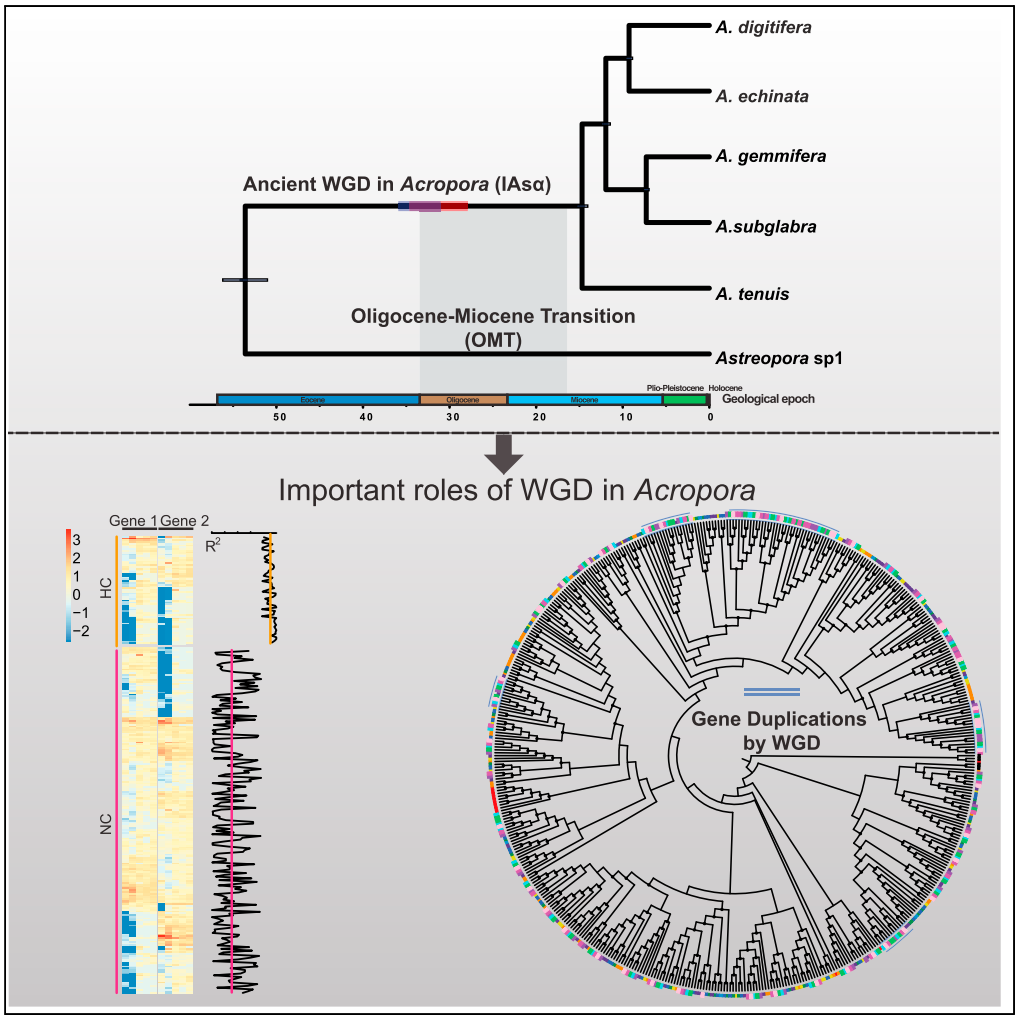
iScience

CellPress

Article

A Likely Ancient Genome Duplication in the Speciose Reef-Building Coral Genus, *Acropora*



Yafei Mao,
Noriyuki Satoh

yafei.mao@oist.jp

HIGHLIGHTS

An ancient genome duplication occurred in the most recent common ancestor of *Acropora*

This WGD event likely occurred between 28 and 36 mya in *Acropora*

The WGD event potentially contributes to the origin and diversification of *Acropora*

Duplications of toxic proteins were found in *Acropora* following the WGD

Mao & Satoh, iScience *13*, 20–32
March 29, 2019
https://doi.org/10.1016/j.isci.2019.02.001

iScience

CellPress

Article

A Likely Ancient Genome Duplication in the Speciose Reef-Building Coral Genus, *Acropora*

Yafei Mao[1,2,*] and Noriyuki Satoh[1]

SUMMARY

Whole-genome duplication (WGD) has been recognized as a significant evolutionary force in the origin and diversification of multiple organisms. *Acropora*, a speciose reef-building coral genus, is suspected to have originated by polyploidy. Yet, there is no genetic evidence to support this hypothesis. Using comprehensive phylogenomic and comparative genomic approaches, we analyzed six *Acroporid* genomes and found that a WGD event likely occurred ~31 million years ago in the most recent common ancestor of *Acropora*, concurrent with a worldwide coral extinction. We found that duplicated genes were highly enriched in gene regulation functions, including those of stress responses. The functional clusters of duplicated genes are related to the divergence of gene expression patterns during development. Some proteinaceous toxins were generated by WGD in *Acropora* compared with other cnidarian species. Collectively, this study provides evidence for an ancient WGD event in corals, which helps explain the origin and diversification of *Acropora*.

INTRODUCTION

Reef-building corals contribute to tropical marine ecosystems that support innumerable marine organisms, but reefs are increasingly threatened because of recent increases in seawater temperatures, pollution, and other stressors (Ainsworth et al., 2016; Renema et al., 2016). The Acroporidae is a family of reef-building corals within the phylum Cnidaria, one of the basal phyla of the animal clade (Richards et al., 2013; Wallace, 2012; Wallace and Rosen, 2006). *Astreopora* (Anthozoa: Acroporidae) is the sister genus in the acroporid lineage according to fossil records and molecular phylogenetic evidence (Fukami et al., 2000; Suzuki and Nomura, 2013; Wallace, 2012). Importantly, *Acropora* (Anthozoa: Acroporidae), one of the most diverse genera of reef-building corals, including more than 150 species in the Indo-Pacific Ocean, is thought to have originated from *Astreopora* ~60 mya with several species turnovers (Edinger and Risk, 1994; Renema et al., 2008; Wallace, 2012; Wallace and Rosen, 2006). Investigating the evolutionary history of this group importantly contributes to our understanding of coral reef biodiversity and conservation. Hybridization among *Acropora* species has been observed in the wild (Vollmer and Palumbi, 2002), and variable chromosome numbers have been determined in different *Acropora* lineages (Kenyon, 1997). In addition, gene duplications have been shown in several *Acropora* gene families (Gacesa et al., 2015; Hamada et al., 2013). Based on their unique lifestyle, variable chromosome numbers, and complicated reticular evolutionary history, Indo-Pacific *Acropora* likely originated via polyploidy (Gacesa et al., 2015; Hamada et al., 2013; Kenyon, 1997; Richards and Hobbs, 2015; Van Oppen et al., 2001; Vollmer and Palumbi, 2002; Willis et al., 2006). However, there is no direct molecular and genetic evidence to support this hypothesis.

Ancient whole (large-scale)-gene/genome duplication (WGD), or paleopolyploidy, has shaped the genomes of vertebrates, green plants, and other organisms, and is usually regarded as an evolutionary landmark in the origin and diversification of organisms (Soltis et al., 2015; Van de Peer et al., 2009; Van De Peer et al., 2017) (Figure S1). Two separate WGD events have been documented in the common ancestors of vertebrates (two rounds of WGD) (Dehal and Boore, 2005) and another major WGD has been reported in the last common ancestor of teleost fish (Christoffels et al., 2004; Glasauer and Neuhauss, 2014). Meanwhile, living angiosperms share an ancient WGD event (Jiao et al., 2011; Tiley et al., 2016), and many other WGD events have been reported in major clades of angiosperms (Soltis et al., 2009; Vanneste et al., 2014; Wang et al., 2018). In addition, two rounds of WGDs in the vertebrates are suggested to have occurred during the Cambrian Period, and some WGDs in plants are believed to have occurred during Cretaceous-Tertiary (Smith et al., 2013; Van De Peer et al., 2017; Vanneste et al., 2014). Thus, WGD is

[1]Marine Genomics Unit, Okinawa Institute of Science and Technology Graduate University, Onna, Okinawa 904-0495, Japan

[2]Lead Contact

*Correspondence: yafei.mao@oist.jp

https://doi.org/10.1016/j.isci.2019.02.001

regarded as an important evolutionary way to reduce the risk of extinction (Van de Peer et al., 2009; Van De Peer et al., 2017; Vanneste et al., 2014). However, the study of WGD in Cnidaria has received less attention (Kenny et al., 2016; Li et al., 2018; Schwager et al., 2017; Van de Peer et al., 2009; Van De Peer et al., 2017).

Duplicated genes created by WGD have complex fates during diploidization (Sémon and Wolfe, 2007; Van de Peer et al., 2009). Usually, one of the duplicated genes is silenced or lost due to redundancy of gene functions, termed "nonfunctionalization." However, retained duplicated genes provide important sources of biological complexity and evolutionary novelty due to subfunctionalization, neofunctionalization, and dosage effects (Conant et al., 2014; Jiao et al., 2011). Duplicated genes may develop complementary gene functions via subfunctionalization, evolve new functions through neofunctionalization, or are retained in complicated regulatory networks with different gene expressions due to dosage effects. For instance, duplicated MADS-Box genes are crucial for flower development and the origin of phenotypic novelty in plants (Van de Peer et al., 2009; Veron et al., 2006). Duplicated homeobox genes provide raw genetic material for vertebrate development (Canestro et al., 2013; Glasauer and Neuhauss, 2014). In addition, toxin diversification following gene duplications has been recognized as a mechanism to enhance adaptation in animals (Kondrashov, 2012; Kordiš and Gubenšek, 2000), especially in snake venoms (Hargreaves et al., 2014; Vonk et al., 2013). Interestingly, toxic proteins are involved in various important processes in corals, including prey capture, protection from predators, wound healing, etc (Armoza-Zvuloni et al., 2016; Ben-Ari et al., 2018), but it is still unclear how gene duplications of toxic proteins evolved in corals.

Isozyme electrophoresis and restriction fragment length polymorphism were used to identify gene duplications in polyploids a few decades ago (Fürthauer et al., 1999; Stuber and Goodman, 1983). In the past 10 years, next-generation sequencing has generated a wealth of genomic data at vastly decreased cost and reduced efforts (Goodwin et al., 2016; Hardwick et al., 2017). Three main methods were developed to identify WGD, based on (1) analysis of the rate of synonymous substitutions per synonymous site (dS) of duplicated genes within a genome (dS-based method) (Blanc et al., 2003; Lynch and Conery, 2000; Vanneste et al., 2014; Mao, 2019), (2) phylogenetic analysis of gene families among multiple genomes (phylogenomic analysis) (Blomme et al., 2006; Jiao et al., 2011), and (3) synteny block identification compared with sister lineages without WGD (synteny analysis) (Bowers et al., 2003; Dehal and Boore, 2005; Zhang et al., 2017). The dS-based method and phylogenomic analysis only require gene family information, without genome assembly. However, too ancient WGD cannot be detected by the dS-based method, while gene tree uncertainty usually causes bias in the phylogenomic analysis. Both methods rely heavily on gene family estimation and clustering. Inaccurate gene predictions (gene models) and rough gene family cluster algorithms can easily fail to detect WGD using either method. In contrast, the synteny analysis relies heavily on genome assembly quality. Poor assembly quality can hide the WGD signals, and some genomes with huge rearrangements cannot be used to detect WGD using synteny block identification. Therefore, the most credible conclusions depend on complementary evidence from different methods (Chen and Birchler, 2013; Soltis and Soltis, 2012; Tiley et al., 2016).

Here, using all three methods, we analyzed a genome of *Astreopora* (*Astreopora* sp1) as an outgroup, and five *Acropora* genomes (*Acropora digitifera*, *Acropora gemmifera*, *Acropora subglabra*, *Acropora echinata*, and *Acropora tenuis*) to address the following questions: (1) whether and when WGD occurred in *Acropora*, (2) what is the fate of duplicated genes in *Acropora* after the event, (3) what are the gene expression patterns of duplicated genes across five developmental stages in *A. digitifera*, and (4) what is the role of WGD in diversification of toxic proteins in *Acropora* (Figure S2).

RESULTS

Calibration of the Acroporid Phylogenomic Tree

We clustered all homologs of the six *Acroporid* species and found that 6,520 gene families are shared among 19,760 gene families in total (Figure S3, see Data S1). Our previous gene family cluster analysis of the five *Acropora* species showed that each *Acropora* genome had very few unique gene families (<100) (Mao et al., 2018). Yet, we found that *Astreopora* sp1 had 836 unique gene families, suggesting that *Astreopora* sp1 is genetically divergent from the five *Acropora* species. A total of 3,461 single-copy orthologs selected from 6,520 shared gene families were concatenated to reconstruct a calibrated phylogenomic tree based on the reported divergence time of *Acropora* (Mao et al., 2018). We found that *Astreopora* sp1 splitted from *Acropora* ~53.6 mya (95% highest posterior density: 51.02–56.21 Ma) (Figures 1 and S4). This result established a timescale to analyze the timing of the subsequent WGD.

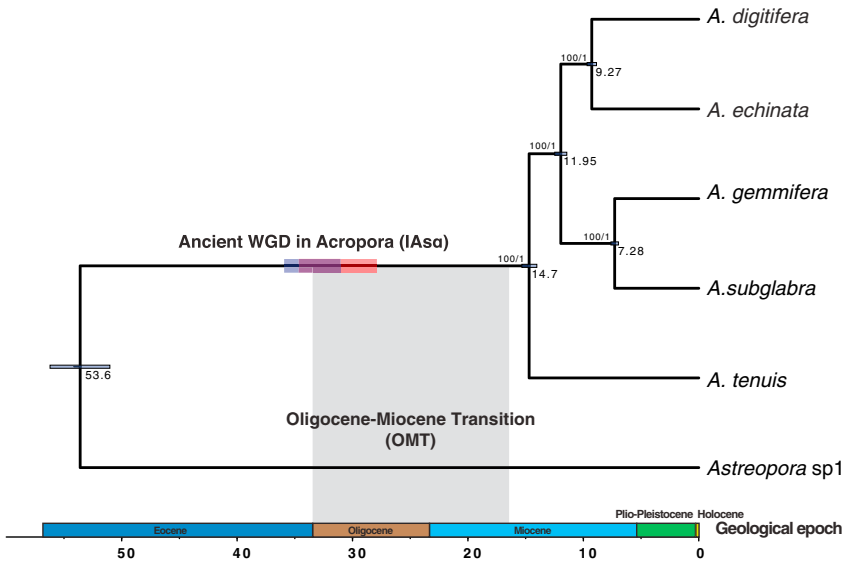


Figure 1. Ancient WGD in the Reef-Building Coral *Acropora* (IAsα)

A calibrated phylogenomic tree of six *Acroporid* species inferred from 3,461 single-copy orthologs using BEAST2. Horizontal bars on branches of the tree represent the timing of WGD in *Acropora*. The timing of IAsα was estimated at 35 mya (95% confidence interval: 31.18–35.7 mya) and 31 mya (95% confidence interval: 27.86–34.77 mya) by the dS-based method (horizontal purple bar) and phylogenomic analysis (horizontal orange bar), respectively. Gray shading represents the timing of one coral species turnover event, the Oligocene-Miocene transition (OMT), suggesting that IAsα is correlated with OMT. See also Figures S1, S3, and S4, Tables S5 and S6.

WGD Identification with the dS-Based Method

Synonymous substitution rate (dS) analysis has been widely used to infer WGD (Vanneste et al., 2012, 2014). We identified over 10,000 paralogous gene pairs, based on their sequence similarities, and over 10,000 anchor gene pairs, based on synteny information of each species (Table S1; see Methods). Using paralogous and anchor gene pairs, we calculated their dS values and reconstructed dS distributions for each species.

An "L-shaped" distribution was evident in both paralogous and anchor gene pair dS distributions of *Astreopora* sp1, illustrating that no WGD occurred in *Astreopora* sp1. However, all five *Acropora* species displayed a similar peak in dS distributions of both paralogous and anchor gene pairs (peak: 0–0.3), suggesting that WGD did occur in *Acropora* (Figures 2A, S5, and S6).

dS values of orthologous gene pairs between two pairs of species (*Astreopora* sp1 and *A. tenuis*; *A. tenuis*, and *A. digitifera*) were estimated as the proxy of speciation time between them according to neutral evolution theory (Berthelot et al., 2014; Zhang et al., 2017). We combined the dS values of paralogous gene pairs of the five *Acropora* species and estimated the peak in the log dS distribution (modal value = −1.82). We also estimated the distribution of orthologous gene pairs between *Astreopora* sp1 and *A. tenuis* (modal value = −0.31) and between *A. tenuis* and *A. digitifera* (modal value = −3.46). The result indicates that the WGD occurred in *Acropora* before the split of *A. tenuis* and *A. digitifera* and after the split of *A. tenuis* and *Astreopora* sp1 (Figure S7). In other words, an ancient WGD event likely occurred in the most recent common ancestor of *Acropora*. Based on speciation time estimated in the calibrated phylogenomic tree and the dS-based method (Vanneste et al., 2014), we estimated that the WGD of *Acropora* occurred ~35 mya (95% confidence interval: 31.18–35.7 Ma) (Table S2, see Methods). Here, we defined this event as invertebrate α event of WGD specifically in *Acropora* (IAsα).

Phylogenomic and Synteny Analysis of IAsα

If the proposed IAsα is deemed correct, the ohnologs of *Acropora* (paralogs created by IAsα) should form two clades from their orthologs in *Astreopora* sp1 by mapping IAsα onto phylogenetic trees (Jiao et al., 2011; Marcet-Houben and Gabaldón, 2015). In other words, the phylogenetic topology would be

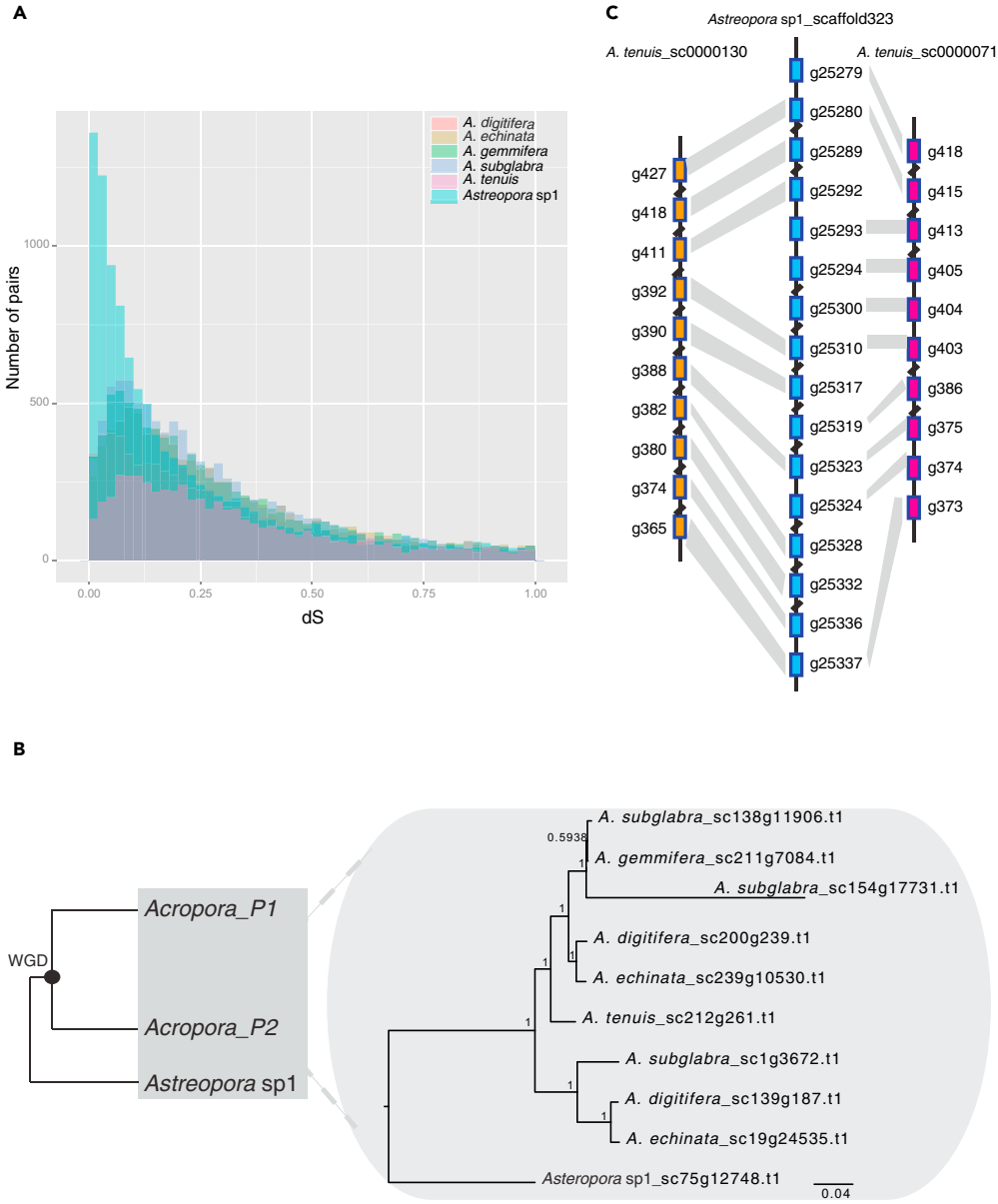


Figure 2. Ancient WGD Identification (IAsα) and Timing of the Event in *Acropora*

(A) Frequency distribution of dS values for paralogous gene pairs in five *Acropora* and one *Astreopora* species showing that a WGD event occurred in *Acropora*. Similar peaks (dS value: 0.1–0.3) in dS distributions of the five *Acropora* lineages indicate that a WGD event occurred in *Acropora*. (Light red, *A. digitifera*; light yellow, *A. echinata*; light green, *A. gemmifera*; light blue, *A. subglabra*; light purple: *A. tenuis*; light cyan, *Astreopora* sp1).

(B) Hypothetical tree topology of duplicated genes in the *Acroporidae* and the phylogeny of one duplicated gene (alpha-protein kinase 1-like). The phylogenetic tree shows gene retention, loss, and duplications following WGD.

(C) Co-linear gene alignments of *Astreopora* sp1 and *A. tenuis* on scaffolds. The gray links show orthologs between *Astreopora* sp1 and *A. tenuis*.

See also Figures S5–S10 and Tables S1–S3.

(((*Acropora clade1*) bootstrap1, (*Acropora clade2*) bootstrap2), *Astreopora* sp1), defined as gene duplication topology (Figure 2B).

We performed a phylogenomic analysis to further evaluate the proposed IAsα. Firstly, we defined orthogroups as clusters of homologous genes in *Acropora* derived from a single gene in *Astreopora* sp1.

Each orthogroup contained at least seven homologous genes, including at least one gene copy in each *Acropora* species and only one gene copy in *Astreopora* sp1. We selected 883 orthogroups from 19,760 gene families, and reconstructed the phylogeny of 883 orthogroups using both maximum likelihood (ML) and Bayesian methods. We found that the phylogenetic topology of 205 orthogroups was consistent with gene duplication topology supporting IAsα. We further defined the 205 orthogroups as core-orthogroups (Table S3, see Data S1).

In particular, we found differential gene loss, retention, and duplication in *Acropora* lineages. For instance, the phylogeny of orthogroup 1370 (alpha-protein kinase 1-like) showed gene retention in *A. subglabra*, *A. digitifera*, and *A. echinata*; gene loss in *A. tenuis*; and an extra gene duplication in *A. subglabra*. This implies that diversification of duplicated genes may contribute to species complexity and evolutionary innovation in *Acropora* (Glasauer and Neuhauss, 2014) (Figure 2B).

To estimate the split time of the two *Acropora* clades that could be regarded as the timing of IAsα, we selected 154 high-quality core-orthogroups, with bootstrap values in both *Acropora* clades >70 in ML phylogeny, to reconstruct a time-calibrated phylogeny from the 205 core-orthogroups using BEAST2 (Jiao et al., 2011). However, we found that it was difficult for the parameters in MCMC to converge in 70 core-orthogroups, and we only successfully dated the phylogenetic trees of 135 high-quality core-orthogroups (see Data S1). Next, we estimated the distribution of inferred node ages between the two *Acropora* clades, and found that the peak value was estimated to be 31 Ma (95% confidence interval: 27.86–34.77 Ma), indicating that IAsα occurred at 31 Ma (Figure S8). This result is in coincidence with the timing of the IAsα estimated by the dS-based method.

Intergenomic co-linearity is often used to directly identify ancient WGD and to reconstruct ancestral karyotypes in vertebrates (Berthelot et al., 2014; Nakatani et al., 2007; Zhang et al., 2017). We performed intergenomic co-linearity and synteny analysis between *Astreopora* sp1 and *A. tenuis* to support IAsα. We found that synteny blocks of 21 scaffolds in *Astreopora* sp1 have at least two duplicated segments in *A. tenuis* (Figure S9). For example, two duplicated segments in scaffold 130 and scaffold 70 of *A. tenuis* corresponded to scaffold 323 in *Astreopora* sp1 (Figure 2C). In addition, we also observed that duplicated segments of the five *Acropora* species, which corresponded to the longest scaffold of *Astreopora* sp1, had good co-linearity (Figure S10).

In summary, we found evidence to support IAsα with the dS-based method, and phylogenomic and synteny analyses. Moreover, we suggest that IAsα probably occurred between 28 and 36 mya (Figure 1).

The Fate of Duplicated Genes Originating from IAsα

Duplicated genes provide substrates for diversification and evolutionary novelty, and most of them are regulators of gene networks in vertebrates and plants (Jiao et al., 2011; Kassahn et al., 2009; Zhang et al., 2017). We examined Gene Ontology (GO) for all genes among the 154 high-quality core-orthogroups to investigate their roles in IAsα and found that their molecular functions are enriched in specific categories: transporter, catalytic, binding, and receptor activity. Some of those are involved in gene regulation (Table 1).

Furthermore, we identified some duplicated genes under subfunctionalization and neofunctionalization, possibly contributing to stress responses of corals. dnaJ homolog subfamily B member 11-like (DNAJB) protein was shown to be involved in heat stress responses in marine organisms (Fujikawa et al., 2010; Wang et al., 2014). Orthogroup 1247 (DNAJB) has two main domains (Ras and Dnaj domains) in *Astreopora* sp1 representing the ancient state. Each of the two domains was independently lost in the duplicated genes, resulting in complementary functions of the duplicated genes after IAsα (Figures 3A and S11). In addition, excitatory amino acid transporters may be related to symbiotic interactions in *Acropora* (Bertucci et al., 2015). Orthogroup 1244 (excitatory amino acid transporter 1-like) was predicted as a six transmembrane protein, and a high number of mutations have accumulated in both untransmembrane and transmembrane regions, suggesting that new functions would be generated (Figures 3B and S12). These examples suggest that IAsα might participate in stress responses and symbiotic interactions of *Acropora*. Together, these results agree with previous patterns of the fate of duplicated genes in vertebrates and plants (Jiao et al., 2011; Soltis et al., 2015; Van De Peer et al., 2017; Zhang et al., 2017), indicating that the IAsα possibly contributes to the species complexity and diversification in *Acropora*.

Annotation Cluster	P Value
Transmembrane	1.90×10^{-6}
Death domain	3.10×10^{-5}
G-protein-coupled receptor	1.20×10^{-4}
VIT domain	3.30×10^{-3}
Protein kinase-like domain	1.90×10^{-2}

Table 1. Functional Annotation Clustering on the GO Terms of 154 High-Quality Core-Orthogroups

Gene Expression Patterns of Duplicated Genes across Five Developmental Stages in *A. digitifera*

To better understand the evolution of duplicated genes, gene expression analysis across five developmental stages in *A. digitifera* (blastula, gastrula, postgastrula, planula, and adult polyps) was carried out (Reyes-Bermudez et al., 2016). We identified 236 ohnologous pairs in *A. digitifera* from 883 ML phylogeny (see Methods) and found that these ohnologous pairs presented an interesting gene expression profiling. We divided 236 ohnologous pairs into two clusters based on pairwise correlation of gene expression during development (high correlation [HC]: $p < 0.05$; no correlation [NC] $p \geq 0.05$; Pearson's correlation test): 25% (25/236) ohnologous pairs in HC and 75% (211/236) ohnologous pairs in NC (Figure 4A). Ohnologous pairs in the HC cluster are enriched in protein kinase, whereas ohnologous pairs in the NC cluster are enriched in membrane transporter and ion-binding proteins (Figure 4B). This result indicates that the two clusters of ohnologous pairs potentially evolved into different gene functions. In addition, we compared dN/dS values to investigate selective pressure on HC and NC clusters (Figure 4C), but there is no significant difference between the two clusters (Mann-Whitney-Wilcoxon test, $p = 0.51$).

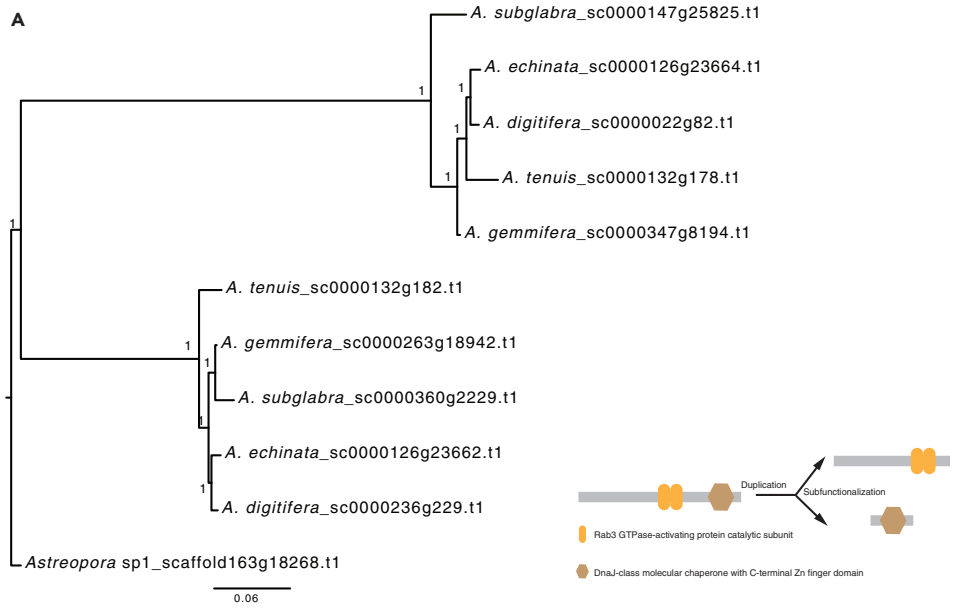
Evolution of Toxic Proteins in Cnidaria

Next, we investigated the role of IAsα in the diversification of toxins in *Acropora*. We identified ~200 putative toxic proteins in each of the five *Acropora* species, and we clustered them with the putative toxic proteins of *Astreopora* sp1 and other six Cnidarian species (*Hydra magnipapillata*, *Nematostella vectensis*, *Montastraea cavernosa*, *Porites australiensis*, *Porites astreoides*, *and Porites lobata*) into 24 gene families (Table S4, See Methods). Based on the gene family phylogeny, each of which contains at least 15 genes (See Data S1), we found that toxic proteins have undergone widespread gene duplications in Cnidaria, and most of the gene duplications occurred in individual species lineages, except for *Acropora* (Figures 5 and S13–S20). Gene duplications occurred in the most recent common ancestor of *Acropora* in 9 of 15 gene families, potentially caused by WGD (IAsα). For example, in gene family 1 (coagulation factor X), each species contains ~50 genes, except *H. magnipapillata* and *P. astreoides*. Gene duplications occurred frequently in individual species lineages: *Astreopora* sp1, *M. cavernosa*, *N. vectensis*, and *P. australiensis*. Yet, five gene duplications are inferrred to have occurred in the most recent common ancestor of *Acropora* by WGD (Figure 5). These results indicated that IAsα potentially contributed to the diversification of proteinaceous toxins in *Acropora*.

DISCUSSION

Ancient WGD is considered as a significant evolutionary factor in the origin and diversification of evolutionary lineages (Soltis et al., 2015; Van De Peer et al., 2017), but much work remains to be done to definitively identify WGD and to understand its consequences in different evolutionary lineages. Staghorn corals of the genus *Acropora*, which constitute the foundation of modern coral reef ecosystems, are hypothesized to have originated through polyploidization (Kenyon, 1997; Renema et al., 2016; Willis et al., 2006). However, there is no genetic evidence to support this assertion. To that end, we analyzed the genomes of one *Astreopora* and five *Acropora* species to address the possibility of WGD in *Acropora* and the functional fate of duplicated genes from that event.

To the best of our knowledge, this is the first study to report genomic-scale evidence of WGD in corals (IAsα). We find that large numbers of ohnologs are retained in *Acropora* species and hundreds of gene families display phylogenetic duplication topology among the five *Acropora* species. Meanwhile, our synteny analysis between *Astreopora* sp1 and *A. tenuis* directly supports IAsα. However, reconstruction of the



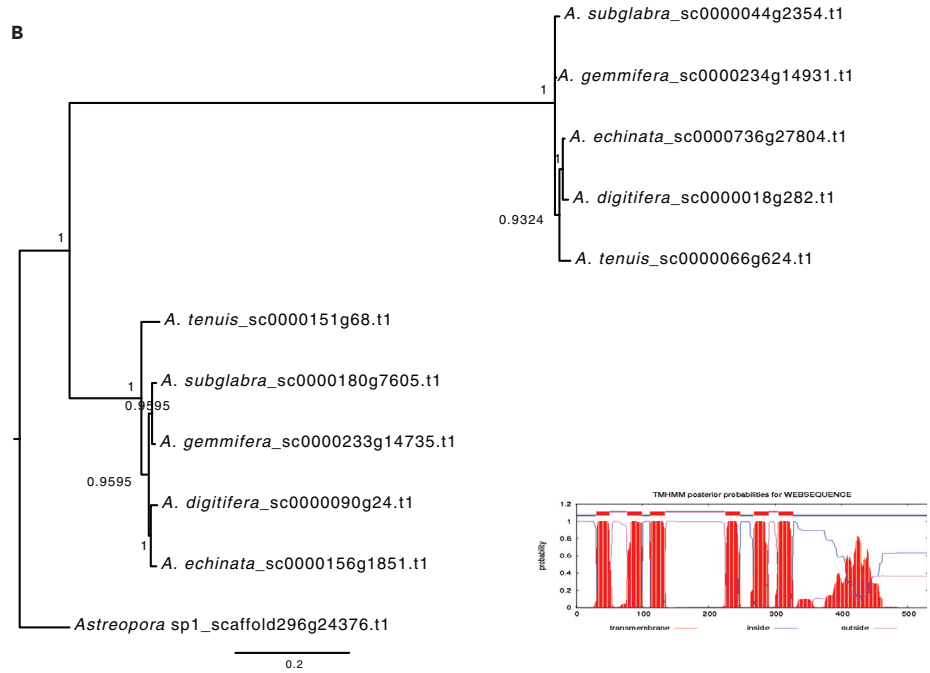


Figure 3. Duplicated Genes under Subfunctionalization or Neofunctionalization are shown with phylogenetic trees.

(A) The phylogeny of orthogroup 1247 (dnaJ homolog subfamily B member 11-like) reconstructed with MrBayes shows a duplicated gene under subfunctionalization. Bayesian posterior probabilities are shown at each node. The bottom right panel shows that two domains are in *Astreopora* sp1, but each domain was independently lost in duplicated genes under subfunctionalization in orthogroup 1247.

Figure 3. *Continued*

(B) The phylogeny of orthogroup 1244 (excitatory amino acid transporter 1-like) reconstructed with MrBayes shows a duplicated gene under neofunctionalization. Bayesian posterior probabilities are shown at each node. Six transmembrane helices prediction is shown at the bottom right.

See also Figures S11 and S12.

ancestral karyotype will necessitate the assemblage of genomes to the chromosome level to fully understand gene fractionation and chromosome rearrangements in *Acropora* under IAsα (Smith and Keinath, 2015; Smith et al., 2013).

Ancient WGD is usually inferred using the dS-based method, but artificial signals in dS distributions have been reported in previous studies, because of dS saturation (dS value >1) or using poorly annotated genomes or allelic variations or low retention rates (Rabier et al., 2014; Tiley et al., 2016; Vanneste et al., 2012; Mao, 2019). There is an extra peak in the dS distribution of anchor gene pairs in *A. digitifera* and *A. tenuis* (Figure S6). One possible explanation is that the extra peak is artifactual because few anchor gene pairs were used in the analysis. However, this could also indicate a second WGD event in *Acropora*. We found few orthogroups with topologies that fit the two proposed WGD events (Figure S21). In addition, a new ML phylogeny modeling approach was recently developed to overcome shortcomings of the dS-based method (Rabier et al., 2014; Tiley et al., 2016). We used it to test whether a second WGD occurred in *Acropora*. The result showed that one WGD event is the best model in *Acropora* and that it occurred 30.69–34.69 mya (Tables S5 and S6; see Methods). Moreover, we applied the dS-based method on two recently released genomes by other groups (*Acropora tenuis* and *Acropora millepora*) (see Methods). We found similar peaks on these two dS distributions of paralogous gene pairs (Figure S22). Together, we have genome-scale evidence to support IAsα, yet, there is no conclusive evidence to support a second WGD in *Acropora*.

It is crucial to accurately estimate the timing of a WGD event to understand its evolutionary consequences (Jiao et al., 2011; Vanneste et al., 2014). Our study has clearly estimated the timing of IAsα using both phylogenomic analysis and the dS-based method. We suggest that IAsα probably occurred between 28 and 36 mya (Figure 1). Interestingly, species turnover events usually occurred with extinctions (Jackson and Sax, 2010), and one species turnover event in corals (Oligocene-Miocene transition [OMT]) was suggested to have occurred between 15.97 and 33.7 mya (Edinger and Risk, 1994). The timing of IAsα may correspond to a massive extinction of corals created by OMT. This finding supports the hypothesis that WGD may enable organisms to escape extinction during drastic environmental changes (Van De Peer et al., 2017) (Figure 1).

The occurrence of IAsα raises the question of what impact it may have had upon coral evolution (Conant et al., 2014; Willis et al., 2006). We performed GO analysis on duplicated genes and examined duplicated gene families, showing that duplicated genes following IAsα indeed provided raw genetic material for *Acropora* to diversify and are potentially crucial for stress responses. In particular, toxin diversification in *Acropora* was mainly generated by WGD. In addition, we focused on expression patterns of duplicated genes in *A. digitifera*, showing that expressions of duplicated protein kinases are likely to be correlated during development. A possible explanation may be that protein kinases are probably retained in complex signal transduction pathways via subfunctionalization or dosage effects (Conant et al., 2014; Glasauer and Neuhauss, 2014). However, expressions of duplicated membrane proteins are likely uncorrelated because these proteins may have developed different functions via neofunctionalization, such as excitatory amino acid transporters (orthogroup 1244). However, more work is needed to be done to investigate molecular mechanisms of duplicated genes in order to examine these hypotheses in the diversification of *Acropora* (Yasuoka et al., 2016). For instance, previous gene functional studies have demonstrated that voltage-gated sodium channel gene paralogs, duplicated in teleosts, contributed to the acquisition of new electric organs via neofunctionalization in both mormyroid and gymnotiform electric fishes (Arnegard et al., 2010; Zakon et al., 2006).

Our previous study proposed that adaptive radiation in *Acropora* was probably driven by introgression (Mao et al., 2018); thus *Acropora* is the first invertebrate lineage reported to have undergone both WGD and introgression. Meanwhile, both introgression and WGD have also been reported in cichlid fish lineages (Berner and Salzburger, 2015), a famous model of adaptive radiation in vertebrates (Berner and Salzburger, 2015; Seehausen et al., 2014). Both WGD and introgression are regarded as significant forces

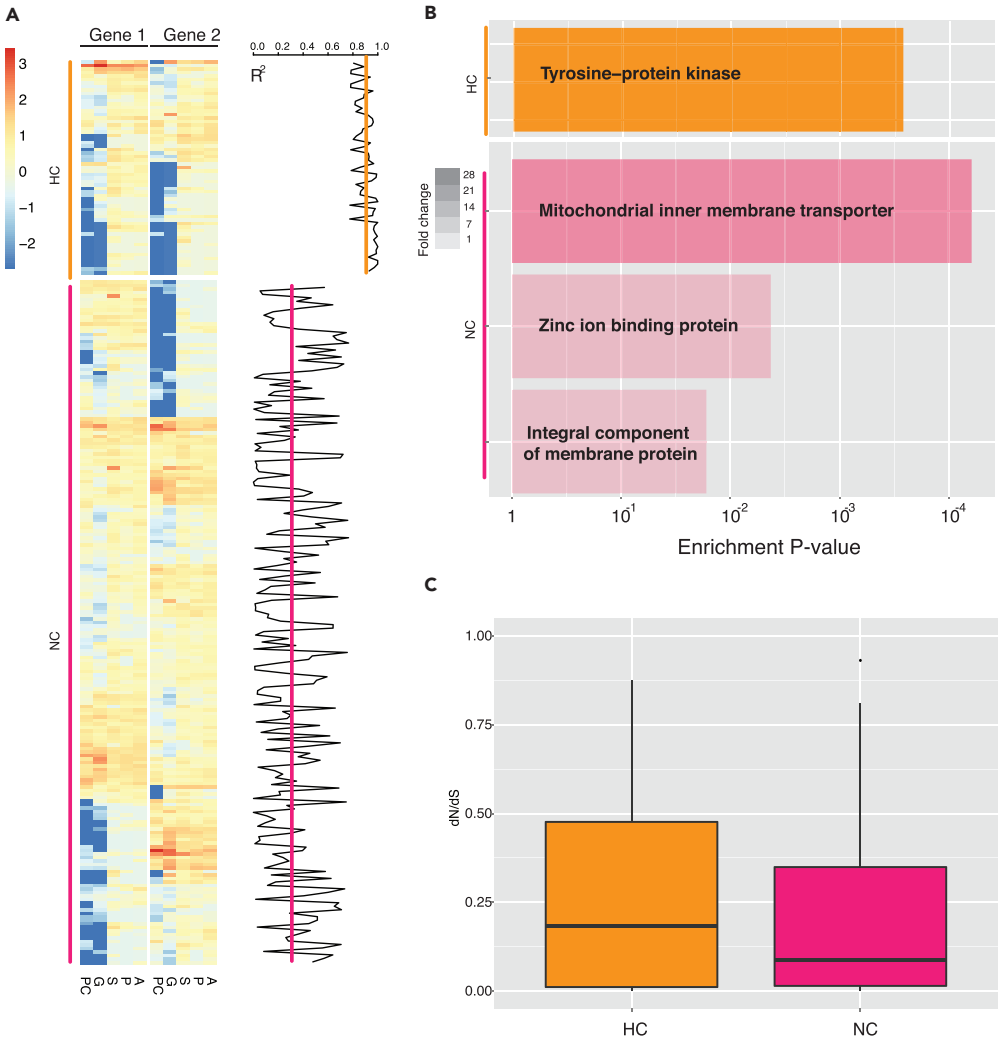


Figure 4. Gene Expression Profiling Reveals Evolution of Duplicated Genes in *A. digitifera*

(A) Gene expression profiling across five developmental stages (blastula, PC; gastrula, G; postgastrula, S; planula, P; and adult polyps, A) in *A. digitifera*. Two clusters of gene expression of ohnologous gene pairs: HC, high correlation: $p < 0.05$; NC, no correlation: $p \geq 0.05$ (Pearson's correlation test). Pearson's correlation coefficients between two ohnologous gene pairs are presented in the right panel, and lines represent average values of correlation coefficients in each cluster.

(B) Significant functional enrichments of two clusters of ohnolog gene pairs ($p < 0.05$, Fisher's exact test) indicate that divergence of gene expression is associated with gene functions. Colors of the bar represent fold change values in enrichments.

(C) Boxplot of dN/dS values of ohnologous gene pairs shows no significant difference between the two clusters ($p = 0.51$, Mann-Whitney test).

in adaptive radiation of organisms (Berner and Salzburger, 2015; Van De Peer et al., 2017), but we still do not understand the relationship between WGD and introgression in adaptive radiations (Soltis and Soltis, 2009).

In conclusion, this study identified an ancient WGD shared by *Acropora* species (IAsα) that not only provides new insights into the evolution of reef-building corals but also expands a new model of WGD in animals.

Limitation of the Study

Small-scale gene duplication continually occurs within the evolution of organisms (Maere et al., 2005), but large-scale gene/genome duplication or entire genome duplication was regarded as a rare evolutionary

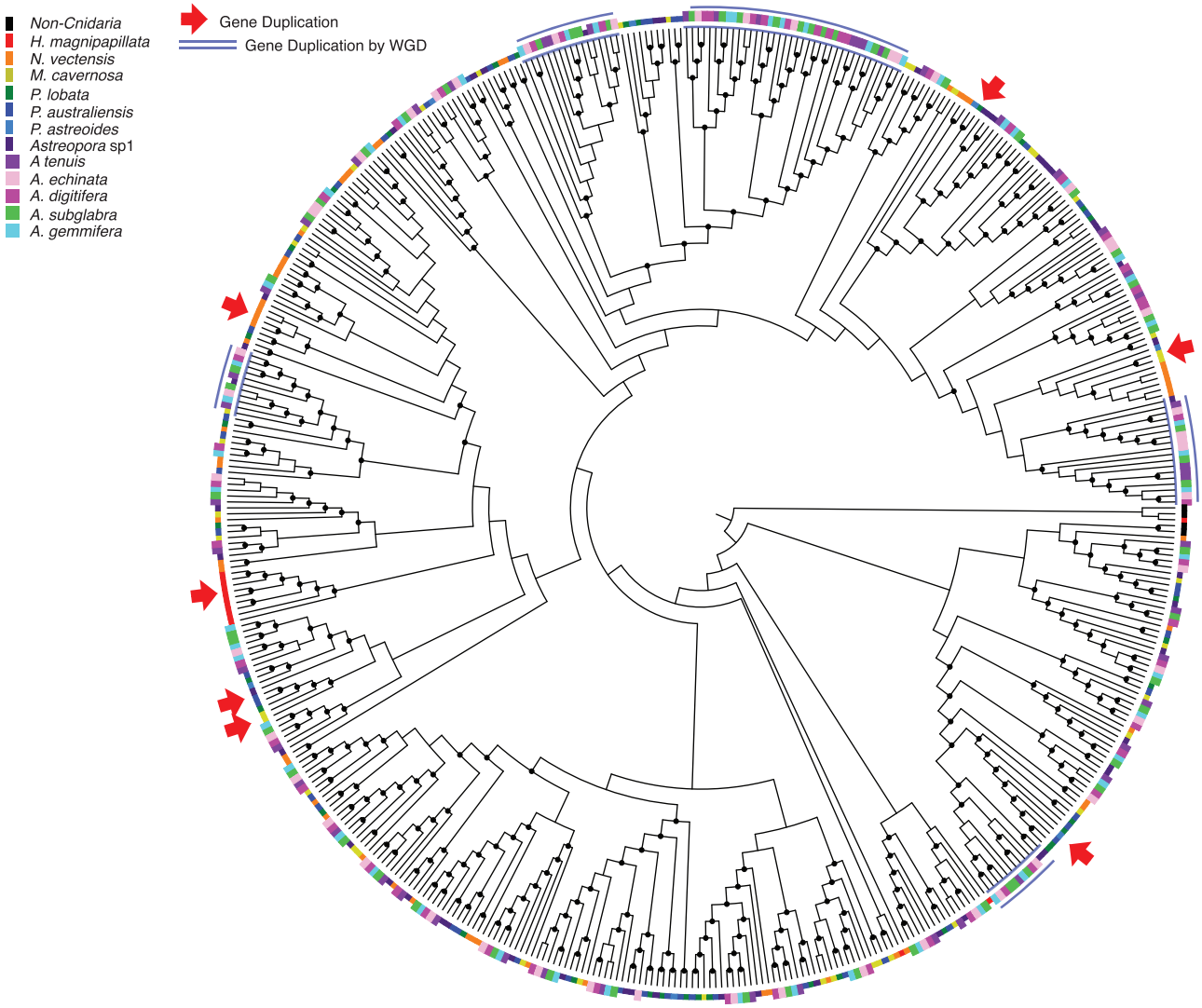


Figure 5. Diversification of Toxic Proteins via Gene Duplications in Cnidaria

Phylogenetic analysis of Coagulation factor X in 12 cnidarian species shows wide gene duplications. Gene duplication occurred in individual species lineages (red arrows), and gene duplications by WGD in *Acropora* are indicated with blue arches. Outer color strips represent 12 cnidarian species, and black strip represents non-cnidarian species. Bootstrap values greater than 50 are shown with black dots at nodes. See also Figures S13–S20 and Table S4.

event in the animals. With advanced increase of genomic data, we observed more and more WGD in animals (Van De Peer et al., 2017), such as vertebrates (Berthelot et al., 2014; Dehal and Boore, 2005; Kenny et al., 2016), insects (Li et al., 2018), and corals (this study). Yet, it is hard to distinguish large-scale gene/genome duplication from entire genome duplication using the dS-based method and phylogenomic and synteny analyses without precise genomic data. For example, the second-round WGD has been argued to be large-scale genome duplication rather than entire genome duplication (Smith and Keinath, 2015). Hence, in this study, we defined WGD as large-scale gene/genome duplication. Our evidence from different analyses supports the fact that WGD occurred in the common ancestor of *Acropora*, but there is still lack of sufficient evidence to support the fact that the WGD is generated by entire genome duplication. In addition, the distribution of inferred age nodes (Figure S8) was not shown as a standard normal distribution, one possible reason for which is that there is a lack of data in the phylogenomic approach.

METHODS

All methods can be found in the accompanying Transparent Methods supplemental file.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Transparent Methods, 22 figures, 6 tables, and one data file and can be found with this article online at https://doi.org/10.1016/j.isci.2019.02.001.

ACKNOWLEDGMENTS

This study was supported by OIST (Internal Fund to N.S.), Japan and was funded by JSPS grant (No. 17J00557 to Y.M.), Japan. We thank Dr. Chuya Shinzato for providing the genomic data in this study. We thank Dr. Evan P Economo, Dr. Douglas E Soltis, and Dr. Pamela S Soltis for their comments and insights into the draft manuscript. We thank Dr. Steven D. Aird and Hong Huat Hoh for editing the manuscript.

AUTHOR CONTRIBUTIONS

Y.M. and N.S. conceived the study. Y.M. performed all analysis in this study. Y.M. and N.S. wrote the initial manuscript and edited the final manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: July 10, 2018
Revised: January 28, 2019
Accepted: January 31, 2019
Published: March 29, 2019

REFERENCES

Ainsworth, T.D., Heron, S.F., Ortiz, J.C., Mumby, P.J., Grech, A., Ogawa, D., Eakin, C.M., and Leggat, W. (2016). Climate change disables coral bleaching protection on the Great Barrier Reef. Science *352*, 338–342.

Armoza-Zvuloni, R., Schneider, A., Sher, D., and Shaked, Y. (2016). Rapid Hydrogen Peroxide release from the coral Stylophora pistillata during feeding and in response to chemical and physical stimuli. Sci. Rep. *6*, 21000.

Arnegard, M.E., Zwickl, D.J., Lu, Y., and Zakon, H.H. (2010). Old gene duplication facilitates origin and diversification of an innovative communication system—twice. Proc. Natl. Acad. Sci. U S A *107*, 22172–22177.

Ben-Ari, H., Paz, M., and Sher, D. (2018). The chemical armament of reef-building corals: inter- and intra-specific variation and the identification of an unusual actinoporin in Stylophora pistilata. Sci. Rep. *8*, 251.

Berner, D., and Salzburger, W. (2015). The genomics of organismal diversification illuminated by adaptive radiations. Trends Genet. *31*, 491–499.

Berthelot, C., Brunet, F., Chalopin, D., Juanchich, A., Bernard, M., Noël, B., Bento, P., Da Silva, C., Labadie, K., and Alberti, A. (2014). The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. Nat. Commun. *5*, 3657.

Bertucci, A., Foret, S., Ball, E.E., and Miller, D.J. (2015). Transcriptomic differences between day and night in *Acropora millepora* provide new insights into metabolite exchange and light-enhanced calcification in corals. Mol. Ecol. *24*, 4489–4504.

Blanc, G., Hokamp, K., and Wolfe, K.H. (2003). A recent polyploidy superimposed on older large-scale duplications in the Arabidopsis genome. Genome Res. *13*, 137–144.

Blomme, T., Vandepoele, K., De Bodt, S., Simillion, C., Maere, S., and Van de Peer, Y. (2006). The gain and loss of genes during 600 million years of vertebrate evolution. Genome Biol. *7*, R43.

Bowers, J.E., Chapman, B.A., Rong, J., and Paterson, A.H. (2003). Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature *422*, 433–438.

Canestro, C., Albalat, R., Irimia, M., and Garcia-Fernàndez, J. (2013). Impact of gene gains, losses and duplication modes on the origin and diversification of vertebrates. Semin. Cell Dev. Biol. *24*, 83–94.

Chen, Z.J., and Birchler, J.A. (2013). Polyploid and Hybrid Genomics (John Wiley & Sons).

Christoffels, A., Koh, E.G.L., Chia, J.M., Brenner, S., Aparicio, S., and Venkatesh, B. (2004). Fugu genome analysis provides evidence for a whole-genome duplication early during the evolution of ray-finned fishes. Mol. Biol. Evol. *21*, 1146–1151.

Conant, G.C., Birchler, J.A., and Pires, J.C. (2014). Dosage, duplication, and diploidization: clarifying the interplay of multiple models for duplicate gene evolution over time. Curr. Opin. Plant Biol. *19*, 91–98.

Dehal, P., and Boore, J.L. (2005). Two rounds of whole genome duplication in the ancestral vertebrate. PLoS Biol. *3*, 1700–1708.

Edinger, E.N., and Risk, M.J. (1994). Oligocene miocene extinction and geographic restriction of caribbean corals - roles of turbidity, temperature, and nutrients. Palaios *9*, 576–598.

Fujikawa, T., Munakata, T., Kondo, S., Satoh, N., and Wada, S. (2010). Stress response in the ascidian Ciona intestinalis: transcriptional profiling of genes for the heat shock protein 70 chaperone system under heat stress and endoplasmic reticulum stress. Cell Stress Chaperon *15*, 193–204.

Fukami, H., Omori, M., and Hatta, M. (2000). Phylogenetic relationships in the coral family Acroporidae, reassessed by inference from mitochondrial genes. Zoolog. Sci. *17*, 689–696.

Fürthauer, M., Thisse, B., and Thisse, C. (1999). Three different noggin genes antagonize the activity of bone morphogenetic proteins in the zebrafish embryo. Dev. Biol. *214*, 181–196.

Gacesa, R., Chung, R., Dunn, S.R., Weston, A.J., Jaimes-Becerra, A., Marques, A.C., Morandini, A.C., Hranueli, D., Starcevic, A., Ward, M., et al. (2015). Gene duplications are extensive and contribute significantly to the toxic proteome of nematocysts isolated from *Acropora digitifera* (Cnidaria: Anthozoa: Scleractinia). BMC Genomics *16*, 774.

Glasauer, S.M.K., and Neuhauss, S.C.F. (2014). Whole-genome duplication in teleost fishes and its evolutionary consequences. Mol. Genet. Genomics *289*, 1045–1060.

Goodwin, S., McPherson, J.D., and McCombie, W.R. (2016). Coming of age: ten years of next-generation sequencing technologies. Nat. Rev. Genet. *17*, 333–351.

Hamada, M., Shoguchi, E., Shinzato, C., Kawashima, T., Miller, D.J., and Satoh, N. (2013). The complex NOD-like receptor repertoire of the

coral *Acropora digitifera* includes novel domain combinations. Mol. Biol. Evol. *30*, 167–176.

Hardwick, S.A., Deveson, I.W., and Mercer, T.R. (2017). Reference standards for next-generation sequencing. Nat. Rev. Genet. *18*, 473.

Hargreaves, A.D., Swain, M.T., Hegarty, M.J., Logan, D.W., and Mulley, J.F. (2014). Restriction and recruitment—gene duplication and the origin and evolution of snake venom toxins. Genome Biol. Evol. *6*, 2088–2095.

Jackson, S.T., and Sax, D.F. (2010). Balancing biodiversity in a changing environment: extinction debt, immigration credit and species turnover. Trends Ecol. Evol. *25*, 153–160.

Jiao, Y.N., Wickett, N.J., Ayyampalayam, S., Chanderbali, A.S., Landherr, L., Ralph, P.E., Tomsho, L.P., Hu, Y., Liang, H.Y., Soltis, P.S., et al. (2011). Ancestral polyploidy in seed plants and angiosperms. Nature *473*, 97–U113.

Kassahn, K.S., Dang, V.T., Wilkins, S.J., Perkins, A.C., and Ragan, M.A. (2009). Evolution of gene function and regulatory control after whole-genome duplication: comparative analyses in vertebrates. Genome Res. *19*, 1404–1418.

Kenny, N.J., Chan, K.W., Nong, W., Qu, Z., Maeso, I., Yip, H.Y., Chan, T.F., Kwan, H.S., Holland, P.W.H., Chu, K.H., et al. (2016). Ancestral whole-genome duplication in the marine chelicerate horseshoe crabs. Heredity *119*, 388.

Kenyon, J.C. (1997). Models of reticulate evolution in the coral genus Acropora based on chromosome numbers: parallels with plants. Evolution *51*, 756–767.

Kondrashov, F.A. (2012). Gene duplication as a mechanism of genomic adaptation to a changing environment. Proc. Biol. Sci. *279*, 5048–5057.

Kordiš, D., and Gubenšek, F. (2000). Adaptive evolution of animal toxin multigene families. Gene *261*, 43–52.

Li, Z., Tiley, G.P., Galuska, S.R., Reardon, C.R., Kidder, T.I., Rundell, R.J., and Barker, M.S. (2018). Multiple large-scale gene and genome duplications during the evolution of hexapods. Proc. Natl. Acad. Sci. U S A *115*, 4713–4718.

Lynch, M., and Conery, J.S. (2000). The evolutionary fate and consequences of duplicate genes. Science *290*, 1151–1155.

Mao, Y. (2019). GenoDup Pipeline: a tool to detect genome duplication using the dS-based method. PeerJ *7*, e6303.

Mao, Y., Economo, E.P., and Satoh, N. (2018). The roles of introgression and climate change in the rise to dominance of Acropora Corals. Curr. Biol. *28*, 3373–3382.

Maere, S., De Bodt, S., Raes, J., Casneuf, T., Van Montagu, M., Kuiper, M., and Van de Peer, Y. (2005). Modeling gene and genome duplications in eukaryotes. Proc. Natl. Acad. Sci. U S A *102*, 5454–5459.

Marcet-Houben, M., and Gabaldón, T. (2015). Beyond the whole-genome duplication: phylogenetic evidence for an ancient interspecies hybridization in the baker's yeast lineage. PLoS Biol. *13*, e1002220.

Nakatani, Y., Takeda, H., Kohara, Y., and Morishita, S. (2007). Reconstruction of the vertebrate ancestral genome reveals dynamic genome reorganization in early vertebrates. Genome Res. *17*, 1254–1265.

Rabier, C.E., Ta, T., and Ane, C. (2014). Detecting and locating whole genome duplications on a phylogeny: a probabilistic approach. Mol. Biol. Evol. *31*, 750–762.

Renema, W., Bellwood, D.R., Braga, J.C., Bromfield, K., Hall, R., Johnson, K.G., Lunt, P., Meyer, C.P., McMonagle, L.B., Morley, R.J., et al. (2008). Hopping hotspots: global shifts in marine biodiversity. Science *321*, 654–657.

Renema, W., Pandolfi, J.M., Kiessling, W., Bosellini, F.R., Klaus, J.S., Korpanty, C., Rosen, B.R., Santodomingo, N., Wallace, C.C., and Webster, J.M. (2016). Are coral reefs victims of their own past success? Sci. Adv. *2*, e1500850.

Reyes-Bermudez, A., Villar-Briones, A., Ramirez-Portilla, C., Hidaka, M., and Mikheyev, A.S. (2016). Developmental progression in the coral *Acropora digitifera* is controlled by differential expression of distinct regulatory gene networks. Genome Biol. Evol. *8*, 851–870.

Richards, Z.T., and Hobbs, J.P.A. (2015). Hybridisation on coral reefs and the conservation of evolutionary novelty. Curr. Zool. *61*, 132–145.

Richards, Z.T., Miller, D.J., and Wallace, C.C. (2013). Molecular phylogenetics of geographically restricted Acropora species: implications for threatened species conservation. Mol. Phylogenet. Evol. *69*, 837–851.

Schwager, E.E., Sharma, P.P., Clarke, T., Leite, D.J., Wierschin, T., Pechmann, M., Akiyama-Oda, Y., Esposito, L., Bechsgaard, J., Bilde, T., et al. (2017). The house spider genome reveals an ancient whole-genome duplication during arachnid evolution. BMC Biol. *15*, 62.

Seehausen, O., Butlin, R.K., Keller, I., Wagner, C.E., Boughman, J.W., Hohenlohe, P.A., Peichel, C.L., Saetre, G.P., Bank, C., Brannstrom, A., et al. (2014). Genomics and the origin of species. Nat. Rev. Genet. *15*, 176–192.

Sémon, M., and Wolfe, K.H. (2007). Consequences of genome duplication. Curr. Opin. Genet. Dev. *17*, 505–512.

Smith, J.J., and Keinath, M.C. (2015). The sea lamprey meiotic map improves resolution of ancient vertebrate genome duplications. Genome Res. *25*, 1081–1090.

Smith, J.J., Kuraku, S., Holt, C., Sauka-Spengler, T., Jiang, N., Campbell, M.S., Yandell, M.D., Manousaki, T., Meyer, A., Bloom, O.E., et al. (2013). Sequencing of the sea lamprey (*Petromyzon marinus*) genome provides insights into vertebrate evolution. Nat. Genet. *45*, 415–421.

Soltis, D.E., Albert, V.A., Leebens-Mack, J., Bell, C.D., Paterson, A.H., Zheng, C., Sankoff, D., Pamphilis, C.W., Wall, P.K., and Soltis, P.S. (2009). Polyploidy and angiosperm diversification. Am. J. Bot. *96*, 336–348.

Soltis, P.S., Marchant, D.B., Van de Peer, Y., and Soltis, D.E. (2015). Polyploidy and genome evolution in plants. Curr. Opin. Genet. Dev. *35*, 119–125.

Soltis, P.S., and Soltis, D.E. (2009). The role of hybridization in plant speciation. Annu. Rev. Plant Biol. *60*, 561–588.

Soltis, P.S., and Soltis, D.E. (2012). Polyploidy and Genome Evolution (Springer).

Stuber, C.W., and Goodman, M.M. (1983). Inheritance, intracellular localization, and genetic variation of phosphoglucomutase isozymes in maize (Zea mays L.). Biochem. Genet. *21*, 667–689.

Suzuki, G., and Nomura, K. (2013). Species boundaries of Astreopora corals (Scleractinia, Acroporidae) inferred by mitochondrial and nuclear molecular markers. Zoolog. Sci. *30*, 626–632.

Tiley, G.P., Ane, C., and Burleigh, J.G. (2016). Evaluating and characterizing ancient whole-genome duplications in plants with gene count data. Genome Biol. Evol. *8*, 1023–1037.

Van de Peer, Y., Maere, S., and Meyer, A. (2009). The evolutionary significance of ancient genome duplications. Nat. Rev. Genet. *10*, 725–732.

Van De Peer, Y., Mizrachi, E., and Marchal, K. (2017). The evolutionary significance of polyploidy. Nat. Rev. Genet. *18*, 411–424.

Van Oppen, M.J., McDonald, B.J., Willis, B., and Miller, D.J. (2001). The evolutionary history of the coral genus Acropora (Scleractinia, Cnidaria) based on a mitochondrial and a nuclear marker: reticulation, incomplete lineage sorting, or morphological convergence? Mol. Biol. Evol. *18*, 1315–1329.

Vanneste, K., Baele, G., Maere, S., and Van de Peer, Y. (2014). Analysis of 41 plant genomes supports a wave of successful genome duplications in association with the Cretaceous-Paleogene boundary. Genome Res. *24*, 1334–1347.

Vanneste, K., Van de Peer, Y., and Maere, S. (2012). Inference of genome duplications from age distributions revisited. Mol. Biol. Evol. *30*, 177–190.

Veron, A.S., Kaufmann, K., and Bornberg-Bauer, E. (2006). Evidence of interaction network evolution by whole-genome duplications: a case study in MADS-box proteins. Mol. Biol. Evol. *24*, 670–678.

Vollmer, S.V., and Palumbi, S.R. (2002). Hybridization and the evolution of reef coral diversity. Science *296*, 2023–2025.

Vonk, F.J., Casewell, N.R., Henkel, C.V., Heimberg, A.M., Jansen, H.J., McCleary, R.J.R., Kerkkamp, H.M.E., Vos, R.A., Guerreiro, I., and Calvete, J.J. (2013). The king cobra genome reveals dynamic gene evolution and adaptation in the snake venom system. Proc. Natl. Acad. Sci. U S A *110*, 20651–20656.

Wallace, C.C. (2012). Acroporidae of the Caribbean. Geologica Belgica *15*, 388–393.

Wallace, C.C., and Rosen, B.R. (2006). Diverse staghorn corals (Acropora) in high-latitude Eocene assemblages: implications for the

evolution of modern diversity patterns of reef corals. Proc. Biol. Sci. *273*, 975–982.

Wang, J.-P., Yu, J.-G., Li, J., Sun, P.-C., Wang, L., Yuan, J.-Q., Meng, F.-B., Sun, S.-R., Li, Y.-X., and Lei, T.-Y. (2018). Two likely auto-tetraploidization events shaped kiwifruit genome and contributed to establishment of the actinidiaceae family. iScience *7*, 230–240.

Wang, W., Hui, J.H.L., Chan, T.F., and Chu, K.H. (2014). De novo transcriptome sequencing of the snail echinolittorina malaccana: identification of genes responsive to thermal stress and development of genetic markers for population studies. Mar. Biotechnol. (NY) *16*, 547–559.

Willis, B.L., van Oppen, M.J.H., Miller, D.J., Vollmer, S.V., and Ayre, D.J. (2006). The role of hybridization in the evolution of reef corals. Annu. Rev. Ecol. Evol. Syst. *37*, 489–517.

Yasuoka, Y., Shinzato, C., and Satoh, N. (2016). The mesoderm-forming gene brachyury regulates ectoderm-endoderm demarcation in the coral *Acropora digitifera*. Curr. Biol. *26*, 2885–2892.

Zakon, H.H., Lu, Y., Zwickl, D.J., and Hillis, D.M. (2006). Sodium channel genes and the evolution of diversity in communication signals of electric fishes: convergent molecular evolution. Proc. Natl. Acad. Sci. U S A *103*, 3675–3680.

Zhang, G.Q., Liu, K.W., Li, Z., Lohaus, R., Hsiao, Y.Y., Niu, S.C., Wang, J.Y., Lin, Y.C., Xu, Q., Chen, L.J., et al. (2017). The Apostasia genome and the evolution of orchids. Nature *549*, 379–383.

ISCI, Volume *13*

Supplemental Information

A Likely Ancient Genome Duplication in the Speciose Reef-Building Coral Genus, *Acropora*

Yafei Mao and Noriyuki Satoh

Supplemental Figures

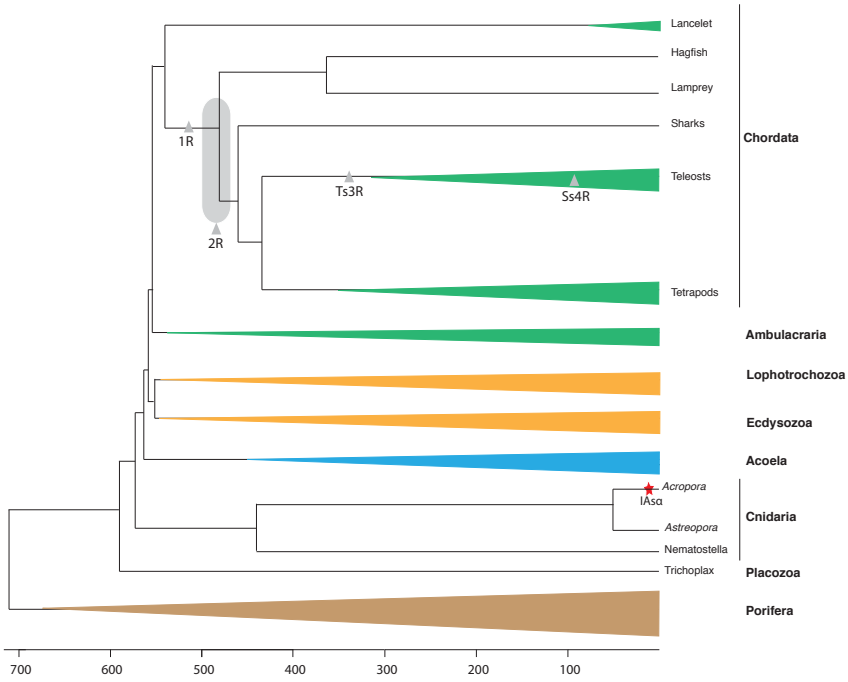


Figure S1. WGD events in evolution of the animal clade. Related to Figure 1. The backbone and divergence time of the tree are based on various sources (e.g., Satoh, 2016). The shaded grey oval represents the uncertain position of two rounds of WGD and colored triangles represent the corresponding divergent groups. Grey triangles represent WGDs and the red star represents invertebrate WGD specific to *Acropora* (IAsα) reported in this study.

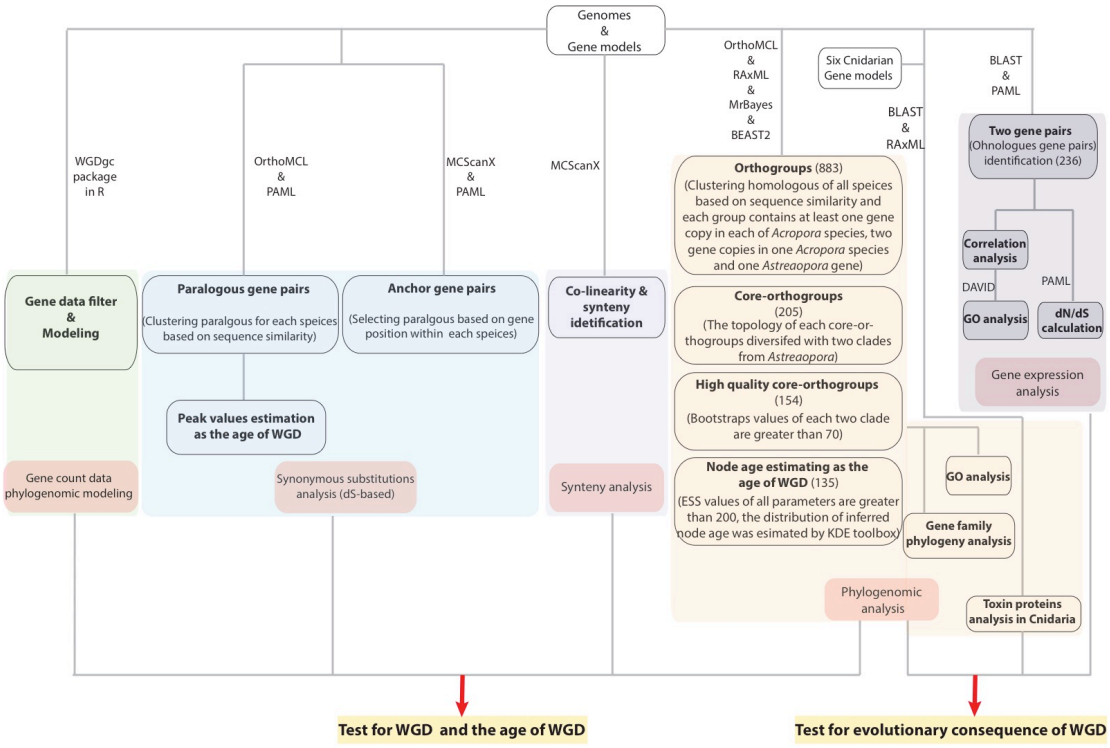
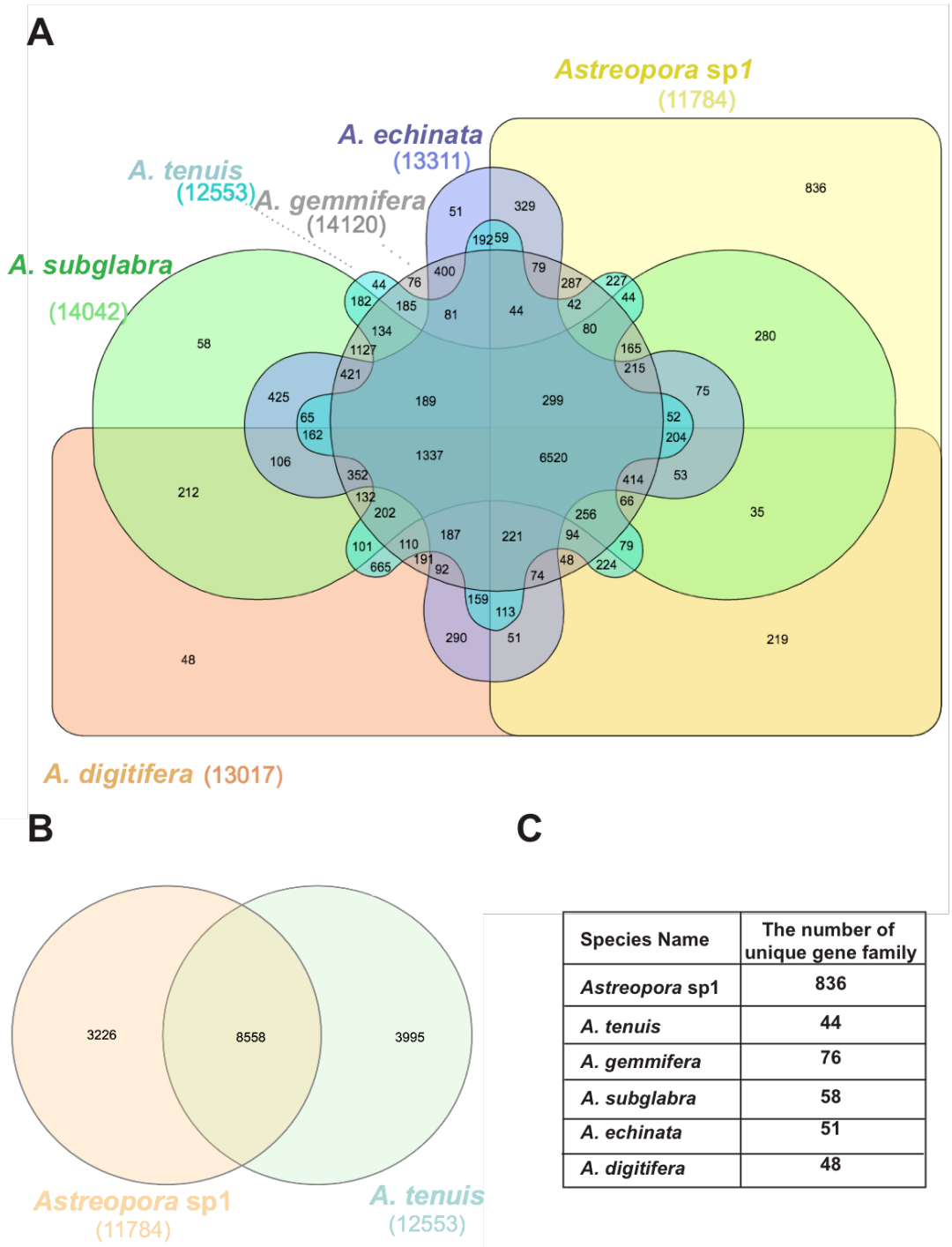


Figure S2. Flowchart of study design. Related to Figures 2-5. The flowchart illustrates methods used to test the main questions in this study.



Species Name	The number of unique gene family
Astreopora sp1	836
A. tenuis	44
A. gemmifera	76
A. subglabra	58
A. echinata	51
A. digitifera	48

Figure S3. Venn diagrams of gene families in six *Acroporid* species. Related to Figure 1. (A). Venn diagram of shared and unique gene families in six *Acroporid* species. (B). Venn diagram of shared and unique gene families between *Astreopora* sp1 and *A. tenuis*. (C). The table of the number of unique gene family in six *Acroporid* species.

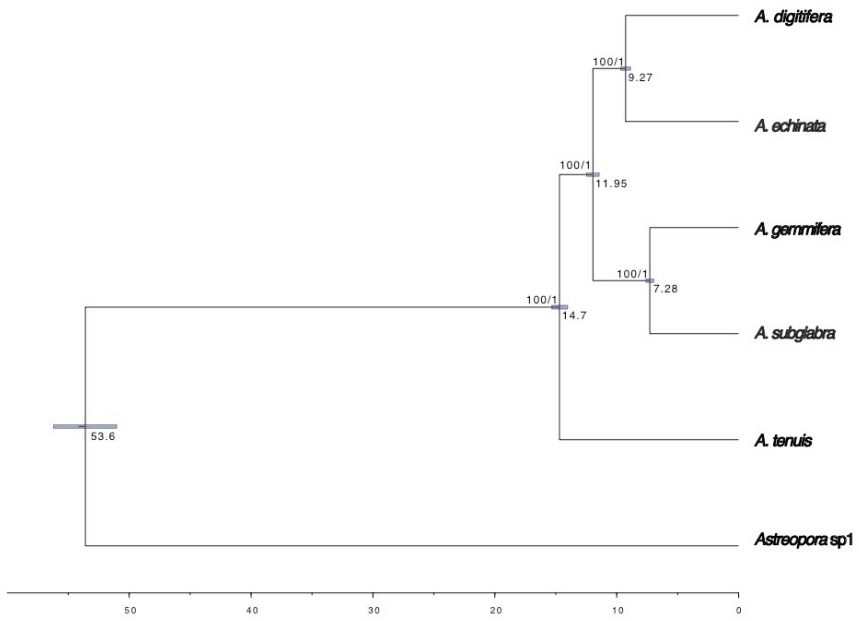


Figure S4. Phylogeny of the Family *Acroporidae.* Related to Figure 1. Time-calibrated phylogenetic tree reconstructed based on fossil calibration and concatenated coding sequences (7,467,066 bp in total) from 3,461 single-copy orthologous genes with BEAST2. Branch lengths are scaled to estimated-divergence time. Posterior 95% CIs of node ages are represented with blue horizontal bars as well as ML bootstrap values and Bayesian posterior probabilities are shown at each node.

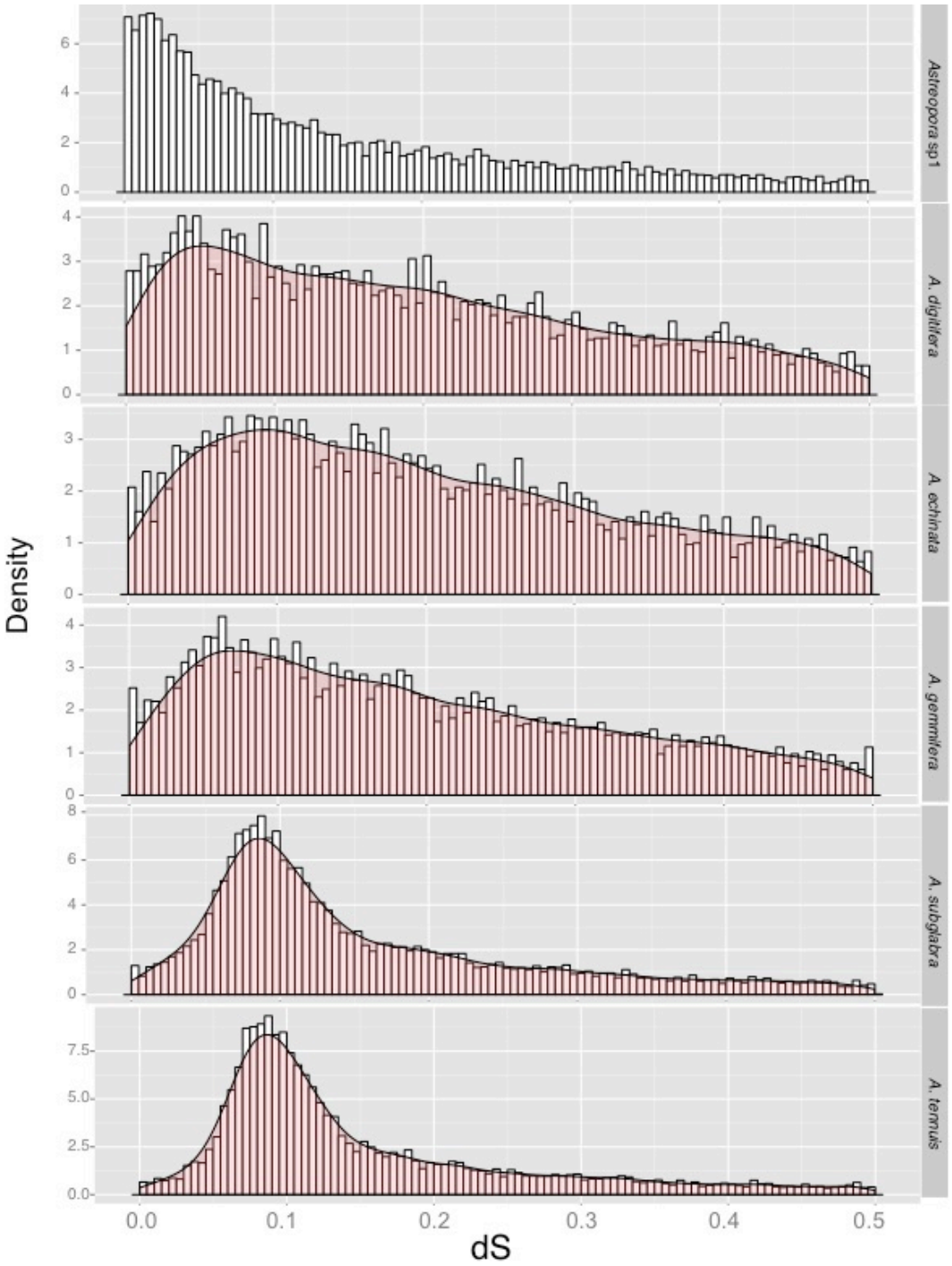


Figure S5. Frequency distributions of dS values for paralogous gene pairs in five *Acropora* and one *Astreopora* species. Related to Figure 2. The distributions of dS values of paralogs, estimating neutral evolutionary divergence since the two paralogs diverged, are plotted with a bin size of 0.005, showing the similar peaks (dS value: 0-0.3) in *Acropora*.

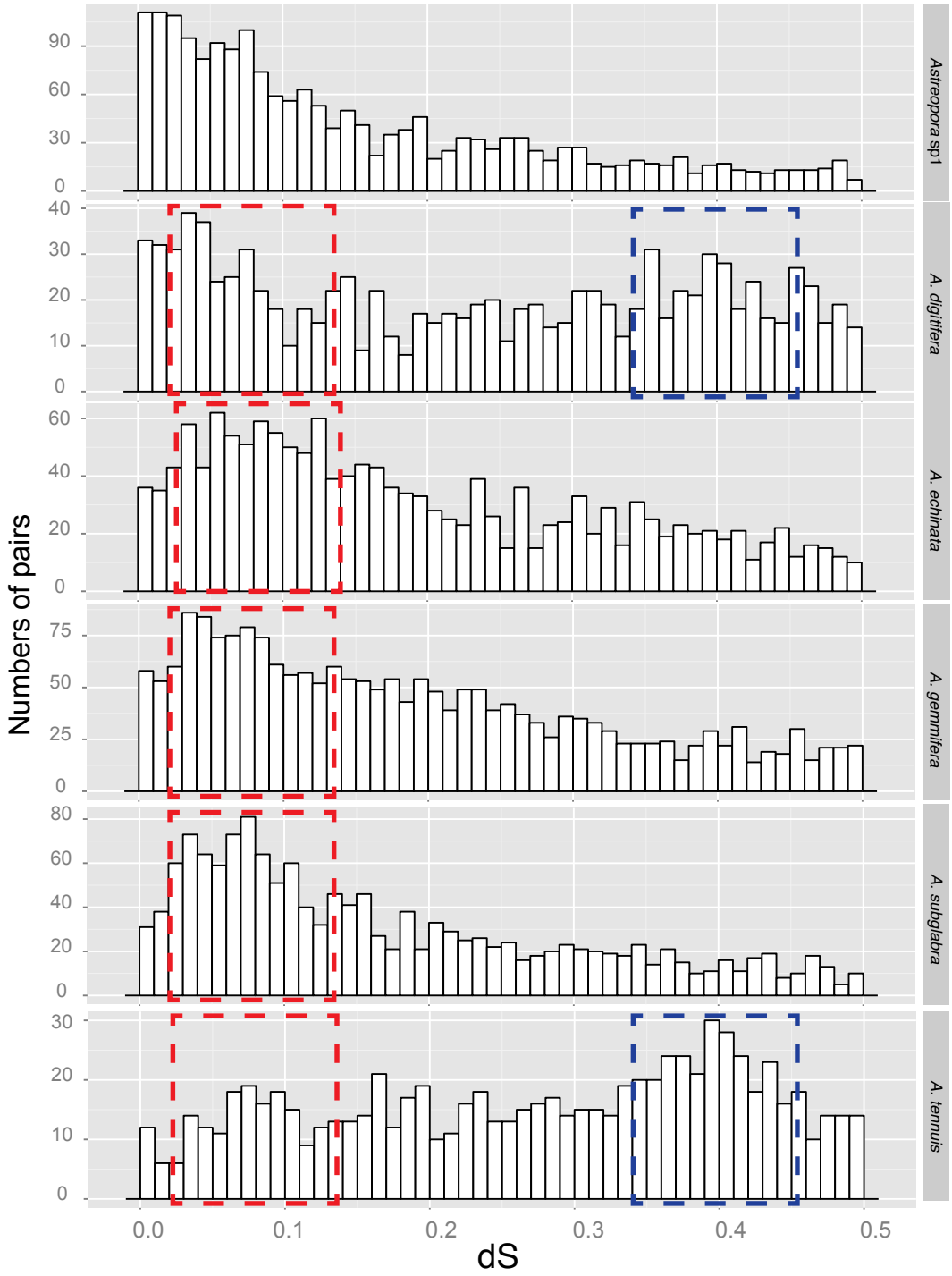


Figure S6. Frequency distributions of dS values for anchor-gene pairs in five *Acropora* and one *Astreopora* species. Related to Figure 2. Distributions of dS values of anchor paralogs, estimating the neutral evolutionary divergence times since the paralogs diverged, are plotted with a bin size of 0.01, showing the similar peaks (dS value: 0-0.3, red boxes) in *Acropora* and extra peaks in *A. digitifera* and *A. tenuis* (dS value: 0.3-0.5, blue boxes).

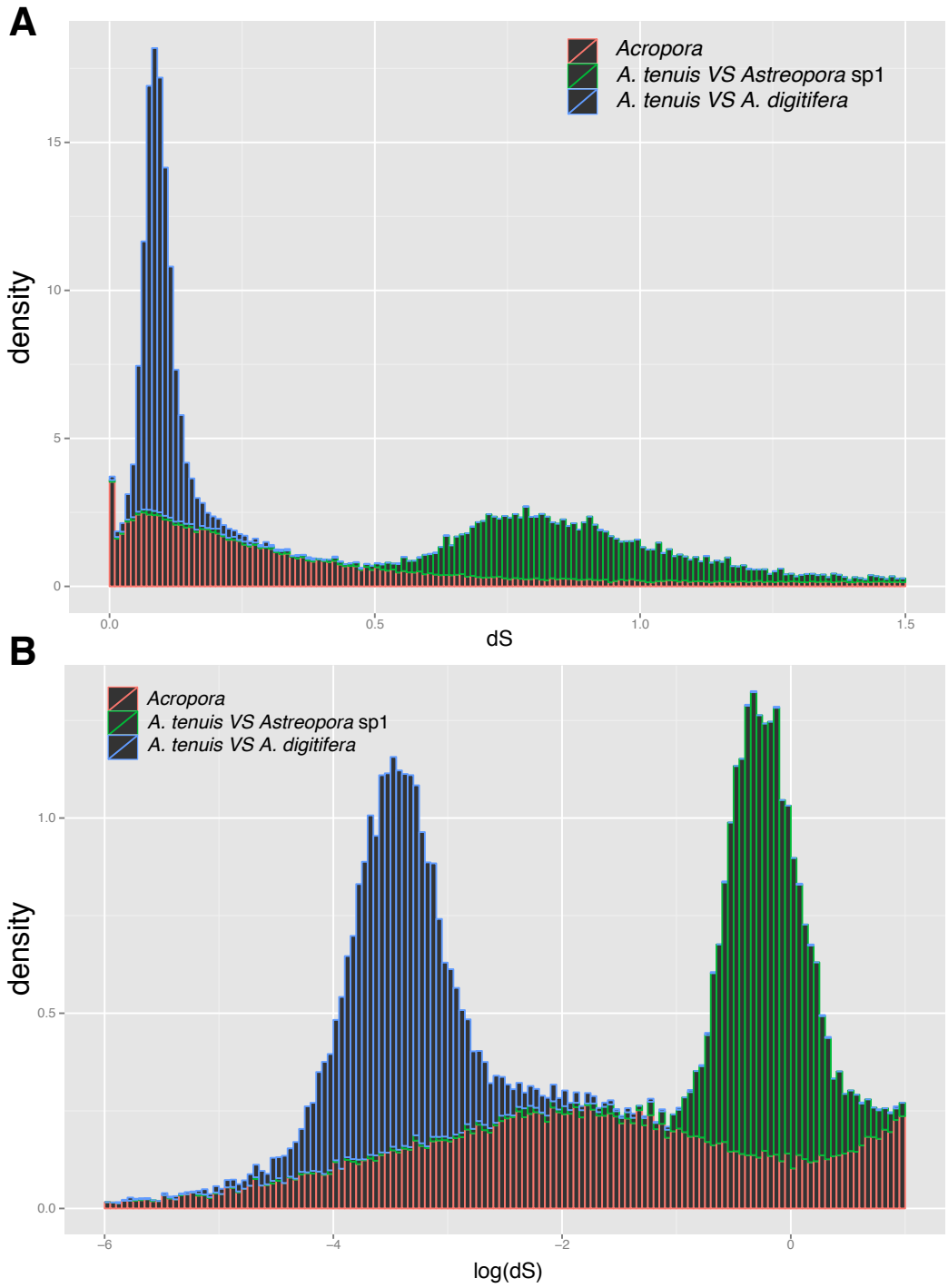


Figure S7. Frequency distributions of dS values for paralogous genes in *Acropora* and for orthologous genes. Related to Figure 2. (A) Frequency distribution of dS values for paralogous genes in *Acropora* and for orthologous genes showing that a WGD event occurred in the most recent common ancestor of *Acropora*. Distributions are plotted with a bin size of 0.01. (B) Frequency distribution of log dS values for paralogous genes in *Acropora* and for orthologous genes. Distributions are plotted with a bin size of 0.05.

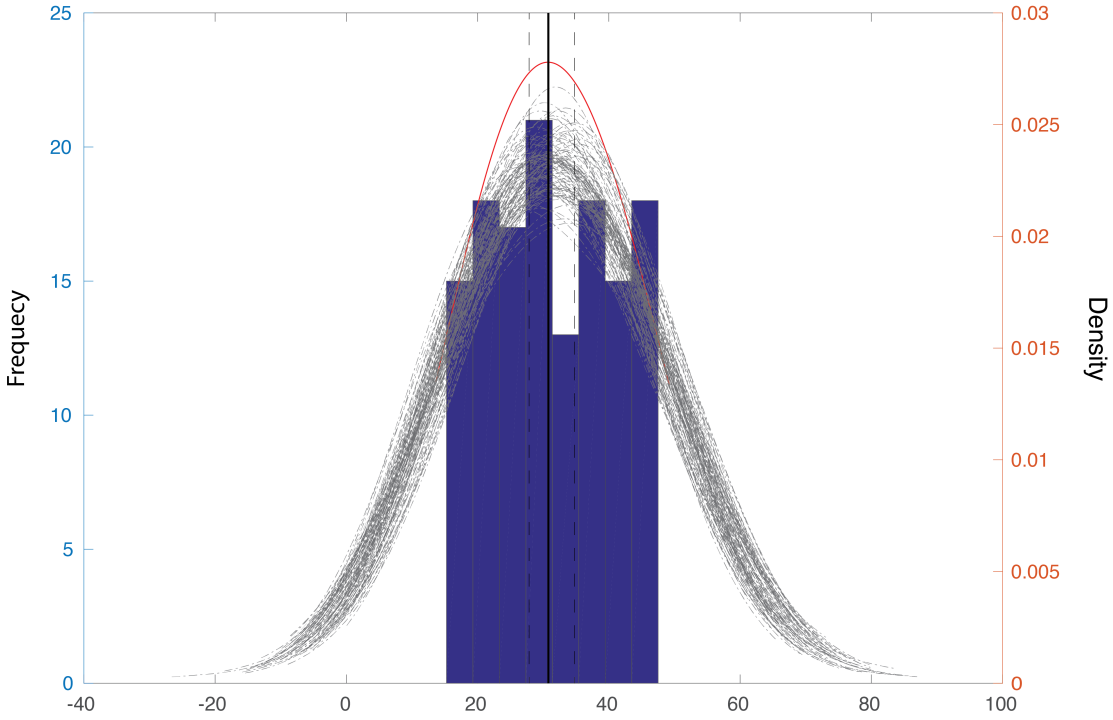


Figure S8. Node age distribution of IAsα. Related to Figure 2. Inferred node ages from 135 phylogenies were analyzed with KDE toolbox to show the peak at 30.78 My, represented by the black solid line. The grey lines represent density estimations from 1000 bootstraps and the black dotted line represents the corresponding 95% confidence interval (27.86 - 34.77 My) from 100 bootstraps.

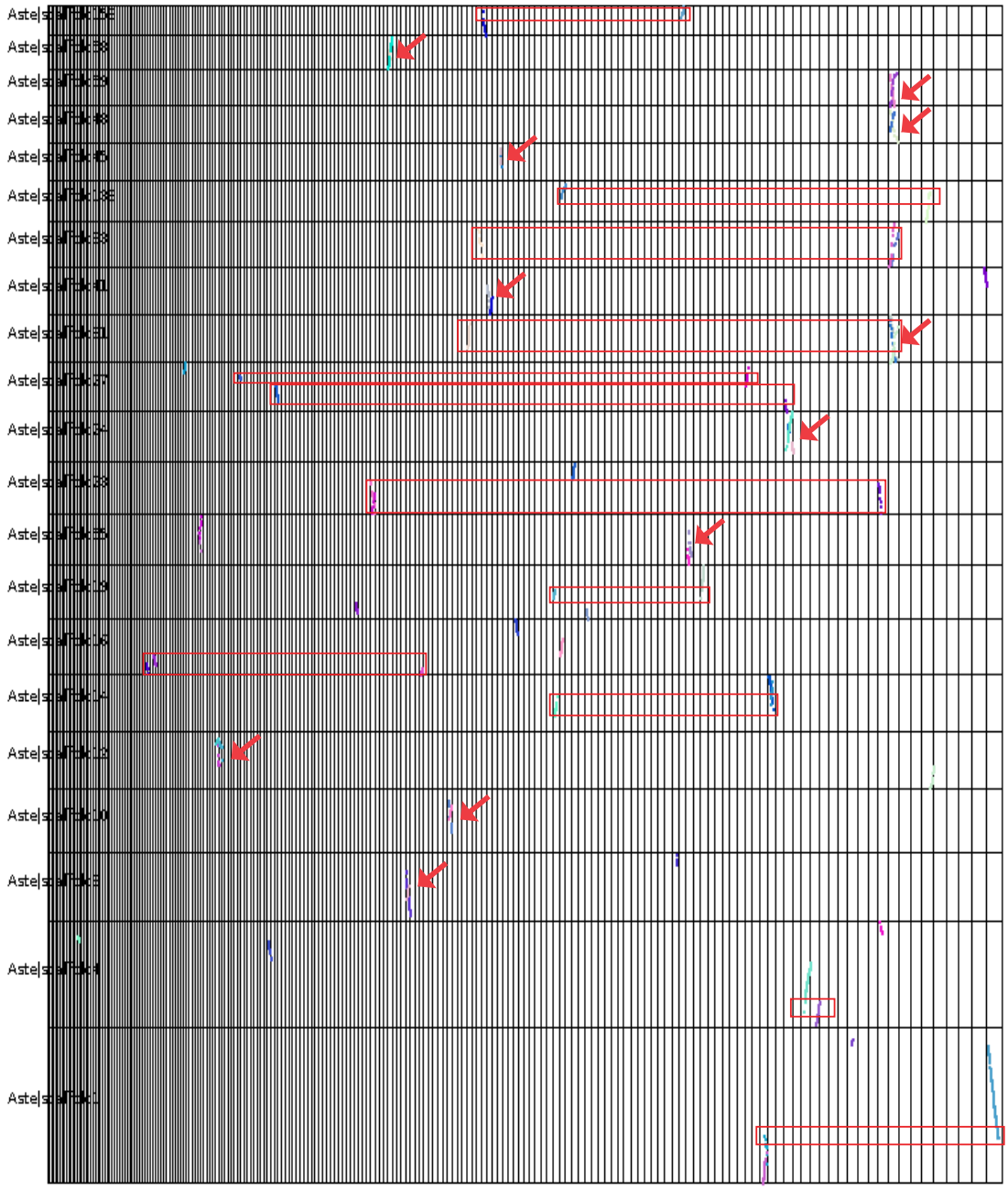


Figure S9. Synteny blocks between *Astreopora* sp1 and *A. tenuis*. Related to Figure 2. Only co-linear segments with at least 8 anchor pairs are shown in between the top length 100 scaffolds of *Astreopora* sp1 (Left side) and the top length 200 scaffolds of *A. tenuis* (Bottom). Only the scaffolds of *Astreopora* sp1 representing duplicated segments with *A. tenuis* are shown. The duplicated segments on different scaffolds are covered with red boxes. The duplicated segments on the same scaffolds are marked with red arrows.

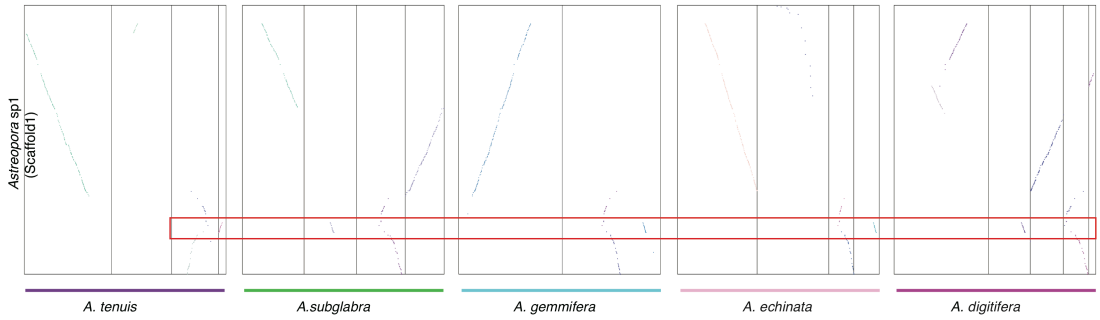


Figure S10. An example of synteny blocks between *Astreopora* sp1 and the five *Acropora* species. Related to Figure 2. Only co-linear segments with at least 8 anchor pairs are shown in between the longest scaffolds of *Astreopora* sp1 (Left side) and scaffolds of other five *Acropora* species (Bottom). The red box represents the retained the duplicated segments.

Figure S11. Alignment of orthogroup 1247 (dnaJ homolog subfamily B member 11-like) showing the independent loss of the domain in duplicates. Related to Figure 3.

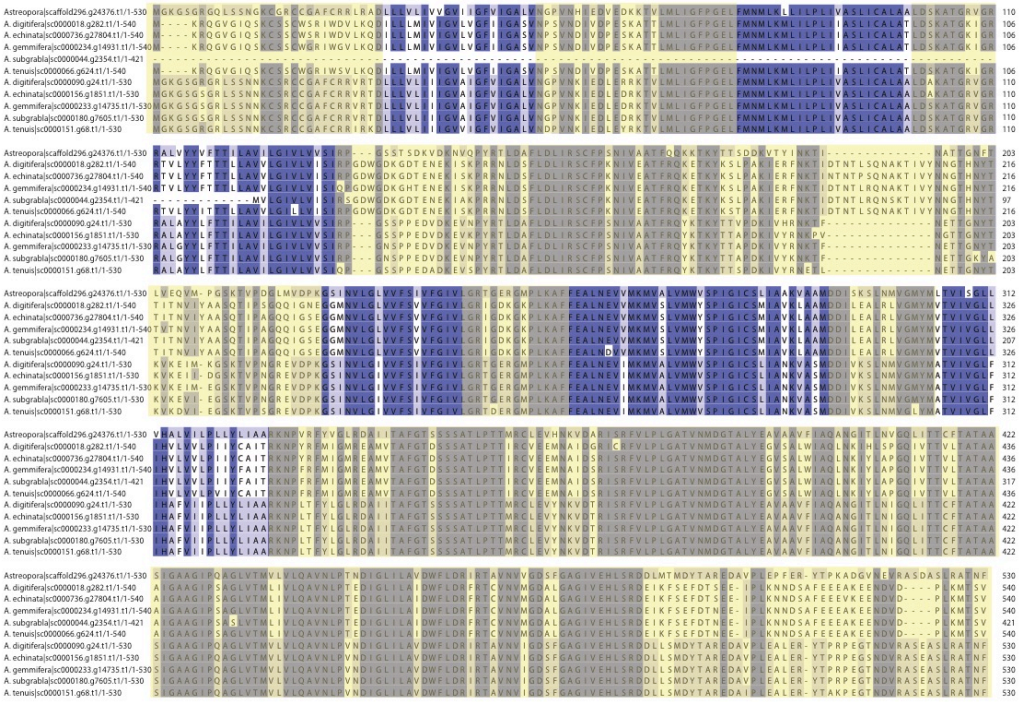


Figure S12. Alignment of orthogroup 1244 (excitatory amino acid transporter 1-like) showing mutations on transmembrane and exposed regions, suggesting that new functions would be generated. **Related to Figure 3.** Exposed regions are shown in yellow.

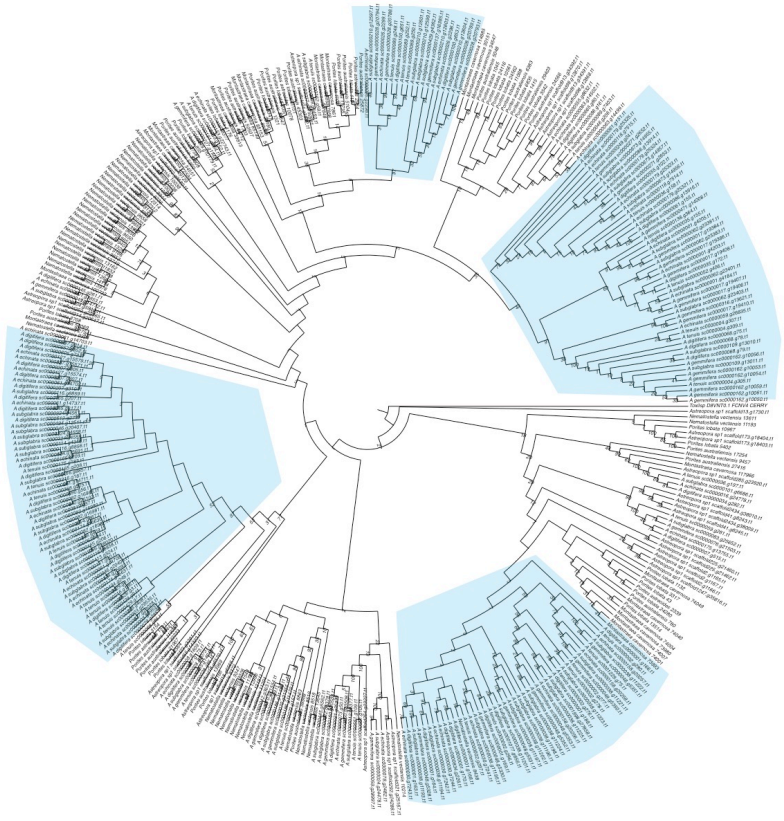


Figure S13. Phylogeny of the toxic protein, Ryncolin-4. Related to Figure 5. The phylogeny was reconstructed using ExaML and bootstrap values are shown at each node. Gene duplications caused by WGD in *Acropora* are shown in cyan.

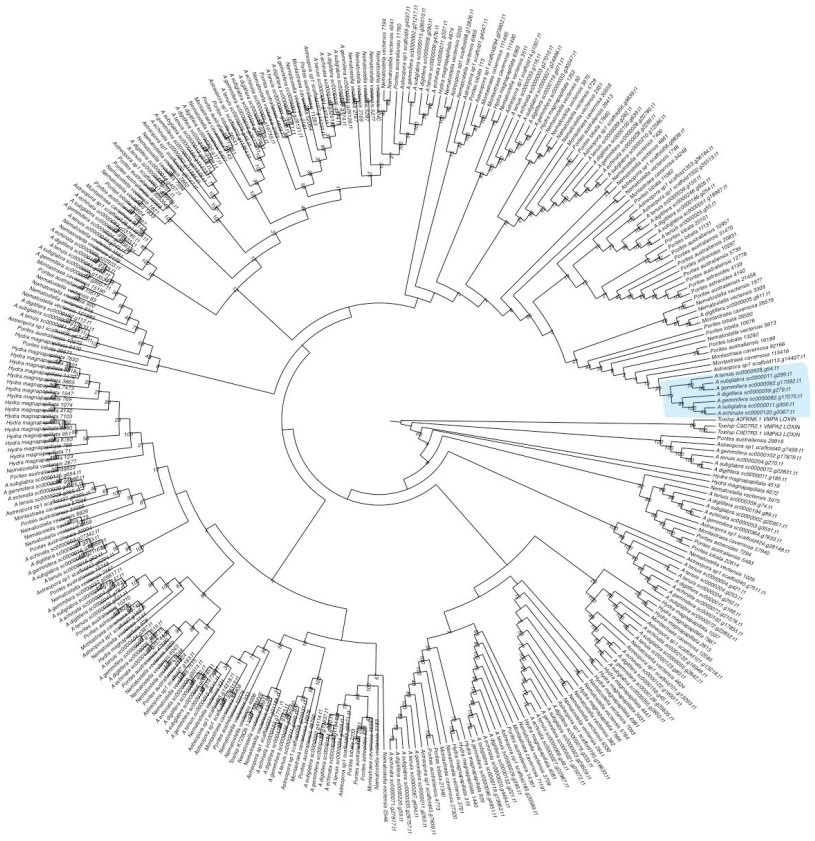


Figure S14. Phylogeny of the toxic Astacin-like metalloprotease. Related to Figure 5. The phylogeny was reconstructed using ExaML and bootstrap values are shown at each node. Gene duplication caused by WGD in *Acropora* is shown in cyan.

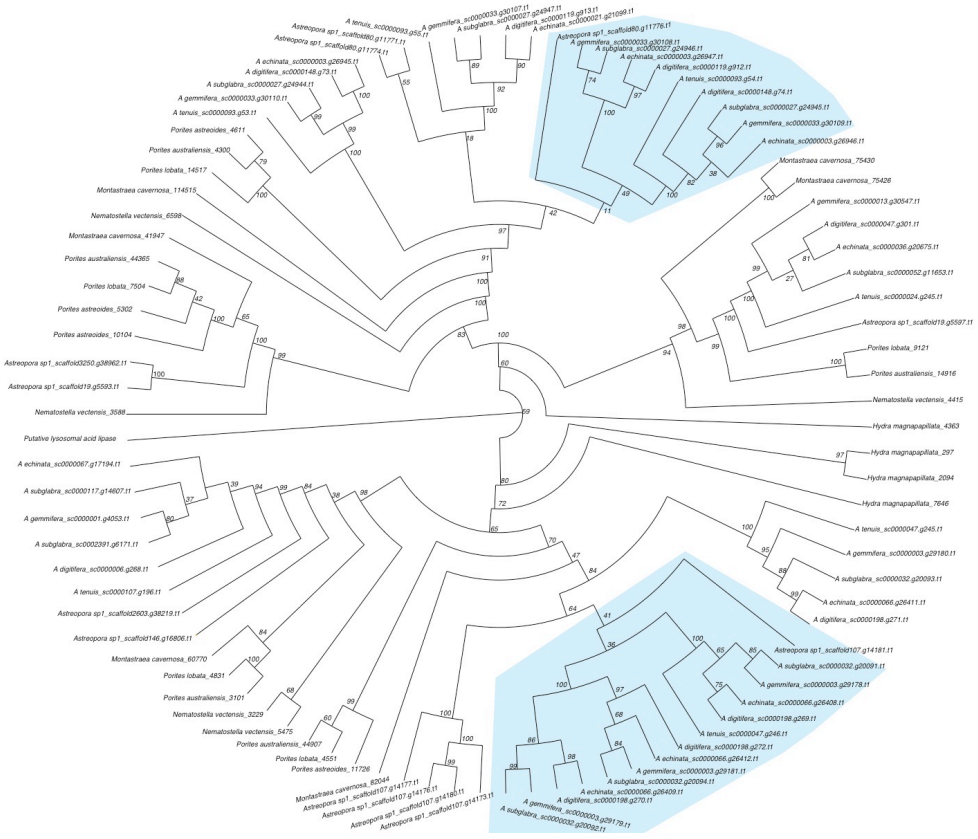


Figure S15. Phylogeny of a toxic protein (Putative lysosomal acid lipase/ cholesteryl ester hydrolase). Related to Figure 5. The phylogeny was reconstructed using RAxML and bootstrap values are shown at each node. Gene duplications caused by WGD in *Acropora* are shown in cyan shadows.

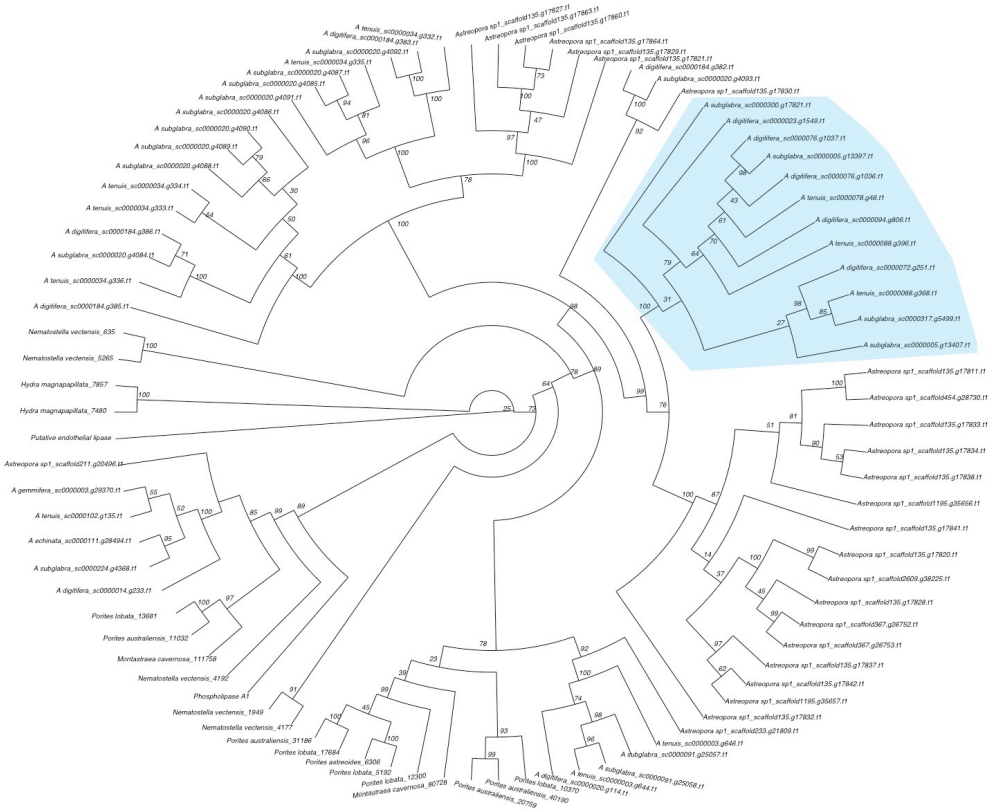


Figure S16. Phylogeny of the toxic putative endothelial lipase. Related to Figure 5. The phylogeny was reconstructed using RAxML and bootstrap values are shown at each node. Gene duplications caused by WGD in *Acropora* is shown in cyan.

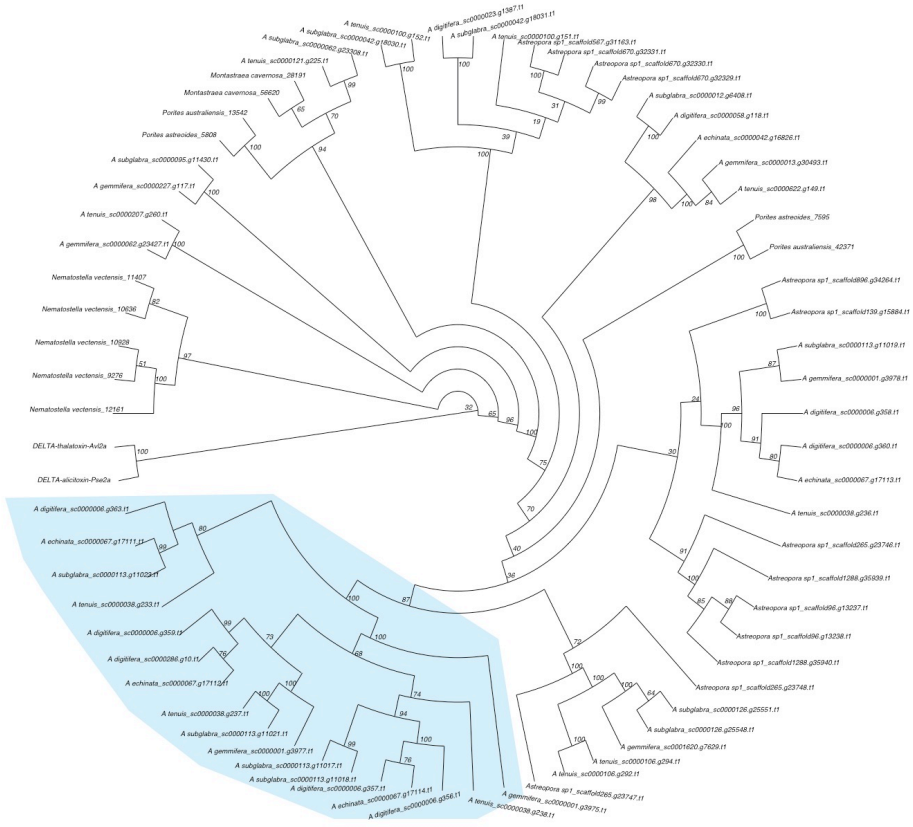


Figure S17. Phylogeny of the toxin protein (DELTA-thalatoxin-Avl2a/DELTA-alicitoxin-Pse2a). Related to Figure 5. The phylogeny was reconstructed using RAxML and bootstrap values are shown at each node. Gene duplication by WGD in *Acropora* is shown in cyan.

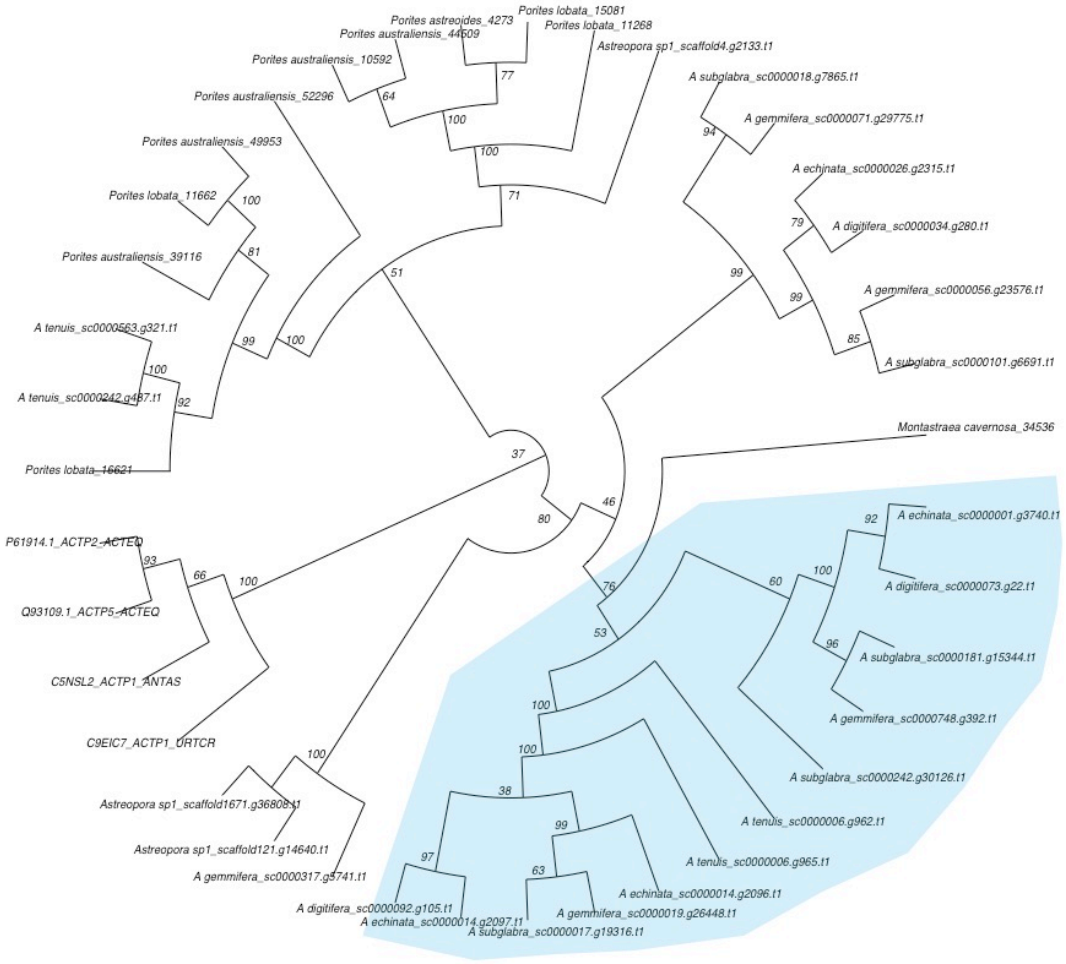


Figure S18. Phylogeny of the toxic protein (DELTA-actitoxin-Aas1a). Related to Figure 5. The phylogeny was reconstructed using RAxML and bootstrap values are shown at each node. Gene duplication by WGD in *Acropora* is shown in cyan.

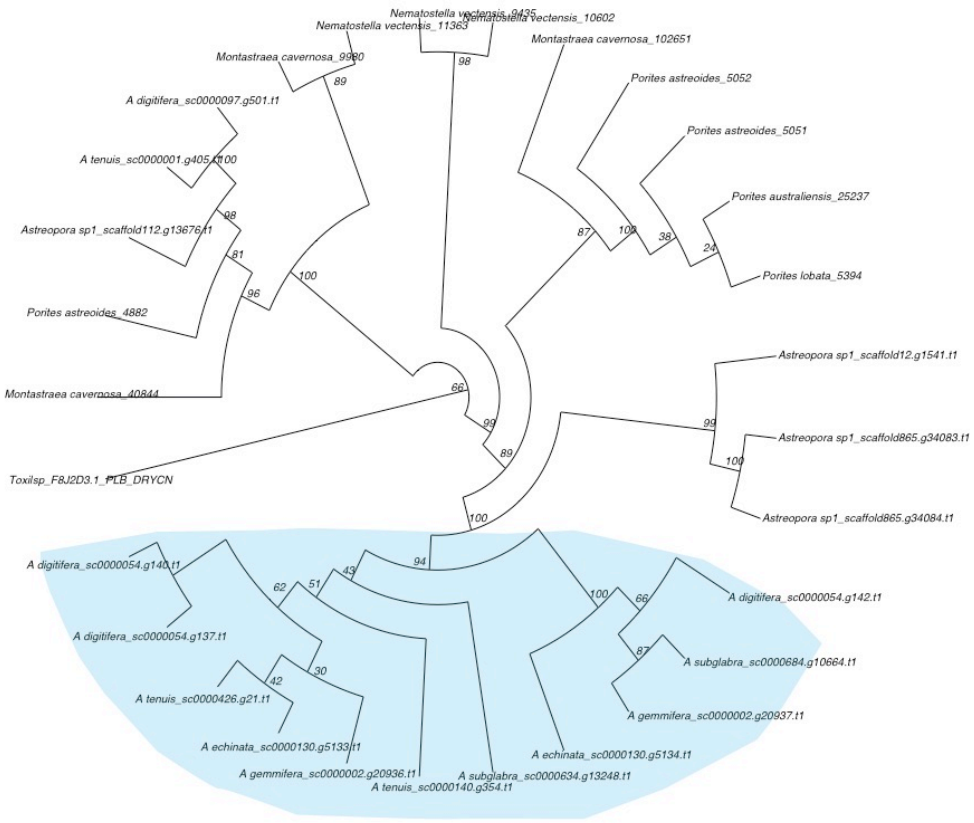


Figure S19. Phylogeny of the toxic protein (Phospholipase-B 81). Related to Figure 5. The phylogeny was reconstructed using RAxML and bootstrap values are shown at each node. Gene duplication by WGD in *Acropora* is shown in cyan.

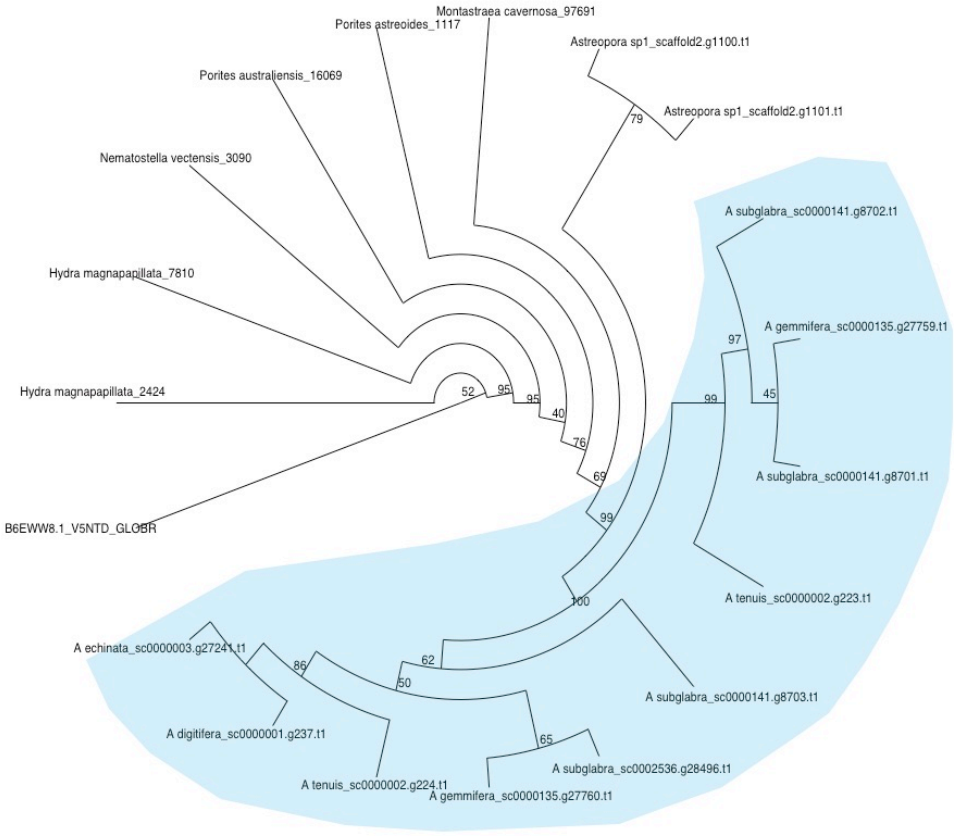


Figure S20. Phylogeny of the toxic snake venom 5'-nucleotidase. Related to Figure 5. The phylogeny was reconstructed using RAxML and bootstrap values are shown at each node. Gene duplication by WGD in *Acropora* is shown in cyan.

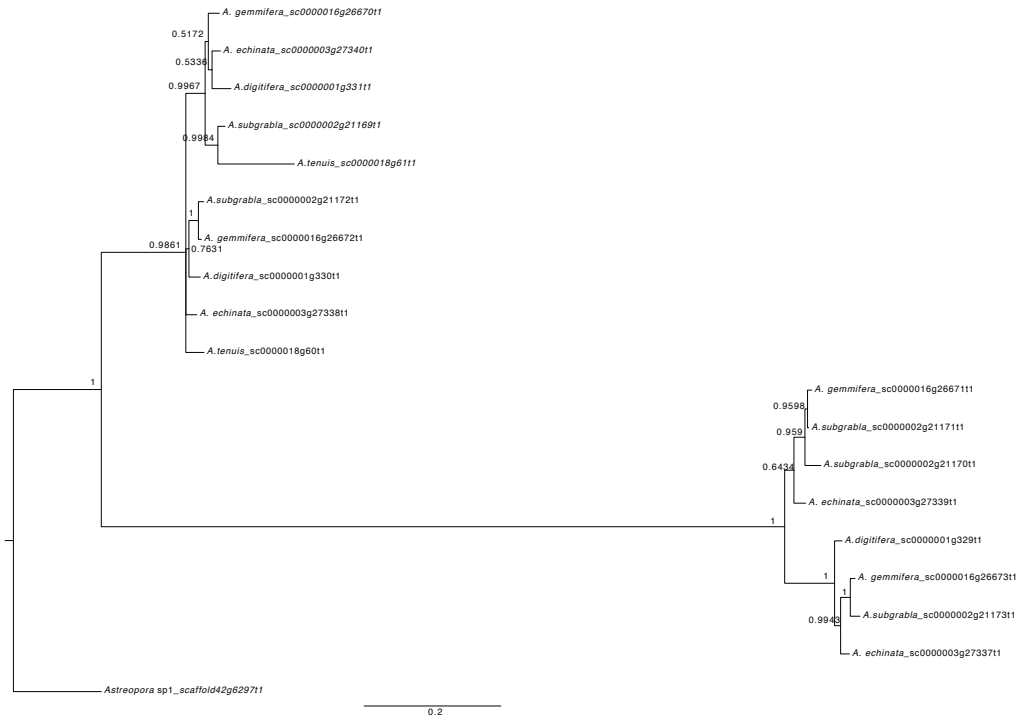


Figure S21. Phylogeny of orthogroup 434 (somatostatin receptor type 5-like) shows duplicates are under two WGD topology. Related to Figure 1. The phylogeny was reconstructed using MrBayes, and Bayesian posterior probabilities are shown at each node.

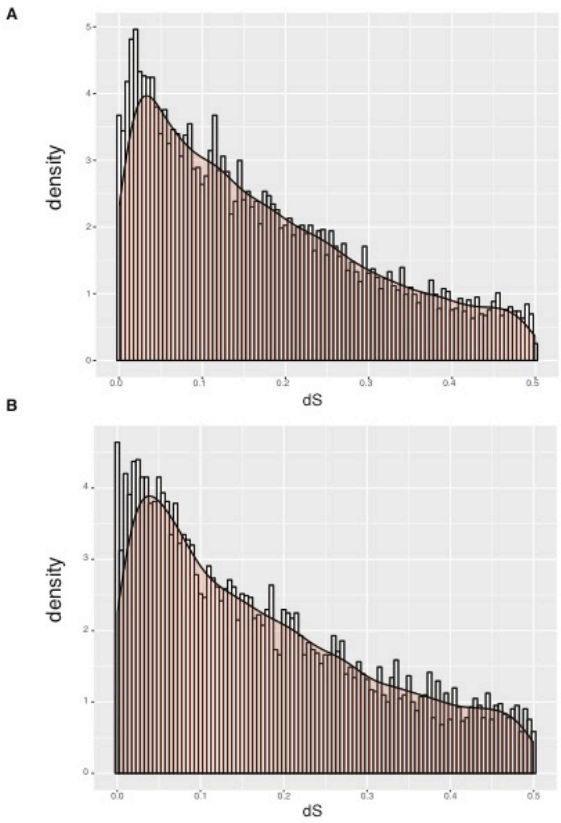


Figure S22. Frequency distributions of dS values for paralogous gene pairs in two new released genomes: (A) *Acropora millepora* and (B) *Acropora tenuis*. Related to Figure 2. The distributions of dS values of paralogs, estimating neutral evolutionary divergence since the two paralogs diverged, are plotted with a bin size of 0.005, showing the similar peaks (dS value: 0-0.3) in *Acropora*.

Supplemental Tables

Table S1. Numbers of gene pairs in the paralogous gene pairs and anchor gene pairs datasets. Related to Figure 2.

	Paralogous gene pairs (<=20 gene families)		Anchor gene pairs (<=20 gene families)	
	Total numbers	Total numbers (0<dS<2)	Total numbers	Total numbers (0<dS<2)
A. digitifera	39827	8249	46559	1958
A.echinata	47299	10948	54956	2530
A. gemmifera	48051	11093	56972	3299
A. subgrabla	50852	12077	44093	2380
A. tenuis	34097	6635	28073	1488
Astreopora sp1	49135	13648	52481	3033

Table S2. Peak value estimations of dS distribution by KDE toolbox. Related to Figure 2.

	Astreopora sp1_*A. tenuis*	*A. tenuis_A. digitifera*	*Acropora_WGD*
Peak age	53.6	14.69	35.01458704
log2(dS_paralog_peak)	-0.314	-3.4596	-1.8165
95%_HDP_log2(dS_paralog_peak)	(-0.22031,-0.33195)	(-3.4008,-3.5141)	(-1.7606,-2.1261)
95%_HDP_Age	NA	NA	(31.18,35.7)

Table S3. Numbers of gene family in orthogroups, core-orthogroups and high-quality core-orthogroups. Related to Figure 2.

	Numbers
Orthogroups	883
Core-orthogroups	205
High-quality core-orthogroups	154

Table S4. The number of putative toxin proteins in 12 Cnidarian species. Related to Figure 5.

Gene family	Query_name	*A. digitifera*	*A.echinata*	*A. gemmifera*	*A. subgla bra*	*A. tenuis*	*Astreo pora sp1*	*H. magni papill ata*	*M. cavern osa*	*N. vecten sis*	*P. astreoi des*	*P. austra liensis*	*P. lobata*
Gene family_1	Coagulation factor X	52	48	52	58	50	49	12	39	56	7	46	34
Gene family_2	Ryncolin-4	48	45	37	62	38	29	0	23	46	5	24	29
Gene family_3	Astacin-like metalloprote ase toxin	28	23	25	26	30	33	36	20	60	6	31	14
Gene family_4	Reticulocalb in	18	15	14	14	18	20	5	16	18	6	19	17
Gene family_5	Putative lysosomal acid lipase/choles teryl ester hydrolase	10	10	10	11	7	13	4	6	5	4	5	5
Gene family_6	Venom carboxyleste rase-6	8	6	9	7	7	17	1	5	14	3	8	4
Gene family_7	Putative endothelial lipase	11	1	1	17	11	25	2	2	5	1	4	5
Gene family_8	DELTA-thalatoxin-Avl2a/DELT A-alicitoxin-Pse2a	9	5	7	12	11	13	0	2	5	2	2	0

Gene family_9	Venom phosphodiesterase 2	5	6	6	6	5	7	1	7	9	3	5	5
Gene family_10	DELTA-actitoxin-Aas1a	3	4	5	5	4	3	0	1	0	1	5	4
Gene family_11		7	5	2	3	2	2	1	1	1	1	2	1
Gene family_12	Phospholipase-B 81	4	2	2	2	3	4	0	3	3	3	1	1
Gene family_13	Venom dipeptidyl peptidase 4	5	2	1	2	2	3	2	4	3	0	0	1
Gene family_14	Hyaluronidase-1	3	2	2	2	2	2	1	2	2	1	2	1
Gene family_15	Snake venom 5'-nucleotidase	1	1	2	4	2	2	2	1	1	1	1	0

Table S5. Likelihood of multiple WGDs hypotheses in *Acropora* using WGDgc method with gene counts data. Related to Figure 1.

WGD event(s)	Likelihood	Likelihood Ratio Test	P_value
0	-38731.86	0 VS 1	7.66E-05
1	-38724.04	1 VS 2	0.01248965
2	-38720.92	2 VS 3	0.05990546
3	-38719.15		

Table S6. Likelihood of different times of WGD under one WGD event in *Acropora* using WGDgc. Related to Figure 1.

Time of WGD	Likelihood
18.697005	-38724.14
22.697005	-38724.09
26.697005	-38724.06
30.697005	-38724.04
34.697005	-38724.04
38.697005	-38724.06
42.697005	-38724.11
46.697005	-38724.2
50.697005	-38724.35

Transparent Methods

Species information, genomic data and gene families cluster

Data can be accessed at: http://marinegenomics.oist.jp, https://przeworskilab.com/data/ and http://comparative.reefgenomics.org/datasets.html (Bhattacharya et al., 2016). The *Acropora* species information in this study was described in our previous paper (Mao et al., 2018). Information about *Astreopora* sp1 was described previously (Suzuki and Nomura, 2013). *Astreopora* sp1 was sampled, sequenced, and assembled in the same way of *Acropora* species (Shinzato et al., in preparation). Transcriptome data of *A. digitifera* across five development stages was described previously (Reyes-Bermudez et al., 2016). Protein sequences of the six species were combined to perform all-against-all BLASTP approach to find all orthologs and paralogs among six species. Then, OrthoMCL was used with default settings to cluster homologs into 19,760 gene families according to sequence similarity (Li et al., 2003).

Single-copy orthologs and reconstruction of a calibrated phylogenomic tree

A custom python script was used to select 3,461 single-copy orthologs with only one gene copy in each species. For each sequence alignment of single-copy orthologs, coding sequences were aligned with MAFFT (Katoh et al., 2002) as described previously (Mao et al., 2018). Then, the concatenated sequences of 3,461 single-copy orthologs were used to reconstruct the phylogenomic tree (species tree) with BEAST2 (Bouckaert et al., 2014). First, we partitioned the concatenated coding sequences by codon position. Molecular clock and trees, except substitution model, were linked together. Then, divergence time was estimated using the HKY substitution model, relaxed lognormal clock model, and calibrated Yule prior with the divergence time estimated in our previous study. We ran BEAST2 three times independently, 50 million Markov chain Monte Carlo (MCMC) generations for each run, then we used Tracer to check the log files and we found that ESS of each of parameters exceeded 200.

Orthogroup selection and detection of a WGD event with dS analysis

(a) dS distributions of paralogous gene pairs

Paralogous gene pairs of each species were identified by all-against-all BLASTP approach and then OrthoMCL was used to cluster paralogs into gene families for each species (Li et al., 2003). Gene families with fewer than 20 genes were used to calculate dS values. Each gene pair within a given gene family was aligned with MAFFT (Katoh et al., 2002) and aligned sequences were used to calculate dS values with Codeml package in PAML with parameters: noisy = 9, verbose = 1, runmode = -2, seqtype = 1, CodonFreq = 2, model = 0, NSsites = 0, icode = 0, fix_kappa = 0, kappa = 1, fix_omega = 0, and omega = 0.5 (Yang, 2007). The dS distribution of each species was plotted in R (Team, 2013). All processes were run in the GenoDup (Mao, 2019). In addition, in order to avoid bias of genomic data, GenoDup (Mao, 2019) was applied to the two recently released genomes (*Acropora tenuis and Acropora millepora*).

(b) dS distributions of anchor gene pairs

We used MCScanX with default settings (except for match_size=3) to find anchor gene pairs based on synteny information for each species (Wang et al., 2012). Each anchor gene pair was aligned with MAFFT (Katoh et al., 2002) and aligned sequences were used to calculate dS values with Codeml package in PAML with parameters: noisy = 9, verbose = 1, runmode = -2, seqtype = 1, CodonFreq = 2, model

= 0, NSsites = 0, icode = 0, fix_kappa = 0, kappa = 1, fix_omega = 0, and omega = 0.5 (Yang, 2007). The dS distribution of each species was plotted in R (Team, 2013). All processes were run in the GenoDup (Mao, 2019).

(c) *dS distributions of orthologous gene pairs*

We used MCScanX with default settings (except for match_size=3) to find orthologous gene pairs based on synteny information between *Astreopora* sp1 and *A. tenuis*, and between *A. tenuis* and *A. digitifera* (Wang et al., 2012). Each orthologous gene pair was aligned with MAFFT (Katoh et al., 2002) and aligned sequences were used to calculate dS values with Codeml package in PAML with parameters: noisy = 9, verbose = 1, runmode = -2, seqtype = 1, CodonFreq = 2, model = 0, NSsites = 0, icode = 0, fix_kappa = 0, kappa = 1, fix_omega = 0, and omega = 0.5 (Yang, 2007). dS distributions of all species were plotted in R (Team, 2013). All processes were run in GenoDup (Mao, 2019).

Detection of a WGD event using phylogenetic analysis

A custom python script was used to select the 883 gene families, including one gene copy in *Astreopora*, one gene copy in each of the five species and at least two ohnologs in one of five *Acropora* species, as orthogroups. Ohnologs are defined as paralogs originating from WGD.

For each of the 883 gene tree reconstructions, we used MAFFT (Katoh et al., 2002) to align amino acid sequences of each single-copy ortholog. We aligned coding sequences with TranslatorX (Abascal et al., 2010) based on amino acid alignments and we excluded the single-copy orthologous genes containing ambiguous 'N'. PartitionFinder (Lanfear et al., 2012) was used to find the best substitution model for RAxML (Version 8.2.2) (Stamatakis, 2014) and MrBayes (Version 3.2.3) (Ronquist et al., 2012), respectively.

Then, 205 orthogroups, for which phylogeny matched the duplication topology (*Astreaopora*, (*Acropora*, *Acropora*)), were selected as core-orthogroups by eyes. The 154 high quality core-orthogroups, for which clades' bootstrap values in ML phylogeny exceeded 70, were used to perform molecular dating with BEAST2 based on the calibrated phylogenomic tree (Bouckaert et al., 2014). Molecular clock and trees, except substitution model, were linked together. Then, divergence time was estimated using the HKY substitution model, relaxed lognormal clock model, and calibrated Yule prior with the divergence time from our previous study. We ran BEAST2 three times independently, 30 million Markov chain Monte Carlo (MCMC) generations for each run. Then we used Tracer to check the log files. 135 time-calibrated phylogenies with ESS values exceeded 200 were selected.

Estimating peak values in dS distributions and inferred node ages' distribution with KDE toolbox

Each distribution was estimated using KDE toolbox in MATLAB, as described previously (Zhang et al., 2017).

(a). Estimating peak values in distributions

To estimate the age of WGD within dS distributions, we assumed the peak value in orthologous gene pair dS distributions as the split time between two species: the split time between *Astreopora* sp1 and *A. tenuis* is 53.6 My, whereas the split time between *A. tenuis* and *A. digitifera* is 14.69 My. Before we used the *kde()* function in KDE toolbox, we first truncated dS distributions to avoid estimation bias due to extreme values: the dS distribution of orthologous gene pairs between *Astreopora* sp1 and *A. tenuis* was truncated with a range from -1 to1 while the dS distribution of

orthologous gene pairs between *A. tenuis* and *A. digitifera* was truncated with a range from -5 to -2. Then, we used the *kde()* function in KDE toolbox to estimate the peak values of these two dS distributions as -0.314 and -3.4596, respectively. Moreover, the distribution of *Acropora* paralogous gene pairs was truncated with a range from -4 to 0 and we estimated the peak value of this distribution as -1.8165. We also used bootstrapping to estimate 95% confidence intervals (CIs) of *Acropora* paralogous gene pairs distribution as -1.7606 to -2.1261 (31.18 to 35.71 My). For bootstrapping, we generated 100 bootstrap samples for each distribution by sampling with replacement from the original data distribution (49,002 samples in the original distribution) with the *sample()* function. We estimated maximum peak values for each 100 bootstrap samples. Then we sorted maximum peak values and values of 6^{th} and 95^{th} rank were used to define the 95% CI.

(b). Estimating peak values in distributions of inferred node age

To estimate the age of WGD in the distribution of inferred node ages, we used the *kde()* function in KDE toolbox to estimate the peak value as 30.78 My, and we used bootstrapping to estimate the 95% CIs as 27.86 to 34.77 My. For bootstrapping, we generated 100 bootstrap samples from the distribution by sampling with replacement from the original data distribution (135 samples in the original distribution) with the *sample()* function. We estimated maximum peak values for each of 100 bootstrap samples. Then, we sorted maximum peak values and values of 6^{th} and 95^{th} rank defined the 95% CI.

Maximum likelihood approach to detect WGD with gene family count data

First, we filtered gene family cluster data generated by OrthoMCL described above (Li et al., 2003). The gene family, including only one *Astreopora* sp1 gene and at least one gene in each of the five *Acropora* species, was counted. Then, we used the WGDgc package in R to estimate log likelihood for parameters (0, 1, 2, 3) of WGD event(s) with setting (dirac=1,conditioning="twoOrMore") (Rabier et al., 2014). Then, we performed likelihood ratio test (pchisq(2*(Likelihood_1-Likelihood_2), df=1, lower.tail=FALSE)) to find the best model and found that one WGD event was the best model to fit the gene family count data. We estimated the age of WGD on 4 My intervals between 18.69 and 38.69 My under a one WGD event model. The lowest log likelihood was shown at the age of WGD: 30.69 and 34.69 My.

Gene expression profiling analysis and dN/dS calculation

We selected 236 gene pairs of *A. digitifera* (ohnologous gene pairs) from 883 orthogroups. We BLASTed these ohnologous gene pairs against the gene expression data across five developmental stages (Reyes-Bermudez et al., 2016) and these data were normalized for each developmental stage. Correlations between two ohnologous genes were performed using Pearson's correlation in R (Team, 2013). Hierarchical clustering was performed using Pheatmap for HC cluster genes and NC cluster genes, respectively. Pairwise dN/dS ratios were calculated with PAML using codeml based on the coding sequence alignment of ohnologous gene pairs with parameters: noisy = 9, verbose = 1, runmode = -2, seqtype = 1, CodonFreq = 2, model = 0, NSsites = 0, icode = 0, fix_kappa = 0, kappa = 1, fix_omega = 0, and omega = 0.5 (Yang, 2007). The dN/dS distribution was plotted with ggplot2 in R and significance tests of differences between dN/dS distributions were evaluated by a Mann-Whitney test in R (Team, 2013).

Evolution analysis of toxic proteins in corals

The 55 toxic proteins of *A. digitifera* identified in the previous study were downloaded from http://www.uniprot.org/ as queries. The protein sequences of *Porites astreoides*, *Porites australiensis*, *Porites lobata*, *Montastraea cavernosa*, *Hydra magnipapillata* and *Nematostella vectensis* were downloaded from http://comparative.reefgenomics.org/datasets.html (Bhattacharya et al., 2016), and combined them with protein sequences of six *Acroporid* species to create a search database.

We identified candidates of toxic proteins by BLASTing the 55 toxins against the combined protein sequences with settings: e-value < $1e^{-20}$ and identity > 30%. Then, we used OrthoMCL to cluster candidates of toxins into 24 gene families and reconstructed their ML gene trees with ExaML (Kozlov et al., 2015) and RAxML. Each gene tree was rooted at a branch or clade of query sequences.

Gene ontology enrichment for duplicated genes of core-orthogroups and protein domains and transmembrane helices prediction

We BLASTed the sequences of 154 high quality core-orthogroups of *Acropora* against the UNIPROT database to find best hits. Identical hits in each ohonlogs group were removed and the remaining hits were used to perform gene enrichment in David (Huang et al., 2009). We also used InterProScan (Zdobnov and Apweiler, 2001) to predict protein domains and used the TMHMM Server (v. 2.0) (Krogh et al., 2001) to predict transmembrane helices from protein sequences.

Supplemental References

Abascal, F., Zardoya, R., and Telford, M.J. (2010). TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. Nucleic acids research *38*, W7-13.

Bhattacharya, D., Agrawal, S., Aranda, M., Baumgarten, S., Belcaid, M., Drake, J.L., Erwin, D., Foret, S., Gates, R.D., Gruber, D.F., et al. (2016). Comparative genomics explains the evolutionary success of reef-forming corals. Elife *5*.

Bouckaert, R., Heled, J., Kuhnert, D., Vaughan, T., Wu, C.H., Xie, D., Suchard, M.A., Rambaut, A., and Drummond, A.J. (2014). BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS computational biology *10*, e1003537.

Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature Protocols *4*, 44-57.

Kajitani, R., Toshimoto, K., Noguchi, H., Toyoda, A., Ogura, Y., Okuno, M., Yabana, M., Harada, M., Nagayasu, E., Maruyama, H., et al. (2014). Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. Genome research *24*, 1384-1395.

Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic acids research *30*, 3059-3066.

Kozlov, A.M., Aberer, A.J., and Stamatakis, A. (2015). ExaML version 3: a tool for phylogenomic analyses on supercomputers. Bioinformatics *31*, 2577-2579.

Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E.L.L. (2001). Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. Journal of Molecular Biology *305*, 567-580.

Lanfear, R., Calcott, B., Ho, S.Y.W., and Guindon, S. (2012). PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Molecular biology and evolution *29*, 1695-1701.

Li, L., Stoeckert, C.J., Jr., and Roos, D.S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome research *13*, 2178-2189.

Mao, Y., 2019. GenoDup Pipeline: a tool to detect genome duplication using the dS-based method. PeerJ, 7, p.e6303.

McKain, M.R., Tang, H., McNeal, J.R., Ayyampalayam, S., Davis, J.I., dePamphilis, C.W., Givnish, T.J., Pires, J.C., Stevenson, D.W. and Leebens-Mack, J.H., (2016). A phylogenomic assessment of ancient polyploidy and genome evolution across the Poales. Genome biology and evolution, 8(4), pp.1150-1164.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Hohna, S., Larget, B., Liu, L., Suchard, M.A., and Huelsenbeck, J.P. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Systematic biology *61*, 539-542.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics *30*, 1312-1313.

Team, R.C. (2013). R: A language and environment for statistical computing.

Wang, Y.P., Tang, H.B., DeBarry, J.D., Tan, X., Li, J.P., Wang, X.Y., Lee, T.H., Jin, H.Z., Marler, B., Guo, H., et al. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic acids research *40*.

Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. Molecular biology and evolution *24*, 1586-1591.

Zdobnov, E.M., and Apweiler, R. (2001). InterProScan–an integration platform for the signature-recognition methods in InterPro. Bioinformatics *17*, 847-848.