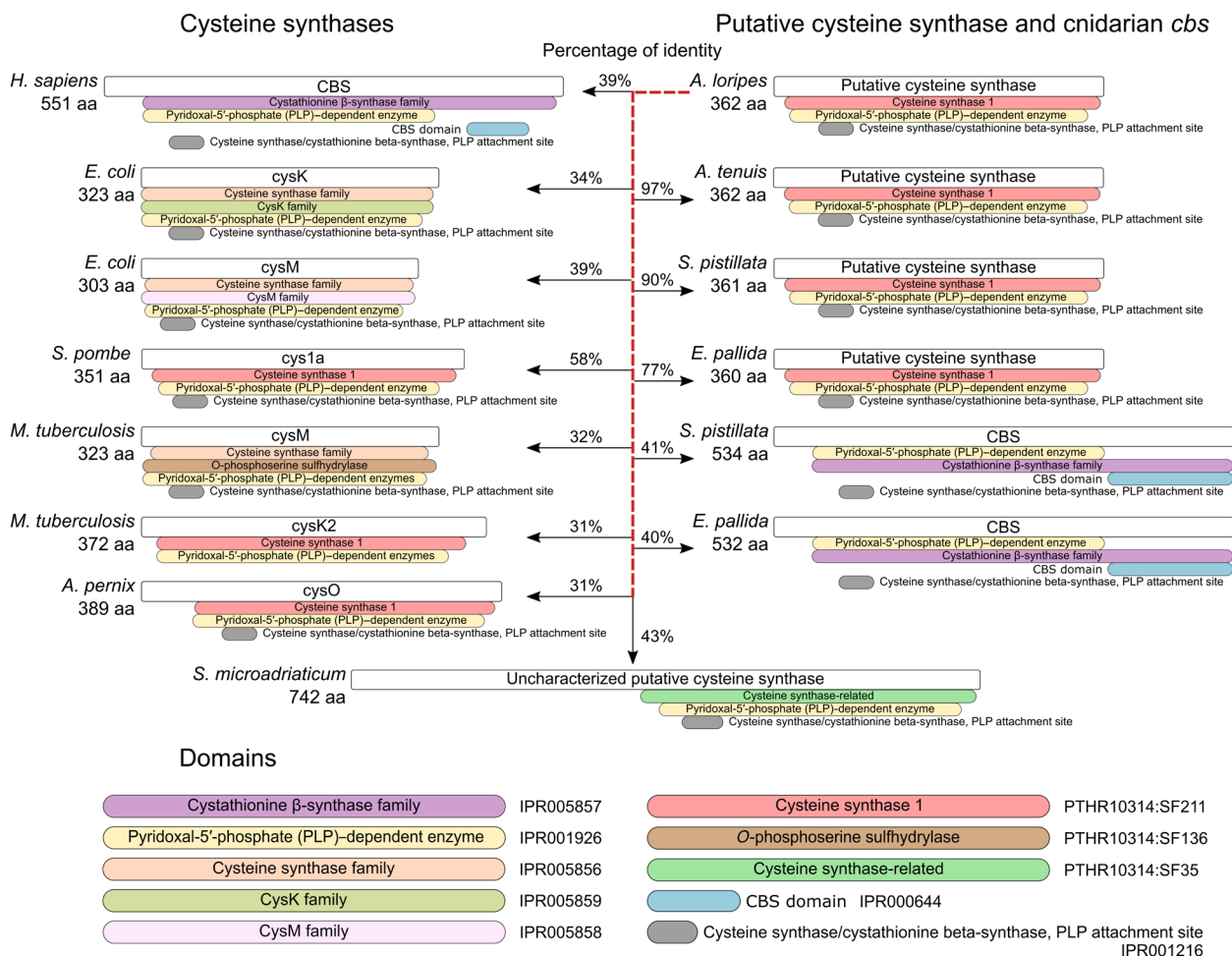
Fig. 3. Cysteine synthases and their domain architectures. Sequence identities are shown between *A. loripes* putative cysteine synthase and other cysteine synthases. Protein domains are identified by their InterPro or PANTHER domains.

(Fig. 6A) and *Acys2* (Fig. 6B) rescued the mutant phenotype and enabled the mutant cells to grow in the absence of cysteine in the media. These results confirm that Acys1a is an *O*-succinylserine sulfhydrylase and Acys2 an L-serine-*O*-succinyl transferase, showing that *A. loripes* has the necessary proteins for the biosynthesis of cysteine through the *O*-succinylserine pathway.

DISCUSSION

It was previously thought that cysteine was exclusively produced through the transsulfuration pathway in animals. Here, we present evidence for the existence of an alternative pathway for cysteine biosynthesis in animals through the analysis of the genome of the complex coral *A. loripes*. We find that the loss of *CBS* in *Acropora* was likely due to the action of transposable elements and identify the existence of an alternative pathway for cysteine biosynthesis through the *O*-succinylserine pathway. This pathway has been recently identified in fungi (*13*) and requires the action of two proteins: Cys2, an L-serine-*O*-succinyltransferase, and Cys1a, an *O*-succinylserine sulfhydrylase. To our surprise, we found the existence of both enzymes not only in the genome of *A. loripes* but also across all animal phyla except for vertebrates, arthropods, and nematodes. Furthermore, these enzymes are

found throughout eukaryotes but appear to be absent in higher plants, suggesting that it is an ancestral pathway that was lost independently in plants and several animal lineages. The fact that *A. loripes Acys2* and A*cys1a* can rescue *S. pombe Cys2* and *Cys1a* mutants from cysteine auxotrophy demonstrates that these proteins are indeed functional and likely capable of producing cysteine in vivo in animals.

Our findings highlight that our current understanding of the metabolic capabilities of animals, and life in general, is still incomplete even regarding biosynthetic capabilities that are thought to be well understood. This emphasizes the importance of exercising caution when generalizing and extrapolating findings to emerging systems. In the case of animals, much of our knowledge is derived from studies in a smaller group of model organisms encompassing mammals, arthropods, and nematodes. As shown in this study, this can be misleading and can generate biased conclusions as the characteristics of certain organisms might not necessarily apply to others. Agreeingly, it was previously considered that animals could not synthesize histidine. However, a complete fungal-like histidine biosynthesis pathway was recently identified in robust corals (*33*). Although the functionality of the proteins in this histidine biosynthesis pathway still has to be confirmed, its presence supports the notion that many biosynthetic capabilities are yet to be found.
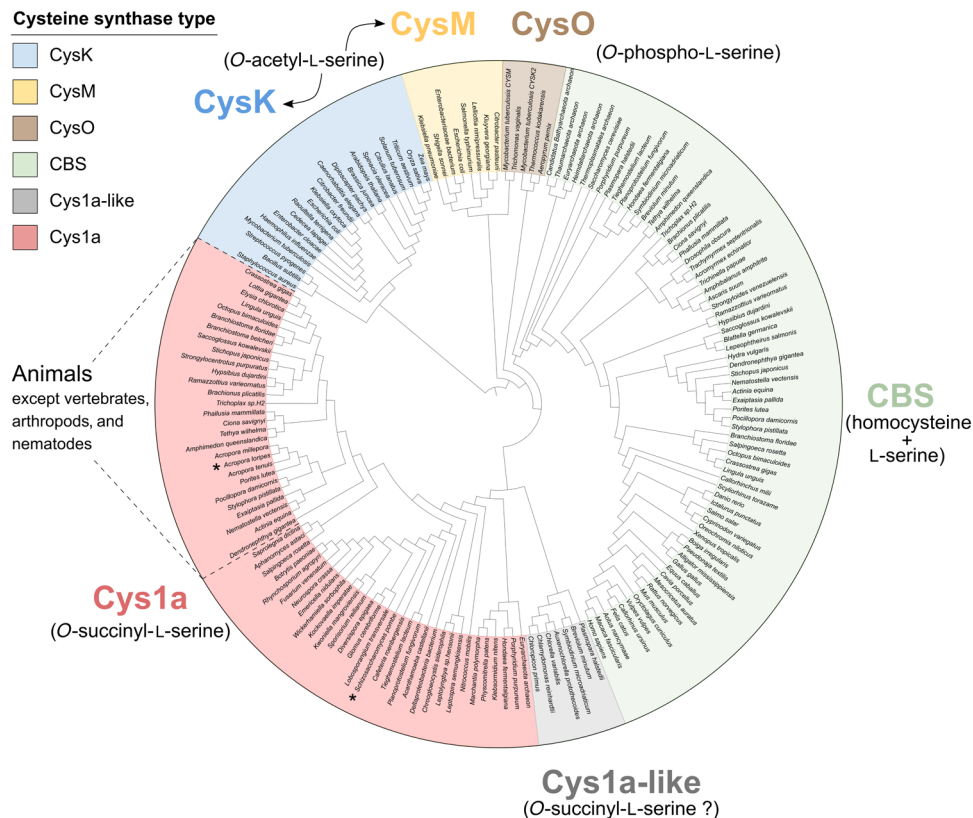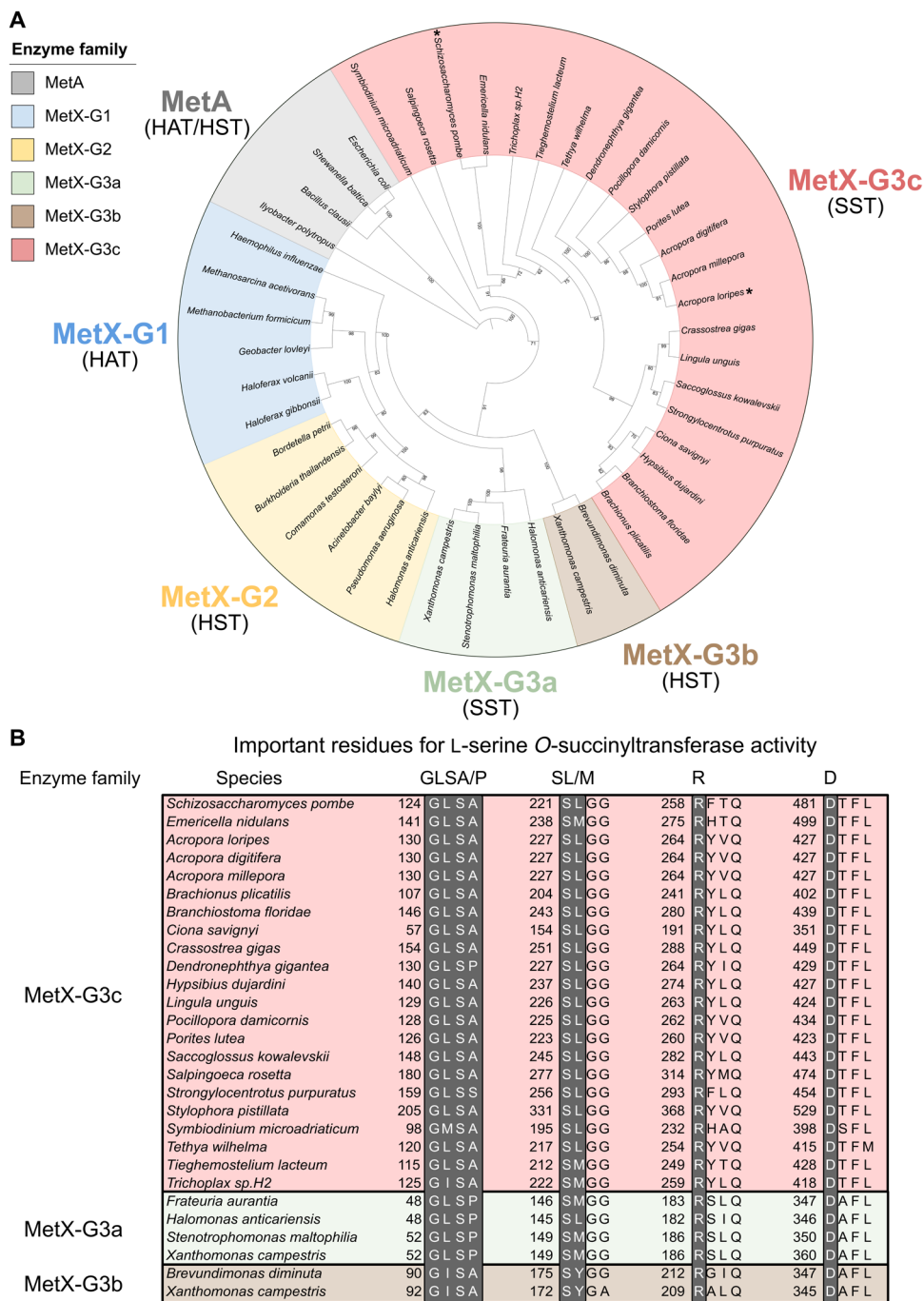
Fig. 4. Phylogenetic reconstruction of cysteine synthases. Protein sequences from cysteine synthases belonging to the transsulfuration pathway (CBS), sulfate assimilation pathway (CysK and CysM), *O*-phosphoserine pathway (CysO), and *O*-succinylserine pathway (Cys1a) were used for the generation of a phylogenetic tree. Primary substrates for each cysteine synthase are shown in parentheses. Sequences similar to *A. loripes* putative cysteine synthase Acys1a and *S. pombe* Cys1a (both denoted by *) were included in the analysis and either grouped within the Cys1a cluster or formed a sister clade classified as Cys1a-like. Acys1a was classified as a Cys1a-type cysteine synthase together with animal sequences from all lineages except for vertebrates, arthropods, and nematodes. The phylogenetic tree was generated with IQ-TREE (*77*) and visualized with iTOL (*79*).

The reason why the enzymes of the *O*-succinylserine pathway were lost in vertebrates, arthropods, and nematodes is yet to be elucidated. However, differences in domain organization and regulation are found in animal CBSs that could provide an explanation for the repeated loss of this pathway. Mammalian CBS enzymes are known to be more highly regulated than other CBSs. They are positively regulated by the binding of *S*-adenosylmethionine to the CBS domain (*34*), and disruption of this domain results in increased activity and the formation of dimers instead of tetramers (*35*, *36*). Conversely, *Drosophila melanogaster* CBS natively forms dimers and is highly active, and deletion of the C-terminal CBS domains leads to decreased activity (*35*, *36*). In addition, CBS enzymes can have an N-terminal heme binding domain (*37*, *38*), with some vertebrates harboring two (*38*). These domains increase thermal stability (*36*) and activity (*38*, *39*) and have not been found outside metazoans (*37*, *38*, *40*, *41*) (fig. S5). Nematode CBSs, on the other hand, follow a different organization, with longer, highly active monomeric proteins that lack both the N-terminal heme binding and C-terminal CBS domains but have two PLP-dependent domains (*40*). The role of the two PLP-dependent domains is unknown, but it was proposed that the first is involved in protein stability and solubilization, while the second domain is involved in the catalytic activity (*40*). These factors indicate a variable and tight regulation of cysteine biosynthesis, particularly in vertebrates, that is not present in other animals. This required tight regulation may have promoted the conservation of only one pathway for cysteine biosynthesis, yet further investigation is required.

We also identify the existence of a group of proteins greatly resembling Cys1a, composed of photosynthetic algae and oomycetes. Coral dinoflagellate symbionts were found in this clade and were additionally found to have homologs of the Cys2 protein, suggesting that they might also have an alternative pathway for cysteine biosynthesis. Further studies are required to identify whether these proteins can produce cysteine through the *O*-succinylserine pathway or through a similar substrate. The evidence presented in this study shows the biosynthesis of cysteine through an alternative pathway to the transsulfuration pathway in animals. It also shows that corals of the genus *Acropora* have the required machinery for cysteine biosynthesis and thus might not depend on symbiosis with Symbiodiniaceae algae for its acquisition. It is yet to be elucidated whether the loss of cysteine biosynthesis through the transsulfuration pathway in *Acropora* contributes to its increased thermal sensitivity, as evidence for the expression responses of genes involved in cysteine biosynthesis upon thermal stress in *Acropora* and other corals is currently inconclusive (*42*, *43*). It would be interesting to further investigate the contribution of each pathway to cysteine pools in species with both pathways and explore the potential causes favoring the presence of one or two pathways for cysteine biosynthesis in an organism.

Fig. 5. Classification of MetX proteins. (**A**) Phylogenetic reconstruction of MetX proteins: MetX-G1 in blue, MetX-G2 in yellow, MetX-G3a in green, MetX-G3b in brown, MetX-G3c in red, and MetA as the outgroup in gray. Each group has its enzymatic activity in parenthesis: HAT, HST, and SST. *A. loripes* putative Cys2 (Acys2) and *S. pombe* Cys2 are denoted by an asterisk (*). Acys2 was classified as a MetX-3Gc SST together with animal sequences from all lineages except for vertebrates, arthropods, and nematodes. The phylogenetic tree was generated with IQ-TREE and visualized with iTOL. (**B**) Identification of important residues in putative Cys2 protein sequences for SST activity. MetX-G3c in red, MetX-G3a in green, and MetX-G3b in brown. Important residues for an SST activity are highlighted in gray. *A. loripes* Cys2 and other eukaryotic putative Cys2 proteins have all the required residues for SST activity.

MATERIALS AND METHODS
DNA extraction, library preparation, and genome size estimation

A. loripes samples used for DNA and RNA extraction were collected from Trunk Reef on the Great Barrier Reef (18°35′S, 146°80′E) (fig. S1). High–molecular weight DNA was extracted by grinding frozen sperm in a glass homogenizer filled with homogenization buffer [100 mM tris (pH 9.0), 100 mM EDTA, 100 mM NaCl, and 1% SDS]. After homogenization, proteinase K was added at a concentration of 530 μg/ml, and the homogenate was placed at 70°C for

Fig. 6. Involvement of *A. loripes* Cys1 and Cys2 proteins in cysteine biosynthesis in vivo. Yeast spot assays in *S. pombe* mutants grown in the presence or absence of L-cysteine for L-cysteine auxotrophy gene complementation with L-leucine as selection marker. (**A**) *A. loripes* Cys1a gene (*Acys1a*) complementation assay. (**B**) *A. loripes* Cys2 gene (*Acys2*) complementation assay. Strains: WT (ED666), *ade6 ura4 leu1*; *cys1aΔ*, *ade6 ura4 leu1 cys1a::kanMX4*; and *cys2Δ*, *ade6 ura4 leu1 cys2::kanMX4*. All assays were carried out with the addition of adenine and uracil as the background strain (WT) is auxotrophic for both. The plasmid pREP3xN, containing *LEU2* as selection marker to complement leucine auxotrophy in the ED666 WT strain, was used as the expression vector and transformed empty, with *A. loripes Acys1a* gene or with *A. loripes Acys2* gene. All assays were grown in EMM medium and done in triplicates, with the figure showing one representative assay.

1 hour. After incubation, 250 μl of KOAc per 1 ml of homogenate was added, mixed thoroughly, incubated for 30 min on ice, and centrifuged at 25,000*g* for 15 min at 4°C. The supernatant was recovered and mixed with 0.8 volumes of isopropanol. The mixture was incubated at room temperature for 15 min and centrifuged at 25,000*g* for 15 min at 20°C to precipitate DNA. The precipitate was washed with 70% ethanol and centrifuged at 25,000*g* for 10 min at 20°C, air-dried, and resuspended overnight at room temperature in 1 ml of resuspension buffer [10 mM tris (pH 7.5) and EDTA 1 mM] with ribonuclease A (100 μg/ml). Genomic DNA was recovered using QIAGEN Genomic-tips following the instructions of the manufacturer and resuspended in resuspension buffer. DNA quality was assessed with NanoDrop (Thermo Fisher Scientific, USA), Qubit 2.0 Fluorometer (Invitrogen, USA), and Fragment Analyzer (Agilent, USA). Two different libraries for genome sequencing were prepared for two different technologies: PacBio and 10x Chromium. The first library was prepared with the SMRTbell Library Preparation Kit (Pacific Biosciences, USA). Fifty-one SMRT cells were run on a PacBio RS II (Pacific Biosciences, USA) system at the Bioscience Core Labs [King Abdullah University of Science and Technology (KAUST), Saudi Arabia], generating 3.2 million reads with an average length of 10.7 kb, resulting in an 85 × coverage. The second library was prepared from 40-kb DNA molecules with the Chromium Genome Library Kit v2 (10x Genomics, USA). The library was sequenced on two lanes of Illumina HiSeq 4000 (Illumina, USA) run in 151–base pair (bp) paired-end mode at the Bioscience Core Labs (KAUST, Saudi Arabia), resulting in 609 million reads and a 210 × coverage.

Genome size and heterozygosity were estimated with GenomeScope (*44*) by measuring abundances of *k*-mers within a set of DNA

sequencing reads. For this analysis, *k*-mer frequencies from 10x Chromium DNA sequencing reads stripped of their barcodes were counted with Jellyfish (*45*), and the histogram of *k*-mer abundances was input to GenomeScope (http://qb.cshl.edu/genomescope/) (fig. S3).

RNA extraction, sequencing, and processing
RNA was extracted from a single nubbin with the Quick-RNA Plant Mini-Prep Kit (R2024) (Zymo Research, USA) following the manufacturer's instructions. RNA quality was assessed with NanoDrop (Thermo Fisher Scientific, USA), Qubit 2.0 Fluorometer (Invitrogen, USA), and Bioanalyzer (Agilent, USA). RNA sequencing (RNA-seq) libraries were prepared with the TruSeq RNA Library Prep Kit v2 (Illumina, USA) according to the manufacturer's instructions. RNA-seq libraries were sequenced on one lane of Illumina HiSeq 4000 (Illumina, USA) under 151-bp paired-end mode, resulting in 113 × coverage (based on genome size). Reads were quality trimmed and removed of adapter sequences with Trimmomatic (*46*).

Genome assembly
SMRT sequencing reads, subsequently referred to as PacBio reads, were error corrected with LoRDEC (*47*) using Chromium 10x linked reads stripped of their barcodes, to resemble Illumina DNA sequencing reads, as reference. Error-corrected PacBio reads were used for genome assembly with Canu (*48*), resulting in a 900-Mb genome. Heterozygosity in the assembly was reduced with Purge Haplotigs (*49*) using parameters -a 41 and -a 66, resulting in two genome versions of 429 (genome_a41) and 541 (genome_a66) Mb, respectively. Assemblies were further corrected and polished with Pilon (*50*) using Minimap2 (*51*) to map error-corrected PacBio reads

and BWA-MEM (52) to map 10x reads. Genomes were then scaffolded with SSPACE-LongRead (53) using PacBio reads, followed by a correction with Pilon, and a further scaffolding with ARCS (54), to make use of the barcode information within the 10x reads. Genomes were gap filled with LR_Gapcloser (55) and further reduced by removing duplicated or heterozygous regions with HaploMerger2 (56). Preliminary annotation of the genomes was performed with BRAKER2 (18, 57, 58) (for details, see the "Genome annotation and quality assessment" section), and their completeness was assessed with BUSCO (17) using the Metazoa v.9 dataset. Missing BUSCO genes from the Metazoa dataset in scaffolds of genome_a41 were searched in genome_a66. Scaffolds from genome_a66 containing missing BUSCO genes in genome_a41 were added to genome_a41. This accounted for only 0.8% of BUSCO genes in the Metazoa dataset in 10 scaffolds comprising 30.9 Mb of genome sequence. From this point in the analysis, only genome_a41 with the added scaffolds was used. To remove duplicated regions after the addition of scaffolds, another round of HaploMerger2 was performed, removing 28.7 Mb of genome sequence, followed by a final step of scaffolding with ARCS, and another round of gap filling with LR_Gapcloser. Genome statistics were computed with stats.sh from BBtools (https://sourceforge.net/projects/bbmap/). A comprehensive explanation and commands used for genome assembly can be found in the Supplementary Materials.

Genome annotation and quality assessment

The final genome assembly was masked for repetitive elements before annotation by generating a de novo repeat library with Repeat-Modeler (59), which was then used to mask the genome with RepeatMasker and RepeatProteinMask (60). To generate hints for genome annotation, RNA-seq data from *A. loripes* was mapped onto the assembled genome with HISAT2 (61). Protein prediction on the masked genome was carried out with BRAKER2 using the mapped RNA-seq reads as hints and using the BUSCO-trained Augustus file resulting from running BUSCO on the genome. The completeness of the predicted proteins was assessed by running BUSCO on the predicted proteins with BRAKER2 against the Metazoa v.10 dataset. The quality of the protein prediction was compared against previously sequenced genomes of the genus *Acropora*: an updated *Acropora digitifera* genome (4) (GenBank accession GCA_000222465.2), the genome of *Acropora tenuis* (62) [Reefgenomics.org (63)], and the genome of *Acropora millepora* (64) (GenBank accession QTZP00000000). Repetitive elements were also identified for the genomes of *A. digitifera* and *A. tenuis* in the same way as for *A. loripes* (tables S2 to S4). Annotation of the genome of *A. loripes* consisted of performing BLAST (18) against three databases: Swiss-Prot, TrEMBL (27), and National Center for Biotechnology Information (NCBI) NR (non-redundant) (65). Protein domains were identified with Inter-ProScan (66), and Gene Ontology terms (67) were identified by InterPro entries and by appropriating them from Swiss-Prot–derived annotations. KEGG (10) orthologies were retrieved from Swiss-Prot annotations and de novo identified with BlastKOALA (68). The complete annotation can be found in data S1.

Synteny analysis of CBS

The *CBS* gene was searched in the genomes of the following anthozoan taxa: complex corals: *A. digitifera*, *A. loripes*, *A. tenuis*, and *P. lutea* v.1.1 (69) (Reefgenomics.org); robust corals: *P. damicornis* v1.0 (70) (Reefgenomics.org) and *S. pistillata* v1.0 (71) (Reefgenomics.org);

Actiniaria: *Actinia equina* v1.0 (72) (Reefgenomics.org), *Exaiptasia pallida* (also known as *Exaiptasia diaphana*) (73) (GenBank accession GCA_001417965.1) (Reefgenomics.org), and *Nematostella vectensis* v1.0 (74) (https://genome.jgi.doe.gov/); and Octocorallia: *Dendronephthya gigantea* (75) (GenBank accession GCA_004324835.1). All genomes were reannotated with Swiss-Prot and InterProScan. Loci harboring *CBS* were manually inspected, and neighboring genes were retrieved. To corroborate the presence or absence of synteny, genes showing strong synteny in stony corals had their proteins searched in all other species using TBLASTn against the genomes and BLASTp against their annotations. Preference was given to manual curation over previous annotations, resulting in the inclusion, fusion, or exclusion of genes with respect to their GFF files when sufficient evidence supported it. For the identification of repetitive elements at the *CBS* locus, the RepeatMasker results derived from the de novo repeat libraries generated for all *Acropora* species were analyzed.

Cysteine biosynthesis pathway analysis

To identify genes that could be involved in the biosynthesis of cysteine in *A. loripes*, its gene annotation was searched for terms related to cysteine. In addition, a BLASTp search was performed with proteins involved in known cysteine biosynthesis pathways against the predicted proteins of *A. loripes*. Genome annotations were also mapped into KEGG pathways (10) and analyzed. One protein was identified to have a cysteine synthase 1 domain. Homologs for this putative cysteine synthase were searched in the other anthozoan species. Domains of the putative cysteine synthase and of other cysteine synthases were analyzed with InterProScan (76). From these analyses, two proteins resembling *S. pombe* Cys1a and Cys2 were identified in *A. loripes*.

Phylogenetic analysis

For the classification of *A. loripes* putative Cys1a (Acys1a) and Cys2 (Acys2) proteins, phylogenetic trees were reconstructed. For the phylogenetic reconstruction of cysteine synthases, protein sequences from other cysteine biosynthesis pathways were recovered from individual anthozoan genome projects, Reefgenomics (63), NCBI, and UniProt (data S3). For the phylogenetic reconstruction of MetX proteins, only protein sequences from enzymes with validated activities were selected (13). For the identification of Cys1a and Cys2 homologs in other species, proteins were recovered through BLASTp in the aforementioned databases. For the identification of phylum-specific sequences, UniProt BLASTp searches were used against the specified phylum. Protein sequences were selected to cover for all possible animal phyla and kingdoms of life (data S3). Phylogenetic reconstructions were carried out with IQ-TREE (77) under a maximum likelihood model with 10,000 ultrafast bootstraps (78), and trees were visualized with iTOL (79).

Yeast complementation assays

S. pombe codon-optimized versions of the *A. loripes* Acys1a and Acys2 genes were synthesized by Genscript. Genes were cloned into the pREP3xN_ccdB vector carrying the *LEU2* selection marker. The pREP3xN_ccdB vector was a gift from G. Akusjärvi and P. Bjerling (Addgene plasmid #79021; http://n2t.net/addgene:79021; RRID: Addgene_79021). Vectors carrying *A. loripes* Acys1a or Acys2 or empty were transformed with the lithium acetate method (80) into the *S. pombe* haploid strains: wild type (ED666), *ade6*-M210 *ura4*-D18 *leu1*-32; *cys1aΔ*, *ade6*-M216 *ura4*-D18 *leu1*-32 *cys1a:: kanMX4*; and

cys2Δ, *ade6*-M216 *ura4*-D18 *leu1*-32 *cys2::kanMX4* (*31*), from Bioneer. Yeast spot assays were carried out in Edinburgh Minimal Media (EMM). In short, *S. pombe* strains were grown in liquid EMM medium with adenine (250 mg/liter) and the corresponding amino acids at a concentration of 250 mg/liter. Strains carrying a construct were grown in the absence of leucine to maintain the plasmid selection. *S. pombe* cells were grown for 2 days at 30°C after which 500 µl was subcultured into 5 ml of liquid EMM. Cultures were allowed to grow for 16 hours and then diluted to an optical density at 600 nm of 0.6 with fresh liquid EMM containing only adenine and without amino acids. For each culture, four 10-fold dilutions were prepared, and 10 µl per sample was spotted in triplicate on EMM plates containing adenine and uracil but with or without the addition cysteine or leucine. Plates were grown for 3 to 5 days before being photographed.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at https://science.org/doi/10.1126/sciadv.abq0304

View/request a protocol for this paper from *Bio-protocol*.

REFERENCES AND NOTES

1. A. J. L. Cooper, Biochemistry of sulfur-containing amino acids. *Annu. Rev. Biochem.* **52**, 187–222 (1983).
2. S. H. Mudd, J. D. Finkelstein, F. Irreverre, L. Laster, Homocystinuria: An enzymatic defect. *Science* **143**, 1443–1445 (1964).
3. S. H. Mudd, F. Skovby, H. L. Levy, K. D. Pettigrew, B. Wilcken, R. E. Pyeritz, G. Andria, G. H. Boers, I. L. Bromberg, R. Cerone, The natural history of homocystinuria due to cystathionine beta-synthase deficiency. *Am. J. Hum. Genet.* **37**, 1–31 (1985).
4. C. Shinzato, E. Shoguchi, T. Kawashima, M. Hamada, K. Hisata, M. Tanaka, M. Fujie, M. Fujiwara, R. Koyanagi, T. Ikuta, A. Fujiyama, D. J. Miller, N. Satoh, Using the *Acropora* digitifera genome to understand coral responses to environmental change. *Nature* **476**, 320–323 (2011).
5. C. Shinzato, K. Khalturin, J. Inoue, Y. Zayasu, M. Kanda, M. Kawamitsu, Y. Yoshioka, H. Yamashita, G. Suzuki, N. Satoh, Eighteen coral genomes reveal the evolutionary origin of *Acropora* strategies to accommodate environmental changes. *Mol. Biol. Evol.* **38**, 16–30 (2020).
6. V. Schoepf, M. Stat, J. L. Falter, M. T. McCulloch, Limits to the thermal tolerance of corals adapted to a highly fluctuating, naturally extreme temperature environment. *Sci. Rep.* **5**, 17639 (2015).
7. Y. Loya, K. Sakai, K. Yamazato, Y. Nakano, H. Sambali, R. van Woesik, Coral bleaching: The winners and the losers. *Ecol. Lett.* **4**, 122–131 (2001).
8. T. Leustek, M. N. Martin, J. A. Bick, J. P. Davies, Pathways and regulation of sulfur metabolism revealed through molecular and genetic studies. *Annu. Rev. Plant. Physiol. Plant. Mol. Biol.* **51**, 141–165 (2000).
9. E. Guédon, I. Martin-Verstraete, Cysteine Metabolism and Its Regulation in Bacteria, in *Amino Acid Biosynthesis ~ Pathways, Regulation and Metabolic Engineering*, V. F. Wendisch, Ed. (Springer Berlin Heidelberg, 2006), pp. 195–218.
10. M. Kanehisa, S. Goto, KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
11. M. Kanehisa, Toward understanding the origin and evolution of cellular organisms. *Protein Sci.* **28**, 1947–1951 (2019).
12. R. Caspi, R. Billington, L. Ferrer, H. Foerster, C. A. Fulcher, I. M. Keseler, A. Kothari, M. Krummenacker, M. Latendresse, L. A. Mueller, Q. Ong, S. Paley, P. Subhraveti, D. S. Weaver, P. D. Karp, The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.* **42**, D471–D480 (2016).
13. K. Bastard, A. Perret, A. Mariage, T. Bessonnet, A. Pinet-Turpault, J. L. Petit, E. Darii, P. Bazire, C. Vergne-Vaxelaire, A. Brewee, A. Debard, V. Pellouin, M. Besnard-Gonnet, F. Artiguenave, C. Médigue, D. Vallenet, A. Danchin, A. Zaparucha, J. Weissenbach, M. Salanoubat, V. de Berardinis, Parallel evolution of non-homologous isofunctional enzymes in methionine biosynthesis. *Nat. Chem. Biol.* **13**, 858–866 (2017).
14. M. Grynberg, J. Topczewski, A. Godzik, A. Paszewski, The *Aspergillus nidulans cysA* gene encodes a novel type of serine *O*-acetyltransferase which is homologous to homoserine *O*-acetyltransferases. *Microbiology (Reading)* **146**, 2695–2703 (2000).
15. Y. Ma, R. Sugiura, M. Saito, A. Koike, S. O. Sio, Y. Fujita, K. Takegawa, T. Kuno, Six new amino acid-auxotrophic markers for targeted gene integration and disruption in fission yeast. *Curr. Genet.* **52**, 97–105 (2007).
16. Y. Fujita, K. Takegawa, Characterization of two genes encoding putative cysteine synthase required for cysteine biosynthesis in *Schizosaccharomyces pombe. Biosci. Biotechnol. Biochem.* **68**, 306–311 (2004).
17. F. A. Simão, R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, E. M. Zdobnov, BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
18. C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, T. L. Madden, BLAST+: Architecture and applications. *BMC Bioinform.* **10**, 421 (2009).
19. J. Cappello, S. M. Cohen, H. F. Lodish, *Dictyostelium* transposable element DIRS-1 preferentially inserts into DIRS-1 sequences. *Mol. Cell. Biol.* **4**, 2207–2213 (1984).
20. M. B. Evgen'ev, H. Zelentsova, N. Shostak, M. Kozitsina, V. Barskyi, D.-H. Lankenau, V. G. Corces, *Penelope*, a new family of transposable elements and its possible role in hybrid dysgenesis in *Drosophila virilis. Proc. Natl. Acad. Sci. U.S.A.* **94**, 196–201 (1997).
21. E. F. Hoff, H. L. Levin, J. D. Boeke, *Schizosaccharomyces pombe* retrotransposon Tf2 mobilizes primarily through homologous cDNA recombination. *Mol. Cell. Biol.* **18**, 6839–6852 (1998).
22. K. Mino, K. Ishikawa, A novel *O*-phospho-L-serine sulfhydrylation reaction catalyzed by *O*-acetylserine sulfhydrylase from *Aeropyrum pernix* K1. *FEBS Lett.* **551**, 133–138 (2003).
23. E. M. Steiner, D. Böth, P. Lössl, F. Vilaplana, R. Schnell, G. Schneider, CysK2 from *Mycobacterium tuberculosis* is an *O*-phospho-L-serine-dependent *S*-sulfocysteine synthase. *J. Bacteriol.* **196**, 3410–3420 (2014).
24. D. Agren, R. Schnell, W. Oehlmann, M. Singh, G. Schneider, Cysteine synthase (CysM) of *Mycobacterium tuberculosis* is an *O*-phosphoserine sulfhydrylase: Evidence for an alternative cysteine biosynthesis pathway in mycobacteria. *J. Biol. Chem.* **283**, 31567–31574 (2008).
25. S. E. O'Leary, C. T. Jurgenson, S. E. Ealick, T. P. Begley, *O*-phospho-L-serine and the thiocarboxylated sulfur carrier protein CysO-COSH are substrates for CysM, a cysteine synthase from *Mycobacterium tuberculosis. Biochemistry* **47**, 11606–11615 (2008).
26. G. D. Westrop, G. Goodall, J. C. Mottram, G. H. Coombs, Cysteine biosynthesis in *Trichomonas vaginalis* involves cysteine synthase utilizing *O*-phosphoserine. *J. Biol. Chem.* **281**, 25062–25075 (2006).
27. UniProt Consortium, UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Res.* **49**, D480–D489 (2020).
28. M. P. Ferla, W. M. Patrick, Bacterial methionine biosynthesis. *Microbiology (Reading)* **160**, 1571–1584 (2014).
29. B. Maurya, M. Colaço, J. Wouters, L. Pochet, S. Misquith, Key amino acid residues in homoserine-acetyltransferase from *M. tuberculosis* give insight into the evolution of MetX family of enzymes – HAT, SAT and HST. *Biochimie* **189**, 13–25 (2021).
30. K. Oda, Y. Matoba, T. Kumagai, M. Noda, M. Sugiyama, Crystallographic study to determine the substrate specificity of an L-serine-acetylating enzyme found in the D-cycloserine biosynthetic pathway. *J. Bacteriol.* **195**, 1741–1749 (2013).
31. D.-U. Kim, J. Hayles, D. Kim, V. Wood, H. O. Park, M. Won, H. S. Yoo, T. Duhig, M. Nam, G. Palmer, S. Han, L. Jeffery, S. T. Baek, H. Lee, Y. S. Shim, M. Lee, L. Kim, K. S. Heo, E. J. Noh, A. R. Lee, Y. J. Jang, K. S. Chung, S. J. Choi, J. Y. Park, Y. Park, H. M. Kim, S. K. Park, H. J. Park, E. J. Kang, H. B. Kim, H. S. Kang, H. M. Park, K. Kim, K. Song, K. B. Song, P. Nurse, K. L. Hoe, Analysis of a genome-wide set of gene deletions in the fission yeast *Schizosaccharomyces pombe. Nat. Biotechnol.* **28**, 617–623 (2010).
32. J. Brzywczy, A. Paszewski, Sulfur amino acid metabolism *in Schizosaccharomyces pombe*: Occurrence of two O-acetylhomoserine sulfhydrylases and the lack of the reverse transulfuration pathway. *FEMS Microbiol. Lett.* **121**, 171–174 (1994).
33. H. Ying, I. Cooke, S. Sprungala, W. Wang, D. C. Hayward, Y. Tang, G. Huttley, E. E. Ball, S. Forêt, D. J. Miller, Comparative genomics reveals the distinct evolutionary trajectories of the robust and complex coral lineages. *Genome Biol.* **19**, 175 (2018).
34. J. W. Scott, S. A. Hawley, K. A. Green, M. Anis, G. Stewart, G. A. Scullion, D. G. Norman, D. G. Hardie, CBS domains form energy-sensing modules whose binding of adenosine ligands is disrupted by disease mutations. *J. Clin. Invest.* **113**, 274–284 (2004).
35. M. Koutmos, O. Kabil, J. L. Smith, R. Banerjee, Structural basis for substrate activation and regulation by cystathionine beta-synthase (CBS) domains in cystathionine β-synthase. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 20958–20963 (2010).
36. T. Majtan, A. L. Pey, R. Fernández, J. A. Fernández, L. A. Martínez-Cruz, J. P. Kraus, Domain organization, catalysis and regulation of eukaryotic cystathionine beta-synthases. *PLOS ONE* **9**, e105290 (2014).
37. R. Evande, S. Ojha, R. Banerjee, Visualization of PLP-bound intermediates in hemeless variants of human cystathionine β-synthase: Evidence that lysine 119 is a general base. *Arch. Biochem. Biophys.* **427**, 188–196 (2004).
38. A. Kumar, A. Wißbrock, N. Goradia, P. Bellstedt, R. Ramachandran, D. Imhof, O. Ohlenschläger, Heme interaction of the intrinsically disordered N-terminal peptide segment of human cystathionine-β-synthase. *Sci. Rep.* **8**, 2474 (2018).

39. J. Oliveriusova, V. Kery, K. N. Maclean, J. P. Kraus, Deletion mutagenesis of human cystathionine beta-synthase—Impact on activity, oligomeric status, and *S*-adenosylmethionine regulation. *J. Biol. Chem.* **277**, 48386–48394 (2002).

40. R. Vozdek, A. Hnizda, J. Krijt, M. Kostrouchova, V. Kozich, Novel structural arrangement of nematode cystathionine β-synthases: Characterization of *Caenorhabditis elegans* CBS-1. *Biochem. J.* **443**, 535–547 (2012).

41. K.-H. Jhee, P. McPhie, E. W. Miles, Yeast cystathionine β-synthase is a pyridoxal phosphate enzyme but, unlike the human enzyme, is not a heme protein. *J. Biol. Chem.* **275**, 11541–11544 (2000).

42. G. Dixon, E. Abbott, M. Matz, Meta-analysis of the coral environmental stress response: Acropora corals show opposing responses depending on stress intensity. *Mol. Ecol.* **29**, 2855–2870 (2020).

43. R. Savary, D. J. Barshis, C. R. Voolstra, A. Cárdenas, N. R. Evensen, G. Banc-Prandi, M. Fine, A. Meibom, Fast and pervasive transcriptomic resilience and acclimation of extremely heat-tolerant coral holobionts from the northern Red Sea. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2023298118 (2021).

44. G. W. Vurture, F. J. Sedlazeck, M. Nattestad, C. J. Underwood, H. Fang, J. Gurtowski, M. C. Schatz, GenomeScope: Fast reference-free genome profiling from short reads. *Bioinformatics* **33**, 2202–2204 (2017).

45. G. Marcais, C. Kingsford, A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764–770 (2011).

46. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

47. L. Salmela, E. Rivals, LoRDEC: Accurate and efficient long read error correction. *Bioinformatics* **30**, 3506–3514 (2014).

48. S. Koren, B. P. Walenz, K. Berlin, J. R. Miller, N. H. Bergman, A. M. Phillippy, Canu: Scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Res.* **27**, 722–736 (2017).

49. M. J. Roach, S. A. Schmidt, A. R. Borneman, Purge haplotigs: Allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinform.* **19**, 460 (2018).

50. B. J. Walker, T. Abeel, T. Shea, M. Priest, A. Abouelliel, S. Sakthikumar, C. A. Cuomo, Q. Zeng, J. Wortman, S. K. Young, A. M. Earl, Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLOS ONE* **9**, e112963 (2014).

51. H. Li, Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).

52. H. Li, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997 (2013).

53. M. Boetzer, W. Pirovano, SSPACE-longRead: Scaffolding bacterial draft genomes using long read sequence information. *BMC Bioinform.* **15**, 211 (2014).

54. S. Yeo, L. Coombe, R. L. Warren, J. Chu, I. Birol, ARCS: Scaffolding genome drafts with linked reads. *Bioinformatics* **34**, 725–731 (2018).

55. G. C. Xu, T. J. Xu, R. Zhu, Y. Zhang, S. Q. Li, H. W. Wang, J. T. Li, LR_Gapcloser: A tiling path-based gap closer that uses long reads to complete genome assembly. *Gigascience* **8**, giy157 (2019).

56. S. F. Huang, M. J. Kang, A. L. Xu, HaploMerger2: Rebuilding both haploid sub-assemblies from high-heterozygosity diploid genome assembly. *Bioinformatics* **33**, 2577–2579 (2017).

57. K. J. Hoff, A. Lomsadze, M. Borodovsky, M. Stanke, Whole-genome annotation with BRAKER. *Methods Mol. Biol.* **1962**, 65–95 (2019).

58. K. J. Hoff, S. Lange, A. Lomsadze, M. Borodovsky, M. Stanke, BRAKER1: Unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics* **32**, 767–769 (2016).

59. A. Smit, R. Hubley, RepeatModeler Open-1.0 (2008-2015); www.repeatmasker.org.

60. A. Smit, R. Hubley, P. Green, RepeatMasker Open-4.0 (2013-2015); www.repeatmasker.org.

61. D. Kim, J. M. Paggi, C. Park, C. Bennett, S. L. Salzberg, Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).

62. I. Cooke, H. Ying, S. Forêt, P. Bongaerts, J. M. Strugnell, O. Simakov, J. Zhang, M. A. Field, M. Rodriguez-Lanetty, S. C. Bell, D. G. Bourne, M. J. van Oppen, M. A. Ragan, D. J. Miller, Genomic signatures in the coral holobiont reveal host adaptations driven by Holocene climate change and reef specific symbionts. *Sci. Adv.* **6**, eabc6318 (2020).

63. Y. J. Liew, M. Aranda, C. R. Voolstra, Reefgenomics.Org—A repository for marine genomics data. *Database (Oxford)* **2016**, 1–14 (2016).

64. H. Ying, D. C. Hayward, I. Cooke, W. Wang, A. Moya, K. R. Siemering, S. Sprungala, E. E. Ball, S. Forêt, D. J. Miller, The whole-genome sequence of the coral *Acropora millepora*. *Genome Biol. Evol.* **11**, 1374–1379 (2019).

65. NCBI Resource Coordinators, Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **44**, D7–D19 (2016).

66. P. Jones, D. Binns, H. Y. Chang, M. Fraser, W. Li, C. McAnulla, H. McWilliam, J. Maslen, A. Mitchell, G. Nuka, S. Pesseat, A. F. Quinn, A. Sangrador-Vegas, M. Scheremetjew,

67. S. Y. Yong, R. Lopez, S. Hunter, InterProScan 5: Genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).

67. M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, G. Sherlock, Gene ontology: Tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).

68. M. Kanehisa, Y. Sato, K. Morishima, BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* **428**, 726–731 (2016).

69. C. R. Voolstra, D. J. Miller, M. A. Ragan, A. Hoffmann, O. Hoegh-Guldberg, D. Bourne, E. E. Ball, H. Ying, S. Foret, S. Takahashi, K. D. Weynberg, M. van Oppen, K. Morrow, C. X. Chan, N. Rosic, W. Leggat, S. Sprungala, M. Imelfort, G. W. Tyson, K. Kassahn, P. Lundgren, P. Beeden, T. Ravasi, M. L. Berumen, E. Abel, T. Fyffe, K. M. Morrow, The ReFuGe 2020 Consortium-using "omics" approaches to explore the adaptability and resilience of coral holobionts to environmental change. *Front. Mar. Sci.* **2**, 1–8 (2015).

70. R. Cunning, R. A. Bay, P. Gillette, A. C. Baker, N. Traylor-Knowles, Comparative analysis of the *Pocillopora damicornis* genome highlights role of immune system in coral evolution. *Sci. Rep.* **8**, 16134 (2018).

71. C. R. Voolstra, Y. Li, Y. J. Liew, S. Baumgarten, D. Zoccola, J. F. Flot, S. Tambutté, D. Allemand, M. Aranda, Comparative analysis of the genomes of *Stylophora pistillata* and *Acropora digitifera* provides evidence for extensive differences between species of corals. *Sci. Rep.* **7**, 17583 (2017).

72. C. S. Wilding, N. Fletcher, E. K. Smith, P. Prentis, G. D. Weedall, Z. Stewart, The genome of the sea anemone *Actinia equina* (L.): Meiotic toolkit genes and the question of sexual reproduction. *Mar Genomics* **53**, 100753 (2020).

73. S. Baumgarten, O. Simakov, L. Y. Esherick, Y. J. Liew, E. M. Lehnert, C. T. Michell, Y. Li, E. A. Hambleton, A. Guse, M. E. Oates, J. Gough, V. M. Weis, M. Aranda, J. R. Pringle, C. R. Voolstra, The genome of *Aiptasia*, a sea anemone model for coral symbiosis. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 11893–11898 (2015).

74. N. H. Putnam, M. Srivastava, U. Hellsten, B. Dirks, J. Chapman, A. Salamov, A. Terry, H. Shapiro, E. Lindquist, V. V. Kapitonov, J. Jurka, G. Genikhovich, I. V. Grigoriev, S. M. Lucas, R. E. Steele, J. R. Finnerty, U. Technau, M. Q. Martindale, D. S. Rokhsar, Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science* **317**, 86–94 (2007).

75. Y. Jeon, S. G. Park, N. Lee, J. A. Weber, H. S. Kim, S. J. Hwang, S. Woo, H. M. Kim, Y. Bhak, S. Jeon, N. Lee, Y. Jo, A. Blazyte, T. Ryu, Y. S. Cho, H. Kim, J. H. Lee, H. S. Yim, J. Bhak, S. Yum, The Draft Genome of an Octocoral, *Dendronephthya gigantea*. *Genome Biol. Evol.* **11**, 949–953 (2019).

76. A. L. Mitchell, T. K. Attwood, P. C. Babbitt, M. Blum, P. Bork, A. Bridge, S. D. Brown, H. Y. Chang, S. el-Gebali, M. I. Fraser, J. Gough, D. R. Haft, H. Huang, I. Letunic, R. Lopez, A. Luciani, F. Madeira, A. Marchler-Bauer, H. Mi, D. A. Natale, M. Necci, G. Nuka, C. Orengo, A. P. Pandurangan, T. Paysan-Lafosse, S. Pesseat, S. C. Potter, M. A. Qureshi, N. D. Rawlings, N. Redaschi, L. J. Richardson, C. Rivoire, G. A. Salazar, A. Sangrador-Vegas, C. J. A. Sigrist, I. Sillitoe, G. G. Sutton, N. Thanki, P. D. Thomas, S. C. E. Tosatto, S. Y. Yong, R. D. Finn, InterPro in 2019: Improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res.* **47**, D351–D360 (2019).

77. B. Q. Minh, H. A. Schmidt, O. Chernomor, D. Schrempf, M. D. Woodhams, A. von Haeseler, R. Lanfear, IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).

78. D. T. Hoang, O. Chernomor, A. von Haeseler, B. Q. Minh, L. S. Vinh, UFBoot2: Improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **35**, 518–522 (2017).

79. I. Letunic, P. Bork, Interactive tree of life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296 (2021).

80. R. Daniel Gietz, R. A. Woods, in *Methods Enzymol.*, C. Guthrie, G. R. Fink, Eds. (Academic Press, 2002), vol. 350, pp. 87–96.

81. R. Hubley, R. D. Finn, J. Clements, S. R. Eddy, T. A. Jones, W. Bao, A. F. A. Smit, T. J. Wheeler, The Dfam database of repetitive DNA families. *Nucleic Acids Res.* **44**, D81–D89 (2016).

82. W. D. Bao, K. K. Kojima, O. Kohany, Repbase update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA* **6**, 11 (2015).

83. M. Stanke, M. Diekhans, R. Baertsch, D. Haussler, Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **24**, 637–644 (2008).

84. M. Stanke, O. Schoffmann, B. Morgenstern, S. Waack, Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics* **7**, 62 (2006).

85. L. Li, C. J. Stoeckert, D. S. Roos, OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Res.* **13**, 2178–2189 (2003).

86. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).

87. The Gene Ontology Consortium, The gene ontology resource: 20 years and still going strong. *Nucleic Acids Res.* **47**, D330–D338 (2019).

88. S. Maere, K. Heymans, M. Kuiper, BiNGO: A cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* **21**, 3448–3449 (2005).

89. P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).

90. F. Supek, M. Bosnjak, N. Skunca, T. Smuc, REVIGO summarizes and visualizes long lists of gene ontology terms. *PLOS ONE* **6**, e21800 (2011).

91. R. C. Edgar, MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).

92. J. Castresana, Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17**, 540–552 (2000).

93. G. Talavera, J. Castresana, Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* **56**, 564–577 (2007).

94. D. Darriba, D. Posada, A. M. Kozlov, A. Stamatakis, B. Morel, T. Flouri, ModelTest-NG: A new and scalable tool for the selection of DNA and protein evolutionary models. *Mol. Biol. Evol.* **37**, 291–294 (2020).

95. A. M. Kozlov, D. Darriba, T. Flouri, B. Morel, A. Stamatakis, RAxML-NG: A fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**, 4453–4455 (2019).

96. Z. L. Fuller, V. J. L. Mocellin, L. A. Morris, N. Cantin, J. Shepherd, L. Sarre, J. Peng, Y. Liao, J. Pickrell, P. Andolfatto, M. Matz, L. K. Bay, M. Przeworski, Population genetics of the coral *Acropora millepora*: Toward genomic prediction of bleaching. *Science* **369**, eaba4674 (2020).

97. N. I. Weisenfeld, V. Kumar, P. Shah, D. M. Church, D. B. Jaffe, Direct determination of diploid genome sequences. *Genome Res.* **27**, 757–767 (2017).

98. R. Kajitani, K. Toshimoto, H. Noguchi, A. Toyoda, Y. Ogura, M. Okuno, M. Yabana, M. Harada, E. Nagayasu, H. Maruyama, Y. Kohara, A. Fujiyama, T. Hayashi, T. Itoh, Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome Res.* **24**, 1384–1395 (2014).

99. G. Zhang, X. Fang, X. Guo, L. Li, R. Luo, F. Xu, P. Yang, L. Zhang, X. Wang, H. Qi, Z. Xiong, H. Que, Y. Xie, P. W. H. Holland, J. Paps, Y. Zhu, F. Wu, Y. Chen, J. Wang, C. Peng, J. Meng, L. Yang, J. Liu, B. Wen, N. Zhang, Z. Huang, Q. Zhu, Y. Feng, A. Mount, D. Hedgecock, Z. Xu, Y. Liu, T. Domazet-Lošo, Y. Du, X. Sun, S. Zhang, B. Liu, P. Cheng, X. Jiang, J. Li, D. Fan, W. Wang, W. Fu, T. Wang, B. Wang, J. Zhang, Z. Peng, Y. Li, N. Li, J. Wang, M. Chen, Y. He, F. Tan, X. Song, Q. Zheng, R. Huang, H. Yang, X. Du, L. Chen, M. Yang, P. M. Gaffney, S. Wang, L. Luo, Z. She, Y. Ming, W. Huang, S. Zhang, B. Huang, Y. Zhang, T. Qu, P. Ni, G. Miao, J. Wang, Q. Wang, C. E. W. Steinberg, H. Wang, N. Li, L. Qian, G. Zhang, Y. Li, H. Yang, X. Liu, J. Wang, Y. Yin, J. Wang, The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* **490**, 49–54 (2012).

100. J. T. Simpson, Exploring genome characteristics and sequence quality without a reference. *Bioinformatics* **30**, 1228–1235 (2014).

101. K. I. Savage, D. P. Harkin, BRCA1, a 'complex' protein involved in the maintenance of genomic stability. *FEBS J.* **282**, 630–646 (2015).

102. S. Kim, W. Zeng, S. Bernard, J. Liao, M. Venkateshwaran, J. M. Ane, Y. Jiang, Publisher correction: Ca2$^+$regulated Ca^{2+} channels with an RCK gating ring control plant symbiotic associations. *Nat. Commun.* **10**, 3703–4607 (2019).

103. Y. P. Zhang, E. L. Pohlmann, J. Serate, M. C. Conrad, G. P. Roberts, Mutagenesis and functional characterization of the four domains of GlnD, a bifunctional nitrogen sensor protein. *J. Bacteriol.* **192**, 2711–2721 (2010).

104. E. Garcia, S. G. Rhee, Cascade control of escherichia-coli glutamine-synthetase - purification and properties of Pii uridylyltransferase and uridylyl-removing enzyme. *J. Biol. Chem.* **258**, 2246–2253 (1983).

105. R. Vozdek, A. Hnízda, J. Krijt, L. Šerá, V. Kožich, Biochemical properties of nematode O-acetylserine(thiol)lyase paralogs imply their distinct roles in hydrogen sulfide homeostasis. *Biochim. Biophys. Acta* **1834**, 2691–2701 (2013).

106. D. K. Ma, R. Vozdek, N. Bhatla, H., CYSL-1 interacts with the O2-sensing hydroxylase EGL-9 to promote H2S-modulated hypoxia-induced behavioral plasticity in C. elegans. *Neuron* **73**, 925–940 (2012).

107. M. W. Budde, M. B. Roth, The response of caenorhabditis elegans to hydrogen sulfide and hydrogen cyanide. *Genetics* **189**, 521–532 (2011).