Okinawa Institute of Science and Technology

Graduate University

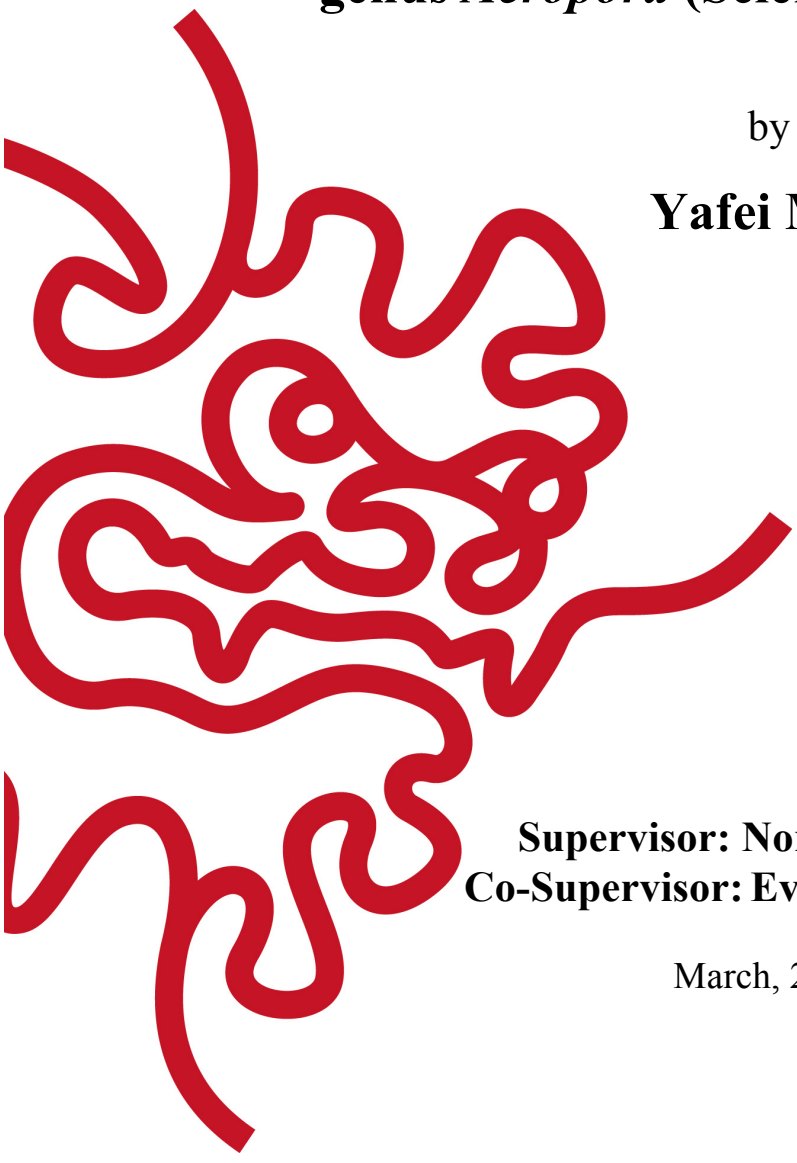
Thesis submitted for the degree

Doctorate of Philosophy

Whole-genome sequence analysis of the evolutionary history of the reef-building coral genus *Acropora* (Scleractinia, Cnidaria)

by

Yafei Mao



Supervisor: Noriyuki Satoh
Co-Supervisor: Evan P. Economo

March, 2019

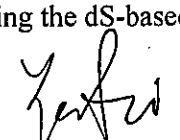
Declaration of Original and Sole Authorship

I, Yafei Mao, declare that this thesis entitled "Whole-genome sequence analysis of the evolutionary history of the reef-building coral genus *Acropora* (Scleractinia, Cnidaria)" and the data presented in it are original and my own work.

I confirm that:

- No part of this work has previously been submitted for a degree at this or any other university.
- References to the work of others have been clearly acknowledged. Quotations from the work of others have been clearly indicated, and attributed to them.
- In cases where others have contributed to part of this work, such contribution has been clearly acknowledged and distinguished from my own work.
- None of this work has been previously published elsewhere, with the exception of the following: (* Corresponding author)

1. **Mao, Y.** *, Economo, E.P. * and Satoh, N. *, 2018. The roles of introgression and climate change in the rise to dominance of *Acropora* corals. *Current Biology*, 28(21), pp.3373-3382.
2. **Mao, Y.** * and Satoh, N., 2019. A Likely Ancient Genome Duplication in the Speciose Reef-Building Coral Genus, Acropora. iScience, 13, pp.20-32.
3. **Mao, Y.** *, 2019. GenoDup Pipeline: a tool to detect genome duplication using the dS-based method. PeerJ, 7, p.e6303.

Signature: 

Date: 6/9/2019

Whole-genome sequence analysis of the evolutionary history of the reef-building coral genus *Acropora* (Scleractinia, Cnidaria)

by
Yafei Mao

Submitted to the Graduate School in partial fulfillment of the requirements for the degree of Doctorate of Philosophy at the Okinawa Institute of Science and Technology Graduate University

Abstract

A major goal of evolutionary biology is to understand the roles of evolutionary and ecological factors in rapid speciation and diversification. Introgression and ancient large-scale/whole genome duplication (paleopolyploidy) have been hypothesized to promote on rapid speciation leading to diversification. In addition, diversification can be promoted by 'ecological opportunity' created by extinction of competitors or the colonization of a new area. Reef-building corals are the foundation of diverse tropical ecosystems, but are currently under threat due to the sensitivity of corals to climate change and anthropogenic factors. *Acropora* (Anthozoa: Acroporidae) is one of the most diverse genera of reef-building corals, including more than 150 species, and based on the fossil record has dominated Indo-Pacific reefs in past 3 Million Years, yet the evolutionary and ecological factors associated with its diversification and the rise to dominance are unclear. Understanding the evolutionary history of this group during its rise to dominance may help understanding their current and future responses to global change. In this dissertation, I used genomic data of *Acropora* generated by Dr. Chuya Shinzato to investigate its evolutionary history and illuminate the roles of introgression, large-scale genome duplication, and ecological opportunity in its diversification and the rise to dominance. In the first chapter, I reviewed recent studies of *Acropora*. In the second chapter, I examined the roles of introgression in *Acropora*. I found that a major introgression event and widespread gene flow occurred in five *Acropora* species, and that introgression genes evolved faster than others. In the third chapter, I examined the roles of climate change in the rise to dominance of *Acropora*. I found that *Acropora* lineages had an experience of population expansion after a climate-driven mass extinction event in the Plio-Pleistocene, suggesting ecological opportunity facilitated the rise to dominance of *Acropora*. In the fourth chapter, I examined evidence for large-scale genome duplication and its consequences in *Acropora*. I found a large-scale genome duplication event likely occurred in *Acropora* and duplicated genes play important roles in the diversification of *Acropora*. Finally, in the fifth chapter, I discussed limitations and future directions arising from this dissertation. Collectively, this dissertation suggests that introgression, climate change, and large-scale genome duplication play important roles in the evolutionary history of *Acropora*.

Acknowledgments

I wish to thank all people who have helped me and encouraged me during my PhD studies.

I wish to thank my supervisor, Noriyuki Satoh, for his support and mentorship. NS provides fantastic resources, absolute freedom and sound advice throughout my PhD studies. I also wish to thank my co-supervisor, Evan P Economo, for his valuable advice and supervision not only in academia but also in the personal career.

I thank for my collaborator, Chuya Shinzato, for kindly providing the genomic data to analyze in this dissertation. I also thank all lab members in Satoh Unit and Economo Unit for their insightful discussions and suggestions.

I appreciate the support and help from my lab administrators, Tomomi Teruya and Shoko Yamakawa, as well as the supports from OIST Graduate School. In addition, thank OIST scientific computing and data analysis section for the help of research computing.

Thank Ricardo Mallarino for providing an opportunity to study in his lab at Princeton University. And thanks to OIST Fellowship and JSPS DC1 Fellowship.

Thanks to my family for their love and support in my life. And many thanks to my dear friends during my PhD studies: Cong Liu, Hong Huat Hoh, Junfeng Shi, Krishnapriya Subramonian Rajasree, Lijun Qiu, Matti Krueger, Mengling Wang, Shuyin Huang, Tingting Tao, and Xi Zhe.

List of Abbreviations

ABAB-BABA test	Patterson's D statistics
ATP6	ATP synthase Fo subunit 6
CFs	concordance factors
CNOX	a HOM/HOX homeobox gene
Cyt-b	cytochrome b
DDC	duplication-degeneration-complementation
dN/dS	nonsynonymous/synonymous substitution ratios
DNAJB	dnaJ homolog subfamily B member 11-like
dS	the rate of synonymous substitution per synonymous site
EAC	escape from adaptive conflict
EST	expressed-sequence-tag
GATK	Genome Analysis Toolkit
GO	gene ontology
HPD	highest posterior density
IAsα	invertebrate α event of GD specifically in *Acropora*,
ILS	incomplete lineage sorting
IMCoalHMM	the coalescent hidden Markov model
Ky	thousand year
MCOL	mini-collagen
MDN	new functions under non-functionality
ML	Maximum likelihood
MPT	mid-Pleistocene transition
Mya	million years ago
NGS	next generation sequencing
OMT	Oligocene-Miocene transition
Pax	paired box gene
PCR	polymerase chain reaction
PSMC	pairwise sequentially Markovian coalescent
rDNA	ribosomal DNA
SNP	single nucleotide polymorphism
TEs	transposon elements
GD	whole (large-scale)-genome duplication

Table of Contents

Abstract
Acknowledgments
Chapter 1. Introduction 1
1.1 Introduction to the Genus *Acropora* 2
1.1.1 Basic information of *Acropora* 2
1.1.2 Previous phylogenetic studies of *Acropora* 6
1.2 Contents of this dissertation 8
Chapter 2. Introgression facilitated the diversification of reef-building coral *Acropora* 9
2.1 Introduction 9
2.2 Methods 13
2.2.1 Genomic data, gene family clustering and single-copy ortholog selection 13
2.2.2 Gene tree and phylogenomic tree reconstruction 13
2.2.3 Bayesian concordance analysis using BUCKy 14
2.2.4 Phylonetwork inferrence from gene trees using Phylonet and SNaQ 14
2.2.5 Genome-wide Patterson's D statistics (ABAB-BABA test) 15
2.2.6 Speciation with isolation and speciation with migration modeling using IMCoalHMM 16
2.2.7 Pairs of single-copy orthologous gene dN/dS ratios calculation 17
2.2.8 Gene ontology (GO) 18
2.3 Analyses and Results 18
2.3.1 Gene family cluster and phylogenomic tree reconstruction 18
2.3.2 Test for introgression by ABBA-BABA test and Phylonetwork theory 20
2.3.3 Syngameon hypothesis identification 25
2.3.4 Evolutionary rates and patterns of selection 27
2.4 Discussion 32
Chapter 3. Climate change provided an ecological opportunity for the rise to dominance of *Acropora* in the Plio-Pleistocene 33
3.1 Introduction 33
3.2 Methods 35
3.2.1 Phylogenomic tree dating with BEAST2 35
3.2.2 Whole genome alignment and mutation rate estimation 36
3.2.3 Demographic history reconstruction using PSMC 37
3.3 Analyses and Results 38
3.3.1 Time-calibrated phylogenomic tree reconstruction 38
3.3.2 Demographic inference with PSMC 40
3.4 Discussion 42
Chapter 4. A likely ancient genome duplication in the speciose reef-building coral genus: *Acropora* 48
4.1 Introduction 48
4.2 Methods 52
4.2.1 Species information, genomic data and gene family cluster 52
4.2.2 Single-copy orthologs and reconstruction of a calibrated phylogenomic tree 53

4.2.3 Orthogroup selection and detection of a GD event with dS analysis 54
4.2.4 Detection of a GD event using phylogenetic analysis 55
4.2.5 Estimating peak values in dS distributions and inferred node ages' distribution with KDE toolbox 56
4.2.6 Maximum likelihood approach to detect GD with gene family count data 57
4.2.7 Gene expression profiling analysis and dN/dS calculation 58
4.2.8 Evolution analysis of toxic proteins in corals 58
4.2.9 Gene ontology enrichment for duplicated genes of core-orthogroups and protein domains and transmembrane helices prediction 59
4.3 Analyses and Results 59
4.3.1 Cluster of gene families and calibration of the acroporid phylogenomic tree 59
4.3.2 GD identification with the dS-based method 61
4.3.3 Phylogenomic and synteny analysis of IAsα 67
4.3.4 The fate of duplicated genes originating from IAsα 72
4.3.5 Gene expression patterns of duplicated genes across five developmental stages in *A. digitifera* 76
4.3.6 Evolution of toxic proteins in Cnidaria 77
4.4 Discussion 79
Chapter 5. Conclusions and limitations of this dissertation 85
5.1 Introgression and gene flow in *Acropora* 85
5.2 Ancient GD shared by *Acropora* 86
5.3 Climate change facilitated the rise to dominance of *Acropora* 87
5.4 Future directions 88
Appendix Genome assembly and annotation statistics of the six coral species **89**
Bibliography 90

List of Figures

Figure 1.1 | Diagram of coral polyp structure
Figure 1.2 | Photographs of five *Acropora* species and an acroporid coral mass spawning
Figure 2.1 | The evolutionary history of *Acropora* inferred from five genomes.
Figure 2.2 | Phylogenomic trees reconstructed by RAxML and MrBayes
Figure 2.3 | The five most common gene tree topologies inferred with MrBayes
Figure 2.4 | Bayesian concordance analysis
Figure 2.5 | Phylonetwork inferred by Phylonet
Figure 2.6 | Phylonetwork inferred by SNaQ
Figure 2.7 | Four taxon ABBA–BABA analysis
Figure 2.8 | Results of speciation with migration model inferred with IMCoalHMM.
Figure 2.9 | Evolutionary rates of introgression and non-introgression genes
Figure 3.1 | Time calibrated phylogenomic tree of *Acropora, Porites* and *Orbicellaa*
Figure 3.2 | Demographic history of *Acropora* lineages
Figure 4.1 | GD events in evolution of the animal kingdom
Figure 4.2 | Venn diagrams of six Acroporid species
Figure 4.3 | Phylogeny of the Family Acroporidae
Figure 4.4 | Frequency distribution of dS values for paralogous gene pairs in five *Acropora* and one *Astreopora* species
Figure 4.5 | Frequency distribution of dS values for anchor-gene pairs in five *Acropora* and one *Astreopora* species
Figure 4.6 | Frequency distribution of dS values for paralogous genes in *Acropora* and for orthologous genes
Figure 4.7 | Hypothetical tree topology of duplicated genes in the Acroporidae and the phylogeny of one duplicated gene (alpha-protein kinase 1-like)
Figure 4.8 | Node age distribution of IAsα
Figure 4.9 | Synteny blocks between *Astreopora* sp1 and *A. tenuis*
Figure 4.10 | Co-linear alignments of *Astreopora* sp1and *A. tenuis*
Figure 4.11 | Ancient GD in the reef-building coral *Acropora* (IAsα)
Figure 4.12 | Phylogenetic trees show duplicated genes under subfunctionalization or neofunctionalization
Figure 4.13 | Alignment of orthogroup 1247 (dnaJ homolog subfamily B member 11-like) showing the independent loss of the domain in duplicates
Figure 4.14 | Alignment of orthogroup 1244 (excitatory amino acid transporter 1-like) showing mutations on transmembrane and exposed regions, suggesting that new functions would be generated
Figure 4.15 | Gene expression profiling reveals evolution of duplicated genes in *A. digitifera*
Figure 4.16 | Diversification of toxic proteins via gene duplications in Cnidaria
Figure 4.17 | Phylogeny of orthogroup 434 (somatostatin receptor type 5-like) shows duplicates are under two GD topology

List of Tables

Table 1.1 | Ecological habitats and morphology of five Acropora species collected in Okinawa
Table 2.1 | Likelihoods and information criteria of Phylonet models fit with different numbers of reticulation events.
Table 2.2 | Statistics of ABBA-BABA test
Table 2.3 | Average IAIC, IMAIC and delta AIC values in species pairs inferred with IMCoalHMM
Table 2.4 | GO enrichment for introgression genes comparing to species tree genes
Table 2.5 | Annotation of non species-tree genes under positive selection
Table 2.6 | Non species-tree genes under selection in the four species pairs
Table 4.1 | Numbers of gene pairs in the paralogous gene pairs and anchor gene pairs datasets
Table 4.2 | Peak value estimations of dS distribution by KDE toolbox
Table 4.3 | Numbers of gene family in orthogroups, core-orthogroups and high-quality core-orthogroups
Table 4.4 | Functional annotation clustering on the GO terms of 154 high-quality core-orthogroups
Table 4.5 | The number of putative toxin proteins in 12 Cnidarian species
Table 4.6 | Likelihood of multiple GDs hypotheses in Acropora using GDgc method with gene counts data.
Table 4.7 | Likelihood of different times of GD under one GD event in Acropora using GDgc
Table A.1 | Raw data and coverage calculation
Table A.2 | Genome statistics and annotation

Chapter 1

Introduction

Understanding the biodiversity is one of the major goals in evolutionary biology (Helfman et al., 2009; Nosil et al., 2017; Schluter, 2000; Schluter and Pennell, 2017; Weber et al., 2017). In the 'Genomic Era', advances on technologies, such as Next Generation Sequencing (NGS), allow us to investigate molecular mechanisms of organismal diversification (Berner and Salzburger, 2015; Metzker, 2010; Neale et al., 2017; Seehausen et al., 2014). In recent decades, studies with large-scale analyses of genomic data found that the most of organism groups under rapid speciation or/and diversification undergo introgression and large-scale genome duplication (GD or paleopolyploidy), such as Darwin's finches, Cichlid fish, and green plants (Berner and Salzburger, 2015; Lamichhaney et al., 2015; Meier et al., 2017; Seehausen, 2015; Van de Peer et al., 2009; Van De Peer et al., 2017).

Coral reef ecosystems have long captivated both scientists and the general public (Ainsworth et al., 2016; Hughes et al., 2017). However, we have much to learn about the evolution of the organisms that form their basis: reef-building corals. Reef-building corals provide the structural basis for one of Earth's most spectacular and diverse—but increasingly threatened—ecosystems (Ainsworth et al., 2016; Hemond and Vollmer, 2010; Hughes et al., 2017; Shinzato et al., 2011). Modern Indo-Pacific reefs are dominated by species of the staghorn coral genus *Acropora* (Anthozoa: Acroporidae), one of the most diverse genera with close to 150 species (Fukami et al., 2008; Fukami et al., 2000; van Oppen et al., 2001; Wallace, 1999; Wallace and Rosen, 2006). Previous studies suggested that introgression probably has a huge impact on the diversification of *Acropora* (Montaggioni and Braithwaite, 2009; van Oppen et al., 2001). Meanwhile, *Acropora* is suspected to originate from polyploidy

(Kenyon, 1997; van Oppen et al., 2001; Vollmer and Palumbi, 2002; Willis et al., 2006). However, as yet there is no genomic evidence to support the hypotheses. Our group sequenced the first coral genome in 2011 (Shinzato et al., 2011) and continues on the genomic projects of reef-corals (see http://marinegenomics.oist.jp/). Hence, in order to understand the diversification and the rise to dominance of *Acropora*, I studied the newly-sequenced genomes of six coral species (five *Acropora* and one *Astreopora*); and investigated what the roles of introgression, climate change and large-scale genome duplication play in the evolutionary history of *Acropora* in genomic perspectives. Next, I will review basic information of *Acropora* and the phylogenic studies of *Acropora*.

1.1 Introduction to the Genus *Acropora*

The Genus *Acropora* (Family Acroporidae, Class Anthozoa, Order Scleractinia, Phylum Cnidaria), comprising at least 150 species and 20 species groups, is one of the most diverse genera of reef-building corals in the Indo-Pacific Ocean (Fukami et al., 2008; Fukami et al., 2000; van Oppen et al., 2001; Wallace, 1999; Wallace and Rosen, 2006). However, due to recent increases in seawater temperatures, seawater acidification, pollution, and overdevelopment, both species and genetic diversity within *Acropora* are rapidly declining (Hemond and Vollmer, 2010). This has severely impacted tropical ecosystems in the Indo-Pacific (Ainsworth et al., 2016; Hemond and Vollmer, 2010; Hughes et al., 2017; Shinzato et al., 2011).

1.1.1 Basic information of Acropora

There are six colony shapes of *Acropora*: corymbose (*A. tenuis*), digitate (*A. digitifera*), hispidose (*A. echinata*), arborescent (*A. formosa*), arborescent table (*A.*

valenciennesi), and plate-like (*A. clathrata*) (Wallace, 1999). Regardless of colony shape, each colony consists of numerous polyps, each of which projects a mouth surrounded by tentacles into the external environment. Polyps have two epithelia, oral and aboral, in cross section (Work et al., 2008). Each epithelium contains two single cell layers: ectoderm and endoderm. Two single cell layers are separated by an acellular layer: mesoglea. In addition, endodermal cell layers encircle the gastric cavity or coelenteron. The oral ectoderm faces seawater and the aboral ectoderm covers the skeleton, forming the calcidodermis (Marshall et al., 2007; Woodley et al., 2016). Importantly, *Acropora* form a mutualistic symbiosis with dinoflagellates (eg, Symbiodinium sp.), which reside in the oral endoderm (gastrodermal cells). In addition, individual polyps are housed in a skeletal casing, the corallite. Adjacent corallites are connected by the coenosteum. While, polyps are connected through the coenenchyme (coenosarc) (Figure 1.1), the structure of which may allow individual polyps to share nutrients with others in the same colony (Marshall et al., 2007; Woodley et al., 2016). Numerous studies have focused on the symbiotic relationship between *Acropora* and dinoflagellates (Lin et al., 2015; Sheppard et al., 2017; Shoguchi et al., 2013), but this is beyond the scope of my study, so I will not discuss it in detail here.

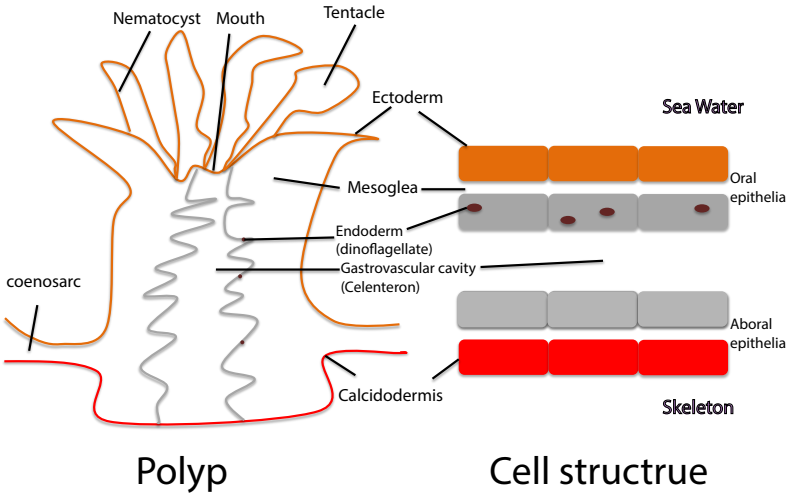


Figure 1.1. **Diagram of coral polyp structure.** Endoderm is in grey and oral ectoderm is in orange, and aboral ectoderm (calcidodermis) is in red. Symbiotic dinoflagellates, which reside in oral endoderm (gastrodermis), are represented by brown dots.

Branches of *Acropora* are typically formed by axial corallites rather than radial corallites. *Acropora* reproduction is unique among corals. Gonads are attached to mesenteries, but mature sperm and eggs are released from polyps, while fertilization and development of zygotes are external (Kojis, 1986; Wallace, 2011). In contrast, in the sister-genus, Isopora, oocytes are borne in the mesenteries, and fertilization and development of zygotes occur in the polyps (Kojis, 1986). It is worth noting that unique polyp characteristics, reproductive biology, and skeletal structures (e.g., septa coenosteum and synapticulate framework) are diagnostic for the Genus *Acropora*. Hence, 20 species groups within *Acropora* are classified according to its ecological habitats and morphology (Kojis, 1986; Renema et al., 2016; Wallace, 2012; Wallace and Rosen, 2006). Noteworthy, previous studies have shown that morphological characteristics of modern *Acropora,* such as skeletogenesis, are heavily influenced by environmental factors in the Indo-Pacific Ocean (Bak, 1983; Faith and Richards, 2012). In detail, the morphology of conspecific individuals varies

according to water depth and depending upon environment stressors, such as increased temperature and pH (Faith and Richards, 2012).

Here, I briefly summarize the ecological characters and morphology of *Acropora* corals studied in this dissertation and all of them were collected in Okinawa (Table 1.1, Figure 1.2).

Table 1.1. Ecological habitats and morphology of five *Acropora* species collected in Okinawa

Spcies	Colony shape	Niches habitat	Spawning time	Species group
A. tenuis	Corymbose	5-20 meter	7-8 pm	*A. selago*
A. digitifera	Digitate	0-10 meter	9-10 pm	*A. humilis*
A. gemmifera	Digitate	0-10 meter	11-12 pm	*A.humilis*
A. echinata	Hispidose	15-30 meter	9-10 pm	*A.echinata*
A. subglabra	Hispidose	15-30 meter	9-10 pm	*A. echinata*

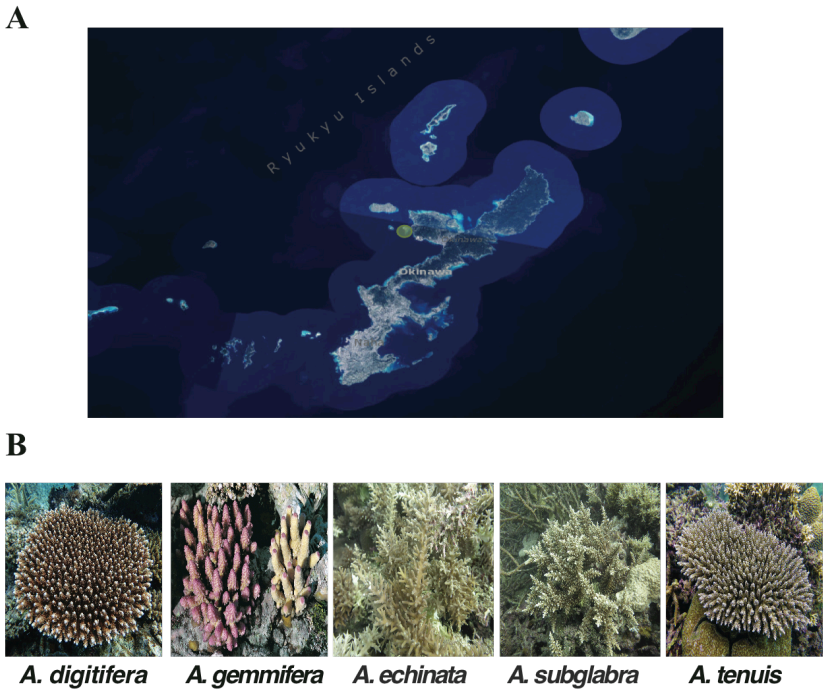


Figure 1.2. ***Acropora* species and sampling location.** (A) Sampling location is marked with a yellow circle in Okinawa map generated by ArcGIS. (B) Five Acropora species photos provided by Dr. Yuna Zayasu (Zayasu and Shinzato, 2016).

1.1.2 Previous phylogenetic studies of Acropora

Acropora is well-represented in the fossil records and the earliest known *Acropora* fossil was discovered in Somalia and dated to the Late Paleocene (54-65 million years ago (Mya)) (Wallace, 2011; Wallace, 2012; Wallace and Rosen, 2006). Based on its rich fossil records, modern *Acropora* have diversified within the past 10 million years and have been the rise to dominance in the past 3 million years through the Indo-Pacific Ocean (Baird et al., 2001; Renema et al., 2016; Vanneste et al., 2014; Wallace, 2012). Considering its physical characteristics, its dominance is facilitated by its 'synapticular' skeletal framework, which allows for rapid growth and efficient skeletogenesis (Renema et al., 2016; Sheppard et al., 2017; Wallace, 2011). In addition, *Acropora* is capable of mass spawning and rapid recolonization (Bak, 1983; Vollmer and Palumbi, 2002).

In order to investigate the evolutionary history of *Acropora* with molecular evidence, various DNA fragments have been used as DNA markers to reconstruct *Acropora* phylogeny (Fukami et al., 2000; Liu et al., 2015; Márquez et al., 2003; Rosser et al., 2017; van Oppen et al., 2002; van Oppen et al., 2001; Wei et al., 2006). Although Cyt-b is a common marker for phylogeny reconstruction, it is extremely conserved in *Acropora*; thus, it is not an informative maker for this genus (van Oppen et al., 2002). Additionally, use of ribosomal RNA sequences to reconstruct phylogeny is also problematic in *Acropora*, as *Acropora* RNA sequences are highly diversified (Wei et al., 2006). By far, *Acropora* phylogenetic trees have been reconstructed based on single markers (MCOL, Cnox2, Calmodulin or the intron of Pax-C) (Faith and Richards, 2012) or microsatellites or single nucleotide polymorphisms (SNPs), such as in *A. palmate*, *A. millepora,* and *A. hyacinthus* (van Oppen and Gates, 2006). Besides, a phylogenetic tree reconstructed using mitochondrial genes (ATP6 and Cyt-

b) calibrated with fossil information suggests that modern *Acropora* is split from other corals about at 6.6 Mya (Fukami et al., 2000). However, another phylogenetic tree reconstructed by nuclear DNA (*Pax-C*) and mitochondrial genes with fossil information calibration showed that modern *Acropora* is split from other corals at 36 Mya (Richards et al., 2013). A recent study showed that *Acropora* is split from other corals at 15 Mya (Richards et al., 2013).

Although previous studies have attempted to determine the phylogeny of *Acropora* to investigate the evolutionary history of *Acropora*, as yet there is no conclusive phylogenetic tree (Faith and Richards, 2012; Richards et al., 2013; van Oppen et al., 2002; van Oppen et al., 2001). First, phylogeny construction of *Acropora* is severely limited by lacking of informative molecular markers (van Oppen and Gates, 2006). Furthermore, phylogenetic relationships of recently diverged species are not easy to be resolved by a few markers (Ohta, 1992). In particular, the Genus *Acropora* is diversified in a short time, so the few available markers do not yield a stable phylogeny. On the other hand, phylogenetic trees reconstructed by different markers are incongruous, suggesting that incomplete lineage sorting (ILS) or introgression may have occurred in *Acropora*. Remarkably, there is no strong evidence to identify ILS and/or introgression for these inconsistent phylogenetic trees, due to limitations of the methods available at that time of tree constructions (Faith and Richards, 2012; Richards et al., 2013; van Oppen et al., 2001). Importantly, although there is no conclusive phylogeny of Acropora, a few studies have shown that there are four major clades in the phylogeny of Acropora (Márquez et al., 2002; Shinzato et al., 2014; van Oppen et al., 2001).

Meanwhile, polyploidy has long been suspected in the evolution of *Acropora*. First, the simultaneous mass spawning of *Acropora* provides a unique fertilize

strategy for hybridization (Baird et al., 2001). And some interspecific fertilize experiments showed that there are some possibilities for different *Acropora* species to generate hybrid offsprings both in wild and in lab (Vollmer and Palumbi, 2002). Secondly, the previous research found that different *Acropora* species have different chromosome numbers (Kenyon, 1997). These studies suggest that *Acropora* may originate from polyploidy (Willis et al., 2006).

1.2 Contents of this dissertation

Our group decoded the first *Acropora* genome (*A. digitifera*) in 2011 and continues working on coral genomic projects (Shinzato et al., 2011). Dr. Chuya Shinzato decoded other four *Acropora* genomes (*A. gemmifera*, *A. subglabra*, *A. echinata* and *A. tenuis*) and an *Astreopora* genome (*Asteropora* sp1) with high coverage recently (Shinzato et al., in preparation; see http://marinegenomics.oist.jp/). Hence, I used these genomic data to investigate the evolutionary history of *Acropora* in this dissertation.

Introgression has been regarded as a crucial way of rapid speciation enhancing the diversification of organisms and it has been a long-standing question in *Acropora* (van Oppen et al., 2001). Thus, in the second chapter, I used phylogenomic and coalescent hidden Markov model approaches to test for the presence and nature of introgression in *Acropora.* In addition, I also investigated the putative adaptive introgression in *Acropora*. Fossil records showed that *Acropora* are originated from 60 Mya but it becomes dominant species in Indo-Pacific Ocean until recent 3 Mya (Renema et al., 2016; Wallace, 2012). Therefore, in the third chapter, I used genomic data to reconstruct the high quality time-calibrated phylogeny of *Acropora* and used demographic inference to examine the roles of ecological opportunity in the rise to dominance of *Acropora*. The origin of Indo-Pacific *Acropora* is suspected from

polyploidy (Willis et al., 2006). Thus, in the fourth chapter, I analyzed the five *Acropora* genomes with an *Astreopora* genome to investigate whether and when large-scale genome duplication occurred in *Acropora* using comprehensive phylogenomic and dS-based approaches, and what the fate befell duplicated genes in *Acropora* after the event(s).

Chapter 2

Introgression facilitated the diversification of reef-building coral *Acropora*

2.1 Introduction

Reef-building corals support one of the most productive and diverse ecosystems on our planet (Bhattacharya et al., 2016; Wallace and Rosen, 2006), but they are increasingly threatened due to recent increases in seawater temperatures, pollution, and rapid sea-level changes (Shinzato et al., 2015; Shinzato et al., 2011). Modern Indo-Pacific reefs are dominated by species of the staghorn coral genus *Acropora* (Anthozoa: Acroporidae), one of the most diverse genera with close to 150 species, but the evolutionary factors associated with its diversification are unclear. Understanding those factors provides critical context for evaluating the resilience of the *Acropora*, and thus reef ecosystems as a whole, to the ongoing global changes of the Anthropocene.

Recent work on evolutionary radiations across a wide range of taxa has demonstrated the importance of introgression in promoting diversification (Meier et al., 2017b; Meyer et al., 2016; Wagner et al., 2012). Introgression can promote diversification by generating the genotypic and phenotypic variance necessary for natural selection and adaptation, and can facilitate the spread of favorable alleles across species (Berner and Salzburger, 2015; Heliconius Genome, 2012; Seehausen, 2004). Given the complexity of morphological variation in corals, problems with resolving phylogenetic relationships, and other evidence, the idea that introgression is important for coral evolution has long been suspected and debated (Grigg, 1995; Montaggioni and Braithwaite, 2009; van Oppen et al., 2001).

Introgression can occur in well-defined hybridization events that transfer a large amount of genetic material between two lineages and creating a "hybrid swarm", or occur continuously at among networks of interconnected populations (i.e. the syngameon) (Meier et al., 2017b; Seehausen, 2004). Either model of introgression could facilitate adaptive evolution and promote the ability to exploit ecological opportunity. For instance, mimicry and divergence of wing patterns in *Heliconius* are caused by adaptive introgression (Heliconius Genome, 2012), and ancient introgressions and massive niche emergence enable the diversification and adaptive radiation of cichlid fish (Meier et al., 2017).

Previous studies have attempted to illustrate introgression in *Acropora*, but there is no direct evidence to identify introgression in *Acropora* because of the lack of strong genetic data and proper methods. Indeed, the phylogeny reconstructed by a few genetic markers is not able to reveal "real" species tree in corals and distinguishing introgression from ILS was also a major challenge (Solís-Lemus and Ané, 2016). Notably, NGS and phylogenetic network theory have progressed rapidly in the past 10 years and thus the developed methods, inferring phylogenetic networks from gene trees, have been successfully applied to empirical data for distinguishing introgression from ILS in concert with coalescent theory (Yu et al., 2014). Meanwhile, the ABBA-BABA test to detect introgression based on the prediction of single nucleotide polymorphism (SNP) patterns has been widely applied to non-model organisms (Durand et al., 2011). Thus, a wealth of genetic data produced by the NGS and available whole-genome genotyping algorithms provide new ways to test the role of introgression in the evolution of *Acropora*.

I selected five *Acropora* species*, Acropora tenuis*, *A. digitifera, A. gemmifera, A. subglabra* and *A. echinata* (Figure 2.1 A). The taxonomy of *Acropora* species

based on adult morphology suggested that (1) of them, *A. tenuis* belongs to a species group named the *A. selago* group, (2) *A. digitifera* and *A. gemmifera* are categorized into the *A. humilis* group, and (3) *A. subglabra* and *A. echinata* belong to the *A. echinata* group, respectively (Wallace, 1999; Wallace, 2012). Previous molecular phylogeny demonstrated that *A. tenuis* is a sister species to other four *Acropora* species (Shinzato et al., 2014) and that *A. digitifera* and *A. gemmifera* are clustered into a group with *A. humilis* (Richards et al., 2013; van Oppen et al., 2001), but *A. subglabra* and *A. echinata* have not been included in molecular phylogeny analysis. In addition, these five species are sampled from the four major clades of phylogeny of *Acropora* in order to reduce bias of sampling limitation. Our research group has challenged coral genome-decoding projects led by Dr. Chuya Shinzato. Our group decoded the genome of *A. digitifera* as first coral genome (~422 Mb, 28,958 gene models) (Shinzato et al., 2011), and then *A. tenuis* (~408 Mb, 26,445 gene models). We have further characterized genomes of *A. gemmifera* (~407 Mb, 30,776 gene models), *A. subglabra* (~432 Mb, 30,992 gene models), and *A. echinata* (~411 Mb, 28,280 gene models) (Shinzato et al., in preparation; see http://marinegenomics.oist.jp/).

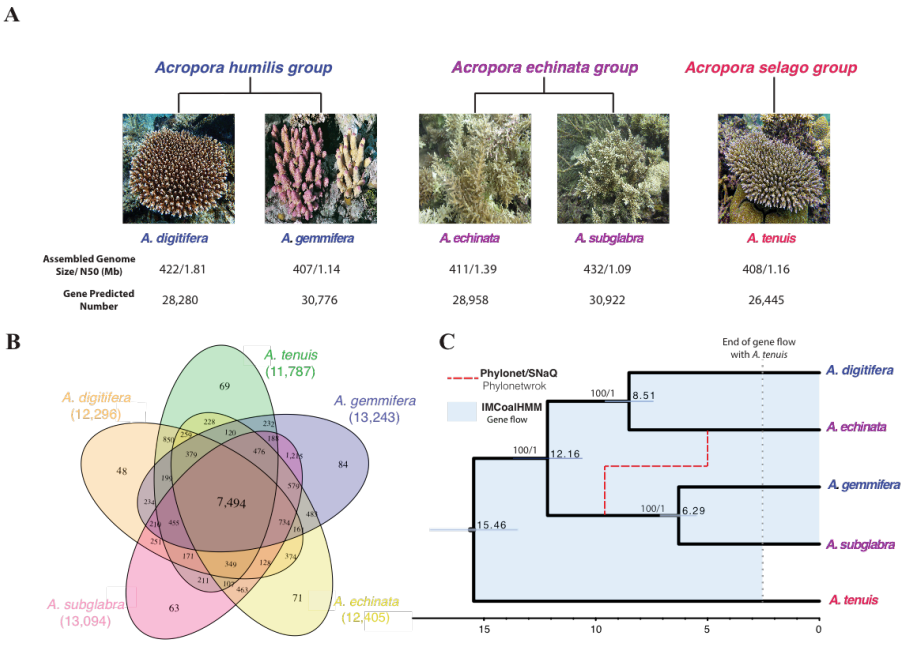


Figure 2.1. **The evolutionary history of *Acropora* inferred from five genomes.** (A) Adult morphology of five *Acropora* species annotated by species group and genome statistics (Photos of *Acropora* provided by Dr. Yuna Zayasu). (B) Venn diagram of shared and unique gene families in five *Acropora* species. (C) Fossil-calibrated phylogenetic tree inferred with 3,361 single-copy orthologs with BEAST2 (black). Phylonetwork analysis inferred a single major introgression event between the stem branch of *A. gemmifera*/*A. subglabra* and the lineage leading to *A. echinata* (red arrow). In addition to this major introgression event, IMCoalHMM inferred background gene flow among all pairs of lineages marked in the blue shade. The gene flow between *A. tenuis* and the other lineages ended 2.5 Mya (gray dotted line).

Here, I used the five *Acropora* genomes to investigate the role of introgression in the diversification of this group. First, using phylogenomic methods, I investigated introgression in the genus and reconstructed a phylogenetic network representing its reticulate evolutionary history. Second, I examined whether introgressed loci are more likely to be evolving faster than non-introgressed loci. Finally, I used a coalescent hidden Markov model approach to test syngameon hypothesis in *Acropora*.

2.2 Methods

2.2.1 Genomic data, gene family clustering and single-copy ortholog selection

All genomic data were downloaded from http://marinegenomics.oist.jp. In brief, each genome was sequenced with HiSeq 2500 in Rapid mode (Illumina) over 100 sequence coverage respectively and genome annotation (gene model) of each species was predicted with *de novo* methods based on repeats-masked genomes and transcriptome information. More detail is given in Shinzato's paper (in preparation) and see http://marinegenomics.oist.jp.

I combined the predicted proteins of each species together and used Blastp (2.2.30+) (Boratyn et al., 2013) to do all-against-all Blast. Then, OrthoMCL was used with the default settings to cluster homologous proteins into 16,885 gene families (Li et al., 2003). I used a custom script to select 4,954 single-copy orthologous gene families, in which only one gene copy is included in each species.

2.2.2 Gene tree and phylogenomic tree reconstruction

Gene tree reconstruction

I used MAFFT (Katoh et al., 2002) to align the amino acid sequences of each single-copy orthologs. I aligned coding sequences with TranslatorX based on amino acid alignments and I excluded the single-copy orthologous genes containing ambiguous 'N' (Abascal et al., 2010). PartitionFinder 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), and gene trees for all 4,954 loci were reconstructed using both programs. For each reconstruction of gene trees, I used the same settings below:

RAxML:

-f a -# autoMRE -m GTRGAMMA -q %s.pat -s %s -p 12345 -x 28754 -n %s

Mrbayes:

unlink Tratio=(all) Revmat=(all) Statefreq=(all) Shape=(all) Pinvar=(all);

prset applyto=(all) ratepr=variable;

mcmcp ngen=50000000 nchain=4 relburnin=yes burninfrac=0.25

printfreq=50000 samplefreq=10000 savebrlens=yes Stoprule=yes

Stopval=0.01;

Phylogenomic tree (species tree) reconstruction

The alignment of 4,954 genes' coding sequences were concatenated into 10,547,082 bp total. The concatenated sequences were used to reconstruct the phylogenomic tree with RAxML and MrBayes under a GTR+CAT+I model or a GTR+Γ +I model, respectively. As well, I applied -autoMRE to generate bootstrap in RAxML and I run MrBayes with setting: ngen=100000000 relburnin=yes burninfrac=0.25 printfreq=50000 samplefreq=10000 savebrlens=yes Stoprule=yes Stopval=0.01. The phylogenomic tree was regarded as the species tree of *Acropora*.

2.2.3 Bayesian concordance analysis using BUCKy

I used BUCKy (1.4.4) to summarize concordance among gene trees generated by MrBayes, by reconstructing the primary concordance tree and estimating concordance factors (CFs) with default setting (Larget et al., 2010) (alpha=1).

2.2.4 Phylonetwork inferrence from gene trees using Phylonet and SNaQ

I selected 4,643 Maximum likelihood (ML) trees with bootstrap support values greater than 50. Each of the trees was rooted with *A. tenuis*, and used to infer the phylonetwork first with the Phylonet ML method (Yu and Nakhleh, 2015).

Reticulation parameters of 0, 1, 2, 3 were applied and run 10 times each. I used likelihood ratio tests to compare models of increasing complexity (i.e. more reticulation events). The likelihood ratio test supported a single reticulation event as the optimal number. I then repeated the analysis 100 further times with reticulation parameter of 1 again and found the results of phylonetwork topology were consistent.

As an additional test, quartet CFs estimated by BUCKy were used to infer the phylonetwork with SNaQ (Solis-Lemus and Ane, 2016). The concatenated phylogenomic tree was used as the initial tree to infer phylonetwork of reticulation equal to 0 and then the result of tree was used to infer phylonetwork with reticulation equal to 1 and so on. The phylonetwork with the reticulation equal to 1 was the only topology inferred by SNaQ under different reticulation settings.

2.2.5 Genome-wide Patterson's D statistics (ABAB-BABA test)

The *A. tenuis* genome was used as the reference for mapping shotgun reads from the other four species using BWA with default settings (Li, 2013). Further, PICARD was used to mask duplications. Then, Samtools was used to index and sort Bam files (Li et al., 2009), while Genome Analysis Toolkit (GATK) was used for insertion/deletion realignment (McKenna et al., 2010). ANGSD was used to perform Genome-wide ABBA-BABA tests with quality control "base quality > 30, mapping quality >60, minimum depth (summing all 4 samples) > 80 and maximum depth (summing all 4 samples) < 600" (-doAbbababa 1 -blockSize 3000000 -anc Aten.fa -doCounts 1 -minQ 30 -minMapQ 60 -P 24 -setMinDepth 80 -setMaxDepth 600) (Korneliussen et al., 2014). The commands were shown below:

bwa mem -R '@RG\tID:H277GBCXX:1\tSM:\tLB:\tPL:illumina1' -t 24 Aten.fa .R1.trimmed .R2.trimmed > .sam

```
samtools view -bS .sam -o .bam
samtools fixmate -O bam .bam _fixmate.bam
rm .sam
samtools sort -@ 24 -O bam -o _sorted.bam -T /tmp/_temp _fixmate.bam
rm _fixmate.bam
java -jar picard-tools-2.1.0/picard.jar MarkDuplicates INPUT=_sorted.bam
OUTPUT=_DM_sorted.bam METRICS_FILE=.bam.metrics
samtools index _DM_sorted.bam
java -jar GenomeAnalysisTK.jar -T RealignerTargetCreator -nt 24 -R Aten.fa -I
_DM_sorted.bam -o _realignment_targets.list
java -jar GenomeAnalysisTK.jar -T IndelRealigner -R Aten.fa -I _DM_sorted.bam
-targetIntervals _realignment_targets.list -o _realigned_reads.bam
samtools index _realigned_reads.bam
```

2.2.6 Speciation with isolation and speciation with migration modeling using IMCoalHMM

Genome alignments

Shotgun reads of each ingroup species were mapped to the *A. tenuis* assembled genome as described above to generate BAM files. Then, the consensus sequence of each species was generated by Samtools with settings: mapping quality greater than 50 and reads quality greater than 30. The consensus sequences of each species on the same scaffolds of *A. tenuis* were considered as whole genome alignments. I selected 238 scaffolds, of which length are greater than 50 Kb, to make pairwise alignments of each species and then these were used in subsequent analysis.

Speciation with isolation and speciation with migration modeling

For each pair of taxa, I fit the data to the speciation with isolation model and speciation with migration model using IMCoalHMM (Mailund et al., 2012). I generated 10 bootstrap samples for each pair by sampling with replacement 238 scaffolds from original 238, and I ran both the speciation with migration and speciation with isolation models on each bootstrap sample.

I calculated AIC values for the speciation with isolation model and speciation with migration model, then, I estimated the delta AIC (delta AIC = speciation with isolation AIC (IAIC) - speciation with migration AIC (IMAIC)). The values less than 0 of delta AIC represented the speciation with isolation model was better otherwise the speciation with migration was better. For isolation period and migration periods parameters estimation under the speciation with migration model, I assumed that I have already known the divergent time between each pair from the time-calibrated phylogenomic tree and calculate them as below:

tua_splitting_period=tua1_isolation_period+tua_migration_period

T_siplitting_age=substitution rate* tua_splitting_period

T1_isoaltion_time= substitution rate * tua1_isolation_period=(T_siplitting_age/ tua_splitting_period)* tua1_isolation_period

2.2.7 Pairs of single-copy orthologous genes dN/dS ratios calculation

Pairwise dN/dS ratio was calculated with PAML using codeml based on the coding sequences alignment of 4,954 single-copy orthologous genes with setting (noisy = 9, verbose = 1, runmode = -2, seqtype = 1, CodonFreq = 2, model = 0, NSsites = 0, icode = 0, fix_kappa = 0, kappa = 1, fix_omega = 0, omega = 0.5) (Yang, 2007). The distribution of dN/dS ratio was plot with ggplot2 in R excluding the value greater than 70 (Team, 2013).

2.2.8 Gene ontology (GO)

I applied the protein sequences to Interproscan's databases (https://www.ebi.ac.uk/interpro/), GO (https://www.uniprot.org/help/gene_ontology), KEGG (https://www.genome.jp/kegg/) and Unipathway (https://www.uniprot.org/database/DB-0170) (Zdobnov and Apweiler, 2001). Then, the protein sequences were used to blast to the Uniport database and the best hits were used to estimate GO enrichments with DAVID 6.7 (Huang et al., 2009). I used the putative introgression genes (1,593 single-copy orthologs) as a gene list and the rest of the single-copy orthologs (3,361 single-copy orthologs) as background.

2.3 Analyses and Results

2.3.1 Gene family cluster and phylogenomic tree reconstruction

Homologous genes were identified across all the five species; 11,787 for *A. tenuis,* 12,296 for *A. digitifera,* 13,243 for *A. gemmifera,* 13094 for *A. subglabra,* and 12,405 for *A. echinata,* respectively (Figure 2.1B). They were clustered into 16,885 gene families in total based on sequence similarity (Figure 2.1B). The five species shared 7,495 gene families, which accounted for 66.89% of predicted proteins (58,887/88,030) (Figure 2.1B). Each *Acropora* genome had very few unique gene families, suggesting that they were closely related to each other. Then, 4,954 single-copy orthologs that were selected from 7,495 shared gene families, were concatenated to reconstruct phylogenomic trees by both Maximum likelihood (ML) and Bayesian methods, with *A. tenuis* as an outgroup. The trees obtained showed the same topology and extremely similar branch length, in which *A. digitifera* and *A. echinata* were sister species, while *A. gemmifera* and *A. subglabra* were sister species (Figure 2.2). All

branches showing the four species relationships were received with 100% bootstrap support (Figure 2.2). In addition, a pair-wise gene family comparison among the five species showed that *A. gemmifera* and *A. subglabra* shared 1,423 gene families, a much higher number of shared gene families than any other species pairs (Figure 2.1B). Next highest was between *A. digitifera* and *A. echinata,* supporting results of phylogenomic analyses.

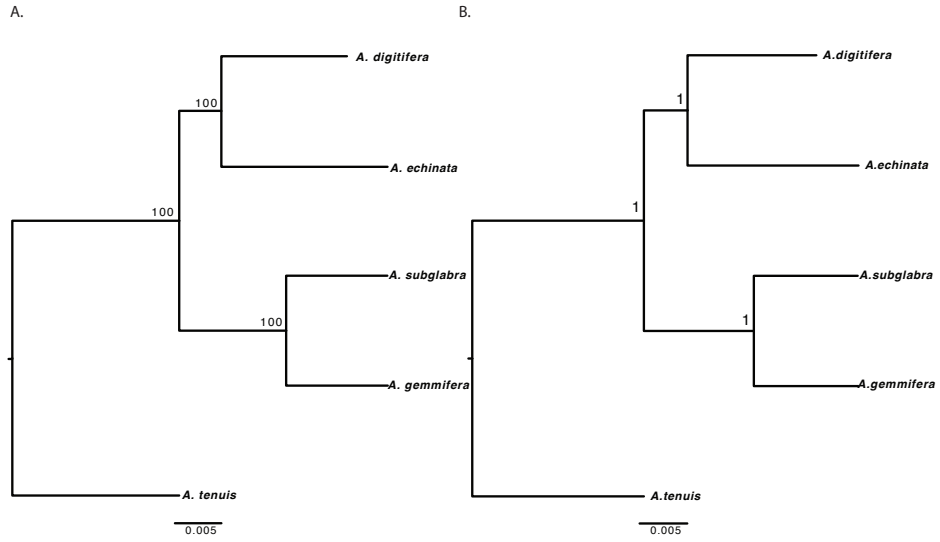


Figure 2.2. **Phylogenomic trees reconstructed by RAxML and MrBayes.** (A) The maximum likelihood phylogenomic tree reconstructed with concatenated sequences under the GTR+CAT+I model in RAxML, nodes numbers reflect bootstrap support. (B) The Bayesian consensus phylogenomic tree reconstructed by concatenated sequences under GTR+GAMMA+I model in MrBayes, with node numbers representing posterior probability.

These results suggest at least two new insights into the diversification of *Acropora* species. First, when two species (*A. digitifera* and *A. gemmifera*) of the *A. humilis* group and two species (*A. echinata* and *A. subglabra*) of the *A. echinata* group were analyzed, *A. digitifera* and *A. echinata* were sister species, and *A. gemmifera* and *A. subglabra* were sister species. Since previous studies did not target

A. echinata and *A. subglabra,* the present result does not deny the previous notion in which *A. digitifera* and *A. gemmifera* are grouped together (Richards et al., 2013; van Oppen et al., 2001). As is evident in Figure 2.1, adult morphology is similar between *A. digitifera* and *A. gemmifera*; and between *A. echinata* and *A. subglabra.* Namely, the relationships based on adult morphology conflict with those based on phylogenomic analysis. This conflict between morphological and genetic relationships suggests that morphological convergence has occurred in these five species.

2.3.2 Test for introgression by ABBA-BABA test and Phylonetwork theory

I then used phylonetwork theory to test for major introgression event(s) in the history of these five lineages. I reconstructed gene trees for each of the 4,954 single-copy orthologs with ML and Bayesian methods, respectively. Of those, half (49%) had a gene tree topology identical to the whole-genome phylogeny, 14% of the loci had a secondary topology, while the remaining 37% loci were distributed across the remaining topologies (Figure 2.3).

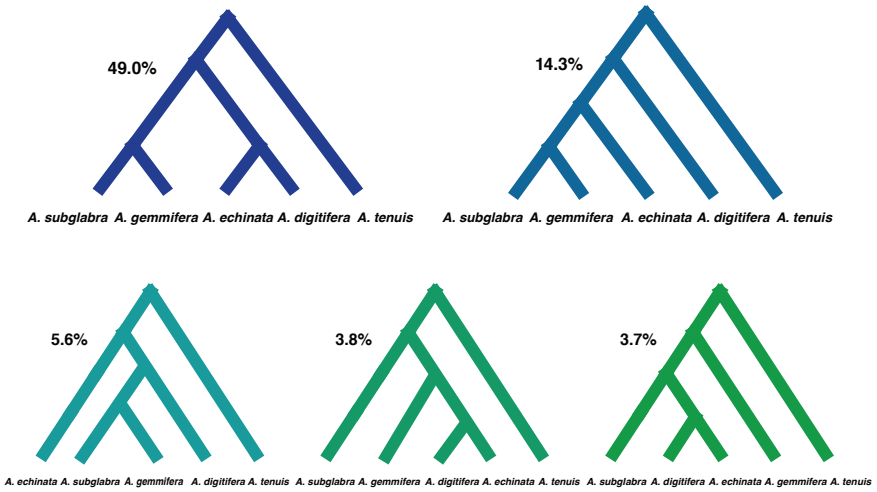


Figure 2.3. **The five most common gene tree topologies inferred with MrBayes**. The frequency of each topology was scored across 4,954 single-copy orthologous gene trees inferred with MrBayes with BUCKy.

Bayesian concordance analysis showed that the concordance factor in the clade of *A. digitifera* and *A. echinata* was less than 0.8, suggesting that the incongruence might be caused by introgression (Cui et al., 2013) (Figure 2.4). In order to distinguish introgression from incomplete lineage sorting, I used the gene trees to infer reticulate evolution with the phylogenetic network ML and pseudo-ML methods (Detail in Methods). Both of results consistently demonstrated the phylonetwork with a single reticulation between the branch of *A. gemmifera* / *A. subglabra* and *A. echinata* was the best model fitting to our gene trees data (Figures. 2.5, 2.6 and Table. 2.1). I also used NeighborNet in SplitsTree to confirm this result and the NeighborNet showed the same result as PhyloNet and SNaQ. In addition, I used gene trees to infer species tree with ASTRAL and MP-EST, both results showed the same species tree as concatenation method in Figure 2.1 C.

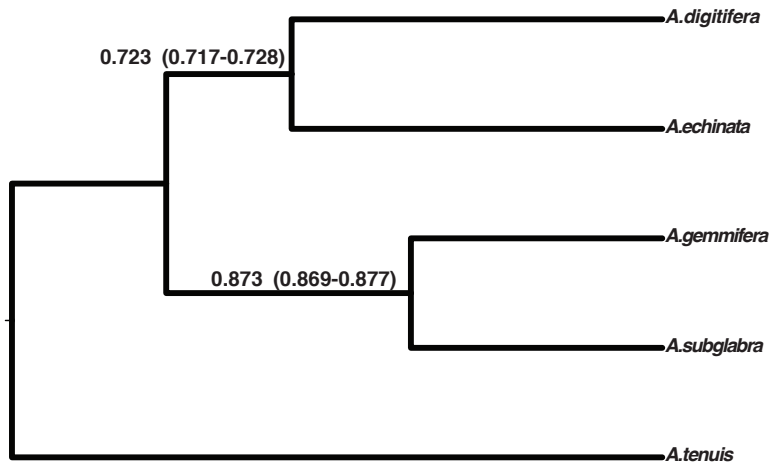


Figure 2.4. **Bayesian concordance analysis.** The primary concordance tree reconstructed with BUCKy inferred from posterior distributions of 4,954 gene trees. Node values represent Bayesian concordance factors (CFs) with 95% confidence interval (CI).

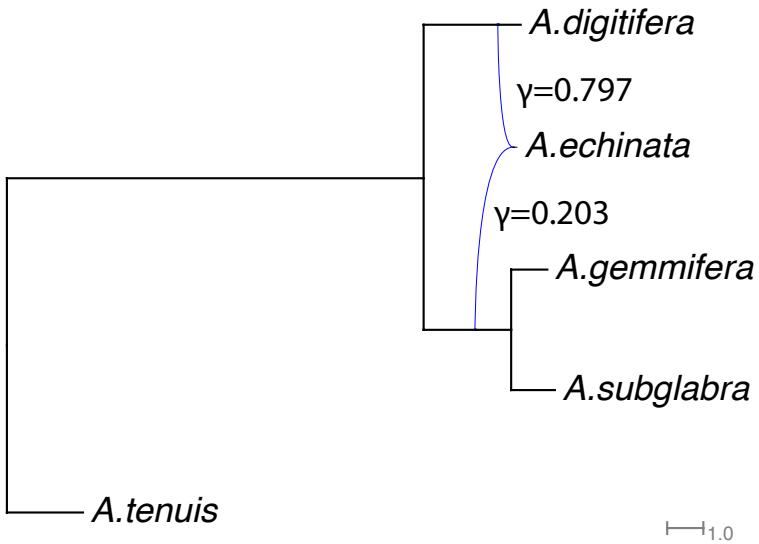


Figure 2.5. **Phylonetwork inferred by Phylonet.** The phylonetwork with highest likelihood was inferred from rooted 4,643 Maximum likelihood (ML) trees with the setting reticulation number to 1. Proportions of introgressed genome (γ) are shown the hybrid branch.

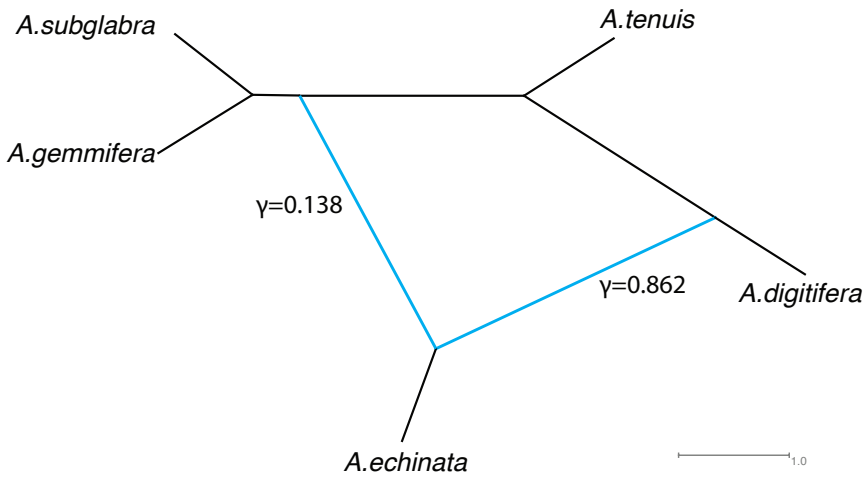


Figure 2.6. **Phylonetwork inferred by SNaQ.** The phylonetwork was inferred from quartet concordance factors (CFs) estimated by BUCKy. Proportions of the introgressed genome (γ) are shown in the hybrid branch.

Table 2.1. Likelihoods and information criteria of Phylonet models fit with different numbers of reticulation events

Reticulate node	Likelihood	AIC	AICc	BIC
0	-5592.38	11190.76591	11190.77108	11210.09525
1 (optimal)	-5200.78	10415.55172	10415.57588	10460.65353
2	-5187.44	10396.87695	10396.93396	10467.75123
3	-5183.82	10397.64593	10397.74966	10494.29267

In addition, I performed ABBA-BABA test to confirm the introgression detected by the phylonetwork approach (Durand et al., 2011). I found that both *A. gemmifera* and *A. subglabra* had a closer genetic relationship with *A. echinata* rather than with *A. digitifera* (Z= -5.15, Z= -5.37, t-test), indicating that introgression had occurred among *A. gemmifera, A. subglabra* and *A. echinata* (Figure 2.7). In contrast, when I tested whether introgression occurred from *A. echinata* or *A. digitifera* to the clade of *A. gemmifera*/*A. subglabra*, I did not find introgression signal among them (Table. 2.2). Therefore, the ABBA-BABA test was strongly consistent with

phylonetwork analysis illustrating one major introgression event between the branch of *A. gemmifera* / *A. subglabra* and *A. echinata* (Figure 1C).

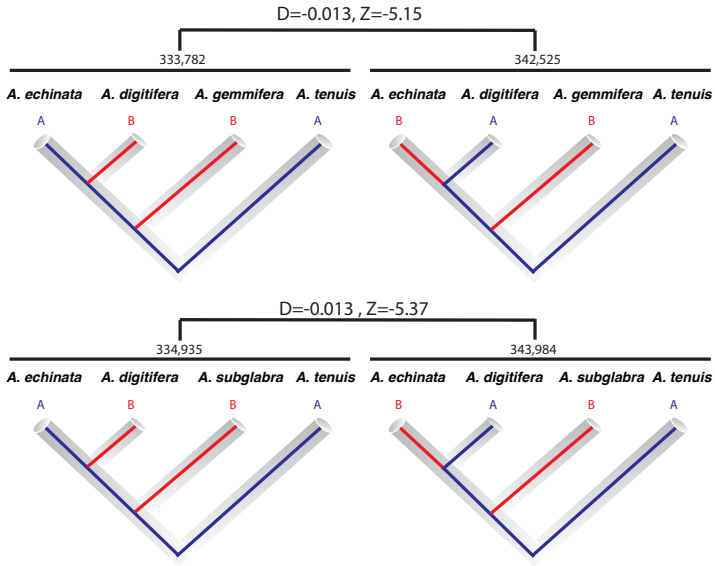


Figure 2.7. **Four taxon ABBA–BABA analysis.** The total numbers of each gene genealogy across the whole genome. Equal numbers of ABBA and BABA gene genealogies are expected under a null hypothesis of no introgression. The D statistics and Z values are calculated for testing the null hypothesis.

Table 2.2. Statistics of ABBA-BABA test

H1	H2	H3 (Hybrid condidate)	nABBA	nBABA	Z
A.echinata	*A.digitifera*	*A.gemmifera*	333782	342525	-5.147
A.echinata	*A.digitifera*	*A.subglabra*	334935	343984	-5.367
A.gemmifera	*A.subglabra*	*A.echinata*	280690	280863	-0.115
A.gemmifera	*A.subglabra*	*A.digitifera*	284045	285194	-0.786

Taken together, results of the Bayesian concordance analysis, phylogenetic network ML, pseudo-ML, and ABBA-BABA tests all support a single reticulation event between the branch of *A. gemmifera*/ *A. subglabra* and *A. echinata* (Figure 2.1C, Figures. 2.4-2.7 and Tables 2.1, 2.2).

2.3.3 Syngameon hypothesis identification

The phylonetwork/ABBA-BABA analysis identified one major introgression event in the history of these five species, but those methods better suited to inferring major episodes of introgression rather than low-level, recurrent migration among lineages (Mailund et al., 2012; Solis-Lemus and Ane, 2016), as would be expected under the syngameon hypothesis. The gene trees analysis showed that 37% gene trees' topologies match neither the species tree topology nor the topology consistent with the inferred introgression event (Figure 2.3). This incongruence between gene tree and species tree can be caused by gene flow or ILS or selection or gene tree reconstruction noise. Yet, here, I hypothesized this is due to continuous gene flow between *Acropora* species, which under the 'syngameon hypothesis' could facilitate adaptation of different morphologies and ecologies (Seehausen, 2004; van Oppen et al., 2001; Wallace, 1999). I used the coalescent hidden Markov model (IMCoalHMM) approach to compare models of speciation with isolation and speciation with migration using whole-genome alignments of all species pairs (Mailund et al., 2012). For all pairs of *Acropora*, a speciation with migration model strongly outperformed a model with isolation. Migration between the sister species to other species, *A. tenuis*, with the other five species apparently ended 2.5 Mya, while migration between all other pairs continues until the present (Figure 2.1C, Figure 2.8, Table. 2.3). In detail, I applied whole-genome alignments of each species pair to speciation-with-isolation model and speciation-with-migration model and then compared the AIC values between the two models. I found that whole-genome alignments of each species pair were better explained by speciation with migration model suggesting that gene flow existed in each species pair among *Acropora* (Table. 2.3). In addition, I estimated that the gene flow between *A. tenuis* and other four species ceased around 2.5 Mya and the

species pairs between the other four species ceased at present (Figure 2.8). Importantly, The analyses showed that gene flow either continue or stop between morphological species groups.

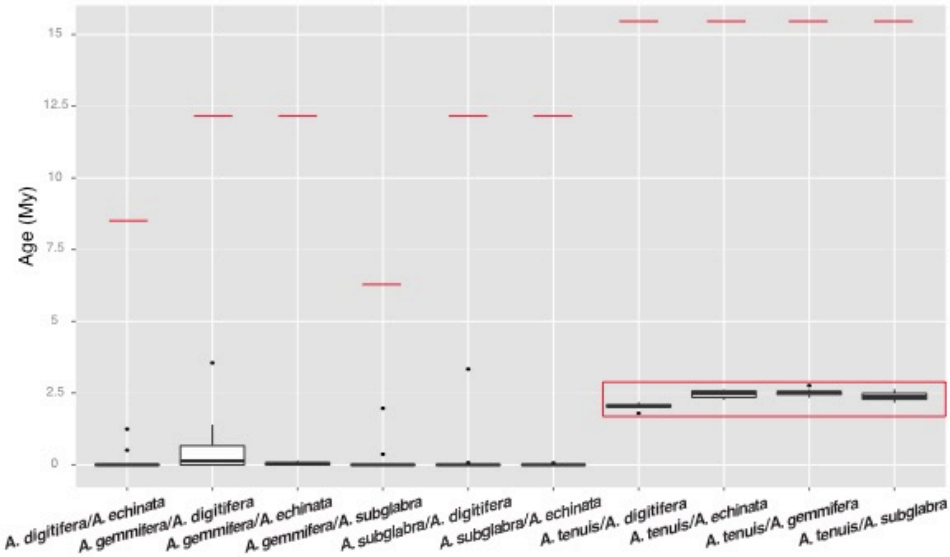


Figure 2.8. **Results of speciation with migration model inferred with IMCoalHMM.** For each species pair, red horizontal lines indicate divergence time (inferred through phylogenomic analysis) and boxes denote distribution of the end of migration over. All pairs inferred migration until essentially the present, except pairs including *A. tenuis*.

Table 2.3. Average IAIC, IMAIC and delta AIC values in species pairs inferred with IMCoalHMM

Species pairs	IAIC (Speciation with isolation)	IMAIC (Speciation with migration)	Delta AIC (IAIC-IMAIC)
A. digitifera/ A. echinata	29960549.44	29959015	1534.438964
A. subglabra/ A. echinata	39807303.24	39795248.68	12054.56131
A. subglabra/ A. digitifera	39188056.05	39181351.53	6704.520982
A. gemmifera/ A. digitifera	39165160.33	39159497.1	5663.228709
A. gemmifera/ A. subglabra	18520537.48	18520270.39	267.0876945
A. gemmifera/ A. echinata	39607985.07	39595195.95	12789.11567
A. tenuis/ A. digitifera	71840011.95	71744856.99	95154.96584
A. tenuis/ A. echinata	70484041.84	70386647.66	97394.18804
A. tenuis / A. subglabra	68279164.9	68186635.16	92529.74031
A. tenuis/ A. gemmifera	68336023.92	68246480.04	89543.87544

2.3.4 Evolutionary rates and patterns of selection

Since introgression has apparently occurred, it raises the question of what role the transfer of genetic material may play in coral evolution. I hypothesized that if introgression were involved with adaptive evolution, loci that were involved in introgression should be evolving faster than those that were not. To test this, I compared evolutionary rates in genes that matched the species tree ("species-tree" genes), with those that have a different topology ("non species-tree" genes). Although a discordant gene tree is not in itself definitive evidence of introgression for a given locus (due to other explanations such as ILS), on the whole genes involved with introgression should be highly overrepresented in this discordant group compared to the group matching the species tree.

I found elevated rates of evolution among the non-species tree genes and the major introgression topology genes relative to species tree genes in the three lineages involved with the major introgression event (*A. gemmifera/A. tenuis, A. subglabra/A. tenuis,* and *A. echinata/A. tenuis,* but not *A. digitifera/A.tenuis*) ($P<0.001$, Mann-Whitney test, Figure. 2.9), which is consistent with a role for adaptive evolution. One interpretation of this is that certain loci that are undergoing adaptive evolution in one lineage may be more likely to be introgressed into another lineage during a major introgression event. However, while these findings are suggestive about the adaptive role of introgression, further work is needed to analyze these processes in more detail.

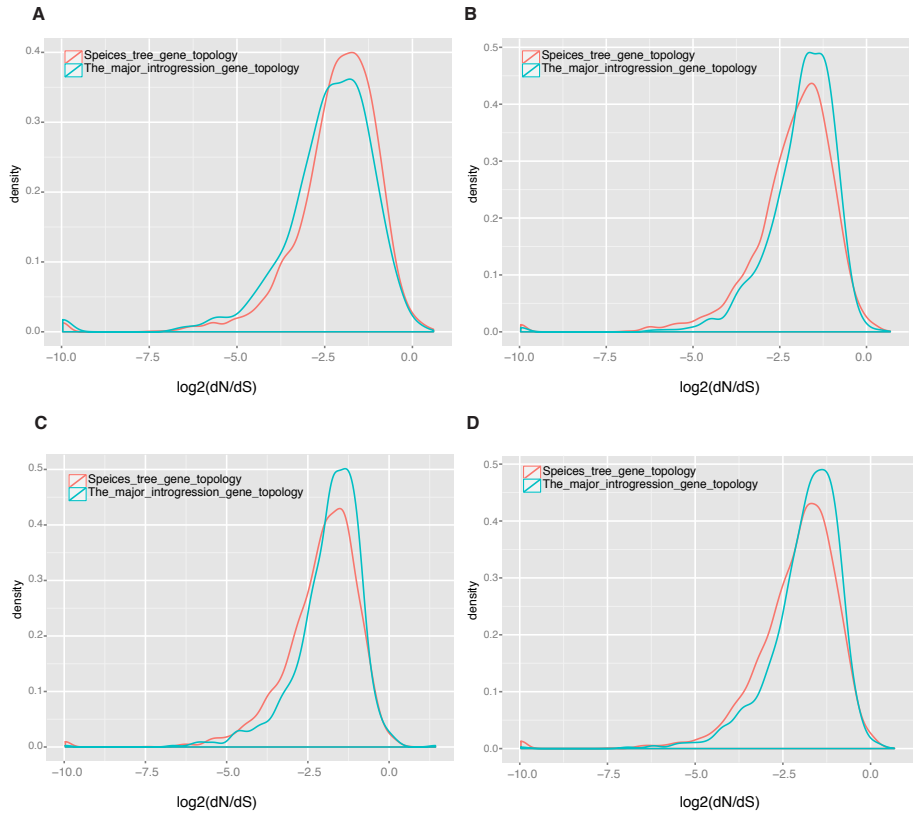


Figure 2.9. **Evolutionary rates of introgression and non-introgression genes.** Distributions of dN/dS value of the 4,954 single-copy orthologs with species tree topology (3,361 single-copy orthologs) or topology indicating introgression (1,593 single-copy orthologs) in (A) *A. digitifera,* (B) *A. echinata,* (C) *A. gemmifera,* and (D) *A. subglabra.* Evolutionary rates of introgression genes evolved significantly faster than species tree genes (P<0.001, Mann-Whitney test) in all lineages except *A. digitifera*, which was not involved with the major introgression event.

I performed Gene Ontology (GO) analysis to examine whether there are any functional differences between species tree genes (gene tree topology matched the species tree topology) and non species-tree genes (gene tree topology mismatched the species tree topology). I found that ontologies including G protein–coupled receptors, binding proteins and transporters in relation to DNA replication, oxidation-reduction reaction, cell apoptosis, iron and amino acid transportation, are significantly more likely to have topologies that do not match the species tree (Barshis et al., 2013)

(Table. 2.4). I also identified ~30 (out of 1,539) of the non species-tree genes that are under positive selection (dN/dS>1) (Table. 2.5). These also included genes involved in the responses to stressful environments according to previous transcriptome analyses (Barshis et al., 2013) (Table. 2.5, Table 2.6).

Table 2.4. GO enrichment for introgression genes comparing to species tree genes

Cluster	Enrichment Score	P-Value	Benjamini
Transmembrane	5.53	2.2×10^{-7}	9.1×10^{-6}
G-protein coupled receptor	3.38	1.2×10^{-5}	1.9×10^{-3}
Immunoglobulin-like fold	1.27	4.5×10^{-2}	6.4×10^{-1}
Ion transport	1.24	2.4×10^{-2}	7.3×10^{-1}
Dopamine neurotransmitter receptor	1.2	4.5×10^{-2}	4.4×10^{-1}
ANK repeat	0.27	5.1×10^{-1}	8.8×10^{-1}
DNA-binding	0.17	5.7×10^{-1}	9×10^{-1}

Table 2.5. Annotation of non species-tree genes under positive selection

Genes	Molecular Function	Stress Response Types in Coral
G-protein coupled receptor 83	G-protein coupled receptor activity	Bleaching
Neuropilin-1-like	growth factor binding	Growth anomaly
Peroxidasin	protein binding	Heating
Carbonic anhydrase 2-like	catalytic activity	Elevated pCO2
Plexin domain-containing protein 2-like	protein binding	Symbiont colonization
RAD51-associated protein 1-like	DNA binding/ protein binding	Ultraviolet radiation
Zinc transporter ZIP1-like	transporter activity	Bleaching

Table 2.6. Non species-tree genes under selection in the four species pairs

Gene family ID	Gene	Gene_ID					**Positive selection on lineages**
		A. digitifera	*A. echinata*	*A. gemmifera*	*A. subglabra*	*A. tenuis*	
led2829	uncharacterized protein LOC107349793	sc0000028.g769.t1	sc0000561.g9556.t1	sc0000180.g7905.t1	sc0000129.g5077.t1	sc0001273.g193.t1	*A. gemmifera*
led3142	uncharacterized protein LOC107339225	sc0000001.g449.t1	sc0000003.g27452.t1	sc0000053.g3531.t1	sc0000002.g21055.t1	sc0000012.g544.t1	*A. digitifera, A. echinata, A. gemmifera*
led3421	uncharacterized protein LOC107345071 isoform X1	sc0000004.g427.t1	sc0000005.g12888.t1	sc0000067.g18552.t1	sc0000197.g21948.t1	sc0000044.g184.t1	*A. subglabra*
led3506	uncharacterized protein LOC107329733	sc0000006.g308.t1	sc0000309.g7413.t1	sc0000056.g23513.t1	sc0000305.g19178.t1	sc0000088.g425.t1	*A. digitifera*
led3514	probable G-protein coupled receptor 83	sc0000006.g328.t1	sc0000067.g17139.t1	sc0000001.g4001.t1	sc0000151.g21614.t1	sc0000038.g263.t1	*A. gemmifera,A. subglabra*
led3579	uncharacterized protein LOC107334990	sc0000006.g552.t1	sc0000092.g16590.t1	sc0000001.g3894.t1	sc0000052.g11586.t1	sc0000042.g72.t1	*A. digitifera*
led3650	carbonic anhydrase 2-like	sc0000008.g168.t1	sc0000011.g477.t1	sc0000002.g21072.t1	sc0000050.g27853.t1	sc0000001.g234.t1	*A. digitifera,A. gemmifera, A. subglabra*
led3873	uncharacterized protein LOC107337495 isoform X1	sc0000011.g165.t1	sc0000072.g21034.t1	sc0000239.g10665.t1	sc0000213.g20843.t1	sc0000246.g212.t1	*A. echinata*
led3909	peroxidasin-like	sc0000012.g37.t1	sc0000002.g19546.t1	sc0000015.g26254.t1	sc0000133.g19590.t1	sc0000075.g416.t1	*A. digitifera,A. echinata, A. gemmifera, A. subglabra*
led4119	plexin domain-containing protein 2-like	sc0000016.g97.t1	sc0000088.g28880.t1	sc0000073.g107.t1	sc0000049.g4415.t1	sc0000179.g556.t1	*A. gemmifera*
led4356	DNA-directed RNA polymerase I subunit RPA1-like	sc0000022.g164.t1	sc0000155.g9051.t1	sc0000007.g24093.t1	sc0000024.g24586.t1	sc0000003.g581.t1	*A. digitifera,A. echinata, A. gemmifera, A. subglabra*
led4505	uncharacterized protein LOC107346997	sc0000025.g740.t1	sc0000182.g17442.t1	sc0000040.g27195.t1	sc0000078.g2109.t1	sc0000305.g95.t1	*A. digitifera*
led4718	neuropilin-1-like	sc0000031.g376.t1	sc0000002.g19305.t1	sc0000065.g7976.t1	sc0000065.g7682.t1	sc0000005.g646.t1	*A. subglabra*
led4907	uncharacterized protein LOC107348150	sc0000036.g1128.t1	sc0000003.g27232.t1	sc0000023.g27936.t1	sc0000134.g14320.t1	sc0000225.g122.t1	*A. digitifera*
led4931	uncharacterized protein LOC107328254 isoform X1	sc0000037.g514.t1	sc0000127.g15551.t1	sc0000026.g1967.t1	sc0000013.g30794.t1	sc0000048.g212.t1	*A. digitifera,A. echinata, A. gemmifera, A. subglabra*
led4977	unknow	sc0000038.g481.t1	sc0000007.g21944.t1	sc0000246.g24751.t1	sc0000114.g4250.t1	sc0000176.g24.t1	*A. gemmifera*
led5051	uncharacterized protein LOC107353140 isoform	sc0000041.g446.t1	sc0000124.g7423.t1	sc0000112.g19888.t1	sc0000120.g3106.t1	sc0000070.g180.t1	*A. subglabra*

	X2						
led5229	uncharacterized protein LOC107358880 isoform X1	sc0000046.g349.t1	sc0000076.g19928.t1	sc0000004.g6284.t1	sc0000205.g2687.t1	sc0000136.g168.t1	*A. digitifera,A. echinata, A. gemmifera, A. subglabra*
led5234	unknow	sc0000046.g393.t1	sc0000126.g23683.t1	sc0000004.g6244.t1	sc0000561.g9562.t1	sc0000279.g146.t1	*A. subglabra*
led5392	uncharacterized protein LOC107346854	sc0000050.g252.t1	sc0000007.g22229.t1	sc0000066.g28394.t1	sc0000019.g26294.t1	sc0000084.g10.t1	*A. gemmifera*
led5557	uncharacterized protein LOC107352608 isoform X2	sc0000057.g179.t1	sc0000013.g28805.t1	sc0000203.g18342.t1	sc0000189.g3240.t1	sc0000003.g571.t1	*A. gemmifera*
led5616	unknow	sc0000060.g395.t1	sc0000041.g23452.t1	sc0000367.g15595.t1	sc0000035.g12890.t1	sc0000076.g209.t1	*A. echinata*
led5664	uncharacterized protein LOC107353328	sc0000063.g246.t1	sc0000013.g28659.t1	sc0000055.g2876.t1	sc0000035.g12814.t1	sc0000273.g74.t1	*A. echinata*
led5750	unknow	sc0000065.g256.t1	sc0000017.g17937.t1	sc0000016.g26856.t1	sc0000069.g24843.t1	sc0000180.g238.t1	*A. gemmifera*
led6047	uncharacterized protein LOC107340678	sc0000080.g76.t1	sc0000017.g17778.t1	sc0000188.g28486.t1	sc0000029.g8957.t1	sc0000167.g326.t1	*A. subglabra*
led6577	uncharacterized protein LOC107342813	sc0000101.g90.t1	sc0000033.g28173.t1	sc0000183.g18699.t1	sc0000005.g13437.t1	sc0000078.g20.t1	*A. echinata, A. gemmifera, A. subglabra*
led7022	RAD51-associated protein 1-like	sc0000133.g68.t1	sc0001517.g5741.t1	sc0000102.g17820.t1	sc0000072.g22890.t1	sc0000004.g385.t1	*A. subglabra*
led7081	uncharacterized protein LOC107349269 isoform X2	sc0000139.g173.t1	sc0000019.g24527.t1	sc0000260.g844.t1	sc0000001.g3664.t1	sc0000051.g226.t1	*A. digitifera, A. gemmifera*
led7139	uncharacterized protein LOC107338839 isoform X1	sc0000143.g282.t1	sc0000098.g9592.t1	sc0000035.g12736.t1	sc0000424.g28001.t1	sc0000150.g286.t1	*A. digitifera*
led7228	uncharacterized protein LOC107345577	sc0000153.g411.t1	sc0000020.g4366.t1	sc0000013.g30648.t1	sc0000055.g2801.t1	sc0000190.g318.t1	*A. digitifera, A. subglabra*
led7350	zinc transporter ZIP1-like	sc0000163.g466.t1	sc0000019.g24584.t1	sc0000075.g12583.t1	sc0000001.g3719.t1	sc0000051.g179.t1	*A. digitifera*
led7525	uncharacterized protein LOC107355338	sc0000187.g198.t1	sc0000087.g19791.t1	sc0000278.g16715.t1	sc0000045.g14904.t1	sc0000041.g232.t1	*A. gemmifera*
led7608	uncharacterized protein LOC107332057	sc0000200.g236.t1	sc0000239.g10528.t1	sc0000211.g7086.t1	sc0000138.g11908.t1	sc0000212.g263.t1	*A. echinata, A. subglabra*

2.4 Discussion

The staghorn corals of the genus *Acropora* constitute the foundation of modern coral reef ecosystems, but much work remains to reconstruct their evolutionary history and identify the processes shaping their diversification. Understanding this is critical for anticipating coral responses to the ongoing multifaceted changes of the Anthropocene (Hemond and Vollmer, 2010; Hughes et al., 2017; Sheppard et al., 2017). Toward that end, the present analysis of the genomes of five *Acropora* species addresses a longstanding issue in coral evolution; the roles of introgression in shaping their histories and diversification. The phylogenomic analysis indicates that, although adult morphology of *A. digitifera* resembles *A. gemmifera* and that of *A. echinata* resembles to *A. subglabra,* these two species pairs are not clustered each other, but *A. digitifera* and *A. subglabra* were clustered together while *A. gemmifera* and *A. echinata* together. Namely, the clustering of adult morphology conflicts with that obtained using phylogenomic analysis. This conflict between morphological and genetic relationships suggests the occurrence of introgression and/or morphological convergence in these five species. Indeed, I find evidence of a major gene flow event between the common ancestor of *A. subglabra* and *A. gemmifera* and *A. echinata*. This study is, to my knowledge, the first to demonstrate genome-scale evidence of introgression in coral evolution using phylogenomic methods. Yet, due to limitation of sampling size, I cannot determine the hybrids in this study but the major goal of this chapter is to distinguish introgression from ILS in *Acropora*.

The evolutionary rates comparisons suggested the adaptive role of introgression in *Acropora*. And GO analysis showed introgression genes are likely involved in the responses to stressful environments. In all, the genome-wide analysis

provides an insight to understand the evolutionary history of *Acropora*: genetic exchange (introgression) probably plays crucial roles in the evolutionary radiation of *Acropora*.

Chapter 3

Climate change provided an ecological opportunity for the rise to dominance of *Acropora* in the Plio-Pleistocene

3.1 Introduction

Global distributions and the rise to dominance of species are usually driven by both biotic and abiotic factors along with population fluctuations or species diversification (Prada et al., 2016; Thomas et al., 2004). Especially, environmental change often has an important influence on species' demography, extinction or/and diversification (Talluto et al., 2017). In Plio-Pleistocene, seawater temperature and sea-level periodically change with glacial-interglacial cycle triggered by the northern hemisphere glaciation around 2.75 Mya (Rohling et al., 2014a). In addition, sea temperature and sea-level periodically change with 41thousand year (Ky) period and then glacial cycles transited from 41 to 100 Ky period during the mid-Pleistocene transition (MPT) from 700 Ky to 1.25 My ago, when the climate underwent fundamental change (Herbert et al., 2010). Fossil record showed that mass extinctions of nearshore marine organisms occurred around 2~3 My probably due to the onset of the northern hemisphere glaciation in Plio-Pleistocene generating massive empty niches (Pimiento et al., 2017; Prada et al., 2016; Talluto et al., 2017). However, interestingly, previous studies showed that *Acropora*, shallow-water reef-building corals, distributed to Indo-Pacific Ocean and became one of the dominant reefs after the onset of the northern hemisphere glaciation. Yet, the diversification of *Acropora* is not observed at that time based on fossil record (Renema et al., 2016). Hence, it is worth considering whether massive empty niches provide a great ecological

opportunity for the rise *Acropora* to global dominance through colonization in empty niches after the northern hemisphere glaciation in Plio-Pleistocene.

Ecological opportunity, the "wealth of evolutionarily accessible resources little used by competing taxa" (Schluter, 2000), provides a favorable selective environment for diversification (Stroud and Losos, 2016). There are several ways to trigger an evolutionary radiation via ecological opportunity (Losos, 2010; Stroud and Losos, 2016): colonization of a new area, mass extinction, and evolution of a key innovation. In particular, mass extinction can remove dominant taxa and generate new resource or/and niches for the species that persist (Stroud and Losos, 2016). In Plio-Pleistocene, seawater temperature and sea-level periodically changed with glacial-interglacial cycle and they were initiated by the northern hemisphere glaciation around 2.75 Mya (Herbert et al., 2010; Rohling et al., 2014a). The fossil record shows that mass extinctions of nearshore marine organisms occurred around 2~3 My probably due to the onset of the northern hemisphere glaciation in Plio-Pleistocene generating massive empty niches (O'dea et al., 2007; Pimiento et al., 2017; Prada et al., 2016; Rohling et al., 2014b; Talluto et al., 2017). Interestingly, the fossils of *Acropora* have been in coral hotspots from the Eocene to the present (Renema et al., 2008; Wallace and Rosen, 2006), however it became one of the dominant reef components after the onset of the northern hemisphere glaciation (Renema et al., 2016; Wallace, 1999; Wallace, 2012; Wallace and Rosen, 2006). This pattern has led some to suggest that the massive empty niches created by the glacial-cycle induced mass extinctions provided a ecological opportunity for the rise of *Acropora* to dominant status (Renema et al., 2016). *Acropora* is also among the most dispersive corals and this has been proposed as a key advantage for them to better cope with rapid sea level changes during the glacial cycles of the Plio-Pleistocene (Renema et al., 2016).

Here, I used the five *Acropora* genomes to investigate questions about the role of ecological opportunity in the rise to dominance of this group. First, I used the new phylogenomic framework to date the age of the group and set the timescale of *Acropora* evolution. Then, using the latter, I examined demographic changes in the coral lineages in the Plio-Pleistocene and evaluate if they correspond to ecological opportunity caused by major shifts in glacial cycles with demographic inference.

3.2 Methods

3.2.1 Phylogenomic tree dating with BEAST2

In order to infer the divergence time of *Acropora* and set the timescale of *Acropora* evolution, I selected 817 single-copy orthologous genes among five *Acropora* and two outgroups, *Orbicellaa* (*Orbicellaa faveolata*) and *Porites (Porites lobata; Porites australiensis and Porites astreoides)*, using OrthoMCL and transcriptome data of *Orbicellaa* and *Porites* (Bhattacharya et al., 2016). Then, I selected 3361 genes with gene trees that were concordant with the species tree (((*A. gemmifera*, *A. subglabra*), (*A. echinata*, *A. digitifera*)), *A. tenuis*). I blasted the 817 single-copy orthologous genes to the 3361 genes (((*A. gemmifera*, *A. subglabra*), (*A. echinata*, *A. digitifera*)), *A. tenuis*), and found 440 single-copy orthologous genes that are shared between all taxa and have gene trees that match the species tree. I concatenated these sequences and used them to infer a time-calibrated phylogeny. First, I 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 in the previous study. *Orbicellaa* and *Porites* split 153 Mya split *Porites* and *Acropora* split at 84 Mya

(Bouckaert et al., 2014; Simakov et al., 2015). I ran BEAST2 three times independently, 50 million Markov chain Monte Carlo (MCMC) generations for each run, then I used Tracer to check the log files and I found that ESS value of each parameter was greater than 200. I chose the highest likelihood tree generated by BEAST2 to present the crown age of these five *Acropora* species to be approximately 15.6 Mya (95% highest posterior density (HPD): 15.39 My~15.87 My). Finally, after inferring the crown age, I used a larger dataset to infer the divergence times for nodes within the *Acropora* clade. For this, I concatenated the 3361 single-copy orthologous genes with gene trees matching the species tree topology, and used them for a BEAST2 analysis with the setting as above, and calibrating the crown age to 15.46 My.

3.2.2 Whole genome alignment and mutation rate estimation

First, I aligned the four species' shotgun data to *A.tenuis* using LASTZ with setting (Harris, 2007) (--seed=12of19 --notransition --chain --gapped --inner=2000 --ydrop=3400 --gappedthresh=6000 --hspthresh=2200 --strand=plus --format=axt). I removed all the gap sites and ambiguous 'N' sites. Then, I calculated the number of consensus sequences and divergent sequences. The mutation rate was calculated as the formula: μ= (counts of divergent loci / (counts of divergent loci+ counts of consensus loci)) / (2*divergence time)*(generation time) (Zhao et al., 2013).

For *A. gemmifera*: $(76154410/(76154410+351440506)/(2 \times 15.5)) \times 5 \times 10^{-6} = 2.87 \times 10^{-8}$

For *A. echinata*: $(67411262/(67411262+318366635)/(2 \times 15.5)) \times 5 \times 10^{-6} = 2.82 \times 10^{-8}$

For *A. subglabra*: $(78384122/(78384122+372773032)/(2 \times 15.5)) \times 5 \times 10^{-6} = 2.80 \times 10^{-8}$

For *A. digitifera*: $(79427941/(79427942+363368171)/(2 \times 15.5)) \times 5 \times 10^{-6} = 2.89 \times 10^{-8}$

Average: $(2.87 \times 10^{-8}+2.82 \times 10^{-8}+2.80 \times 10^{-8}+2.89 \times 10^{-8})/4 = 2.9 \times 10^{-8}$

3.2.3 Demographic history reconstruction using PSMC

Shotgun reads of each species were mapped to their own assembled genomes as described above to generate BAM files. Then the consensus sequence of each species was generated by Samtools with settings: mapping quality greater than 50 and reads quality greater than 30. The demographic history of each species was reconstructed using the PSMC model with settings (Li and Durbin, 2011) (-N25 -t15 -r5 -p "4+25*2+4+6"). The neutral mutation rate was estimated using the divergent time and sequence divergence estimated by the LASTZ as described above (Harris, 2007). Generation time was assumed to be 5 years for each species (Hemond and Vollmer, 2010). Bootstrapping of demographic inference was generated for each of species following previous study (Zhao et al., 2013).

```
samtools mpileup -q 50 -Q 30 -uf .fa _realigned_reads.bam | bcftools call -c |
perl vcfutils.pl vcf2fq -d 16 -D 96 |gzip> .fq.gz
fq2psmcfa -q20 .fq.gz > .psmcfa
psmc -N25 -t15 -r5 -p "4+25*2+4+6" -o .psmc .psmcfa
psmc2history.pl .psmc
perl  utils/psmc_plot.pl -g 5 -u 3e-8 _out .psmc
utils/splitfa .psmcfa > _split.psmcfa
seq 100 | xargs -i echo psmc -N25 -t15 -r5 -b -p "4+25*2+4+6" -o _round-
{}.psmc _split.psmcfa | sh
cat .psmc _round-*.psmc > _combined.psmc
psmc_plot.pl -p -g 5 -u 2.9e-8 _combined _combined.psmc
```

3.3 Analyses and Results

3.3.1 Time-calibrated phylogenomic tree reconstruction

In part due to the phylogenetic difficulties introduced by incongruent loci, it has been a challenge to infer the timescale of *Acropora* evolution using molecular data, with average crown ages ranging from 6-36 Mya in previous studies (Richards et al., 2013; van Oppen et al., 2002).

In order to know the divergent time of *Acropora* without effects of introgressed genes, I selected 3,361 single copy genes, of which topology is same as the species tree. Then I filtered out 440 genes, which could find the single copy orthologous hits with *Orbicellaa* and *Porites* using Blast and OrthoMCL, to reconstruct time-calibrated phylogenomic tree with the known divergent times (Simakov et al., 2015). The result showed that the four species were one monophyletic lineage diversified with *A. tenuis* at 15.5 Mya (95% highest posterior density (HPD): 13.5My~17.4My) (Figure 3.1).

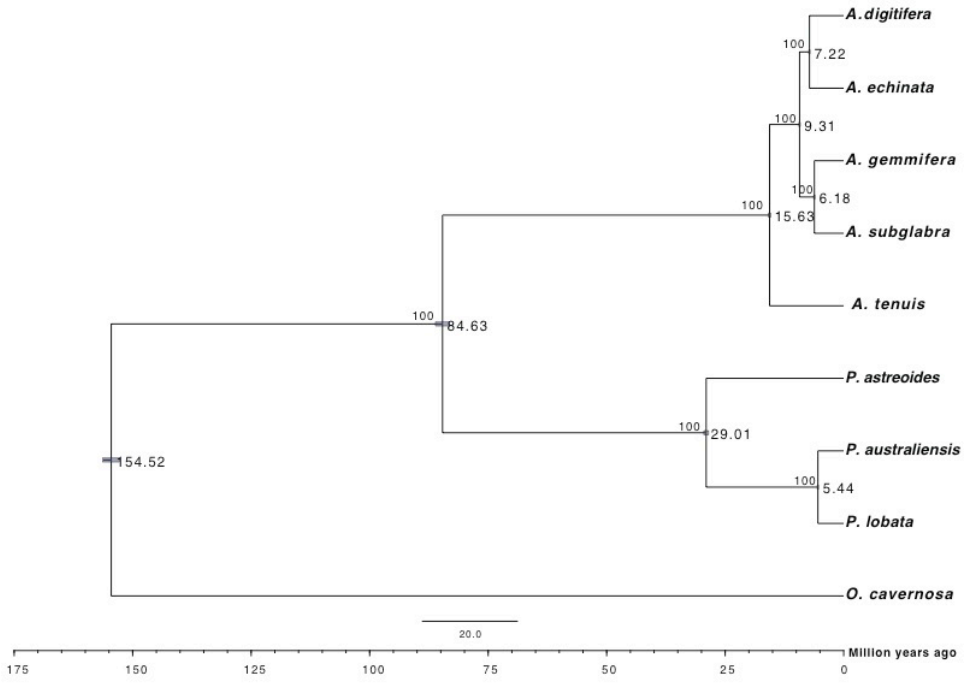


Figure 3.1. **Time calibrated phylogenomic tree of *Acropora*, *Porites* and *Orbicellaa*.** Posterior 95% CIs of node ages are represented with blue horizontal bars as well as ML bootstrap values are shown at each node.

I used this nodes information to date the concatenated sequence of the 3,361 single copy genes among the five *Acropora* species to present the time-calibrated species tree without the effects of introgression genes (Figure 2.1C). Among the five *Acropora* species, the four species of one monophyletic lineage split from *A. tenuis* at 12.16 Mya (95% highest posterior density (HPD): 10.58 My~13.71My); *A. digitifera* and *A. echinata* were split at 8.51 Mya (95% highest posterior density (HPD): 7.41 My~9.60 My); *A. gemmifera* and *A. subglabra* split at 6.29 Mya (95% highest posterior density (HPD): 5.49 My~7.11 My) (Figure 2.1C). In all, I inferred a crown age of 15 Mya *Acropora* with the remaining splits in the tree occurring before 6 Mya (Figure 2.1C and Figure 3.1). This sets a timescale for interpreting the results of the rest of the analyses.

3.3.2 Demographic inference with PSMC

Using the timescale of *Acropora* evolution established by the phylogenomic analysis, I evaluated demographic changes in *Acropora* lineages and link them to Earth's geologic history. I estimated the average mutation rate of *Acropora* as 2.9×10^{-8} per site per generation (see Methods) and then the demographic history was respectively simulated with each of their local density of heterozygotes using the pairwise sequentially Markovian coalescent (PSMC) model (Li and Durbin, 2011). The PSMC analysis showed the five species' demographic histories from 4 Mya to 10 Kya (Figure 3.2 and Figure 3.3). Generally, the five species had similar demographic history with a population expansion from 2 Mya and then decline after 900 Kya during the Mid-Pleistocene Transition (MPT, 0.75-1.25 Mya) (Figure 3.2). The MPT in particular—a period where the amplitude of glaciation-driven sea-level oscillations increased dramatically (Elderfield et al., 2012)—has been identified as a period of local extinction in corals (Getty et al., 2001).

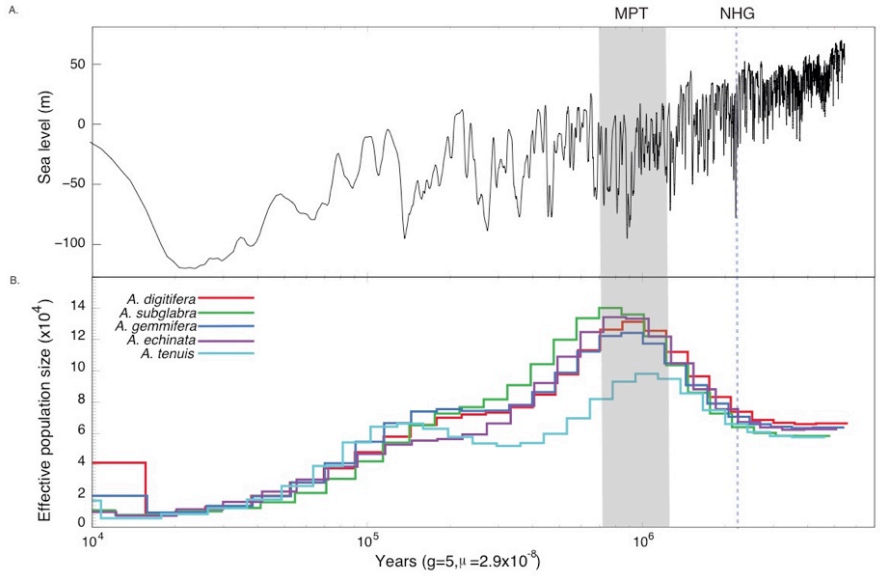


Figure 3.2. **Demographic history of *Acropora* lineages.** (A) Sea-level changes in past 5 My indicated with the onset of northern-hemisphere glaciation (NHG, dashed line) and the Mid-Pleistocene Transition (MPT, gray shade). The onset of NHG and ensuing sea-level fluctuation are associated with mass extinction in the fossil record. (B) Demographic history inference of five *Acropora* species. Effective population size (*Ne)* over time were estimated from patterns of heterozygosity with generation time (g= 5) and average neutral mutation rate per generation ($\mu = 2.9 \times 10^{-8}$) for each species using the pairwise sequentially Markovian coalescent (PSMC) model.

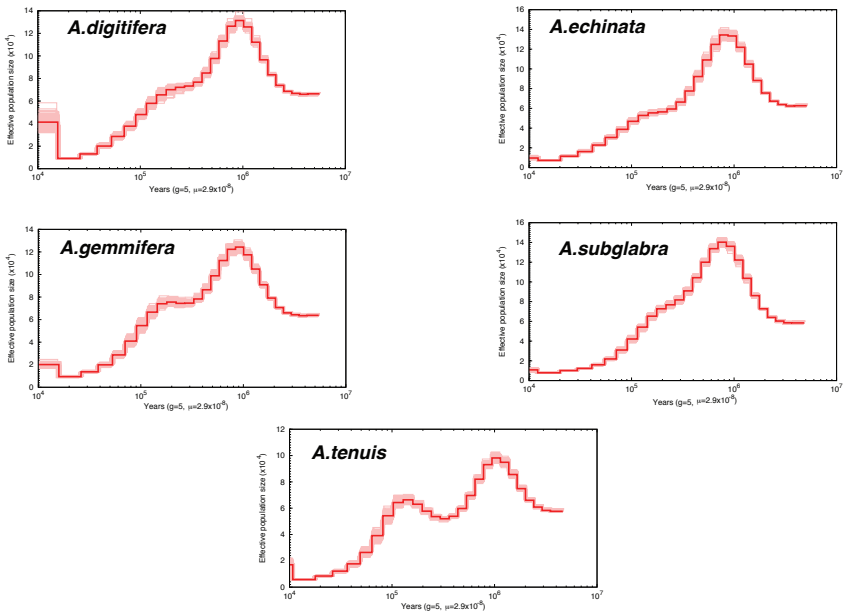


Figure 3.3. **Demographic histories of each species inferred with Pairwise Sequential Markovian Coalescent analysis.** The historical effective population size (Ne) and time scales are estimated from patterns of heterozygosity with generation time (g= 5) and neutral mutation rate per generation ($\mu = 2.9 \times 10^{-8}$) for each species with PSMC model. Thick lines correspond to the PSMC inferences and thin light lines correspond to PSMC inferences on 100 bootstraps.

3.4 Discussion

The staghorn corals of the genus *Acropora* constitute the foundation of modern coral reef ecosystems, but much work remains to identify the process shaping their rise to dominance. Understanding this question is critical for anticipating coral responses to the ongoing multifaceted changes of the Anthropocene.

After accounting for lack of congruence introduced by introgression, I inferred the age of the common ancestor of extant *Acropora* (using only non-introgressed loci) to be within the Miocene (95% highest posterior density (HPD): 13.5My~17.4My). This set a timescale for *Acropora* evolution that I applied to the demographic analysis. Although the five species diverged over six million years ago, they all show relatively similar demographic expansion and contraction in the last 3 My. The fossil record

shows that mass extinctions of nearshore marine organisms occurred around 2~3 My probably due to onset of the northern hemisphere glaciation in Plio-Pleistocene resulting in massive empty niches (O'dea et al., 2007; Pimiento et al., 2017; Prada et al., 2016) and the timing of the demographic expansion matches predictions of the hypothesis that glaciation driven mass extinction opened niche space for *Acropora*, which could better cope with rapid sea level changes since the onset of northern hemisphere glaciation (Figure 3.3). On the other hand, the reasons for the demographic decline of all five lineages since the MPT are more enigmatic, but it could be indicative of increased niche-filling and competition due to either radiation of new *Acropora* lineages or the recovery of other coral lineages as they adapt to more rapid sea-level changes and increase in abundance (Renema et al., 2016; Richards et al., 2013). Notably, the demographic history of *Acropora*, the dominant coral of the Indo-Pacific, is remarkably similar to the demographic pattern recently inferred in species of stony corals in the Caribbean (Prada et al., 2016), and matches broader dynamics inferred from the fossil record (Renema et al., 2016). This suggests that the demographic expansion of certain coral species following a glaciation-driven mass extinction was a generalized global event, and not limited to a single taxonomic group or region. This shaped the composition of the surviving reef communities, preferentially favoring rapidly dispersing and growing groups such as *Acropora*.

In addition to being consistent with the fossil record, the results are also consistent with other recent demographic studies of corals. In particular, the results are strikingly similar to findings in a recent study on stony coral in the Carribbean, a finding which was also supported by the Plio-Pleistocene fossil record. In addition, a recent study of the demography of *A. millepora* over the past 500Ky with a different approach to the one used here. Their result showed the demography of *A. millepora*, a

species not included in the study, and has an overall similar demographic history to the results for *A. tenuis* (Matz et al., 2017). That said, I do acknowledge the possibility that introgression (or population structure) could in principle have an influence in representing *Ne* change of a single lineage with PSMC (Hawks, 2017; Mazet et al., 2015; Mazet et al., 2016). As far as I know, there are no methods that fully account for hybridization in the calculation of demographic history that could be applied to the data, and such analyses are commonly used in the presence of hybridization in other studies(Árnason et al., 2018; Foote et al., 2016) However, I do not believe this to be the most likely explanation for the patterns in the data for the following reasons. First, the basic pattern I found—increase to a peak during the MPT followed by decline—was found in both the different putative "syngaemon" groups, including *A. tenuis* and the other including the rest of the species. Second, different lineages have different levels of introgression, for example *A. echinata* is the recipient of the major introgression event, but all show a similar demographic pattern. Third, previous population-level analysis on *A. millepora*, which was limited to the past 500Ky for methodological reasons, matched the demographic results from PSMC (Matz et al., 2017). Thus, while I cannot completely rule out a role of hybridization in the demographic analysis, it seems unlikely such an effect would cause the analyses to be biased in a way that matches the specific *a priori* predictions based on previous studies.

If the recent dominance of the staghorn corals and other species with similar life histories can be attributed to their ability to cope with the rapid sea level changes of the Plio-Pleistocene, it is tempting to reason that modern reefs should be well-suited to keep up with the climate-driven rapid sea level changes of the Anthropocene. However, if reefs need fast dispersers and rapid growers to keep up with sea level

changes, this apparent strength could prove to be an ecosystem-level weakness. Many taxa with life histories adapted for fast growth and high dispersal rates are more vulnerable to stressors including disease, predators, and environmental perturbations (Darling et al., 2012; Kittel, 2013). Indeed, among the corals *Acropora* are known to be one of the most sensitive to the common Anthropocene disturbances, have fast growth rates and among the most prone to bleaching (Darling et al., 2012; Goreau and Goreau, 1959; Renema et al., 2016). Their global diminishment would undermine the ability of coral reef communities to keep up with rapid sea-level changes, and further threaten the persistence of ecosystems critical for two thirds of marine species (Pimiento et al., 2017).

After accounting for lack of congruence introduced by introgression, I inferred the age of the common ancestor of extant *Acropora* (using only non-introgressed loci) to be within the Miocene. This set a timescale for *Acropora* evolution that I applied to the demographic analysis. Although the five species diverged over six million years ago, they all show relatively similar demographic patterns in the last 3 My. Notably, there was an increase in effective population size beginning near the onset of northern-hemisphere glaciation 2 Mya and reaching a peak around the end of the MPT (800 Kya). Since then, all species have declined toward their present day effective population sizes (1/8-1/12x peak abundance). The timing of the demographic expansion matches predictions of the hypothesis that glaciation driven mass extinction opened niche space for *Acropora*, which could better cope with rapid sea level changes since the onset of northern hemisphere glaciation. The fact that effective population size peaked after the onset of the highest amplitude sea level changes is also consistent with this hypothesis. The reasons for the demographic decline of all five lineages since the MPT are more enigmatic, but it could be

indicative of increased niche-filling and competition due to either radiation of new *Acropora* lineages or the recovery of other coral lineages as they adapt to more rapid sea-level changes and increase in abundance (Renema et al., 2016; Richards et al., 2013).

The demographic history of *Acropora*, the dominant coral of the Indo-Pacific, is remarkably similar to the pattern recently inferred in species of stony corals in the Caribbean (Prada et al., 2016), and matches broader dynamics inferred from the fossil record. This implies that the demographic expansion of certain coral species following a glaciation-driven mass extinction was a generalized global event, and not limited to a single taxonomic group or region.

If the recent dominance of the staghorn corals and other species with similar life histories can be attributed to their ability to cope with the rapid sea level changes of the Plio-Pleistocene, it is tempting to reason that modern reefs should be well-suited to likely climate-driven rapid sea-level changes of the Anthropocene. However, if reefs need fast dispersers and rapid growers to keep up with sea level changes, this apparent strength could prove to be an ecosystem-level weakness. Many taxa with life histories adapted for fast growth and high dispersal rates are more vulnerable to stressors including disease, predators, and environmental perturbations (Darling et al., 2012; Kittel, 2013). Indeed, among the corals *Acropora* are known to be one of the most sensitive to the common Anthropocene disturbances and among the most prone to bleaching (Renema et al., 2016; Woodley et al., 2016). Their global diminishment would undermine the ability of coral reef communities keep up with rapid sea-level changes, and further threaten the persistence of ecosystems critical for two thirds of marine species (Pimiento et al., 2017).

In all, this demographic inferences provide an insight into the rise to dominance of *Acropora* in past 3 My.

Chapter 4

A likely ancient genome duplication in the speciose reef-building coral genus: *Acropora*

4.1 Introduction

Reef-building corals contribute to tropical marine ecosystems that support innumerable marine organisms, but reefs are increasingly threatened due to 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 in 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 of *Acropora* 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 Indo-Pacific Ocean, is thought to have originated from *Astreopora* almost 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). Additionally, gene duplications have been shown in several *Acropora* gene families (Gacesa et al., 2015; Hamada et al., 2013). Thus, 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)-genome duplication ((W/LS)-GD), or paleopolyploidy, has shaped in 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 4.1). Two separate GD events have been documented in the common ancestors of vertebrates (Two-rounds of GD) (Dehal and Boore, 2005) and another major GD 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 GD event (Jiao et al., 2011; Tiley et al., 2016), and many other GD events have been reported in major clades of angiosperms (Soltis et al., 2009; Vanneste et al., 2014). In addition, two-rounds of GDs in the vertebrates are suggested to have occurred during the Cambrian Period, and some GDs 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, GD is regarded as an important evolutionary way to reduce the risk of extinction or the advantages of WGD increases success to survive (Van de Peer et al., 2009; Van De Peer et al., 2017; Vanneste et al., 2014). However, the study of GD in Cnidaria has received less attention (Kenny et al., 2017; Li et al., 2018; Schwager et al., 2017; Van de Peer et al., 2009; Van De Peer et al., 2017).

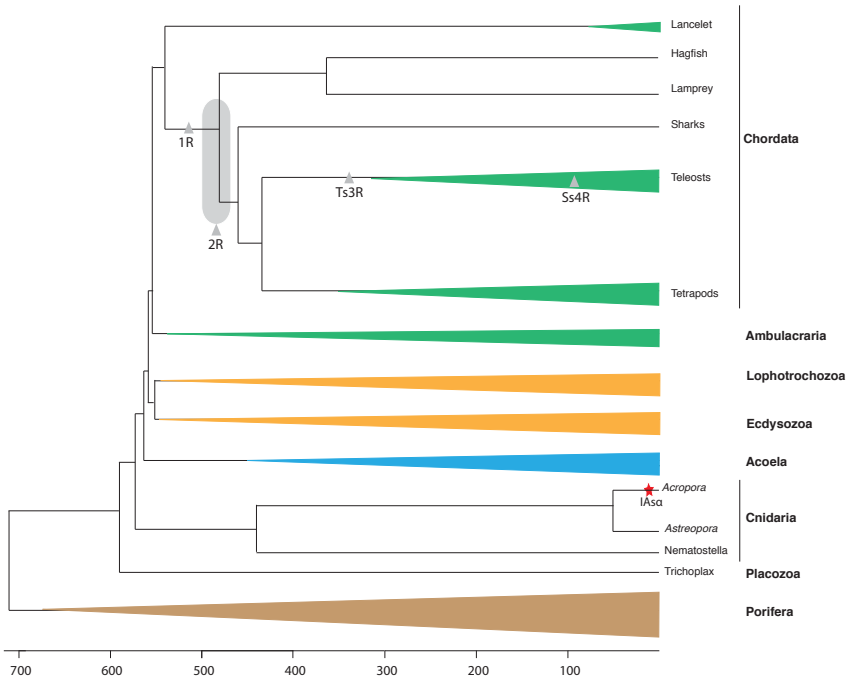


Figure 4.1. **GD events in evolution of the animal clade.** The backbone and divergence time of the tree are based on various sources (e.g., Satoh, 2016). The shaded grey oval represents the uncertain position of two rounds of GD and colored triangles represent the corresponding divergent groups. Grey triangles represent GDs and the red star represents invertebrate GD specific to *Acropora* (IAsα) reported in this study.

Duplicated genes created by GD have complex fates during time to 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, or 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 by 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 (RFLP) were used to identify gene duplications in polyploids a few decades ago (Fürthauer et al., 1999; Stuber and Goodman, 1983). In the past ten years, NGS 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 GD: 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); 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 GD (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, the dS-based method cannot detect ancient GD, and 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 GD using either method. In contrast, the synteny analysis relies heavily on the genome assembly quality. Poor assembly quality can hide the GD signals, and some genomes with huge rearrangements cannot be used to detect GD using synteny block identification. Thus, the most credible conclusions depend on complementary evidence from different methods (Chen and Birchler, 2013; Soltis and Soltis, 2012; Tiley et al., 2016).

Here, I analyzed a genome of *Astreopora* (*Astreopora* sp1) as an outgroup, and five *Acropora* genomes (*A. digitifera*, *A. gemmifera*, *A. subglabra*, *A. echinata* and *A. tenuis*) to address the following questions using all three methods; (I) whether and when GD occurred in *Acropora*, (II) what is the fate of duplicated genes in *Acropora* after the event, (III) what the gene expression patterns of duplicated genes cross five developmental stages in *A. ditigifera*, and (IV) what roles of GD were involved in the diversification of toxic proteins in *Acropora*.

4.2 Methods

4.2.1 Species information, genomic data and gene family cluster

Data can be accessed at: http://marinegenomics.oist.jp and http://comparative.reefgenomics.org/datasets.html (Bhattacharya et al., 2016). The *Acropora* species information in this study will be described in the 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. In detail, coral samples were collected in Okinawa, Japan and the sperms of the single colony were used to isolate high-molecular weight DNAs. PCR-free shotgun libraries were prepared for genome sequencing with HiSeq

2500 in Rapid mode (Illumina). *Astreopora* sp1 were assembled with Platanus assembler (Kajitani et al., 2014). Then, I performed genome annotation of *Astreopora sp1* with de novo methods based on repeats-masked genomes. 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). In addition, the chromosome number of *Acropora* is 2n=28, but there is no report about the chromosome number of *Astreopora* sp1.

4.2.2 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. 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, I 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 the previous study (Mao et al., 2018). I ran BEAST2 three times independently, 50 million Markov chain Monte Carlo (MCMC) generations for each run, then I used Tracer to check the log files and I found that ESS of each of parameters exceeded 200.

4.2.3 Orthogroup selection and detection of a GD 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 to 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 with bins=0.02 in R (Team, 2013). All processes were run in GenoDup (Mao and Satoh, 2018).

(b) dS distributions of anchor gene pairs

I 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 with bins=0.02 in R (Team, 2013). All processes were run in the GenoDup (Mao and Satoh, 2018).

(c) *dS distributions of orthologous gene pairs*

I 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 with bins=0.02 in R (Team, 2013).

4.2.4 Detection of a GD 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 GD.

For each of the 883 gene tree reconstructions, I used MAFFT (Katoh et al., 2002) to align amino acid sequences of each single-copy ortholog. I aligned coding sequences with TranslatorX (Abascal et al., 2010) based on amino acid alignments and I 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 the previous study (Mao et al., 2018). I ran BEAST2 three times independently, 30 million Markov chain Monte Carlo (MCMC) generations for each run. Then I used Tracer to check the log files. 135 time-calibrated phylogeny with ESS values exceeded 200 were carried out by BEAST2.

4.2.5 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 GD within dS distributions, I 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 I used the *kde()* function in KDE toolbox, I 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, I 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 I

estimated the peak value of this distribution as -1.8165. I 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, I 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. I estimated maximum peak values for each 100 bootstrap samples. Then I 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 GD in the distribution of inferred node ages, I used the *kde()* function in KDE toolbox to estimate the peak value as 30.78 My, and I used bootstrapping to estimate the 95% CIs as 27.86 to 34.77 My. For bootstrapping, I 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. I estimated maximum peak values for each of 100 bootstrap samples. Then, I sorted maximum peak values and values of 6^{th} and 95^{th} rank defined the 95% CI.

4.2.6 Maximum likelihood approach to detect GD with gene family count data

First, I 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, I used the GDgc package in R to estimate log likelihood for parameters (0, 1, 2, 3) of GD event(s) with setting (dirac=1,conditioning="twoOrMore") (Rabier et al., 2014). Then, I performed likelihood ratio test (pchisq(2*(Likelihood_1-Likelihood_2), df=1,

lower.tail=FALSE)) to find the best model and found that one GD event was the best model to fit the gene family count data. I estimated the age of GD on 4 My intervals between 18.69 and 38.69 My under a one GD event model. The lowest log likelihood was shown at the age of GD: 30.69 and 34.69 My.

4.2.7 Gene expression profiling analysis and dN/dS calculation

I selected 236 gene pairs of *A. digitifera* (ohnologous gene pairs) from 831 orthogroups. I 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).

4.2.8 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.

I 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, I 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.

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

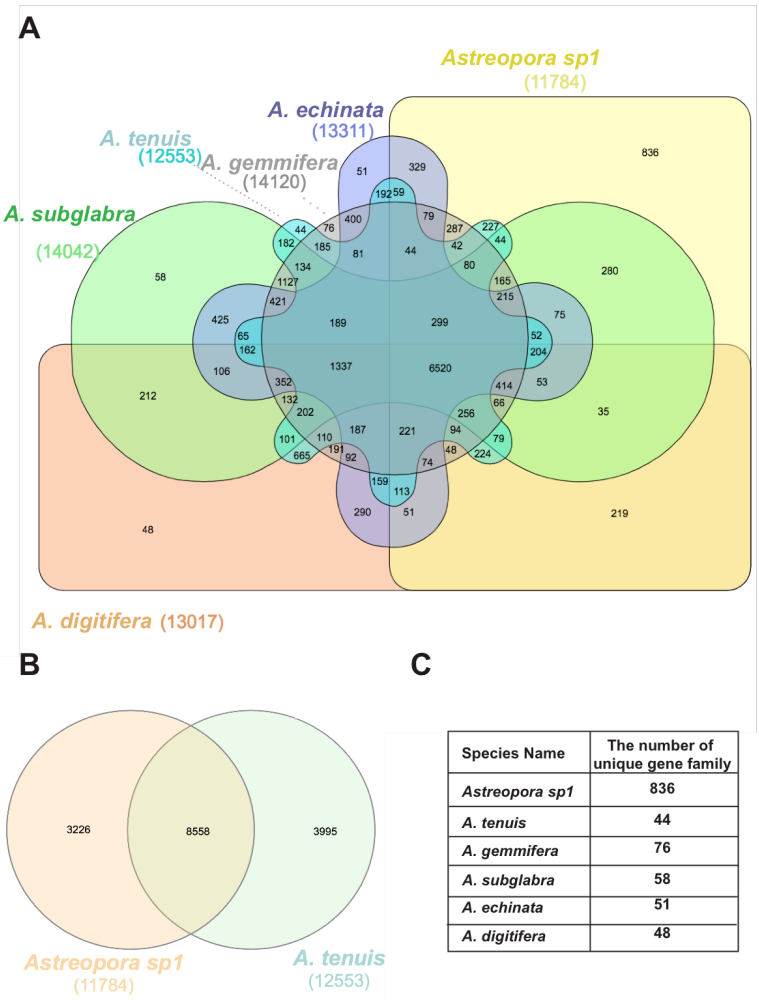
I BLASTed the sequences of 154 high quality core-orthogroups of *Acropora* against the UNIPROT database to find best hits. Identical hits in each ohonlogs group were removed and the remaining hits were used to perform gene enrichment in David (Huang et al., 2009). I also used InterProScan (Zdobnov and 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.

4.3 Analyses and Results

4.3.1 Cluster of gene families and calibration of the acroporid phylogenomic tree

I clustered all homologs among the six *Acroporid* species into 19,760 gene families, and they shared 6,520 gene families (Figure 4.2). My previous gene family cluster analysis of the five *Acropora* species showed that each *Acropora* genome had very few unique gene families (Mao et al., 2018). Interestingly, I found the same pattern in *Acropora* when integrating with the data of *Astreopora* sp1, but *Astreopora*

sp1 had 218 unique gene families, suggesting that *Astreopora* sp1 is genetically divergent from the five *Acropora* species. 3,461 single-copy orthologs were selected from 6,520 shared gene families. These were concatenated to reconstruct a calibrated phylogenomic tree based on the reported divergence time of *Acropora* (Mao et al., 2018). I found that *Astreopora* sp1 split from *Acropora* ~ 53.6 Mya (95% highest posterior density (HPD): 51.02 - 56.21 My) (Figure 4.3). This result established a timescale to analyze the timing of the subsequent GD.



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 4.2. **Venn diagrams of shared and unique gene families in six *Acroporid* species.** (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 families in six *Acroporid* species.

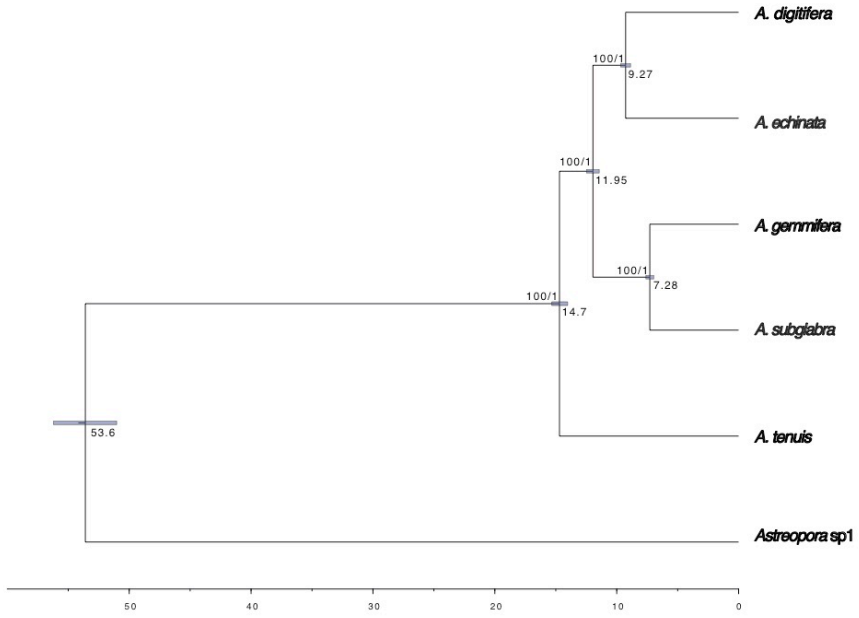


Figure 4.3. **Phylogeny of the Family *Acroporidae.*** Time-calibrated phylogenetic tree reconstructed based on fossil calibration and concatenated coding sequences (7,467,066 bp in total) from 3,461 single-copy orthologous genes with BEAST2. Branch lengths are scaled to estimated-divergence time. Posterior 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.

4.3.2 GD identification with the dS-based method

Synonymous substitution rate (dS) analysis has been widely used to infer GD (Vanneste et al., 2014; Vanneste et al., 2012). I identified over 10,000 paralogous gene pairs, based on their sequence similarities as well as I identified over 10,000 anchor gene pairs, based on synteny information from each species (Table 4.1; See Methods). Then I calculated dS values from paralogous gene pairs and anchor gene pairs for each species.

Table 4.1. Numbers of gene pairs in the paralogous gene pairs and anchor gene pairs datasets

	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. subglabra	50852	12077	44093	2380
A. tenuis	34097	6635	28073	1488
Astreopora sp1	49135	13648	52481	3033

An 'L-shaped' distribution was evident in both paralogous and anchor gene pair dS distributions of *Astreaopora* sp1, illustrating that no GD occurred in *Astreaopora* 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 GD did occur in *Acropora* (Figures. 4.4--4.5).

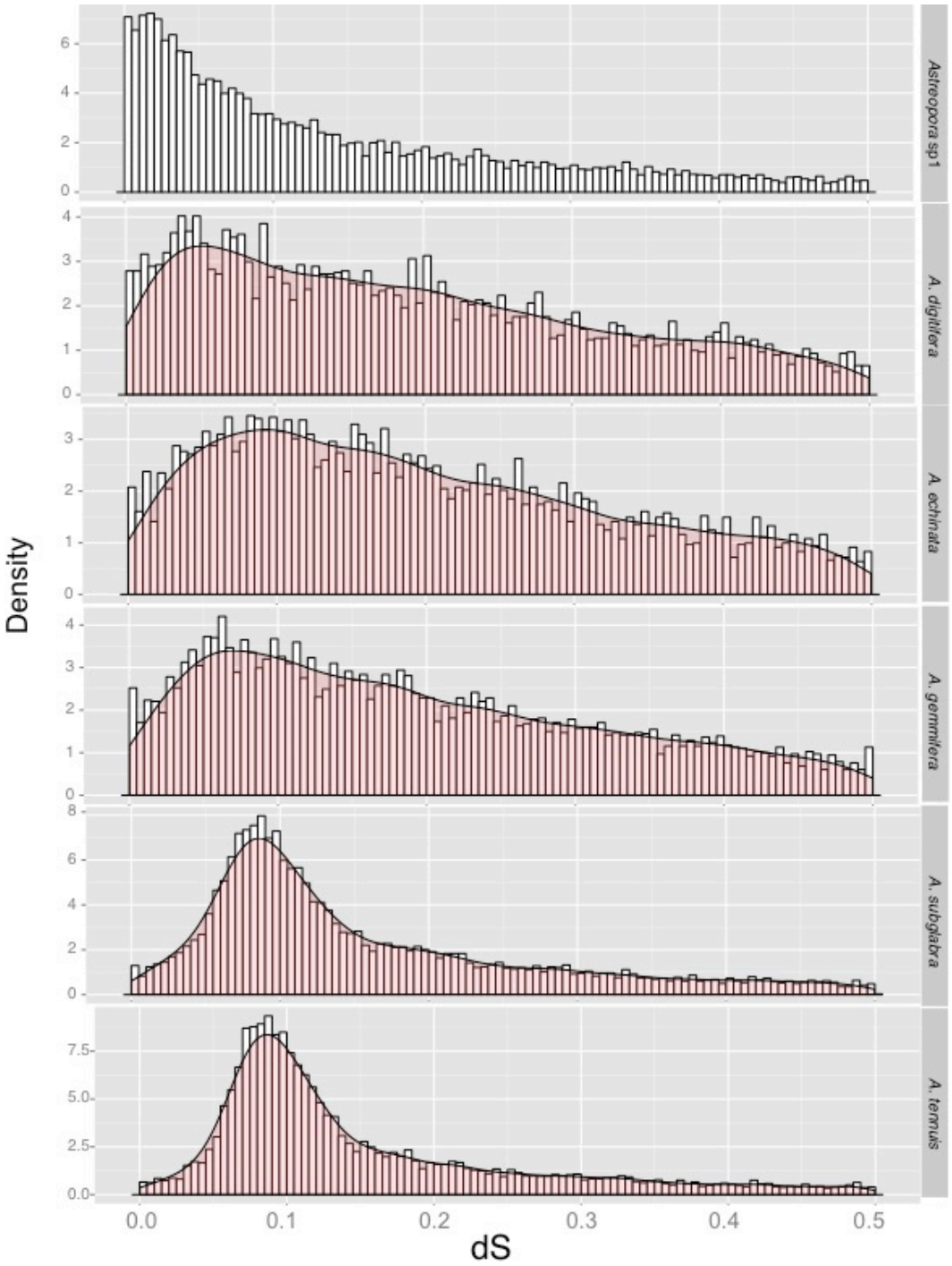


Figure 4.4. **Frequency distribution of dS values for paralogous gene pairs in five *Acropora* and one *Astreopora* species.** The distributions of dS values of paralogs, estimating neutral evolutionary divergence since the two paralogs diverged, are plotted with a bin size of 0.005, showing the similar peaks (dS value: 0-0.3) in *Acropora*.

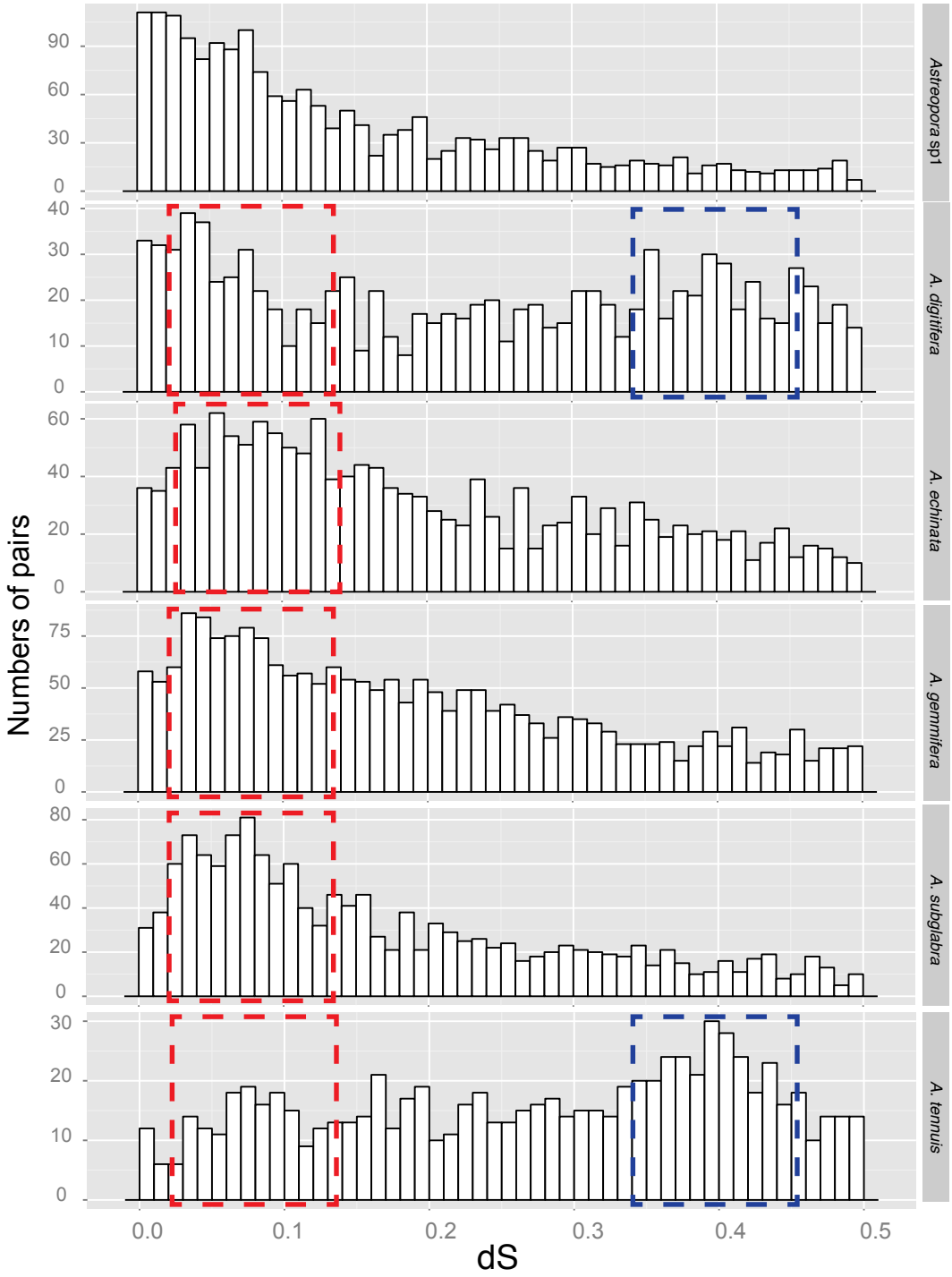


Figure 4.5. **Frequency distribution of dS values for anchor-gene pairs in five *Acropora* and one *Astreopora* species.** Distributions of dS values of anchor paralogs, estimating the neutral evolutionary divergence times since the paralogs diverged, are plotted with a bin size of 0.01, showing the 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).

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 speciation time between them according to neutral evolution theory (Berthelot et al., 2014; Zhang et al., 2017). I combined the dS values of paralogous gene pairs for the five *Acropora* species and estimated the peak in the log dS distribution (modal value = -1.82). Also, I estimated the distribution of orthologous gene pairs between *Astreopora* sp1 and *A. tenuis* (modal value = -0.31) and the distribution of orthologous gene pairs between *A. tenuis* and *A. digitifera* (modal value = -3.46). The result indicates that the GD occurred in *Acropora* after the split of *Astreopora* sp1 and *A. tenuis* (Figure 4.6). In other words, an ancient GD event likely occurred in the most recent common ancestor of *Acropora*. Based on speciation time estimated in the calibrated phylogenomic tree and assuming a constant dS rate (Vanneste et al., 2014), I estimated that the GD of *Acropora* occurred ~ 35 Mya (95% confidence interval: 31.18 - 35.7 My) (Table 4.2, See Methods). Here, I defined this event as invertebrate α event of GD specifically in *Acropora* (IAsα).

Table 4.2. Peak value estimations of dS distribution by KDE toolbox

	Astreopora sp1_*A. tenuis*	*A. tenuis_A. digitifera*	*Acropora_GD*
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)

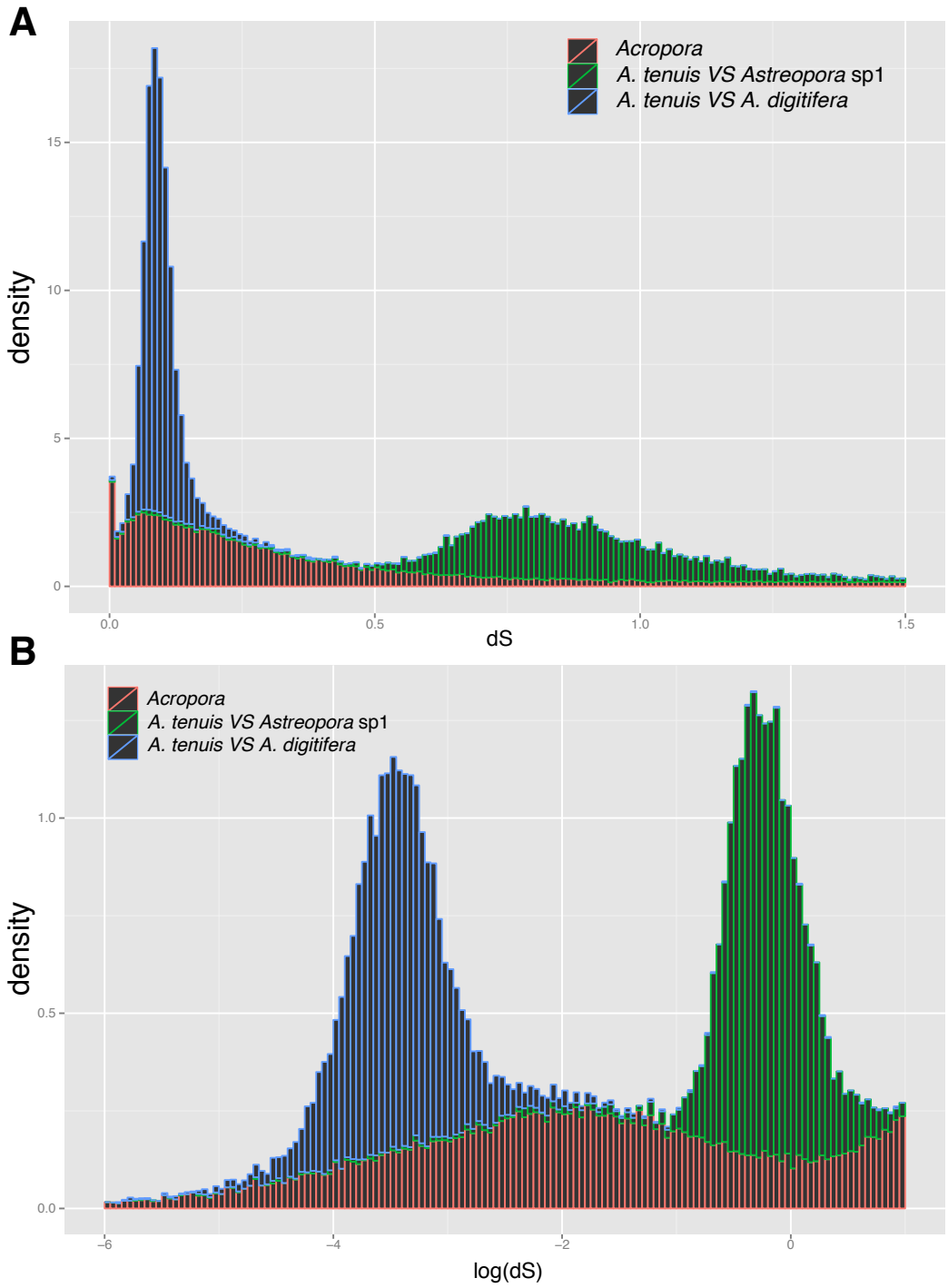


Figure 4.6. **Frequency distribution of dS values for paralogous genes in *Acropora* and for orthologous genes.** (A) Frequency distribution of dS values for paralogous genes in *Acropora* and for orthologous genes showing that a GD event occurred in the most recent common ancestor of *Acropora*. Distributions 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.

4.3.3 Phylogenomic and synteny analysis of IAsα

If the existence of IAsα is correct, then 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 (((*Acropora clade1*) bootstrap1, (*Acropora clade2*) bootstrap2), *Astreopora* sp1), defined as gene duplication topology (Figure 4.7).

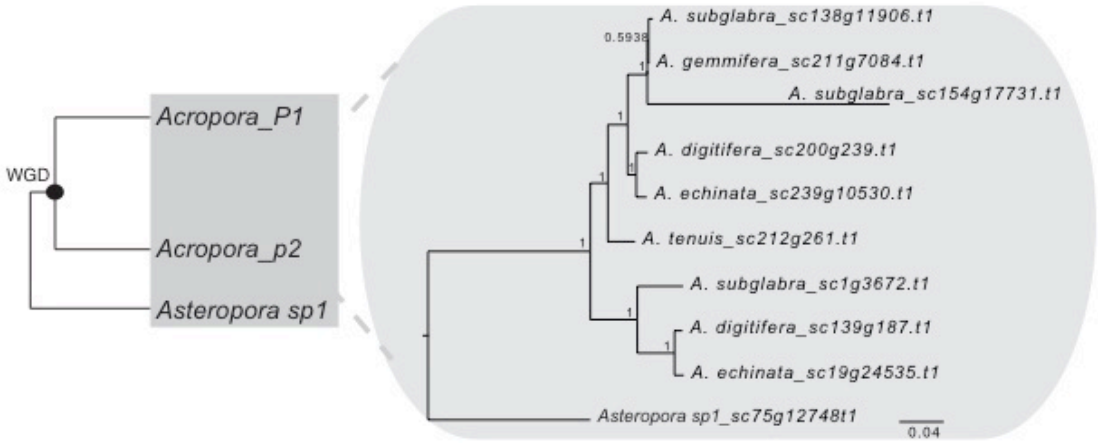


Figure 4.7. **Hypothetical tree topology of duplicated genes in the Acroporidae and the phylogeny of one duplicated gene (alpha-protein kinase 1-like).** The phylogenetic tree shows gene retention, loss, and duplications following with GD.

I performed a phylogenomic analysis to confirm the presence of IAsα. First, I 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 one gene copy in *Astreopora* sp1. I selected 883 orthogroups from 19,760 gene families, and reconstructed the phylogeny of 883 orthogroups using both Maximum likelihood (ML) and Bayesian methods. I found that the phylogeny of 205 orthogroups was consistent

with gene duplication topology supporting IAsα. I further defined the 205 orthogroups as core-orthogroups (Table 4.3).

Table 4.3. Numbers of gene family in orthogroups, core-orthogroups and high-quality core-orthogroups

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

In particular, I 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 4.7).

In order to estimate the split time of the two *Acropora* clades that could be regarded as the timing of IAsα, I selected 154 high-quality core-orthogroups, with both 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, I found that it is difficult for the parameters in MCMC to converge in 70 core-orthogroups, and I successfully dated the phylogenetic trees of only 135 high-quality core-orthogroups. Next, I estimated the distribution of inferred node ages between the two *Acropora* clades and the peak value was estimated as 30.78 My (95% confidence interval: 27.86-34.77 My), indicating that IAsα occurred at 30.78 My (Figure 4.8). This result strongly supports the timing of the IAsα estimated using the dS-based method.

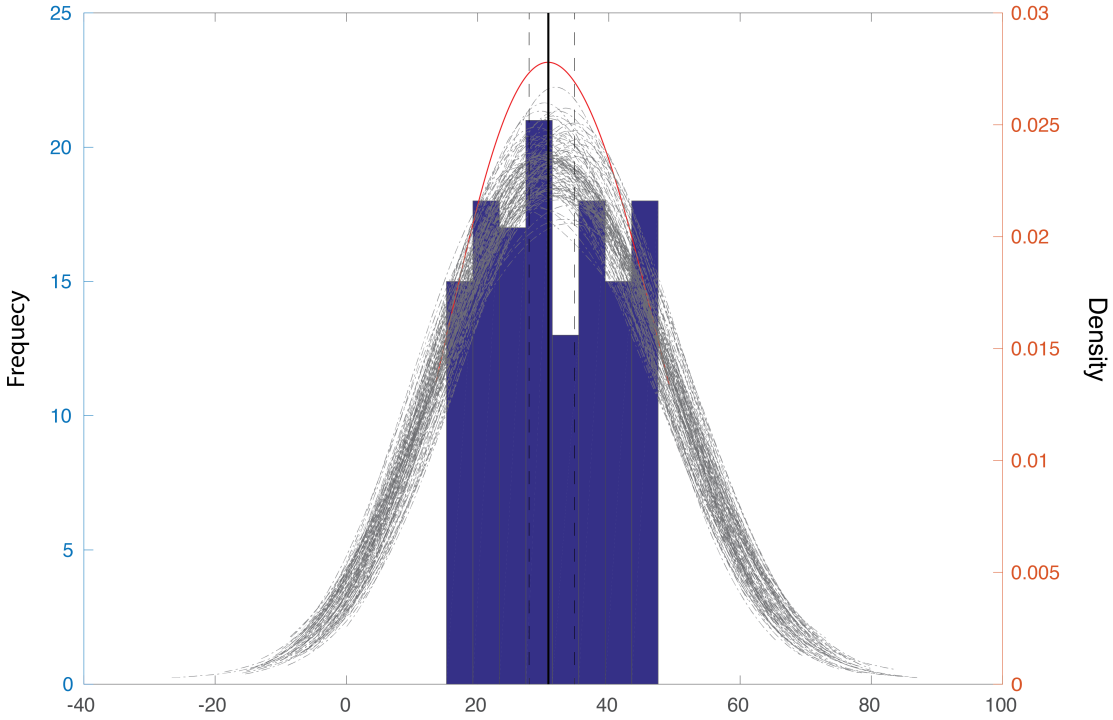


Figure 4.8. **Node age distribution of IAsα.** Inferred node ages from 135 phylogenies were analyzed with KDE toolbox to show the peak at 30.78 My, represented by the black solid line. The grey lines represent density estimations from 1000 bootstraps and the black dotted line represents the corresponding 95% confidence interval (27.86 - 34.77 My) from 100 bootstraps.

Intergenomic co-linearity is often used to directly identify ancient GD and to reconstruct ancestral karyotypes in vertebrates (Berthelot et al., 2014; Nakatani et al., 2007; Zhang et al., 2017). I performed intergenomic co-linearity and synteny analysis between *Astreopora* sp1 and *A. tenuis* to support IAsα. First, I found great co-linearity between *Astreopora* sp1 and *A. tenuis* (Data not shown). Second, I found synteny blocks in 21 scaffolds in *Astreopora* sp1 have at least 2 duplicated segments in *A. tenuis* (Figure 4.9). For example, two duplicated segments in scaffold 130 and scaffold 70 of *A. tenuis* corresponded to a scaffold 323 in *Astreopora* sp1 (Figure 4.10).

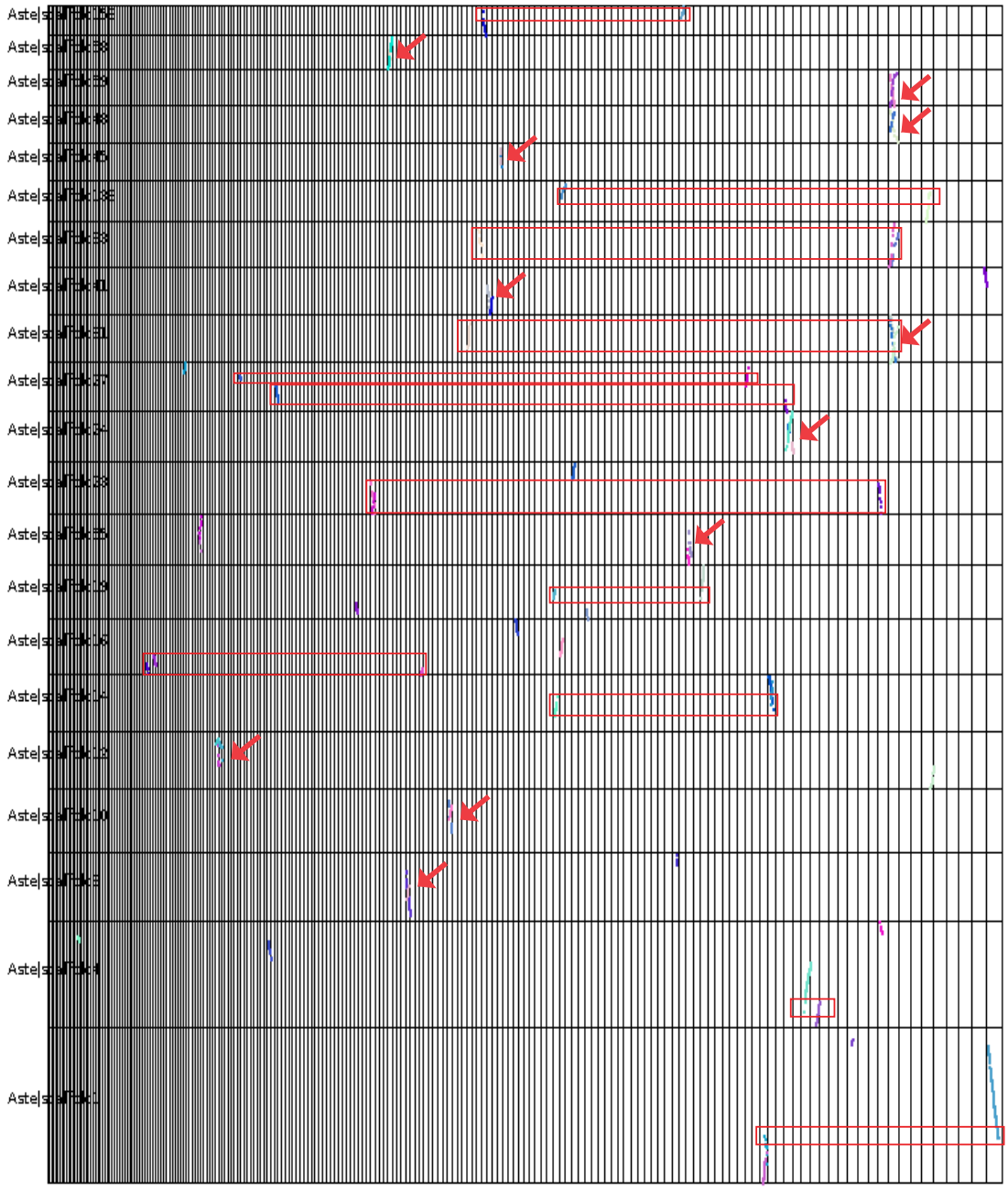


Figure 4.9. **Synteny blocks between *Astreopora* sp1 and *A. tenuis*.** Only co-linear segments with at least 10 anchor pairs are shown in between the top length 100 scaffolds of *Astreopora* sp1 (Left side) and the top length 200 scaffolds of *A. tenuis* (Bottom). Only the scaffolds of *Astreopora* 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.

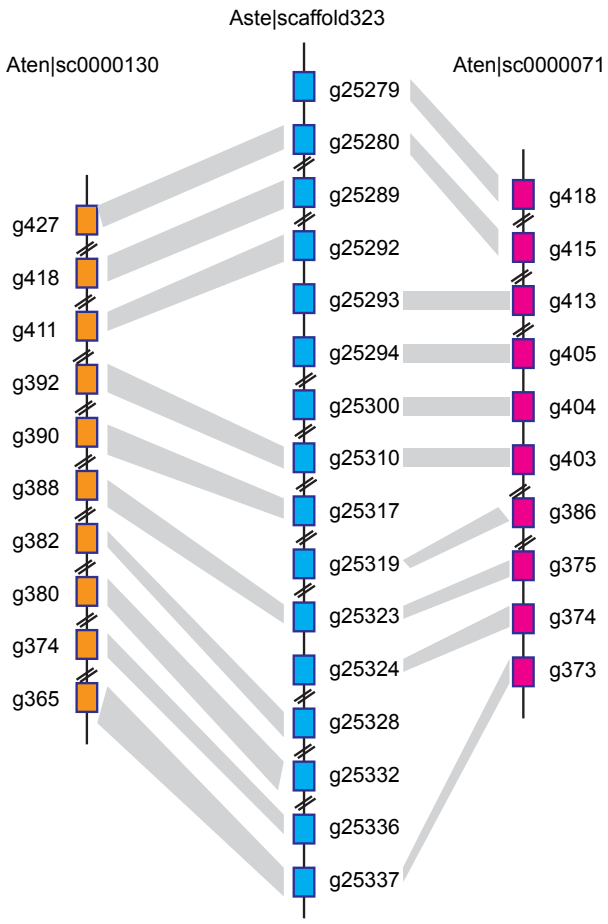


Figure 4.10. **Co-linear gene alignments of *Astreopora sp1*and *A. tenuis* on scaffolds.** The grey links show orthologs between *Astreopora sp1*and *A. tenuis*. Gene order of scaffold 323 in *Astreopora* sp1 is placed in the middle and the duplicated segments in *A. tenuis* are placed in the left and right. The duplicated segments are located in scaffold 130 and scaffold 71 in *A. tenuis*, respectively.

In summary, I clearly established the presence of IAsα using the dS-based method, phylogenomic and synteny analyses. Moreover, I suggest that IAsα probably occurred between 28 and 36 Mya (Figure 4.11).

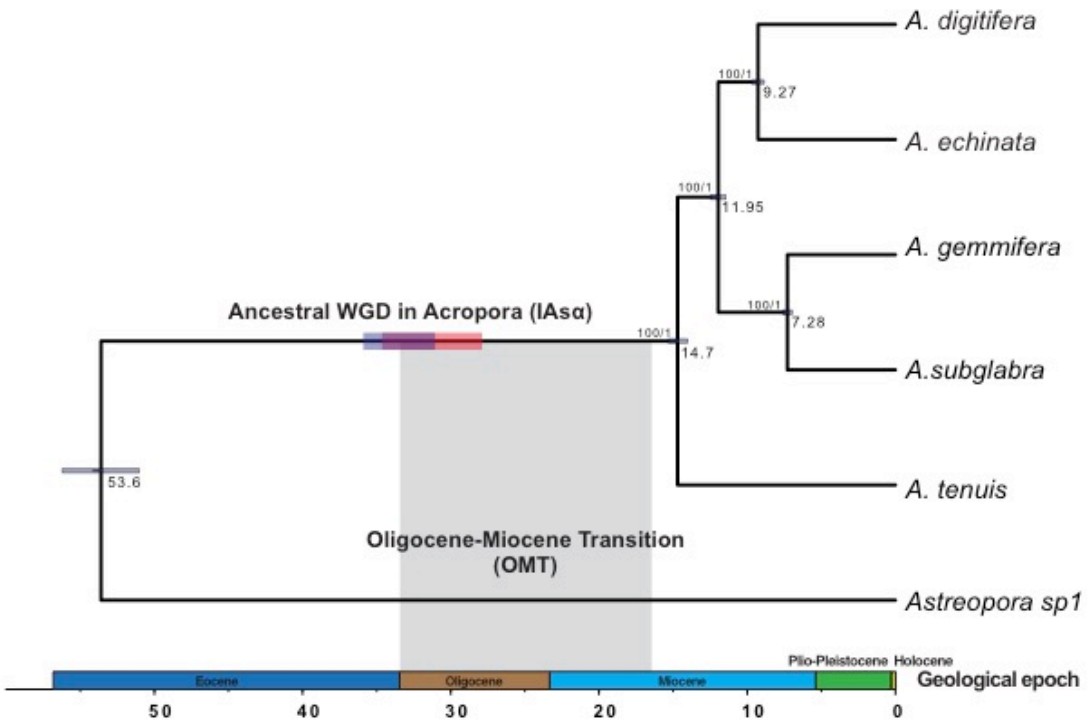


Figure 4.11. **Ancient GD in the reef-building coral *Acropora* (IAsα).** A calibrated phylogenomic tree of six *Acroporid* species inferred from 3,461 single-copy orthologs using BEAST2. Horizontal bars on branches of the tree represent the timing of GD in *Acropora*. The timing of IAsα was estimated at 35 Mya (95% confidence interval: 31.18-35.7 Mya) by dS-based analysis (horizontal blue bar) and 30.78 Mya (95% confidence interval: 27.86-34.77 Mya) by phylogenomic analysis (horizontal orange bar). Grey shading represents the timing of one coral species turnover event, the Oligocene-Miocene transition (OMT), suggesting that IAsα is correlated with OMT.

4.3.4 The fate of duplicated genes originating from IAsα

Duplicated genes provide substrates for diversification and evolutionary novelty, and most of them are regulators of complex gene networks in vertebrates and plants (Jiao et al., 2011; Kassahn et al., 2009; Zhang et al., 2017). I 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 have been

enriched in specific categories; transporter, catalytic, binding, and receptor activity, most of which are involved in gene regulation (Table 4.4).

Table 4.4. Functional annotation clustering on the GO terms of 154 high-quality core-orthogroups

Annotation cluser	P_Value
Transmembrane	1.90E-06
Death domain	3.10E-05
G-protein coupled receptor	1.20E-04
VIT domain	3.30E-03
Protein kinase-like domain	1.90E-02

Further, I 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). Orthogroups 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 (Figure 4.12A and Figure 4.13). In addition, excitatory amino acid transporters may be related to symbiotic interactions in *Acropora* (Bertucci et al., 2015). Orthogroups 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 (Figure 4.12B and Figure 4.14). These examples suggest that IAsα participates in both stress responses and symbiotic interactions in *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*.

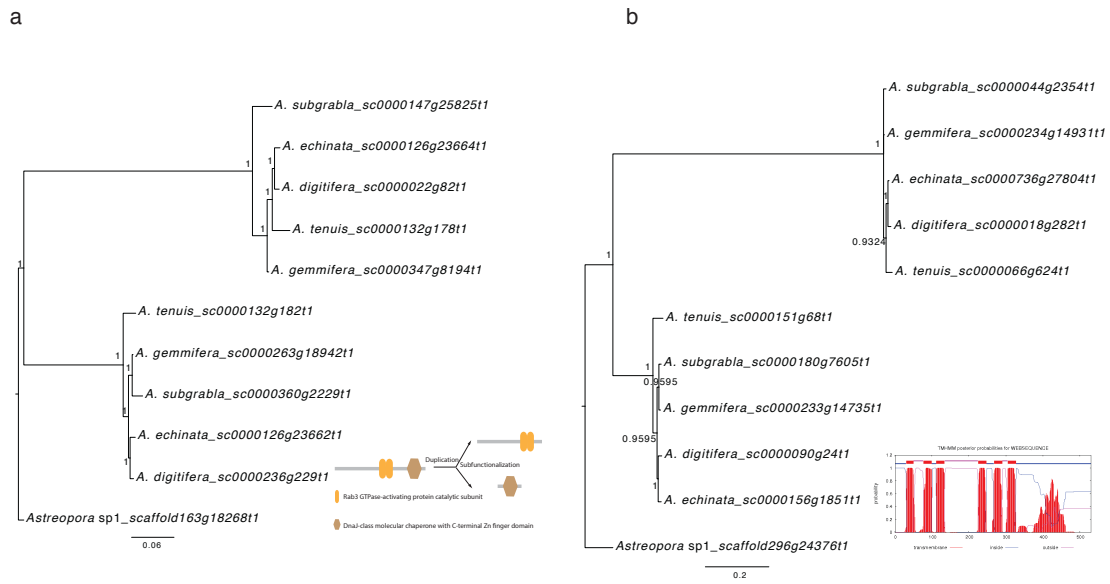


Figure 4.12. **Phylogenetic trees show duplicated genes under subfunctionalization or neofunctionalization**. (A). The phylogeny of orthogroup 1247 (dnaJ homolog subfamily B member 11-like) reconstructed with MrBayes shows a duplicated gene under subfunctionalization. Bayesian posterior probabilities 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 orthogroups 1247. (B). The phylogeny of orthogroup 1244 (excitatory amino acid transporter 1-like) reconstructed with MrBayes show a duplicated gene under neofunctionalization. Bayesian posterior probabilities are shown at each node. Six transmembrane helices prediction is shown in the bottom right.

Figure 4.13. **Alignment of orthogroup 1247 (dnaJ homolog subfamily B member 11-like) showing the independent loss of the domain in duplicates.**

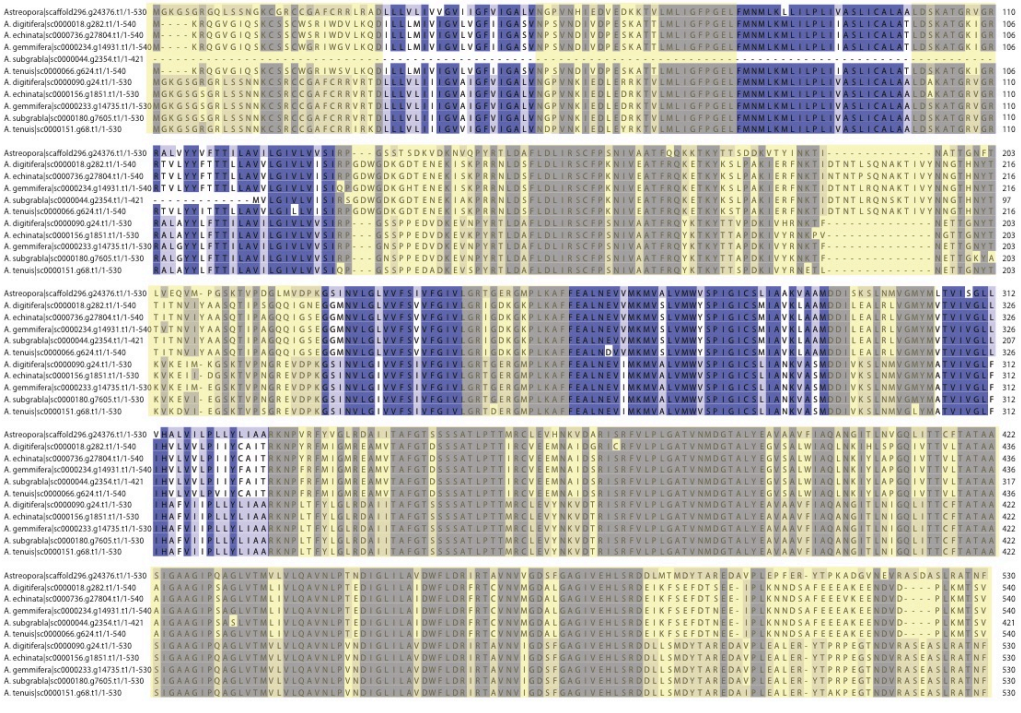


Figure 4.14. **Alignment of orthogroup 1244 (excitatory amino acid transporter 1-like) showing mutations on transmembrane and exposed regions, suggesting that new functions would be generated**. Exposed regions are shown in yellow.

4.3.5 Gene expression patterns of duplicated genes across five developmental stages in *A. digitifera*

To better to understand evolution of duplicated genes, gene expression analysis across five developmental stages in *A. digitifera* (blastula, gastrula, postgastrula, planula, and adult polyps) was carried out based on previous transcriptome data (Reyes-Bermudez et al., 2016). I identified 236 ohnologous pairs in *A. digitifera* from 883 ML phylogeny (See Methods) and found that these ohnologous pairs present an interesting gene expression profiling. I divided 236 ohnologous pairs into two clusters based on the pairwise correlation of gene expression during development (high correlation or HC: $P<0.05$; no correlation or NC: $P>=0.05$; Pearson's correlation test); 25% (25/236) ohnologous pairs in HC and 75% (211/236) ohnologous pairs in NC (Figure 4.15A). Ohnologous pairs in the HC cluster are enriched in protein kinase, while ohnologous pairs in the NC cluster are enriched in membrane transporter and ion binding proteins (Figure 4.15B). This result indicates that the two clusters of ohnologous pairs potentially evolved into different gene functions. Additionally, I compared dN/dS values in order to investigate selective pressure between HC and NC clusters (Figure 4.15C), but there is no significant difference between the two clusters (Mann-Whitney-Wilcoxon Test, $P = 0.51$).

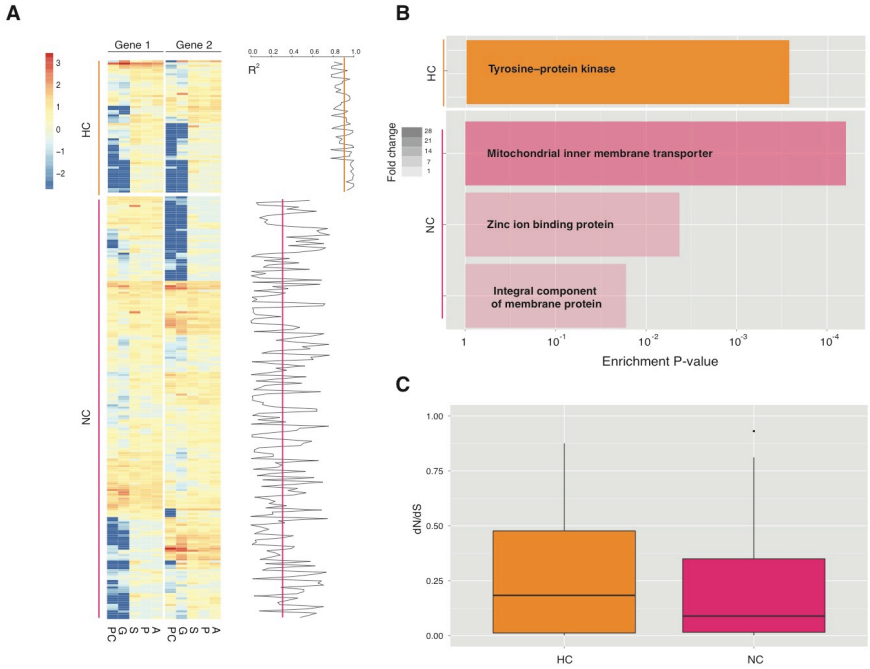


Figure 4.15. **Gene expression profiling reveals evolution of duplicated genes in *A. digitifera*.** (A). Gene expression profiling across five developmental stages (blastula: PC, gastrula: G, postgastrula: S, planula: P, and adult polyps: A) in *A. digitifera*. Two clusters of gene expression of ohnologous gene pairs: HC: high correlation, P<0.05; NC: no correlation, P>=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 ohnologues 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).

4.3.6 Evolution of toxic proteins in Cnidaria

Next, I investigated the role of IAsα in the diversification of toxins in *Acropora*. I identified ~200 putative toxic proteins in each of the five *Acropora* species, and then I clustered them with 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 4.5, See Methods). Based on the gene family phylogeny, each of which contains at least 15 genes, I found that toxic proteins have undergone widespread gene duplications in Cnidaria, and most of gene duplications occurred in individual species lineages, except for *Acropora* (Figure 4.16, other trees not shown). Interestingly, gene duplications occurred in the most recent common ancestor of *Acropora* in 9 over 15 gene families, potentially caused by GD (IAsα). For example, in gene family-1 (Coagulation factor X), each species contains ~50 genes, except *H. magnipapillata* and *P. astreoides*, and gene duplications occurred frequently in individual species lineages: *Astreopora* sp1, *M. cavernosa*, *N. vectensis*, and *P. australiensis*. However, five gene duplications were inferred to have occurred in the most recent common ancestor of *Acropora* by GD (Figure 4.16). These results indicated that IAsα potentially contributed to the diversification of proteinaceous toxins in *Acropora*.

Table 4.5. The number of putative toxin proteins in 12 Cnidarian species

Gene family	Query_name	*A. digitifera*	*A.echinata*	*A. gemmifera*	*A. subglabra*	*A. tenuis*	*Astreopora* sp1	*H. magnipapillata*	*M. cavernosa*	*N. vectensis*	*P. astreoides*	*P. australiensis*	*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 metalloprotease toxin	28	23	25	26	30	33	36	20	60	6	31	14
Gene family_4	Reticulocalbin	18	15	14	14	18	20	5	16	18	6	19	17
Gene family_5	Putative lysosomal acid lipase/cholesteryl ester hydrolase	10	10	10	11	7	13	4	6	5	4	5	5
Gene family_6	Venom carboxylesterase-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/DELTA-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	NA	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

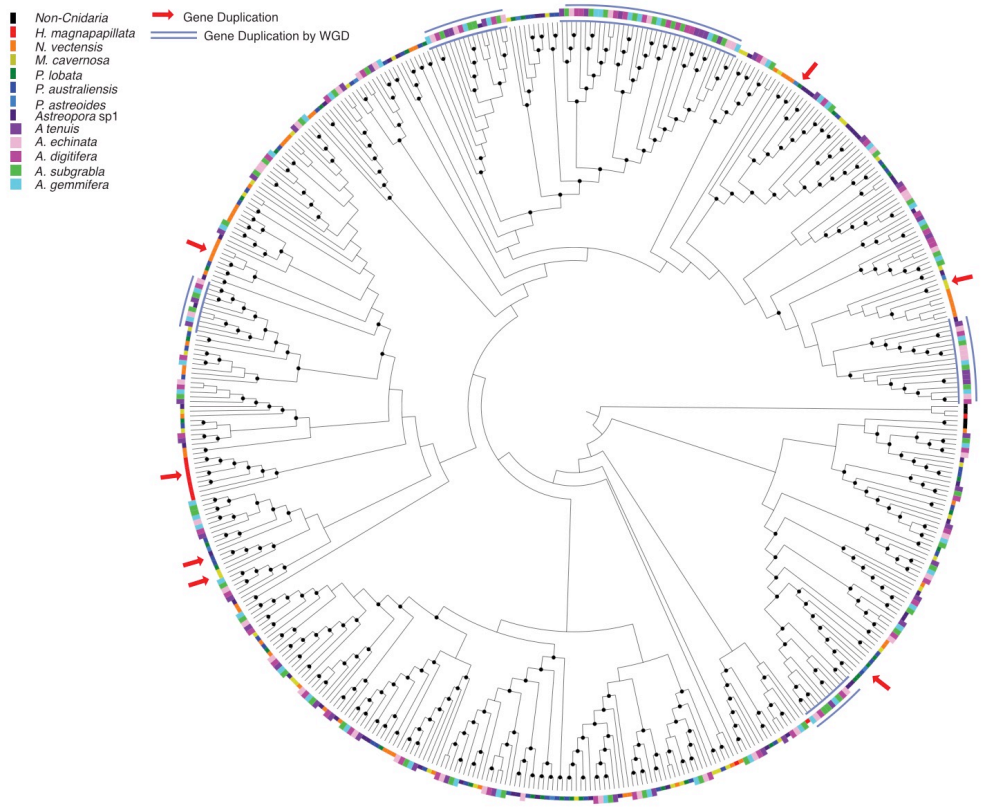


Figure 4.16. **Diversification of toxic proteins via gene duplications in Cnidaria.** Phylogenetic analysis of Coagulation factor X in 12 Cnidarian species shows wide gene duplications. Gene duplication occurred in individual species lineages (red arrows) and gene duplications by GD in *Acropora* are 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.

4.4 Discussion

Ancient GD 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 definitively identify GD 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, I analyzed genomes of one *Astreopora* and five *Acropora* species to address the possibility of GD in *Acropora* and the functional fate of duplicated genes from that event.

To the best of my knowledge, this is the first study to report genomic-scale evidence of GD in corals (IAsα). I 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, the synteny analysis between *Astreopora*. sp1 and *A. tenuis* directly supports IAsα. However, reconstruction of the ancestral karyotype will necessitate genomes assembled to the chromosome level to fully understanding gene fractionation and chromosome arrangements in *Acropora* under IAsα (Smith and Keinath, 2015; Smith et al., 2013).

Ancient GD 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 because of using poorly annotated genomes (Rabier et al., 2014; Tiley et al., 2016; Vanneste et al., 2012). There is an extra peak in the dS distribution of anchor gene pairs in *A. digitifera* and *A. tenuis* (Figure 4.5). One possible explanation is that the extra peak is artifactitious because few anchor gene pairs were used in the analysis. However, this could also indicate a second GD event in *Acropora*. I found few orthogroups with topologies that fit the two proposed GDs events (Figure 4.17). If a second GD event occurred, the reason that the second GD signal appeared among anchor gene pairs rather than among paralogous gene pairs may be that the paralogs generated by the second GD have been largely lost; thus, few

of them are only retained in conserved order. In addition, a new maximum likelihood phylogeny modeling approach was recently developed to overcome difficulties of the dS-based method (Rabier et al., 2014; Tiley et al., 2016). I used it to test whether a second GD occurred in *Acropora*. The result showed that one GD event is the best model in *Acropora* and it occurred 30.69 to 34.69 Mya (Table 4.6, Table 4.7; See Methods). Thus, I have supportive genome-scale evidence to support IAsα, but as yet, there is no conclusive evidence to support a second GD in *Acropora*. In addition, the distribution shapes were quite different in Figure 4.4, one possibility is that the *A. tenuis* has better gene model compared to *A. gemmifera* and *A. echinata*.

Table 4.6. Likelihood of multiple GDs hypotheses in *Acropora* using GDgc method with gene counts data

GD 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 4.7. Likelihood of different times of GD under one GD event in *Acropora* using GDgc

Time of GD	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

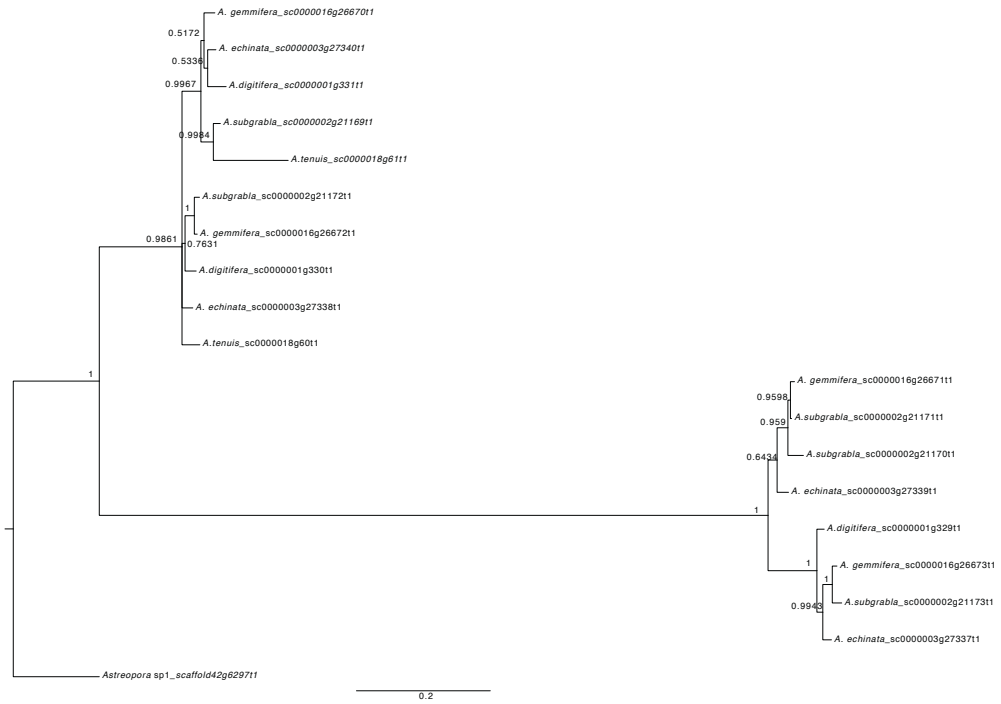


Figure 4.17. **Phylogeny of orthogroup 434 (somatostatin receptor type 5-like) shows duplicates are under two GD topology.** The phylogeny was reconstructed using MrBayes, and Bayesian posterior probabilities are shown at each node.

It is crucial to accurately estimate the timing of a GD event to understand its evolutionary consequences (Jiao et al., 2011; Vanneste et al., 2014). The study has clearly estimated the timing of IAsα using both phylogenomic analysis and the dS-based method. I suggest that IAsα probably occurred between 28 and 36 Mya (Figure 4.11). 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 from 15.97 to 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 GD may enable organisms to escape extinction during drastic environmental changes (Van De Peer et al., 2017) (Figure 4.11).

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). I performed GO analysis on duplicated genes and examined several duplicated gene families, showing that duplicated genes following by 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 GD. In addition, I 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 probably because these proteins may have developed different functions via neofunctionalization, such as excitatory amino acid transporters (orthogroups 1244). However, there is still much work needed to investigate molecular mechanisms of duplicated genes to examine these hypotheses in the diversification of *Acropora* (Yasuoka et al., 2016), especially, more functional analyses are needed for putative subfunctionalization and neofunctionalization of duplicated genes. 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).

The previous study proposed that adaptive radiation in *Acropora* was probably driven by introgression (Mao et al., 2018); thus, *Acropora* is the first invertebrates lineage reported to have undergone both GD and introgression. Meanwhile, both introgression and GD have also been reported in cichlid fish lineages (Berner and Salzburger, 2015), a famous model for adaptive radiation in vertebrates (Berner and

Salzburger, 2015; Seehausen et al., 2014). Both GD and introgression are regarded as significant forces in adaptive radiation of organisms (Berner and Salzburger, 2015; Van De Peer et al., 2017), but I still do not understand the relationship between GD and introgression in adaptive radiations (Soltis and Soltis, 2009).

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

Chapter 5

Conclusions and Limitations of this dissertation

A major goal of evolutionary biology is to understand the processes leading to speciation and diversification, and myriad paths have led to diversification in different group organisms (Helfman et al., 2009; Nosil et al., 2017; Schluter, 2000; Schluter and Pennell, 2017; Weber et al., 2017). In particular, introgression and genome duplication (GD) are regarded as important evolutionary forces on speciation and diversification (Meier et al., 2017b; Meyer et al., 2016; Van De Peer et al., 2017; Wagner et al., 2012).

Reef building corals provide the structural basis for one of Earth's most spectacular and diverse—but increasingly threatened—ecosystems (Bhattacharya et al., 2016; Wallace and Rosen, 2006). Modern Indo-Pacific reefs are dominated by species of the staghorn coral genus *Acropora* (Anthozoa: Acroporidae), one of most diverse genera with close to 150 species, but the evolutionary and ecological factors associated with their diversification and rise to dominance are unclear. Hence, in my dissertation, I analyze the genomes of one *Astreopora*, sister genus of *Acropora*, and five species of *Acropora* to examine the roles of introgression, GD and ecological opportunity in the diversification and the rise to dominance of *Acropora*.

5.1 Introgression and gene flow in *Acropora*

I found strong evidence for a history marked by a major introgression event and introgression genes are evolving faster than others, consistent with a role for introgression in spreading adaptive genetic variations with phylogenomic and comparative genomics approaches.

Although I have shown a major introgression event in corals, it is not easy to examine the timing of the introgression event. In addition, we still have less knowledge of what the diversification rates are in *Acropora*. Namely, it is interesting to investigate the relationship between introgression and diversification rates in *Acropora*. Due to limitation of sampling, it is not easy to determine the permitted hybrids in this study. With more sampling and clear geographic distributions of *Acropora* species, it would be better to find hybrid zones or determine hybrids.

Moreover, the evolutionary rate analysis showed that the non-species tree genes evolved faster than species tree genes in the species involved in the major introgression event as well as the selection occurred before the introgression. Yet, it is still unclear what the mechanisms for this pattern are and whether it is "true" for all organisms under introgression. Besides, the new technologies (e.g. Crisps-Cas9) have been applied into *Acropora* embryo study (Cleves et al., 2018), it becomes possible to explore the functional roles of introgression (adaptive introgression) in the evolution of corals and will help us to understand coral conservation.

5.2 Ancient GD shared by *Acropora*

I used one *Astreopora* genome as outgroup along with five *Acropora* genomes to elucidate that one ancient GD event shared by *Acropora* occurred around 27.9 to 35.7 Million years ago (Mya) potentially in correlation with the Oligocene-Miocene transition of corals using comprehensive phylogenomic and dS-based approaches. I also found that duplicated genes, originating from the ancient GD, were under complicated fates and highly enriched in molecular functions of gene regulation important to the diversification of *Acropora*. This study, reporting the first GD event

in corals, provides new insights into the evolution of reef-building corals as well as expands a new empirical model for polyploidy 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 rare evolutionary events in the animals. With advanced increasing of genomic data, we observed more and more GD in the animals (Van De Peer et al., 2017), such as vertebrates (Berthelot et al., 2014; Dehal and Boore, 2005; Kenny et al., 2017), insects (Li et al., 2018), and corals (this study). Yet, it is hard to distinguish the large-scale gene/genome duplication from entire genome duplication using the dS-based method, phylogenomic and synteny analysis without precise genomic data. For example, the second round WGD in vertebrates was a large-scale genome duplication rather than an entire genome duplication (Smith and Keinath, 2015). Hence, in this dissertation, I defined the GD as large-scale gene/genome duplication. The evidence from different analysis support the GD occurred in the common ancestor of *Acropora*, but it still lacks enough evidence to support the GD is generated by entire genome duplication. Even so, it is still unclear that this duplication is from autopolyploidy or allopolyploidy.

5.3 Climate change facilitated the rise to dominance of *Acropora*

I found that *Acropora* lineages profited from climate-driven mass extinctions in the Plio-Pleistocene with demographic inferences, indicating that *Acropora* exploited ecological opportunity opened by a new climatic regime favoring species that could cope with rapid sea-level changes.

The effective population size simulations highly support the hypothesis that mass extinction provides the ecological opportunity for *Acropora*. Yet, it is worth to

mention that this hypothesis is still needed more evidence to support. In addition, it is also interesting to investigate whether the glacial cycles facilitated introgression/gene flows in corals (Montaggioni and Braithwaite, 2009). In other words, I am curious if the oscillatory change of sea-level with glacial cycles is a factor to generate the chance for coral population re-connections. Moreover, due to limitation of sample size in Acropora, there is a possibility that extra genome duplications might can not be detected on specific lineages.

5.4 Future directions

With advancements in sequencing technologies, bioinformatics and molecular biology, it is a perfect time for us to study large-scale phylogeography of *Acropora* and to study molecular mechanisms of adaptive introgression and to study functions of duplicated genes in "evo-devo" perspectives. In the short-term goal, it would be a good idea to collect more *Acropora* species samples around the world cooperating with other coral researchers for investigating the origination and diversification rate of the whole genus using RNA sequencing or DNA-Barcoding. In the long-term goal, it would be interesting to identify the functions of genes, which present a pairwise correlation of their expression across different developmental stages, as well as to investigate the functions of the duplicated non-coding elements.

In all, my work gives a big picture on coral evolution while addressing open questions in general evolutionary theory. The dissertation raises a number of questions and avenues for future work, while also providing relevant historical context to understanding the current and future challenges to coral reefs, a topic of major concern to scientists and the general public.

Appendix

Genome assembly and annotation statistics of the six coral species

Table A.1Raw data and coverage calculation

Species	Pair-end libraries	Total Sequences	Read length (bp)	Total data	In Total	Coverage
A. echinata	Paired-End (Illumina)	97,853,562	290	28,377,532,980	59,150,230,780	144
	Mate Pair (Illumina)	118,356,530	260	30,772,697,800		
A. digitifera	Paired-End (Illumina)	297,802,374	290	86,362,688,460	127,063,684,380	301
	Mate Pair (Illumina)	156,542,292	260	40,700,995,920		
A. gemmifera	Paired-End (Illumina)	97,047,284	290	28,143,712,360	63,708,948,560	157
	Mate Pair (Illumina)	136,789,370	260	35,565,236,200		
A. subglabra	Paired-End (Illumina)	91,677,722	290	26,586,539,380	63,941,389,380	148
	Mate Pair (Illumina)	143,672,500	260	37,354,850,000		
A. tenuis	Paired-End (Illumina)	543,347,386	120	65,201,686,320	77,510,501,080	190
	Mate Pair (Illumina)	111,898,316	110	12,308,814,760		
Astreopora sp1	Paired-End (Illumina)	131634697	290	38,174,062,130	72,357,203,710	154
	Mate Pair (Illumina)	117872902	290	34,183,141,580		

Table A. 2 Genome statistics and annotation

Species		*A. digitifera*	*A. echinata*	*A. gemmifera*	*A. subglabra*	*A. tenuis*	*Astreopora* sp1
Genome (Mb)	Repetitive DNA (%)	30.43	34.09	32.98	31.92	34.58	36.9
	N50 (Mb)	1.81	1.39	1.14	1.09	1.16	0.674
	L50	63	84	103	110	103	176
	GC content	38.93	38.95	38.93	38.91	38.93	40.63
	Gap (%)	8.8	15.27	9.75	13.43	7.51	8.2
	Reads Coverage	309	144	158	145	188	154
	Assembled size (Mb)	422	411	407	432	408	468
Gene	Gene Number	28,958	28,280	30,776	30,922	26,445	40,430
	Average gene length (bp)	1,330	1,585	1,321	1,306	1,569	1,254

Bibliography

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.

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.

Árnason, Ú., Lammers, F., Kumar, V., Nilsson, M.A., and Janke, A. (2018). Whole-genome sequencing of the blue whale and other rorquals finds signatures for introgressive gene flow. *Science advances* 4, eaap9873.

Baird, A., Sadler, C., and Pitt, M. (2001). Synchronous spawning of Acropora in the Solomon Islands. *Coral Reefs* 19, 286-286.

Bak, R.P.M. (1983). Neoplasia, regeneration and growth in the reef-building coral Acropora palmata. *Marine Biology* 77, 221-227.

Barshis, D.J., Ladner, J.T., Oliver, T.A., Seneca, F.O., Traylor-Knowles, N., and Palumbi, S.R. (2013). Genomic basis for coral resilience to climate change. *Proceedings of the National Academy of Sciences* 110, 1387-1392.

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

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.

Boratyn, G.M., Camacho, C., Cooper, P.S., Coulouris, G., Fong, A., Ma, N., Madden, T.L., Matten, W.T., McGinnis, S.D., Merezhuk, Y., et al. (2013). BLAST: a

more efficient report with usability improvements. *Nucleic acids research* 41, W29-W33.

Bouckaert, R., Heled, J., Kühnert, 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.

Cleves, P.A., Strader, M.E., Bay, L.K., Pringle, J.R., and Matz, M.V. (2018). CRISPR/Cas9-mediated genome editing in a reef-building coral. *Proceedings of the National Academy of Sciences* 115, 5235-5240.

Cui, R., Schumer, M., Kruesi, K., Walter, R., Andolfatto, P., and Rosenthal, G.G. (2013). Phylogenomics reveals extensive reticulate evolution in Xiphophorus fishes. *Evolution* 67, 2166-2179.

Darling, E.S., Alvarez-Filip, L., Oliver, T.A., McClanahan, T.R., and Cote, I.M. (2012). Evaluating life-history strategies of reef corals from species traits. *Ecol Lett* 15, 1378-1386.

Durand, E.Y., Patterson, N., Reich, D., and Slatkin, M. (2011). Testing for ancient admixture between closely related populations. *Molecular biology and evolution* 28, 2239-2252.

Elderfield, H., Ferretti, P., Greaves, M., Crowhurst, S., McCave, I.N., Hodell, D., and Piotrowski, A.M. (2012). Evolution of Ocean Temperature and Ice Volume Through the Mid-Pleistocene Climate Transition. *Science* 337, 704-709.

Faith, D.P., and Richards, Z.T. (2012). Climate change impacts on the tree of life: changes in phylogenetic diversity illustrated for Acropora corals. *Biology* 1, 906-932.

Foote, A.D., Vijay, N., Ávila-Arcos, M.C., Baird, R.W., Durban, J.W., Fumagalli, M., Gibbs, R.A., Hanson, M.B., Korneliussen, T.S., and Martin, M.D. (2016). Genome-culture coevolution promotes rapid divergence of killer whale ecotypes. *Nature communications* 7, 11693.

Fukami, H., Chen, C.A., Budd, A.F., Collins, A., Wallace, C., Chuang, Y.-Y., Chen, C., Dai, C.-F., Iwao, K., and Sheppard, C. (2008). Mitochondrial and nuclear genes suggest that stony corals are monophyletic but most families of stony corals are not (Order Scleractinia, Class Anthozoa, Phylum Cnidaria). *PloS one* 3, e3222.

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

Getty, S.R., Asmerom, Y., Quinn, T.M., and Budd, A.F. (2001). Accelerated Pleistocene coral extinctions in the Caribbean Basin shown by uranium-lead (U-Pb) dating. *Geology* 29, 639-642.

Goreau, T.F., and Goreau, N.I. (1959). The physiology of skeleton formation in corals. II. Calcium deposition by hermatypic corals under various conditions in the reef. *The Biological Bulletin* 117, 239-250.

Veron, J.E.N (1995). Corals in Space and Time the Biogeography and Evolution of the Scleractinia. 269, 1893-1894.

Harris, R.S. (2007). Improved pairwise alignment of genomic DNA. (The Pennsylvania State University).

Hawks, J. (2017). Introgression Makes Waves in Inferred Histories of Effective Population Size. *Hum Biol* 89, 67-80.

Helfman, G., Collette, B.B., Facey, D.E., and Bowen, B.W. (2009). The diversity of fishes: biology, evolution, and ecology. (John Wiley & Sons).

Heliconius Genome, C. (2012). Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature* 487, 94-98.

Hemond, E.M., and Vollmer, S.V. (2010). Genetic Diversity and Connectivity in the Threatened Staghorn Coral (Acropora cervicornis) in Florida. *PloS one* 5.

Herbert, T.D., Peterson, L.C., Lawrence, K.T., and Liu, Z.H. (2010). Tropical Ocean Temperatures Over the Past 3.5 Million Years. *Science* 328, 1530-1534.

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

Hughes, T.P., Kerry, J.T., Alvarez-Noriega, M., Alvarez-Romero, J.G., Anderson, K.D., Baird, A.H., Babcock, R.C., Beger, M., Bellwood, D.R., Berkelmans, R., et al. (2017). Global warming and recurrent mass bleaching of corals. *Nature* 543, 373-+.

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.

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

Kittel, T.G.F. (2013). The Vulnerability of Biodiversity to Rapid Climate Change. *Climate Vulnerability: Understanding and Addressing Threats to Essential Resources; Pielke, RA, Ed*, 185-201.

Kojis, B.L. (1986). Sexual reproduction in Acropora (Isopora)(Coelenterata: Scleractinia). *Marine Biology* 91, 311-318.

Korneliussen, T.S., Albrechtsen, A., and Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. *BMC bioinformatics* 15, 356.

Lamichhaney, S., Berglund, J., Almen, M.S., Maqbool, K., Grabherr, M., Martinez-Barrio, A., Promerova, M., Rubin, C.J., Wang, C., Zamani, N., et al. (2015). Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature* 518, 371-375.

Larget, B.R., Kotha, S.K., Dewey, C.N., and Ane, C. (2010). BUCKy: gene tree/species tree reconciliation with Bayesian concordance analysis. *Bioinformatics* 26, 2910-2911.

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

Li, H., and Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature* 475, 493-496.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078-2079.

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

Lin, S., Cheng, S., Song, B., Zhong, X., Lin, X., Li, W., Li, L., Zhang, Y., Zhang, H., and Ji, Z. (2015). The Symbiodinium kawagutii genome illuminates dinoflagellate gene expression and coral symbiosis. *Science* 350, 691-694.

Liu, S.-Y.V., Chan, C.-L.C., Hsieh, H.J., Fontana, S., Wallace, C.C., and Chen, C.A. (2015). Massively parallel sequencing (MPS) assays for sequencing

mitochondrial genomes: the phylogenomic implications for Acropora staghorn corals (Scleractinia; Acroporidae). *Marine Biology* 162, 1383-1392.

Losos, J.B. (2010). Adaptive Radiation, Ecological Opportunity, and Evolutionary Determinism. *Am Nat* 175, 623-639.

Mailund, T., Halager, A.E., Westergaard, M., Dutheil, J.Y., Munch, K., Andersen, L.N., Lunter, G., Prufer, K., Scally, A., Hobolth, A., et al. (2012). A New Isolation with Migration Model along Complete Genomes Infers Very Different Divergence Processes among Closely Related Great Ape Species. *PLoS genetics* 8.

Márquez, L.M., Miller, D.J., MacKenzie, J.B., and van Oppen, M.J.H. (2003). Pseudogenes contribute to the extreme diversity of nuclear ribosomal DNA in the hard coral Acropora. *Molecular biology and evolution* 20, 1077-1086.

Márquez, L.M., Van Oppen, M.J.H., Willis, B.L., Reyes, A., and Miller, D.J. (2002). The highly cross‐fertile coral species, Acropora hyacinthus and Acropora cytherea, constitute statistically distinguishable lineages. Molecular ecology 11, 1339-1349.

Marshall, A.T., Clode, P.L., Russell, R., Prince, K., and Stern, R. (2007). Electron and ion microprobe analysis of calcium distribution and transport in coral tissues. *Journal of Experimental Biology* 210, 2453-2463.

Matz, M.V., Treml, E.A., Aglyamova, G.V., van Oppen, M.J.H., and Bay, L.K. (2017). Potential for rapid genetic adaptation to warming in a Great Barrier Reef coral. *bioRxiv*.

Mazet, O., Rodriguez, W., and Chikhi, L. (2015). Demographic inference using genetic data from a single individual: Separating population size variation from population structure. *Theor Popul Biol* 104, 46-58.

Mazet, O., Rodriguez, W., Grusea, S., Boitard, S., and Chikhi, L. (2016). On the importance of being structured: instantaneous coalescence rates and human evolution-lessons for ancestral population size inference? *Heredity* 116, 362-371.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., and Daly, M. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research* 20, 1297-1303.

Meier, J.I., Marques, D.A., Mwaiko, S., Wagner, C.E., Excoffier, L., and Seehausen, O. (2017). Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nature communications* 8, 14363.

Metzker, M.L. (2010). Sequencing technologies—the next generation. *Nature reviews genetics* 11, 31.

Meyer, B.S., Matschiner, M., and Salzburger, W. (2016). Disentangling Incomplete Lineage Sorting and Introgression to Refine Species-Tree Estimates for Lake Tanganyika Cichlid Fishes. *Systematic biology*.

Montaggioni, L.F., and Braithwaite, C.J.R. (2009). Quaternary coral reef systems: history, development processes and controlling factors. (Elsevier).

Neale, D.B., Martínez-García, P.J., De La Torre, A.R., Montanari, S., and Wei, X.-X. (2017). Novel insights into tree biology and genome evolution as revealed through genomics. *Annu Rev Plant Biol* 68, 457-483.

Nosil, P., Feder, J.L., Flaxman, S.M., and Gompert, Z. (2017). Tipping points in the dynamics of speciation. *Nature Ecology & Evolution* 1.

O'dea, A., Jackson, J.B.C., Fortunato, H., Smith, J.T., D'Croz, L., Johnson, K.G., and Todd, J.A. (2007). Environmental change preceded Caribbean extinction by 2 million years. *Proc. Natl. Acad. Sci* 104, 5501-5506.

Ohta, T. (1992). The nearly neutral theory of molecular evolution. *Annual Review of Ecology and Systematics* 23, 263-286.

Pimiento, C., Griffin, J.N., Clements, C.F., Silvestro, D., Varela, S., Uhen, M.D., and Jaramillo, C. (2017). The Pliocene marine megafauna extinction and its impact on functional diversity. *Nature Ecology & Evolution*, 1.

Prada, C., Hanna, B., Budd, A.F., Woodley, C.M., Schmutz, J., Grimwood, J., Iglesias-Prieto, R., Pandolfi, J.M., Levitan, D., and Johnson, K.G. (2016). Empty Niches after Extinctions Increase Population Sizes of Modern Corals. *Current Biology* 26, 3190-3194.

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? *Science advances* 2, e1500850.

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.

Rohling, E.J., Foster, G.L., Grant, K.M., Marino, G., Roberts, A.P., Tamisiea, M.E., and Williams, F. (2014a). Sea-level and deep-sea-temperature variability over the past 5.3 million years. *Nature* 508, 477-+.

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.

Rosser, N.L., Thomas, L., Stankowski, S., Richards, Z.T., Kennington, W.J., and Johnson, M.S. (2017). Phylogenomics provides new insight into evolutionary relationships and genealogical discordance in the reef-building coral genus Acropora. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 284.

Schluter, D. (2000). The ecology of adaptive radiation. (OUP Oxford).

Schluter, D., and Pennell, M.W. (2017). Speciation gradients and the distribution of biodiversity. *Nature* 546, 48.

Seehausen, O. (2004). Hybridization and adaptive radiation. *Trends in ecology & evolution* 19, 198-207.

Seehausen, O. (2015). Process and pattern in cichlid radiations - inferences for understanding unusually high rates of evolutionary diversification. *New Phytol* 207, 304-312.

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. *Nature reviews Genetics* 15, 176-192.

Sheppard, C., Davy, S., Pilling, G., and Graham, N. (2017). The biology of coral reefs. (Oxford University Press).

Shinzato, C., Mungpakdee, S., Arakaki, N., and Satoh, N. (2015). Genome-wide SNP analysis explains coral diversity and recovery in the Ryukyu Archipelago. *Scientific reports* 5.

Shinzato, C., Shoguchi, E., Kawashima, T., Hamada, M., Hisata, K., Tanaka, M., Fujie, M., Fujiwara, M., Koyanagi, R., Ikuta, T., et al. (2011). Using the Acropora digitifera genome to understand coral responses to environmental change. *Nature* 476, 320-323.

Shinzato, C., Yasuoka, Y., Mungpakdee, S., Arakaki, N., Fujie, M., Nakajima, Y., and Satoh, N. (2014). Development of novel, cross-species microsatellite markers for Acropora corals using next-generation sequencing technology. *Frontiers in Marine Science* 1, 11.

Shoguchi, E., Shinzato, C., Kawashima, T., Gyoja, F., Mungpakdee, S., Koyanagi, R., Takeuchi, T., Hisata, K., Tanaka, M., Fujiwara, M., et al. (2013). Draft Assembly of the Symbiodinium minutum Nuclear Genome Reveals Dinoflagellate Gene Structure. *Current Biology* 23, 1399-1408.

Simakov, O., Kawashima, T., Marletaz, F., Jenkins, J., Koyanagi, R., Mitros, T., Hisata, K., Bredeson, J., Shoguchi, E., Gyoja, F., et al. (2015). Hemichordate genomes and deuterostome origins. *Nature* 527, 459-+.

Solis-Lemus, C., and Ane, C. (2016). Inferring Phylogenetic Networks with Maximum Pseudolikelihood under Incomplete Lineage Sorting. *PLoS genetics* 12, e1005896.

Solís-Lemus, C., and Ané, C. (2016). Inferring phylogenetic networks with maximum pseudolikelihood under incomplete lineage sorting. *PLoS genetics* 12, e1005896.

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

Stroud, J.T., and Losos, J.B. (2016). Ecological Opportunity and Adaptive Radiation. *Annual Review of Ecology, Evolution, and Systematics*, Vol 47 47, 507-532.

Talluto, M.V., Boulangeat, I., Vissault, S., Thuiller, W., and Gravel, D. (2017). Extinction debt and colonization credit delay range shifts of eastern North American trees. *Nature Ecology & Evolution* 1, 0182.

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

Thomas, C.D., Cameron, A., Green, R.E., Bakkenes, M., Beaumont, L.J., Collingham, Y.C., Erasmus, B.F.N., de Siqueira, M.F., Grainger, A., Hannah, L., et al. (2004). Extinction risk from climate change. *Nature* 427, 145-148.

Van de Peer, Y., Maere, S., and Meyer, A. (2009). OPINION The evolutionary significance of ancient genome duplications. *Nature Reviews Genetics* 10, 725-732.

Van De Peer, Y., Mizrachi, E., and Marchal, K. (2017). The evolutionary significance of polyploidy. *Nature Reviews Genetics* 18, 411-424.

van Oppen, M.J., Catmull, J., McDonald, B.J., Hislop, N.R., Hagerman, P.J., and Miller, D.J. (2002). The mitochondrial genome of Acropora tenuis (Cnidaria; Scleractinia) contains a large group I intron and a candidate control region. *Journal of molecular evolution* 55, 1-13.

van Oppen, M.J., and Gates, R.D. (2006). Conservation genetics and the resilience of reef-building corals. *Molecular ecology* 15, 3863-3883.

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? *Molecular biology and evolution* 18, 1315-1329.

Vanneste, K., Maere, S., and Van de Peer, Y. (2014). Tangled up in two: a burst of genome duplications at the end of the Cretaceous and the consequences for

plant evolution. *Philosophical Transactions of the Royal Society of London B: Biological Sciences,* 369.

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

Wagner, C.E., Harmon, L.J., and Seehausen, O. (2012). Ecological opportunity and sexual selection together predict adaptive radiation. *Nature* 487, 366-U124.

Wallace, C. (1999). Staghorn corals of the world: a revision of the genus Acropora. (CSIRO publishing).

Wallace, C.C. (2011). Acropora. In Encyclopedia of Modern Coral Reefs (Springer), pp. 3-9.

Wallace, C.C. (2012). Acroporidae of the Caribbean. *Geol Belg* 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. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 273, 975-982.

Weber, M.G., Wagner, C.E., Best, R.J., Harmon, L.J., and Matthews, B. (2017). Evolution in a community context: on integrating ecological interactions and macroevolution. *Trends in ecology & evolution* 32, 291-304.

Wei, N.-W.V., Wallace, C.C., Dai, C.-F., Pillay, K.R.M., and Chen, C.A. (2006). Analyses of the ribosomal internal transcribed spacers (ITS) and the 5.8 S gene indicate that extremely high rDNA heterogeneity is a unique feature in the scleractinian coral genus Acropora (Scleractinia; Acroporidae). *Zoological Studies* 45, 404-418.

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* S 37, 489-517.

Woodley, C.M., Downs, C.A., Bruckner, A.W., Porter, J.W., and Galloway, S.B. (2016). Diseases of coral. (John Wiley & Sons).

Work, T.M., Aeby, G.S., and Coies, S.L. (2008). Distribution and morphology of growth anomalies in Acropora from the Indo-Pacific. *Diseases of aquatic organisms* 78, 255-264.

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

Yu, Y., Dong, J., Liu, K.J., and Nakhleh, L. (2014). Maximum likelihood inference of reticulate evolutionary histories. *Proceedings of the National Academy of Sciences* 111, 16448-16453.

Yu, Y., and Nakhleh, L. (2015). A maximum pseudo-likelihood approach for phylogenetic networks. *BMC genomics* 16 Suppl 10, S10.

Zayasu, Y. and Shinzato, C. (2016). Hope for coral reef rehabilitation: massive synchronous spawning by outplanted corals in Okinawa, Japan. *Coral Reefs*, 35(4), pp.1295-1295.

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

Zhao, S.C., Zheng, P.P., Dong, S.S., Zhan, X.J., Wu, Q., Guo, X.S., Hu, Y.B., He, W.M., Zhang, S.N., Fan, W., et al. (2013). Whole-genome sequencing of giant pandas provides insights into demographic history and local adaptation. *Nature genetics* 45, 67-U99.