

# The genetic basis of fluorescence in the stony coral, *Acropora digitifera*

Shiho Kariyazono

Doctor of Philosophy

Department of Evolutionary Studies of Biosystems

School of Advanced Sciences

SOKENDAI (The Graduated University for Advanced Studies)

2017

# Summary

Coral reefs harbor the most biologically diverse ecosystems in warm shallow water. Coral reefs are the calcium carbonate structures built by mainly stony corals (Scleractinia), and corals belong to the genus *Acropora* are major components of coral reef in the Indo-Pacific. The algal symbionts in stony corals allow corals to build extensive structures in oligotrophic environments. However, the symbiotic relationship between corals and algal symbionts is sensitive to climate changes and local stressors. One of the candidate molecules associated with stresses of corals is fluorescent protein (FP). Many biological roles of FPs including stress response such as photo-protection had been proposed.

When I started my doctoral thesis, the fluorescence of corals and its genetic determinants, *FP* genes have been one of the enigmatic problems in corals. Coral FPs have been studied since the discovery of the homologs of jellyfish GFP (green fluorescent protein) in corals, but the biological role remained unclear. One of the main factors making the difficulty to analyze the role of FPs was missing information of the genetic basis of FPs in stony corals.

In chapter 2, I described the multi-*FP* gene family in *Acropora digitifera*. In previous studies, many *FP* genes were cloned from anthozoan species and it was suggested that multi-copies of these genes are present in their genomes. However, the full complement of *FP* genes in any single coral species remained unidentified. In chapter 2, I analyzed the *FP* genes in two stony coral species, *A. digitifera* and *A. tenuis*. *FP* cDNA sequences from the two species revealed the presence of a multi-gene family with an unexpectedly large number of genes, separated into short-to-middle-wavelength emission (S/MWE), middle-to-long-wavelength emission (M/LWE), and chromoprotein (CP) clades. *FP* gene copy numbers in the genomes of

four *A. digitifera* colonies were estimated as 16–22 in the S/MWE, 3–6 in the M/LWE, and 8–12 in the CP clades, and, in total, 35, 31, 33, and 33 *FP* gene copies per individual shown by quantitative PCR. These are the largest sets of *FP* genes per genome among organisms that the *FP* genes were isolated in their genomes. The fluorescent light produced by recombinant protein products encoded by the newly isolated genes explained the fluorescent range of live *A. digitifera*, suggesting that the high copy multi-*FP* gene family generates coral fluorescence. The functionally diverse multi-*FP* gene family must have existed in the ancestor of *Acropora* species, as suggested by molecular phylogenetic and evolutionary analyses. The persistence of a diverse function and high copy number multi-*FP* gene family may indicate the biological importance of diverse fluorescence emission and light absorption in *Acropora* species.

In chapter 3, I reported the genetic mechanism underlying fluorescent polymorphisms in *A. digitifera*. Despite the biological roles of FPs have been proposed based on natural variation in fluorescence in corals, their roles and the genetic basis of FP-mediated color polymorphisms in *Acropora* remain unclear. Using a high-throughput sequencing approach, I found that *FP* gene sequences in the multi-gene family exhibit presence–absence polymorphism among individuals. A few particular sequences in S/MWE and M/LWE clades were highly expressed in adults, and different sequences were highly expressed in larvae. These highly expressed sequences were absent in the genomes of individuals with low total *FP* gene expression. In adults, presence–absence differences of the highly expressed *FP* sequences in the genome were consistent with measurements of emission spectra of corals, suggesting that presence–absence polymorphisms of these *FP* sequences contribute to the fluorescent polymorphisms. The functions of recombinant FPs encoded by highly expressed sequences in

adult and larval stages were different, suggesting that expression of *FP* sequences with different functions may depend on the life-stage of *A. digitifera*. The highly expressed *FP* sequences exhibited presence–absence polymorphisms in subpopulations of *A. digitifera*, suggesting that the presence–absence status was maintained during the evolution of *A. digitifera* subpopulations.

In chapter 2 and 3, I documented the genetic basis of fluorescence and fluorescent polymorphisms in *A. digitifera*. The difference in FPs between adults and larvae, and the polymorphisms of highly expressed *FP* sequences may provide key insight into the biological roles of FPs in corals.

In addition, I reported genetic variation and expression differences among individuals of *A. digitifera* in chapter 4. Despite the importance of characterizing genetic variation among coral individuals for the conservation of coral reefs, the correlation between genetic and functional variation is still poorly understood. In this chapter 4, I detected a high frequency of genes showing presence–absence polymorphisms (PAPs) for single-copy genes in *Acropora digitifera*. Among single-copy genes, some proportion exhibited PAPs, including transposable element (TE)-related genes. Among genes showing PAPs (PAP genes), roughly half were expressed in adults and/or larvae, and the PAP status was associated with differential expression among individuals. Although most of PAP genes were uncharacterized or had ambiguous annotations, these genes were specifically distributed in cnidarian lineages, suggesting that these genes have functional roles related to traits specific to cnidarians or the family Acroporidae. Some proportion of *A. digitifera* PAP genes were also PAPs in *A. tenuis*, the basal lineage in the genus *Acropora*, indicating that PAPs were shared among species in

*Acropora*. Expression differences caused by a high frequency of PAP genes may be a novel genomic feature in the genus *Acropora*.

In this thesis, I documented the genetic mechanisms of coral fluorescence, fluorescent polymorphisms, and gene expression differences among individuals of *A. digitifera*. These findings will contribute to coral conservation by improving our understanding of the genetic basis of individual differences in stony corals.

# Contents

<b>Chapter 1</b> .....	6
General Introduction	
<b>Chapter 2</b> .....	18
Fluorescent protein gene family has persisted during the evolution of <i>Acropora</i> species	
<b>Chapter 3</b> .....	78
Presence-absence polymorphisms of highly expressed <i>FP</i> sequences	
<b>Chapter 4</b> .....	161
Presence–absence polymorphisms of single-copy genes	
<b>Chapter 5</b> .....	218
Perspectives	
<b>Acknowledgements</b> .....	221

# **Chapter 1**

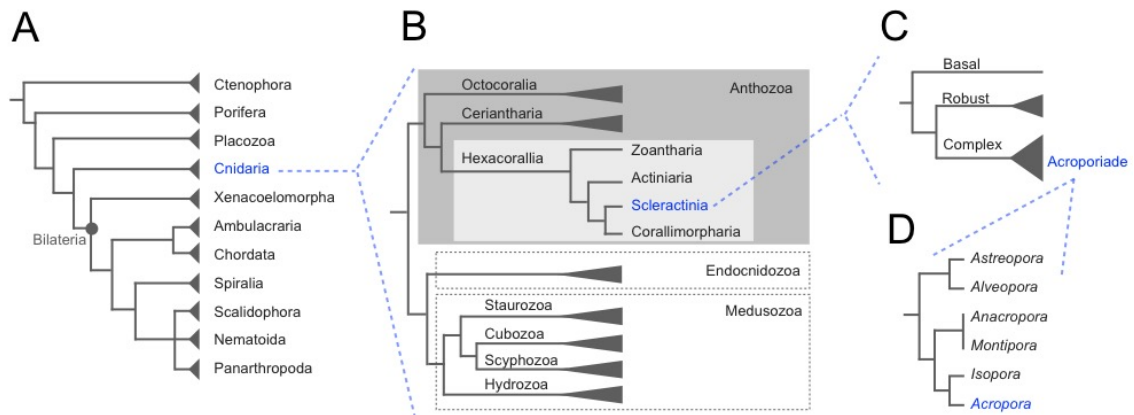
## **General introduction**

***Taxonomy and evolution of stony corals***

The stony corals belong to the order Scleractinia, are an essential group in reef-building corals. Scleractinians belong to the phylum Cnidaria that is the sister group to the Bilateria including most animals such as vertebrates (Figure 1.1A) (Dunn, et al. 2014). The phylum Cnidaria is a group of marine and fresh water invertebrates composed of over 10,000 species (Appeltans, et al. 2012). The specialized cellular structures termed “cnidae” define the phylum Cnidaria. Cnidae comprise coiled tubule and contain venom against prey, predators, and competitors (Jouiaei, et al. 2015). It has been accepted that Cnidaria comprises two groups, Medusozoa including Scyphozoa (true jellyfish); Cubozoa (box jellies); Staurozoa (stalked jellyfish); and Hydrozoa (e.g., hydroids), and Anthozoa (e.g., corals) (Zapata, et al. 2015). However, a recent study supported that Endocnidozoa including microscopic parasitic species of Myxozoa and Polypodiozoa was also included in Cnidaria (Chang, et al. 2015). Cnidarian phylogeny remains under discussion, because inconsistent results were supported by phylogenetic analyses of mitochondrial DNAs (paraphyly of Anthozoa and Scyphozoa) (Kayal, et al. 2013) and nuclear genomes (monophyly of Anthozoa and Scyphozoa) (Zapata, et al. 2015). A recent study supports that a monophyletic Anthozoa clade forms a sister group with a clade containing Medusozoa and Endocnidozoa (Figure 1.1B) (Kayal, et al. 2017). One of the subclasses in Anthozoa, Hexacorallia comprises four orders, and Scleractinia (stony corals) is one of these orders (Kayal, et al. 2017).

In the fossil record, Scleractinian corals appeared at the Triassic period about 237 million years ago (Stanley and Fautin 2001; Stanley Jr 2003). Recent studies indicate the order Scleractinia may be a monophyletic group and this group is separated into “basal”, “robust”,

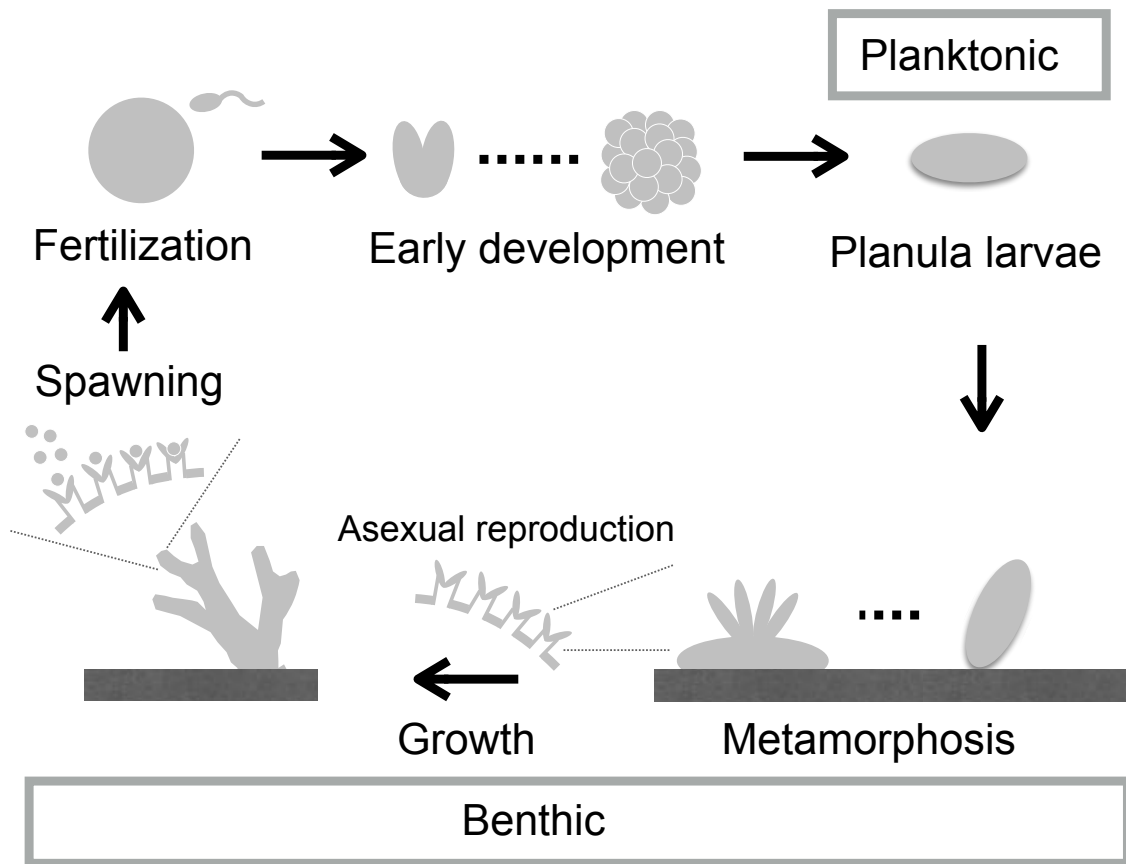
and “complex” clades. The family Acroporidae belongs to complex clade (Figure 1.1C) (Kitahara, et al. 2016). Among the genera in the family Acroporidae (Figure 1.1D), the genus *Acropora* comprises of 134 recognized species (Richards, et al. 2013), and is major components of coral reefs in the Indo-Pacific (Wallace 1999).



**Figure 1.1.** Phylogenetic position of the phylum Cnidaria modified from Dunn *et al.*, 2014 (A), the order Scleractinia modified from Kayal *et al.*, 2017 (B), the family Acroporidae modified from Kitahara *et al.*, 2014 (C), and the genus *Acropora* modified from Kitahara *et al.*, 2016 (D).

***Life cycle of corals***

Most Scleractinia corals are hermaphrodites and proliferate by both sexual- and asexual-reproductions. In sexual reproduction, fertilized gametes become planula larvae after early development, thus Scleractinia corals have two different phases; planktonic larval phase and benthic adult phase in their life cycle (Figure 1.2). Planktonic larvae facilitate dispersal and generate new populations (Harrison 2011). To produce larvae, corals use two reproductive strategies; spawning and brooding. In spawning, gametes are released into the ocean and fertilized, and majority of the species in Scleractinia corals use this strategy. Synchronized spawning within a species to fertilized gametes effectively is known in spawning species. In brooding, after planula larvae are produced by internal fertilization, they are released into the ocean (Shikina and Chang 2016). In asexual reproduction, genetically identical clones are generated by budding of polyps, colony fragmentation, and polyp expulsion that is referred to as “polyp bail-out” (Shikina and Chang 2016).



**Figure 1.2.** The general life cycle of corals

***Symbiosis in stony corals***

Why are the stony corals able to build the extensive calcium carbonate structures that are the basis of coral reefs in an oligotrophic environment? Most of Scleractinia corals depend about 90% of the energy on photosynthetic products that is provided by dinoflagellate symbionts in the genus *Symbiodinium* (Muscatine, et al. 1981; Tremblay, et al. 2012). The breakdown of the symbiotic relationship between corals and their algal symbionts results in pale color of corals that is described as “bleached”, and often leads corals to die (Spalding and Brown 2015). The latitudinal coral distribution is restricted by the amount of photosynthetically available solar radiation over winter (Muir, et al. 2015). In corals belonged to the genus *Acropora*, the maximum latitudinal limit is around 30°, and the maximum depth limit is 26 m at the equator and 9—11 m in latitudinal extremes (Muir, et al. 2015).

The algal symbionts in the genus *Symbiodinium* are classified into nine genetically diverged lineages, termed clades (Pochon and Gates 2010), and these clades include putative species referred as subclade types. The physiological traits are varied among these types (Blackall, et al. 2015). For example, the amount of photosynthetic products translocation from algal symbionts to the coral hosts is different between two *Symbiodinium* clades (Baker, et al. 2013). Most Scleractinia corals associate with more than one *Symbiodinium* type (Silverstein, et al. 2012) and can change the type of *Symbiodinium* (Cunning, et al. 2015). Changing of *Symbiodinium* types with different physiological traits are thought to be one of the strategy to adapt to living environments for corals. In addition to symbiosis with algal symbionts, association with other microorganisms are important for corals (Rosenberg, et al. 2007). For example, bacteria living in coral tissues can provide nitrogen to a coral host (Ceh, et al. 2013).

***Impact of environmental change on coral reef***

Coral reefs harbor the most biologically diverse ecosystems in warm shallow water (Hinrichsen 1997; Knowlton, et al. 2010; Roberts, et al. 2002) and are also important for human activities associated with coral reefs. Coral reefs support fisheries and tourisms, and shield human from storms (Spalding and Brown 2015). However, corals are sensitive to climate change such as ocean acidification, increasing seawater temperatures, and local stressors such as pollution (Hoegh-Guldberg, et al. 2007; Spalding and Brown 2015). Especially, increasing seawater temperatures result from global warming have triggered mass bleaching of corals that is caused by the loss of algal symbionts or photosynthetic pigments in algal symbionts (Heron, et al. 2016). Recently, breaching of corals become more severe, therefore immediate action to stop global warming (Hughes, et al. 2017) and understanding of the molecular mechanisms of stress response of corals are necessary to reduce coral breaching.

When I started my doctoral thesis, fluorescent proteins of corals were one of the enigmatic problems in corals. Many biological functions of fluorescent proteins including stress response such as photo-protection had been proposed (Salih, et al. 2000), but had not been analyzed in detail. In addition to the possible biological roles of fluorescent proteins, the fluorescence from corals was biologically very attractive for me when I dive into the ocean, sometimes I saw fluorescence of corals in different depths and different locations. These knowledge and experience were the starting point of my doctoral thesis.

## Reference

- Appeltans W, et al. 2012.** The magnitude of global marine species diversity. *Current Biology* 22: 2189-2202.
- Baker DM, Andras JP, Jordán-Garza AG, Fogel ML 2013.** Nitrate competition in a coral symbiosis varies with temperature among *Symbiodinium* clades. *The ISME journal* 7: 1248-1251.
- Blackall LL, Wilson B, Oppen MJ 2015.** Coral—the world's most diverse symbiotic ecosystem. *Molecular Ecology* 24: 5330-5347.
- Ceh J, et al. 2013.** Nutrient cycling in early coral life stages: *Pocillopora damicornis* larvae provide their algal symbiont (*Symbiodinium*) with nitrogen acquired from bacterial associates. *Ecology and Evolution* 3: 2393-2400.
- Chang ES, et al. 2015.** Genomic insights into the evolutionary origin of Myxozoa within Cnidaria. *Proc Natl Acad Sci U S A* 112: 14912-14917. doi: 10.1073/pnas.1511468112
- Cunning R, Silverstein RN, Baker AC 2015.** Investigating the causes and consequences of symbiont shuffling in a multi-partner reef coral symbiosis under environmental change. *Proc. R. Soc. B* 282: 20141725.
- Dunn CW, Giribet G, Edgecombe GD, Hejnol A 2014.** Animal phylogeny and its evolutionary implications. *Annual review of ecology, evolution, and systematics* 45: 371-395.
- Harrison PL. 2011.** Sexual reproduction of scleractinian corals. In. *Coral reefs: an ecosystem in transition*: Springer. p. 59-85.
- Heron SF, Maynard JA, van Hooidonk R, Eakin CM 2016.** Warming Trends and Bleaching Stress of the World's Coral Reefs 1985-2012. *Sci Rep* 6: 38402. doi: 10.1038/srep38402
- Hinrichsen D 1997.** Coral reefs in crisis. *Bioscience* 47: 554-558.

**Hoegh-Guldberg O, et al. 2007.** Coral reefs under rapid climate change and ocean acidification. *Science* 318: 1737-1742.

**Hughes TP, et al. 2017.** Global warming and recurrent mass bleaching of corals. *Nature* 543: 373-377.

**Jouiaei M, et al. 2015.** Ancient Venom Systems: A Review on Cnidaria Toxins. *Toxins (Basel)* 7: 2251-2271. doi: 10.3390/toxins7062251

**Kayal E, et al. 2017.** Comprehensive phylogenomic analyses resolve cnidarian relationships and the origins of key organismal traits. *PeerJ Preprints*.

**Kayal E, Roure B, Philippe H, Collins AG, Lavrov DV 2013.** Cnidarian phylogenetic relationships as revealed by mitogenomics. *BMC Evolutionary Biology* 13: 5.

**Kitahara MV, Fukami H, Benzoni F, Huang D. 2016.** The new systematics of Scleractinia: integrating molecular and morphological evidence. In. *The Cnidaria, Past, Present and Future*: Springer. p. 41-59.

**Knowlton N, et al. 2010.** Coral reef biodiversity. *Life in the World's Oceans: Diversity Distribution and Abundance*: 65-74.

**Muir PR, Wallace CC, Done T, Aguirre JD 2015.** Limited scope for latitudinal extension of reef corals. *Science* 348: 1135-1138.

**Muscatine L, R McCloskey L, E Marian R 1981.** Estimating the daily contribution of carbon from zooxanthellae to coral animal respiration. *Limnology and Oceanography* 26: 601-611.

**Pochon X, Gates RD 2010.** A new Symbiodinium clade (Dinophyceae) from soritid foraminifera in Hawai'i. *Mol Phylogenet Evol* 56: 492-497. doi: 10.1016/j.ympev.2010.03.040

**Richards ZT, Miller DJ, Wallace CC 2013.** Molecular phylogenetics of geographically restricted *Acropora* species: Implications for threatened species conservation. *Molecular*

Phylogenetics and Evolution 69: 837-851. doi: 10.1016/j.ympev.2013.06.020

**Roberts CM, et al. 2002.** Marine biodiversity hotspots and conservation priorities for tropical reefs. *Science* 295: 1280-1284.

**Rosenberg E, Koren O, Reshef L, Efrony R, Zilber-Rosenberg I 2007.** The role of microorganisms in coral health, disease and evolution. *Nat Rev Microbiol* 5: 355-362. doi: 10.1038/nrmicro1635

**Salih A, Larkum A, Cox G, Kühl M, Hoegh-Guldberg O 2000.** Fluorescent pigments in corals are photoprotective. *Nature* 408: 850-853.

**Shikina S, Chang C-F. 2016.** Sexual Reproduction in Stony Corals and Insight into the Evolution of Oogenesis in Cnidaria. In. *The Cnidaria, Past, Present and Future*: Springer. p. 249-268.

**Silverstein RN, Correa AM, Baker AC 2012.** Specificity is rarely absolute in coral–algal symbiosis: implications for coral response to climate change. *Proceedings of the Royal Society of London B: Biological Sciences* 279: 2609-2618.

**Spalding MD, Brown BE 2015.** Warm-water coral reefs and climate change. *Science* 350: 769-771.

**Stanley GD, Fautin DG 2001.** The origins of modern corals. *Science* 291: 1913-1914.

**Stanley Jr GD 2003.** The evolution of modern corals and their early history. *Earth-Science Reviews* 60: 195-225.

**Tremblay P, Grover R, Maguer JF, Legendre L, Ferrier-Pagès C 2012.** Autotrophic carbon budget in coral tissue: a new <sup>13</sup>C-based model of photosynthate translocation. *Journal of Experimental Biology* 215: 1384-1393.

**Wallace C. 1999.** Staghorn corals of the world: a revision of the genus *Acropora*: CSIRO

publishing.

**Zapata F, et al. 2015.** Phylogenomic Analyses Support Traditional Relationships within Cnidaria. PLoS One 10: e0139068. doi: 10.1371/journal.pone.0139068

# Chapter 2

**Fluorescent protein gene family has persisted  
during the evolution of *Acropora* species**

**Abstract**

Fluorescent proteins (FPs) are well known and broadly used as bio-imaging markers in molecular biology research. Many *FP* genes were cloned from anthozoan species and it was suggested that multi-copies of these genes are present in their genomes. However, the full complement of *FP* genes in any single coral species remained unidentified. In this chapter, I analyzed the *FP* genes in two stony coral species, *Acropora digitifera* and *Acropora tenuis*. *FP* cDNA sequences from the two species revealed the presence of a multi-gene family with an unexpectedly large number of genes, separated into short-to-middle-wavelength emission (S/MWE), middle-to-long-wavelength emission (M/LWE), and chromoprotein (CP) clades. *FP* gene copy numbers in the genomes of four *A. digitifera* colonies were estimated as 16–22 in the S/MWE, 3–6 in the M/LWE, and 8–12 in the CP clades, and, in total, 35, 31, 33, and 33 *FP* gene copies per individual shown by quantitative PCR. To the best of my knowledge, these are the largest sets of *FP* genes per genome. The fluorescent light produced by recombinant protein products encoded by the newly isolated genes explained the fluorescent range of live *A. digitifera*, suggesting that the high copy multi-*FP* gene family generates coral fluorescence. The functionally diverse multi-*FP* gene family must have existed in the ancestor of *Acropora* species, as suggested by molecular phylogenetic and evolutionary analyses. The persistence of a diverse function and high copy number multi-*FP* gene family may indicate the biological importance of diverse fluorescence emission and light absorption in *Acropora* species.

## Introduction

Fluorescent proteins (FPs), especially green fluorescent protein (GFP), are well known and broadly used as bio-imaging markers in molecular biology research. FP was initially isolated from jellyfish *Aequorea victoria* (Field, et al. 2006; Shimomura 1979). FPs are excited by environmental light and emit longer wavelength fluorescence than the excitation light wavelength (Johnsen 2012). FP emission light is determined by its amino acid sequence (Field, et al. 2006). Subsequently, divergent jellyfish FP homologs were cloned from anthozoan species (Matz, et al. 1999), marine crustacean copepods (Shagin, et al. 2004), and deuterostome chordate amphioxus species (Baumann, et al. 2008; Bomati, et al. 2009; Deheyn DD, et al. 2007; Yue, et al. 2016). These FPs are classified into four groups based on the color of the emitted light: cyan (CFP), green (GFP), yellow (YFP), and red (RFP) (Alieva, et al. 2008; Labas, et al. 2002). Non-fluorescent chromoprotein (CP) is also classified as an *FP* gene family member, on account of amino acid sequence similarity (Labas, et al. 2002). The FPs of the same fluorescence class have emerged repeatedly and independently during coral evolution. It was reported that CFPs, YFPs, and RFPs have evolved several times in phylogenetically distinct lineages (Alieva, et al. 2008). All known FPs from *Acropora* species were included in one of the three clades in the FPs from all corals (Alieva, et al. 2008). The center of the light emission determinant of fluorescent proteins comprises tripeptide “–X–Y–G–”, termed chromophore, where X varies in different FPs (Henderson and Remington 2005). FPs comprise a high proportion, 4.5–14%, of the total soluble protein content of FP-expressing anthozoan tissues (Leutenegger, et al. 2007; Oswald, et al. 2007). Because of this high cellular content, many biological roles have been proposed for FPs. They are considered essential for viability and may

play photo-protective and antioxidant roles (Palmer, et al. 2009; Roth and Deheyn 2013). However, the detailed biological function of FPs in corals remains elusive.

Multiple *FP* genes have been reported for a single species. In the amphioxus genome (*Branchiostoma floridae*), 16 unique *GFP*-like genes were present and this is the largest known *FP* gene family in a single organism to date (Bomati, et al. 2009). Among anthozoans, four to seven separate genetic loci that encode *CFP*, *GFP*, and *RFP* genes were predicted in *Montastraea cavernosa*. These genes may have undergone gene conversion events (Kelmanson and Matz 2003). In the genus *Acropora*, one of the most abundant coral genera in coral reefs of the Indo-Pacific region (Veron 2000), several *FP* sequences were reported for a single species, as follows: in *A. pulchra*, two *FP* genes (one *CFP* and one *CP*) (D'Angelo, et al. 2008); in *A. nobilis*, three *FP* genes (one *GFP* and two *CFP*); in *A. aculeus*, three *FP* genes (two *GFP* and one *CP*); and in *A. hyacinthus*, one *FP* gene (*CP*) (Alieva, et al. 2008). In the studies of *A. millepora*, four (two *GFP*, one *CFP*, one *RFP*) (D'Angelo, et al. 2008) and four (each of one *GFP*, *CFP*, *RFP* and *CP*) (Alieva, et al. 2008) *FP* sequences were isolated, but a homologous-paralogous relationship between sequences from the same emission light groups has not been identified. Eight copies of *RFP* genes were deduced based on exon 3 sequences in *A. millepora* (Gittins, et al. 2015), but full-length coding regions were not determined. Analysis of the entire *A. digitifera* genomic nucleotide sequence revealed ten *FP*-like genes, but nine of these were truncated compared with a typical *FP*-coding sequence (Shinzato, et al. 2012). Studies of anthozoan *FP* sequences indicated the possibility of high *FP* gene copy number in their genomes; however, the full *FP* gene complement in a single coral species is still unidentified.

In this chapter, I analyzed the *FP* genes from two stony coral species, *A. digitifera* that the genome sequences were decoded, and *A. tenuis*. *A. tenuis* is distantly related to *A. digitifera* and is located at the basal lineage in the genus *Acropora* (Fukami, et al. 2000; Richards, et al. 2013) with different habitat, especially in the depth from *A. digitifera* (Suzuki, et al. 2008). I found that *A. digitifera* encodes the largest set of *FP* genes and the multi-*FP* gene family has persisted during the evolution of *Acropora* species.

## Materials and Methods

### *Specimen collection and species identification*

Collections of samples in this thesis were approved by the Aquaculture Agency of Okinawa Prefecture (permits numbers 26–9 and 28–31). Five colonies of *A. digitifera* and one colony of *A. tenuis* were collected from field and subsequently maintained in Sesoko Station aquarium (Tropical Biosphere Research Center, University of the Ryukyus). Among five *A. digitifera* colonies, I did not find clear color differences at photograph level. Additionally, a small piece of one *A. digitifera* colony was collected from field and preserved in RNAlater (Waltham, MA USA). These two species were identified based on morphology. I collected planula larvae from each of five colonies of *A. digitifera* and *A. tenuis* that were kept in separate aquariums. All specimen information is given in Appendix Table S2.1.

### *Live coral fluorescence measurements in the field*

I measured light emission, including reflectance and fluorescence, from five *A. digitifera* colonies in the vicinity of Sesoko Station (Tropical Biosphere Research Center, University of the Ryukyus). I used two excitation illumination light sources: LED source, with spectrum peak at 448 nm, and laser source, with spectrum peak at 452 nm (Figure 2.1A and 2.1D). The distances of illumination of excitation lights and measurement probe from objects were 6 cm in LED and 1 cm in laser measurements. Measurements were performed in the dark, at night, to avoid the effect of sunlight. Each spectrum was recorded by Jaz Spectrometer (Ocean Optics, Dunedin, FL, USA). In all measurements, emission light longer than 660 nm was not considered as coral fluorescence, because chlorophyll *a* from the symbiotic dinoflagellate algae living

within the coral tissues emits light spectra with a primary peak around 685 nm and a secondary peak at 730 nm (Mazel 2003; Moisan and Mitchell 2001).

### ***RNA extraction and sequencing***

Total RNAs were extracted from adult specimens of *A. digitifera* and *A. tenuis*, and a single *A. digitifera* larva, each of ~50 individuals of *A. digitifera* and *A. tenuis* larvae, using TRIzol reagent (Thermo Fisher Scientific, MA, USA). RNA libraries were constructed from adult specimens of *A. digitifera* and *A. tenuis*, and each of ~50 individuals of *A. digitifera* and *A. tenuis* larvae with NEBNext Poly(A) mRNA Magnetic Isolation Module and NEBNext Ultra RNA Library Prep Kit for Illumina (New England Bio Labs, MA, USA). Short DNA sequences (paired-end 100 bp) were determined from libraries by Illumina HiSeq2000 platform (RNA-seq). After removal of the adaptor sequences and low-quality reads, RNA-seq read assembly was performed using CLC genomic workbench (<https://www.qiagenbioinformatics.com/>) with auto-word size. Assembled sequences were used for searching a distantly related *FP* sequence to known *Acropora FP* genes.

### ***Identification and cloning of FP-like cDNA sequences***

To identify *FP*-like sequences, I selected the longest *FP* sequences (*AdiFP2*, *AdiFP8*, and *AdiFP10*, accession numbers BR000963, AB698751, and BR000970, respectively) from three different clades of *A. digitifera FP*-like sequences (Shinzato, et al. 2012); however, *AdiFP8* and *AdiFP10* were shorter than a typical *FP*-coding sequence. To identify the possible previously undetermined 5' terminal gene sequence, *AdiFP8* and *AdiFP10* were used to search RNA-seq

reads for short sequences with high similarity (>90%) to 5' termini of these two *FP*-like sequences. These short sequences were aligned with *AdiFP8* and *AdiFP10*, and 5' termini of the two *FP*-like genes were subsequently extended (Appendix Figure S2.3). I named these extended sequences *AdiFP8L* and *AdiFP10L*. To design primers, I mapped RNA-seq reads from *A. digitifera* and *A. tenuis* to reference sequences, *AdiFP2*, *AdiFP8L*, and *AdiFP10L*. Several primer sets were designed in the conserved regions that were identified based on the mapping results to PCR-amplify all *FP*-like sequences identified in RNA-seq reads.

cDNAs were synthesized from total RNA extracted from adult (n = 1) and larva (n = 1) specimens of *A. digitifera*, and adult (n = 1) and larvae (n = ~50) specimens of *A. tenuis*, using PrimeScript II 1st Strand cDNA Synthesis Kit (Takara, Shiga, Japan). *GFP*-like cDNAs were amplified using PrimeSTAR Max DNA Polymerase (Takara, Shiga, Japan) and primers FP2\_5UHind\_F1 and FP2\_3UBam\_R1 for *A. digitifera*, and three sets of primers, *AdiFP1KpnI\_F* and *AdiFP1XbaI\_R*, *AdiFP2KpnI\_F* and *AdiFP2XbaI\_R*, and *AdiFP4KpnI\_F* and *AdiFP4XbaI\_R*, for *A. tenuis*. *RFP*-like cDNAs were cloned using primers FP10\_5UHind\_F1 and FP10\_3UBam\_R1 for *A. digitifera*, and primers *AdiFP10KpnI\_F* and *AdiFP10XbaI\_R* for *A. tenuis*. *CP*-like cDNAs were cloned using primers FP8\_5U\_F2 and FP8\_5U\_R1 for *A. digitifera*, and primers *AdiFP8KpnI\_F* and *AdiFP8XbaI\_R* for *A. tenuis*. PCR was performed with GeneAmp PCR System 9700 (Applied Biosystems, Foster City, CA, USA). All primer positions and sequences are given in Appendix Figure S2.1A and S2.1B. PCR conditions for amplification of full-length cDNA were as follows: denaturation step for 3 min at 94 °C, followed by 30 cycles of denaturation for 1 min at 94 °C, annealing for 30 sec at 55 °C, and extension for 30 sec at 72 °C. PCR products were used as templates in a second-round PCR

reaction when the quantity of first-round PCR product was insufficient for cloning. PCR products were cloned into T-Vector pMD20 vector (Takara), and the sequences were verified using Applied Biosystems Automated 3130xl Sequencer (Applied Biosystems). These *FP*-like sequences aligned them with the known *FP* sequences from *A. millepora* using ClustalW in MEGA ver. 6 (Tamura, et al. 2013). I identified the mismatched nucleotides in the aligned *FP* sequences. When mismatches from the aligned *FP* sequences occurred less than twice in RNA-seq reads, I defined those mismatches as PCR errors. Sequences containing PCR errors were excluded from further analysis.

To verify the absence of *FP* sequences sharing low similarity with the known *Acropora FP* genes from RNA-seq reads, I mapped RNA-seq reads from *A. digitifera* and *A. tenuis* to several cnidarian *FP* sequences (eqFP611: AY130757, hcriCP: AF363776, hcriGFP: AF420592, pporRFP: DQ206380, KO: AB128820, efasGFP: DQ206385, pporGFP: DQ206391, meleCFP: DQ206382, meleRFP: DQ206386), and *AdiFP2*, *AdiFP8L*, and *AdiFP10L*. Reads showing similarity (>80%, >80 bp) were mapped to reference sequences. The same cnidarian *FP* sequences were used as queries to blastx-search (Gish and States 1993) to verify the absence of *FP* sequences with low similarity to known *Acropora FPs* in each of the assembled contigs from *A. digitifera* and *A. tenuis* adults and larvae.

### ***Cloning, purification, and spectroscopic analysis of recombinant FP proteins***

To construct recombinant *FP* proteins, vectors bearing cloned *FP* sequences verified by sequencing were used as templates for subcloning into expression vectors. *GFP*-like full-length cDNAs were amplified using PrimeSTAR Max DNA Polymerase (Takara), with

AdiFP2KpnI\_F\_L or AdiFP2KpnI\_F2\_L forward primers, and AdiFP2XbaI\_R\_L reverse primers. *RFP*-like full-length cDNAs were amplified using AdiFP10L\_KpnIF\_L forward primer and AdiFP10L\_type1\_R, AdiFP10L\_type2\_R, or AdiFP10L\_type3\_R reverse primers. *CP*-like full-length cDNAs were amplified using AdiFP8L\_KpnIF\_L forward primer, and AdiFP8XbaI\_R1\_L or AdiFP8XbaI\_R2\_L reverse primers. Full-length *FP* cDNAs were subcloned into pCold I expression vector (Takara) and then used to transform BL21 *Escherichia coli* cells (Takara). Each clone was grown in 20 mL of LB medium, supplemented with ampicillin and IPTG, overnight, and the recombinant proteins were extracted by sonication and purified using TALON beads with poly-histidine tags (Takara).

Emission spectra of purified recombinant FP proteins in 50 mmol/l phosphate buffer solution with 500 mmol/l Imidazole, pH 7.0 were measured with USB-4000 Spectrometer (Ocean Optics). Absorption spectra of purified recombinant FP proteins were measured using UV-1800 spectrophotometer (SHIMAZU, Kyoto, Japan). Both measurements were performed three times for each FP protein.

### ***Phylogenetic and sequence comparison analyses***

Coding sequences of *FP* genes were translated into amino acids and aligned with the known *FP* sequences (AY646067, AY646070, AY646073, AY646075, EU709808, EU709809, EU709810, EU709811, JX258845, JX258846, KC349891, KC411499, and KC411500) from *A. millepora* using ClustalW in MEGA ver. 6 (Tamura, et al. 2013). Phylogenetic analysis was performed using the p-distance method with 1,000 bootstrap replications. Three major clades in phylogenetic tree were termed based on the fluorescent emissions of FPs encoded by *FP*

sequences as follows; S/MWE clade comprised short-wavelength emission (SWE) and middle-wavelength emission (MWE1) clades; M/LWE clade comprised middle-wavelength emission (MWE2) and long-wavelength emission (LWE) clade; and chromoprotein (CP) clade. Wavelength of these categories was defined as follows: an emission peak less than 500 nm (short), an emission peak from 500 to 530 nm (middle), an emission peak over 570 nm (long), and absorbance with no emission (CP). Ancestral sequences (see Appendix text) were estimated using the maximum likelihood method with a pre-set tree topology by MEGA ver. 6 (Tamura, et al. 2013). DNA fragments of estimated ancestral sequences were then constructed by *in vitro* mutagenesis (Ho, et al. 1989), and recombinant proteins were purified as described above. The aligned *FP* sequences were used in four-gamete tests to detect recombination events, using an in-house computer program. The nucleotide sequences were deposited in GenBank under accession numbers LC125047–LC125121, LC177540–LC177542 and in the DDBJ Sequenced Read Archive under accession numbers DRX049620–DRX049623.

***Estimation of A. digitifera FP gene copy numbers and identification of high coverage regions***

*FP* gene copy numbers in *A. digitifera* genome were estimated by quantitative PCR (qPCR) from four *A. digitifera* specimens. qPCR was performed with Thermal Cycler Dice TP800 (Takara). I designed three sets of primers specific to S/MWE, M/LWE, and CP clade sequences to amplify all *FP* sequences detected in RNA-seq reads. I amplified partial sequences of *AdiFP2*, *AdiFP8L*, *AdiFP10L*, and elongation factor 1 (*Ef1*) gene. All PCR products were cloned into one pMD20 vector (Takara), in tandem, using In-Fusion HD Cloning Kit (Takara). This plasmid DNA was used as a control for all genes. Primers for the amplification of each gene were as

follows: MWE\_qPCR\_F3 and MWE\_qPCR\_R1 for *AdiFP2* (S/MWE); MLWE\_qPCR\_F3 and MLWE\_qPCR\_R1 for *AdiFP10L* (M/LWE); CP\_qPCR\_F1 and MiA\_CP\_e3\_R1 for *AdiFP8L* (CP); and EF1a-qPCR\_F1 and EF1a-qPCR\_R1 for *EF1*. qPCR reactions were performed with the same primers that were used for amplification of partial sequences using SYBR Premix Ex Taq II (Takara). *A. digitifera* genomic DNA was quantified by Qubit 2.0 Fluorometer (Thermo Fisher Scientific) and used as qPCR template. The number of genome copies in a reaction was calculated from the weight of genomic DNA and genome size (420 Mb) of *A. digitifera* (Shinzato, et al. 2011). A series of diluted plasmid DNA (10 pg, 1 pg, 0.1 pg, 10 fg, and 1 fg, per  $\mu$ L) was used to construct standard curves to estimate *FP* gene copy numbers/reaction. From these two numbers, the number of *FP* gene copies per genome was calculated. Differences between the amplification efficiencies of genomic and plasmid DNA were calibrated using a single copy gene, *EF1*, as standard. qPCR reactions were performed three times for each genomic DNA sample and control plasmid DNA, with all primer sets.

In addition to ten *FP*-like genes (Shinzato, et al. 2012), I identified *FP*-like gene loci in the *A. digitifera* genome (Shinzato, et al. 2011) by blastn search (Altschul, et al. 1990), using *AdiFP2*, *AdiFP8L*, and *AdiFP10L* as query sequences. After locating all *FP* gene exon sequences within ca. 3 kb genomic regions (exons 1 to 5), I defined those regions as *FP* genes. In addition to genes containing complete exon sets, some exon sequences were missing in several genes due to un-assembled genomic regions because of the difficulty associated with high copy number gene assembly (Mariano, et al. 2015). I defined those partial genes also as *FP* genes. To estimate the copies of un-assembled *FP* genes, I mapped short reads of *A. digitifera* (DRX000980 and DRX000981) to its genome and extracted high coverage regions ( $p < 0.0001$ )

using CLC Genomics Workbench.

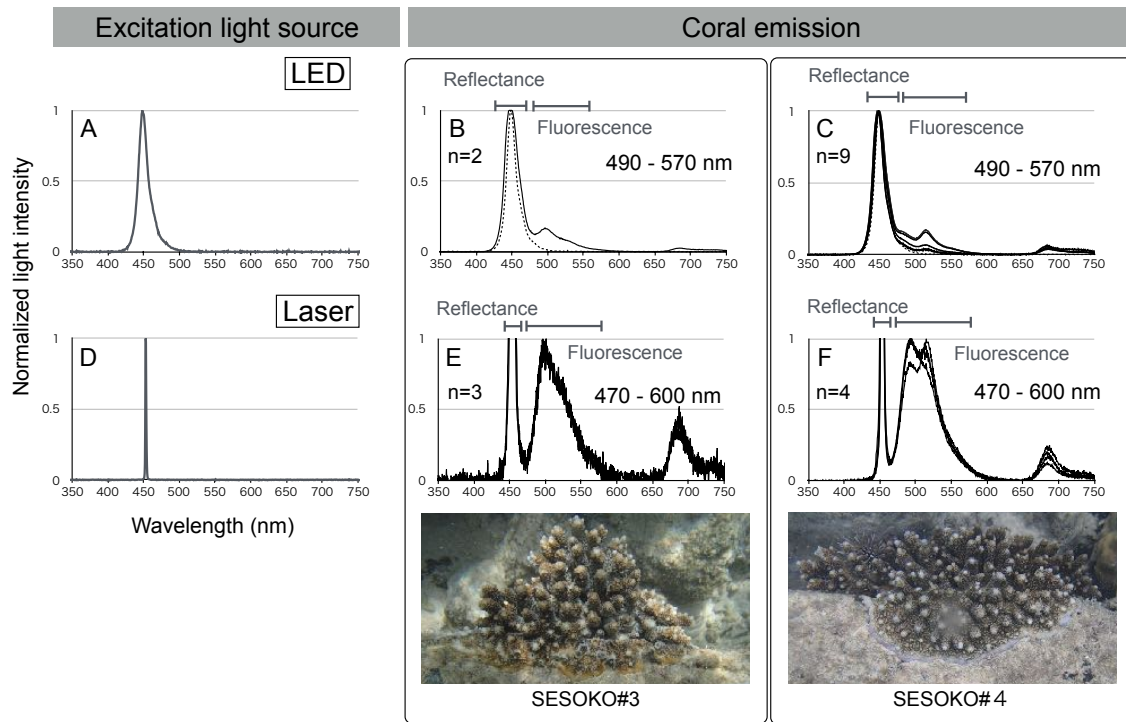
### ***Nucleotide divergence between two subclades in S/MWE clade***

I estimated nucleotide divergence between two subclades in the S/MWE clade. Each subclade comprised *FP* sequences from both *A. digitifera* and *A. tenuis*. I defined the average number of nucleotide differences per site (p-distance) from all pairwise comparisons between the different groups as the divergence between two groups. The analysis involved 34 *FP* sequences, namely, for subclade 1, FP4KX\_Tl\_8, FP2KX\_Tl\_16, FP2KX\_Tl\_4, FP4KX\_Tl\_1, FP2KX\_Tl\_3, FP2KX\_Tl\_7, FP4KX\_Tl\_7, FP2KX\_Tl\_5, FP2KX\_Tl\_15, FP2KX\_Tl\_13, FP2KX\_Tl\_14, FP2KX\_Tl\_9, FP4KX\_Tl\_3, FP4\_Tl\_11, FP2KX\_Tl\_6, FP4\_Tl\_16, FP2KX\_Tl\_2, FP2KX\_Ta\_10, FP2\_BH\_38, FP2\_BH\_41, FP2R1\_Dl1\_5, FP2\_R1\_13, and FP2R1\_Dl1\_9; and, for subclade 2, FP2\_BH\_3, FP2\_BH\_39, FP2\_BH\_1, FP2\_BH\_4, FP1KX\_Tl\_2, FP1\_Tl\_9, FP1\_Tl\_12, FP1KX\_Tl\_1, FP1KX\_Tl\_3, FP2KX\_Ta\_7, and FP2\_S1603\_BH4. Genetic difference between each sequence pair was calculated by MEGA 6 (Tamura, et al. 2013). Mean nucleotide difference (p-distance) between *A. digitifera* and *A. tenuis* was estimated based on sequence pairs of *A. digitifera* genome and *A. tenuis* assembled RNA-seq sequences. I employed reciprocal blastn hit pairing between *A. digitifera* scaffold and *A. tenuis* contigs with e-value  $< 1e^{-50}$  (Altschul, et al. 1990). I discarded *A. tenuis* contigs with a second hit to *A. digitifera* scaffolds with e-value  $< 1e^{-50}$  to avoid putative orthologous pairs with single-multicopy relationships.

## Results

### *A. digitifera* emits a wide range of fluorescence in the sea

Because its entire genome DNA sequence is available, I first focused on one *Acropora* species, *A. digitifera*, to reveal the full complement of its *FP* genes. Before analyzing *FP* and *FP* sequences from this species, I measured the fluorescence emitted from colonies of *A. digitifera* in the sea. I measured the emission light (including reflectance and fluorescence) from five *A. digitifera* colonies (Appendix Table S2.1) excited by two excitation light sources, LED (448 nm spectrum peak; Figure 2.1A) and laser (452 nm spectrum peak; Figure 2.1D). As shown in Figures 2.1B, 2.1C, and Appendix Figure S2.2A–S2.2C, *A. digitifera* fluorescence spanned 490–570 nm, as estimated by the subtraction spectrum from excitation (LED) and emission lights. When laser was used as an excitation light, the boundaries between excitation and fluorescence were clear because of the narrower band of laser light in comparison with LED (Figure 2.1A and 2.1D). The fluorescence spanned 470–600 nm (Figures 2.1E, 2.1F and Appendix Figure S2.2D–S2.2F).



**Figure 2.1.** Live coral fluorescence measurements in the field

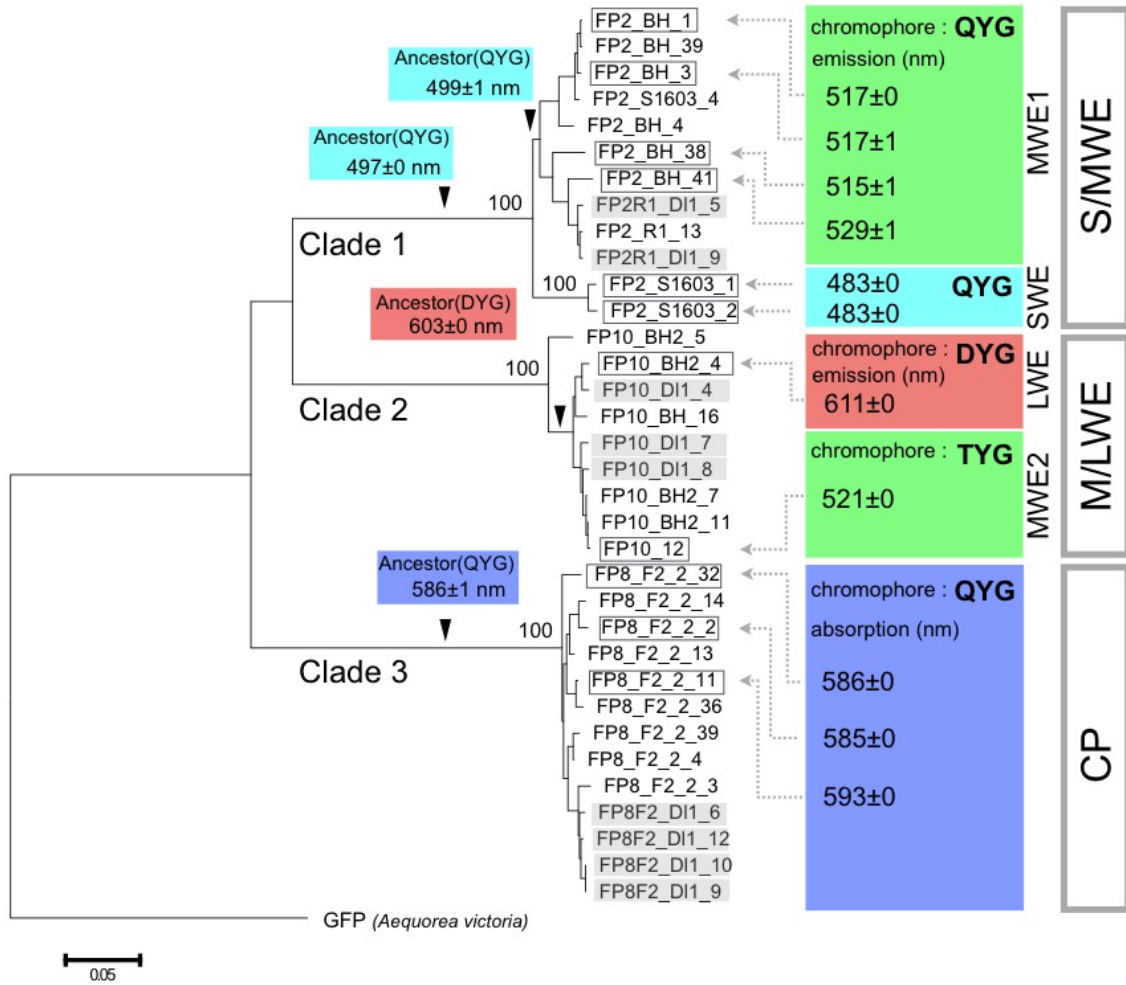
Excitation spectra of different light sources, LED (A) and laser (D). Horizontal and vertical axes indicate wavelengths (nanometers) and normalized light intensities, respectively. A. *digitifera* emission spectra (solid line) excited by LED (B, C) and laser (E, F) lights from two colonies with different fluorescence emissions (SESOKO#3 and SESOKO#4). LED excitation spectrum is shown by a dotted line (B, C).

***FP-like sequences have diversified in A. digitifera***

To isolate all *FP* sequences expressed in different developmental stages (larva and adult), RNA sequences (9.4 and 8.0 Gbp, respectively) were determined by Illumina HiSeq2000 platform. To design primers in the conserved regions among mapped-reads, I mapped paired-end sequences from *A. digitifera* adult and larvae to *AdiFP2*, *AdiFP8L*, and *AdiFP10L* to PCR-amplify all *FP* sequence types. The average coverage was 37–194 for the adult and 18,032–397,662 for larval reads for the three mapped sequences (Appendix Table S2.2). Using these primers, in total 22 and 9 *FP*-like sequences were identified in adult and larva specimen cDNAs, respectively. No *FP* sequences were identical in the adult and larva. Compared with the known *Acropora FP* genes, the new sequences did not contain any insertion/deletion frame shifts or premature stop codons.

To verify the phylogenetic relationships between sequences identified in this chapter and the known *FP* and *FP*-like sequences, I included *A. millepora FP* sequences and *A. digitifera FP*-like sequences identified in its genome (Shinzato, et al. 2012) in the present analysis. As shown in Appendix Figure S2.4A, the newly isolated *FP*-like sequences from clades 1, 2, and 3 clustered with *FP* sequences that encode proteins with emission peak values 484–512 nm, or 516–599 nm, or with *FP* sequences that encode proteins with only absorption, respectively. All *FP*-like sequences identified in *A. digitifera* genome data were also included in the three clades (Appendix Figure S2.4B), even though the number of positions used for phylogenetic tree construction was reduced from 657 to 342 because of sequence truncation. No *FP*-like sequences clustering with the known *CFP* sequences encoding proteins with emission peak values 485–495 nm (Alieva, et al. 2008) were identified in *A. digitifera* adult or larva

cDNA. I additionally extracted RNA from one *A. digitifera* colony (ID: S1603) that emitted fluorescence with an emission peak value less than 500 nm in the field. Two sequences that were identified from cDNA of this colony were clustered with the known *CFP* sequences (Appendix Figure S2.4C). I constructed a phylogenetic tree using all the sequences identified in this chapter. *A. digitifera* *FP*-like sequences formed three different monophyletic clades (clades 1 to 3, Figure 2.2). To verify the absence of *FP* sequences that shared low similarity with the known *Acropora* *FP* genes in RNA-seq reads, I mapped each RNA-seq read from *A. digitifera* (adult and larva) and *A. tenuis* (adult and larvae) that shared similarity (>80%, >80 bp) with several cnidarian *FP* sequences. Large number of reads (minimum 9,820 reads, maximum 2,776,230 reads) mapped to *Acropora* *FP* sequences (*AdiFP2*, *AdiFP8L*, and *AdiFP10L*), whereas six reads mapped to other cnidarian *FP* sequences. Also, in the assembled contigs from *A. digitifera* and *A. tenuis* RNA-seq reads, I only found *FP* genes highly similar to *AdiFP2*, *AdiFP8L*, and *AdiFP10L*. These results suggest that *A. digitifera* only possesses *FP* genes with high similarity to the known *FP* genes of *Acropora* species.



**Figure 2.2.** Phylogenetic relationships between *A. digitifera* FP sequences and FP-encoded functions

The figure shows FP nucleotide sequence-based phylogenetic tree. The tree was constructed using the neighbor-joining method with the p-distance method. Bootstrap probability for each clade was obtained by 1,000 replicates and is shown next to each node. *A. victoria* GFP was used as an outgroup. The scale bar at the bottom represents 0.05 substitutions per site. FP sequences used for recombinant analyses are enclosed in rectangles, and FP sequences from larva are shown in gray. The arrowheads indicate positions of the estimated ancestral sequences. Maximum values for emission or absorption spectra of each recombinant protein are shown at the tails of dotted arrows, in colored branch boxes (ancestral sequences). Chromophores, in colored boxes, are defined by three amino acids.

### ***Functional analysis of FP sequences***

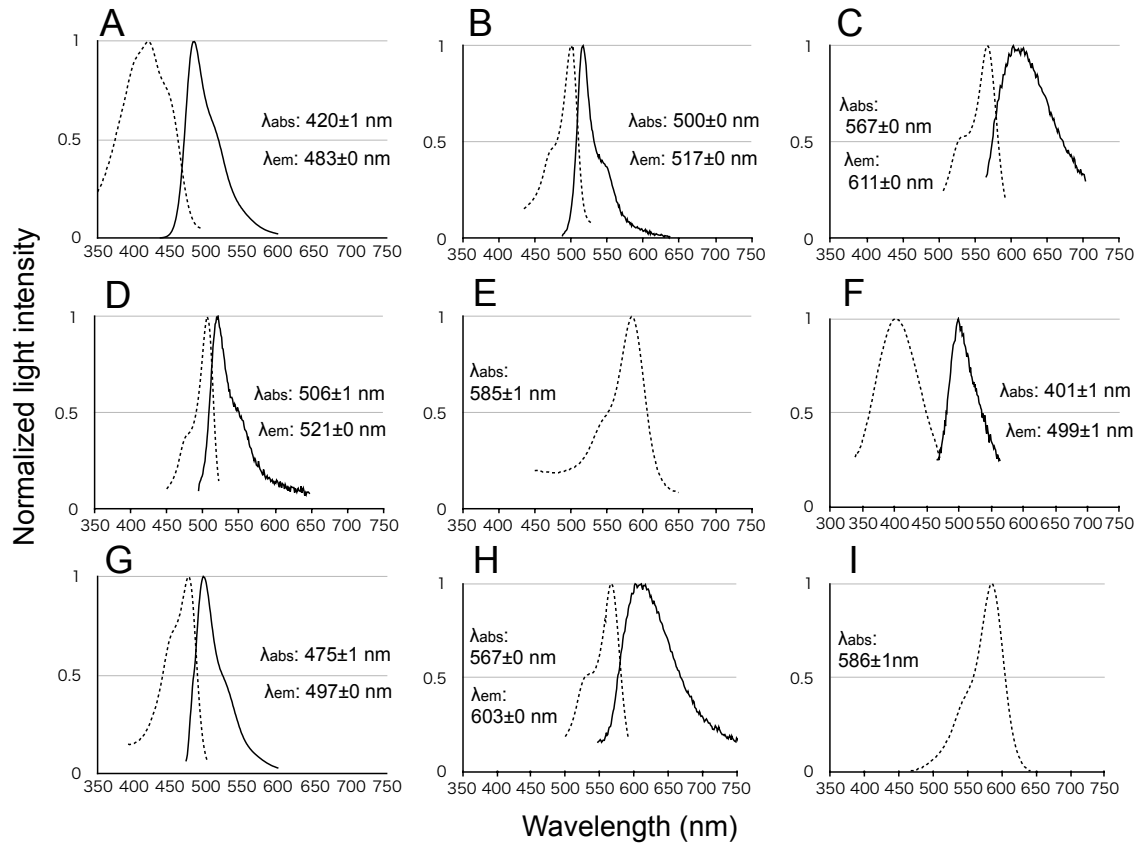
I purified proteins encoded by *FP*-like sequences and measured their absorption and emission spectra. Clade 1 proteins were characterized by SWE (483 nm) (Figures 2.2, 2.3A, and Appendix Figure S.2.5A) or MWE spectra (515–529 nm) (Figure 2.3B and Appendix Figure S.2.5B–D) and were split into two subclades (SWE and MWE1). Clade 3 proteins showed only absorbance (586–593 nm), similarly to the known CPs (Figures 2.2 and 2.3E, Appendix Figures S.2.5H and S.2.5I). Clade 2 was split into two subclades (LWE and MWE2) with DYG and TYG chromophores (Appendix Figure S2.6A), respectively, and proteins encoded by sequences from each clade (FP10\_BH2\_4 and FP10\_12) were characterized by LWE (611 nm, FP10\_BH2\_4) or MWE (521 nm, FP10\_12) spectra (Figures 2.2, 2.3C, and 2.3D). Because *FP*-like sequences from the three clades were highly similar to one other (>95%), I anticipated that all the newly isolated *FP*-like sequences would encode fluorescence or CP functions. Based on the emission and absorption spectra, I termed clade 1 the S/MWE clade, clade 2 the M/LWE clade, and clade 3 the CP clade. Twelve of the newly isolated *A. digitifera* sequences belonged to the S/MWE clade, other nine belonged to the M/LWE clade, and 13 to the CP clade.

I mutated the first amino acid, T, of TYG-chromophore sequence to D (T66D, FP10\_12) in a protein from the MWE2 clade (Appendix Figure S2.6B) and did not detect light emission. An amino acid difference between FP10\_BH2\_4 and FP10\_12 sequences that resulted in amino acid polarity change was located at position 191 (Appendix Figure S2.6B), and, subsequently, an additional mutation (S191P) was introduced in the mutant protein. The fluorescence of the double mutant shifted toward the long wavelengths (Appendix Figure S2.6C,

$\lambda_{\text{em}} = 593 \text{ nm}$ ), but a single change at position 191 (S191P) did not affect the fluorescence emission (Appendix Figure S2.6D).

### ***Assessment of ancestral FP sequences***

I estimated the ancestral sequences for the MWE1 clade, the S/MWE clade, the two subclades of M/LWE clade, and the CP clade (shown by arrowheads in Figure 2.2). Light emission and absorption of purified proteins encoded by these ancestral sequences were measured (Figure 2.3F–3I). Emission or absorption of the ancestral proteins from MWE1, S/MWE, M/LWE, and CP clades were categorized as SWE ( $\lambda_{\text{em}} = 499 \text{ nm}$ ), SWE ( $\lambda_{\text{em}} = 497 \text{ nm}$ ), LWE ( $\lambda_{\text{em}} = 603 \text{ nm}$ ), and CP ( $\lambda_{\text{abs}} = 586 \text{ nm}$ ) spectra, respectively.



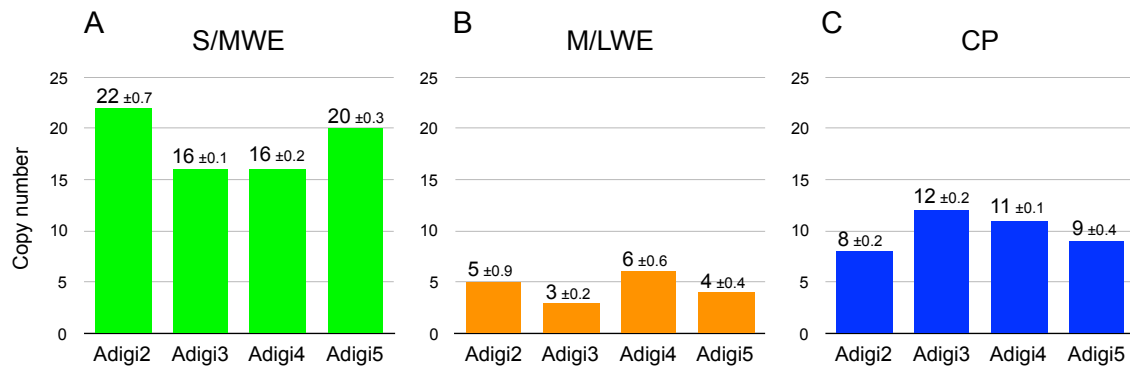
**Figure 2.3.** Emission and absorption spectra of FP recombinant proteins

Absorption (dotted line) and emission (solid line) spectra of recombinants belonging to S/MWE clade (A, FP2\_S1603\_BH1; and B, FP2\_BH\_1), M/LWE clade (C, FP10\_BH2\_4; and D, FP10\_12), CP clade (E, FP8\_F2\_2), MWE1 clade ancestor (F), S/MWE clade ancestor (G), two subclades in M/LWE clade ancestor (H), and CP clade ancestor (I) are shown. Horizontal and vertical axes indicate wavelengths (nanometers) and normalized light intensities, respectively.

***FP genes are present in high copies in the *A. digitifera* genome***

*FP* gene copy numbers in the genome were quantified with qPCR from four adult specimens of *A. digitifera*. The copy numbers of *FP* genes from each clade were 16–22 in the S/MWE clade, 3–6 in the M/LWE clade, and 8–12 in the CP clade (Figure 2.4A–2.4C).

Using three sequences from each of the three *FP* gene clades (*AdiFP2*, *AdiFP10L*, and *AdiFP8L*) as query sequences, I searched the *A. digitifera* genome. I treated partial *FP* sequences with incomplete exon sets as *de facto* *FP* genes because of many un-assembled nucleotide regions in these partial *FP* genes. When I used *AdiFP2* as the query, 12 *FP* genes were identified that shared high similarity with the S/MWE clade sequences on three scaffolds (Appendix Figure S2.7A). Similarly, five and eight *FP* genes were identified as highly similar with M/LWE on three scaffolds, and CP on three scaffolds (Appendix Figure S2.7B and S2.7C). I identified 25 *FP* genes in the *A. digitifera* genome. In addition, I mapped short sequence reads determined from the genome of *A. digitifera* to its genomic sequence to identify high coverage regions. I found that eight S/MWE, five M/LWE, and seven CP genes, out of 25 *FP* genes, reside with high coverage regions (Appendix Figure S2.7D–S2.7F). This suggests the possibility that additional copies of *FP* genes that had not been identified on account of unassembled genomic DNA sequences may be present in the genome.

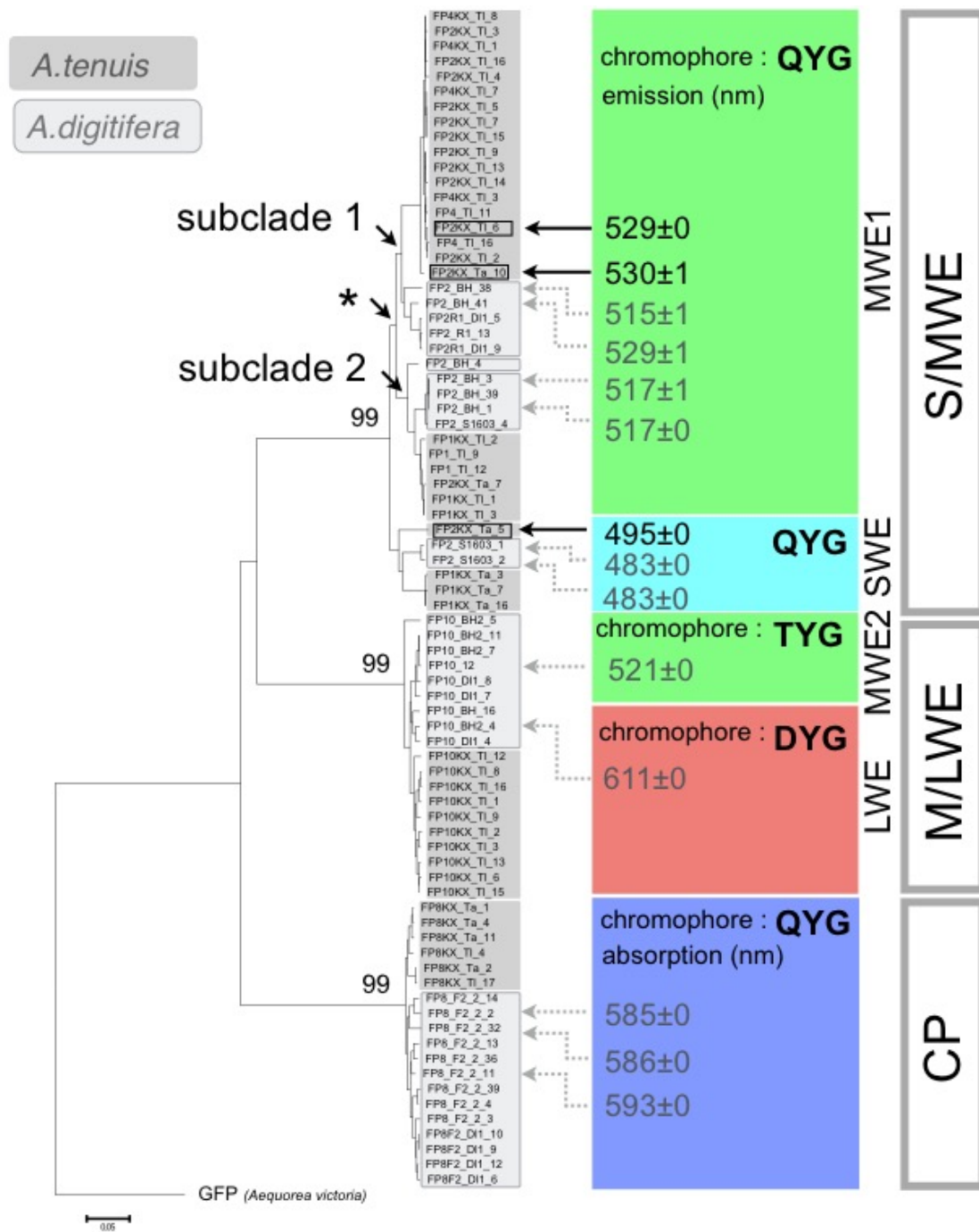


**Figure 2.4.** Individual clade *FP* gene copy numbers

*FP* gene copy numbers in S/MWE (A), M/LWE (B), and CP (C) clades from four colonies of *A. digitifera* are given. Adigi2-Adigi4 indicate colony IDs.

***FP genes have also diversified in A. tenuis***

To examine whether *FP* genes have also diversified within genomes of other *Acropora* species, I determined *FP* sequences of *A. tenuis*. I designed primers (see mapping coverage in Appendix Table S2.2), as described for *A. digitifera*, and amplified *FP* sequences from cDNAs of adult and larval specimens of *A. tenuis*. After accounting for PCR errors, 10 and 34 *FP*-like sequences were identified in the adult and larvae, respectively. I then constructed a phylogenetic tree comprising *FP* genes from *A. digitifera* and *A. tenuis*. All sequences were clustered in one of the three major clades (Figure 2.5). I purified proteins encoded by the S/MWE clade sequences and measured their emission and absorbance spectra. SWE (Appendix Figure S2.5E) and MWE (Appendix Figure S2.5F and S2.5G) spectra were recorded (Figure 2.5). I identified 28 *A. tenuis* sequences in the S/MWE clade, ten sequences in the M/LWE clade, and six sequences in the CP clade. Sequences encoding TYG chromophore were confirmed in the adult and larva *A. tenuis* RNA-seq data; however, I was unable to clone and analyze these sequences because their expression was lower than that of other *FP* sequences in the M/LWE clade. Together with the cloning from *A. digitifera*, the *FP* sequence clustering with known *CFP* sequences was not identified from neither *A. digitifera* nor *A. tenuis* larval specimens.



**Figure 2.5.** Phylogenetic relationships of *A. tenuis* FP sequences and FP-encoded functions

This figure contains FP sequence-based phylogenetic tree. The tree was constructed using the neighbor-joining method with the p-distance method. Bootstrap probability for each clade was obtained by 1,000 replicates and is shown next to each node. The scale bar at the bottom represents 0.05 substitutions per site. *A. victoria* GFP was used as an outgroup. FP sequences from *A. tenuis* and *A. digitifera* are shown in dark and light gray, respectively. *A. tenuis* FP sequences used for recombinant analyses are indicated by black arrows. Maximum values of

emission spectra of each recombinant protein are shown at the tails of black arrows. Asterisk indicates the divergence of subclades 1 and 2 in S/MWE clade. Chromophores, in colored boxes, are defined by three amino acids.

***Recombination between FP sequences from each major clade***

I aligned *A. digitifera* and *A. tenuis* FP sequences from MWE1, M/LWE, and CP clades using ClustalW in MEGA ver. 6 (Tamura, et al. 2013). FP sequences from *A. digitifera* and *A. tenuis* shared nucleotide changes at several synonymous sites in each clade (Appendix Figure S2.8A–S2.8C). Recombination of sequences within extant species and between species, meaning recombination of sequences within common ancestral species (hereafter, between species), was detected in each major clade by four-gamete tests (Appendix Figure S2.9).

***The nucleotide divergence between two subclades in MWE1 clade***

The nucleotide divergence between two subclades in the MWE1 clade (marked in Figure 2.5 by an asterisk) was 0.068. This value was slightly higher than the mean nucleotide divergence between *A. digitifera* and *A. tenuis* (0.065) calculated from sequence pairs between the *A. digitifera* genomic sequence and *A. tenuis* assembled RNA-seq sequences.

## Discussion

### *Multi-member FP gene family underlies A. digitifera fluorescence*

Before analyzing *A. digitifera* FP and *FP* sequences, I measured the fluorescence emitted by the colonies of this species in the sea, because, to the best of my knowledge, these data have not been available. I applied a very narrow band laser light for excitation, and the fluorescence emitted by corals was determined to span 470–600 nm, with clear separation from excitation light.

How many FPs contribute to this fluorescence? To uncover the genetic basis of *A. digitifera* fluorescence, I determined the sequences of *FP* genes from adult and larva cDNAs. No sequence was identical between them, suggesting life stage-specific *FP* gene expression. The sequence difference between the adult and larva may reflect these life stage-specific fluorescence patterns. However, I cannot exclude the possibility that individual, specimen-related differences were responsible for these sequence differences. The types of FPs expressed in adults and larvae of both species might be different; FPs in SWE, MWE1, MWE2, LWE, and CP clades in adults and in MWE1, MWE2, LWE, and CP clades from larvae. The lack of FPs in the SWE clade in larvae was similar to the blue shift of fluorescence from larvae to adults observed in *Seriatopora hystrix* (Roth, et al. 2013). The difference of the number of *FP* sequences isolated from a single larva of *A. digitifera* and multiple larvae of *A. tenuis* originated from five colonies may reflect the variation of the expression difference of *FP* genes or that of *FP* genes among individuals of *A. tenuis* larvae.

Based on the number of *FP* sequences from an adult *A. digitifera* cDNA and the fact that it is a diploid organism (Shinzato, et al. 2011), I estimated that a minimum of four MWE,

three M/LWE, and five CP genes are present in the genome of an adult individual, considering that all adult sequences are allelic variations. Next, I estimated *FP* gene copy numbers in four adult *A. digitifera* specimens by qPCR. The total gene numbers per *A. digitifera* genome were two times higher than the previously reported largest set of *FP* genes, 16 *GFP*-like genes, in the amphioxus genome (Bomati, et al. 2009). These high copy numbers in *A. digitifera* genome could account for the number of *FP* sequences determined from cDNAs. The differences between copy numbers, with low standard deviation between experimental replicates (Figure 2.4; <1 copy in each clade), indicated the presence of copy number variations between individuals. The S/MWE clade contained the highest number of *FP* genes, which was in agreement with the major fluorescence emission from FPs in the S/MWE clade, with a peak around 500 nm, from live *A. digitifera* (Figure 2.1), although the excitation light (LED: 448 nm spectrum peak and laser: 452 nm spectrum) matched with FPs in the S/MWE clade ( $\lambda_{\text{abs}} = 461\text{-}508$  nm). Hence, the *A. digitifera* *FP* genes form a multi-gene family that is larger than previously reported for other organisms, and this *FP* multi-gene family could comprise the genetic basis of coral fluorescence.

My homology search for *FP* genes in the *A. digitifera* genomic sequence data supports the existence of a multi-gene *FP* family as well, since multiple genes were detected in the assembled genome sequences. High copy number genes with high similarity are difficult to assemble (Mariano, et al. 2015). Therefore, many genes identified in the assembled genomic sequence data were incomplete due to un-assembled genomic regions, suggesting a possibility that some *FP* genes were missing from the assembly. This possibility can be evaluated by focusing on high coverage regions in short read mapping. If a proportion of multi-copy genes is

not individually assembled, and at least one gene copy is identified in the genomic sequence, the corresponding genomic region should be mapped with high coverage by short reads. Indeed, high coverage regions were detected in all three major clades, suggesting higher copy numbers of each major clade gene than those estimated from the assembled genomic sequence. The copy numbers of *FP* genes in the M/LWE clade estimated by qPCR were equal to those in the genome sequence data. The high coverage regions in M/LWE clade genes may stem from many un-assembled regions in those genes.

Fluorescence spectra measured for purified FPs revealed that the newly isolated *FP* sequences could be categorized into three clades, S/MWE, M/LWE, or CP. The FPs in S/MWE and M/LWE clades of fluorescence can cover the *A. digitifera* fluorescence. The purified proteins encoded by *CP* sequences absorbed the light, and proteins encoded by CP clade may affect the excitation and emission light of FPs in S/MWE and M/LWE clades by absorbing the light. Accordingly, a multi-*FP* gene family could generate the summative coral fluorescence.

***Multi-member FP gene family has evolved with functional diversity in the genus Acropora***

To examine whether *FP* genes of other *Acropora* species also comprise a multi-gene family, I cloned and determined *FP* sequences from *A. tenuis* cDNA. As shown in Figure 2.5, I identified 28 sequences in the S/MWE clade, ten sequences in the M/LWE clade, and six sequences in the CP clade from an adult and larval *A. tenuis* cDNAs. This result suggests that *FP* genes from *A. tenuis* also form a multi-gene family composed of three major clades. Phylogenetically, *A. tenuis* is located at the basal lineage in the genus *Acropora* (Fukami, et al. 2000; Richards, et al. 2013), suggesting that a multi-member *FP* gene family of three major clades existed in the

common ancestor of *Acropora* species. The estimation of eight copies of *A. millepora RFP* gene using exon 3 sequence (Gittins, et al. 2015) also supports the existence of a multi-*FP* gene family in the genus *Acropora*. Subclades 1 and 2 in MWE1 clade comprised *A. digitifera* and *A. tenuis* sequences, which indicates that an emergence of these two subclades occurred before the divergence of these species. The nucleotide divergence between the two subclades (0.068) was slightly greater than the mean nucleotide divergence between the two species (0.065), and is in agreement with the older divergence of *FP* genes than the species divergence.

Since the *FP* genes are tandemly arrayed in the *A. digitifera* genome (Shinzato, et al. 2012), I tested the possibility of recombination (unequal crossing over) between the *FP* sequences. I analyzed synonymous mutations in *FP* sequences in both species to avoid the effect of functional convergence. For closely related lineages, independent synonymous substitutions are generally thought as a rare event due to very low mutation rate. Sharing patterns of synonymous mutations in each clade were different between sequences of the two species, suggesting recombination events. Recombination of sequences between species may have occurred in the genome of the common ancestral *Acropora* species. Recombination events were also supported by four-gamete tests.

To reveal the function of FPs in the ancestral lineages, I estimated the ancestral sequences from MWE1, S/MWE and CP clades, and a common ancestor of the two subclades in M/LWE clade in *A. digitifera*. Fluorescence spectra of the purified ancestral proteins were categorized into SWE, CP, and LWE spectra, suggesting that different *FP* functions have been already acquired in the ancestral lineages of each clade. Since the ancestral lineages of each clade could be traced back to the common ancestor of the *Acropora* genus, the ancestral species

of this genus may have already possessed a functionally diverse *FP* gene set. *FP* fluorescence is mainly determined by three amino acid residues known as chromophores (Henderson and Remington 2005). M/LWE group sequences were split into two clades, an ancestral DYG chromophore-coding type and a derived TYG-chromophore-coding type. Fluorescence spectra of DYG- and TYG-chromophore sequences were categorized into LWE ( $\lambda_{em} = 611$  nm) and MWE ( $\lambda_{em} = 521$  nm), respectively. Mutating two amino acids, in the chromophore (T66D) and position 191 (S191P), shifted the *FP* emission peak from 521 to 593 nm, suggesting that *FP* sequences with TYG-type chromophores have evolved into MWE from an ancestral LWE. Although their expression was too low to allow cloning, I verified the existence of TYG-chromophore-type sequences in RNA-seq reads from *A. tenuis*, indicating that TYG chromophore has emerged before *A. digitifera* and *A. tenuis* divergence and persisted in these species. Hence, the multi-member *FP* gene family has evolved to maintain a functional diversity in *Acropora* species. The adults and larvae of *A. digitifera* and *A. tenuis* expressed *FP* sequences from each of three major clades. *FPs* from three major clades might possess an important biological function such as photo-protective and antioxidant roles in adults (Palmer, et al. 2009; Roth and Deheyn 2013) and possible roles in larval settlement behavior and long-range dispersal in larvae (Kenkel et al. 2011, Strader et al. 2016).

Similar to multi *FP* gene family in genus *Acropora*, at least five amphioxus *FP* genes in two species of genus *Branchiostoma* have maintained during their evolution (Yue, et al. 2016), indicating the importance of gene copy multiplicity of *FP* genes. Multi-gene families are present in genomes of many organisms, and the functional importance of gene copy multiplicity is well known (Walsh 2008). Compared with previous studies of multi-gene families, the most

enigmatic issues concerning the multi-*FP* gene family are (1) the purpose for multiple *FP* genes in corals, and (2) the biological role of these multiple copies. Considering the known roles for multi-gene families, the importance of multi-*FP* gene family may be of dual roles. The first role is associated with the amount of FPs in the tissues of *Acropora* species. Increased gene copy number can increase transcript levels, as, for example, for the gene family encoding ribosomal RNA (Weider, et al. 2005). Indeed, FPs comprise a high proportion of the total soluble protein content in anthozoan tissues (Leutenegger, et al. 2007; Oswald, et al. 2007). Production of the large amount of FPs may have been essential for survival during the evolution of *Acropora* species. The second role is linked with the distinct FP function. As shown in this thesis, the emissions of short-, middle- and long-wavelength light by FPs in S/MWE and M/LWE clades and absorption of the light by FPs in CP clade have been maintained during the evolution of *Acropora* species. These different functions encoded by *FP* genes in the genomes of two analyzed *Acropora* species may have been important during their evolution.

In this chapter, I identified the complement of *A. digitifera* *FP* genes. While the association between the amount of FP proteins, *FP* gene copy number, and the precise roles of these proteins remains unresolved, knowing the numbers of genes in each major FP clade and their different functions will facilitate the understanding of the biological roles of FPs in future studies.

This chapter is based on Takahashi-Kariyazono *et al.* (2016) published in Genome Biology and Evolution.

## References

- Alieva NO, et al. 2008.** Diversity and evolution of coral fluorescent proteins. PLoS One 3: e2680. doi: 10.1371/journal.pone.0002680
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ 1990.** Basic local alignment search tool. J Mol Biol. 215: 403-410.
- Baumann D, et al. 2008.** A family of GFP-like proteins with different spectral properties in lancelet *Branchiostoma floridae*. Biology direct 3: 28. doi: 10.1186/1745-6150-3-28
- Bomati EK, Manning G, Deheyn DD 2009.** Amphioxus encodes the largest known family of green fluorescent proteins, which have diversified into distinct functional classes. BMC Evol Biol 9: 77. doi: 10.1186/1471-2148-9-77
- D'Angelo C, et al. 2008.** Blue light regulation of host pigment in reef-building corals. Marine Ecology Progress Series 364: 97-106. doi: 10.3354/meps07588
- Deheyn DD, et al. 2007.** Endogenous Green Fluorescent Protein (GFP) in Amphioxus. . Biol Bull. 213: 95-100.
- Field SF, Bulina MY, Kelmanson IV, Bielawski JP, Matz MV 2006.** Adaptive evolution of multicolored fluorescent proteins in reef-building corals. J Mol Evol 62: 332-339. doi: 10.1007/s00239-005-0129-9
- Fukami H, Omori M, Hatta M 2000.** Phylogenetic Relationships in the Coral Family Acroporidae, Reassessed by Inference from Mitochondrial Genes. ZOOLOGICAL SCIENCE 17: 689-696
- Gish W, States DJ 1993.** Identification of protein coding regions by database similarity search. Nat Genet. 3: 266-272.
- Gittins JR, D'Angelo C, Oswald F, Edwards RJ, Wiedenmann J 2015.** Fluorescent protein-

mediated colour polymorphism in reef corals: multicopy genes extend the adaptation/acclimatization potential to variable light environments. *Mol Ecol* 24: 453-465. doi: 10.1111/mec.13041

**Henderson JN, Remington SJ 2005.** Crystal structures and mutational analysis of amFP486, a cyan fluorescent protein from *Anemonia majano*. *Proc Natl Acad Sci U S A* 102: 12712-12717. doi: 10.1073/pnas.0502250102

**Ho SN, Hunt HD, Horton RM, Pullen JK, Pease LR 1989.** Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* 77: 51-59. doi: [http://dx.doi.org/10.1016/0378-1119\(89\)90358-2](http://dx.doi.org/10.1016/0378-1119(89)90358-2)

**Johnsen S. 2012.** The optics of life. New Jersey: Princeton University Press.

**Kelmanson IV, Matz MV 2003.** Molecular basis and evolutionary origins of color diversity in great star coral *Montastraea cavernosa* (Scleractinia: Faviida). *Mol Biol Evol* 20: 1125-1133. doi: 10.1093/molbev/msg130

**Labas YA, et al. 2002.** Diversity and evolution of the green fluorescent protein family. *Proc Natl Acad Sci U S A* 99: 4256-4261. doi: 10.1073/pnas.062552299

**Leutenegger A, et al. 2007.** It's cheap to be colorful. *FEBS Journal* 274: 2496-2505. doi: 10.1111/j.1742-4658.2007.05785.x

**Mariano DC, et al. 2015.** MapRepeat: an approach for effective assembly of repetitive regions in prokaryotic genomes. *bioinformatics* 11: 276-279.

**Matz MV, et al. 1999.** Fluorescent proteins from nonbioluminescent Anthozoa species. *Nature Biotechnology* 17: 969 - 973.

**Mazel CH 2003.** Contribution of fluorescence to the spectral signature and perceived color of corals. *Limnology and Oceanography* 48: 390-401.

- Moisan TA, Mitchell BG 2001.** UV absorption by mycosporine-like amino acids in *Phaeocystis antarctica* Karsten induced by photosynthetically available radiation. *Marine Biology* 138: 217–227.
- Oswald F, et al. 2007.** Contributions of host and symbiont pigments to the coloration of reef corals. *FEBS J* 274: 1102–1109. doi: 10.1111/j.1742-4658.2007.05661.x
- Palmer CV, Modi CK, Mydlarz LD 2009.** Coral fluorescent proteins as antioxidants. *PLoS One* 4: e7298. doi: 10.1371/journal.pone.0007298
- Richards ZT, Miller DJ, Wallace CC 2013.** Molecular phylogenetics of geographically restricted *Acropora* species: Implications for threatened species conservation. *Molecular Phylogenetics and Evolution* 69: 837–851. doi: 10.1016/j.ympev.2013.06.020
- Roth MS, Deheyn DD 2013.** Effects of cold stress and heat stress on coral fluorescence in reef-building corals. *Scientific Reports* 3. doi: 10.1038/srep01421
- Roth MS, Fan TY, Deheyn DD 2013.** Life history changes in coral fluorescence and the effects of light intensity on larval physiology and settlement in *Seriatopora hystrix*. *PLoS One* 8: e59476. doi: 10.1371/journal.pone.0059476
- Shagin DA, et al. 2004.** GFP-like proteins as ubiquitous metazoan superfamily: evolution of functional features and structural complexity. *Molecular Biology and Evolution* 21: 841–850. doi: 10.1093/molbev/msh079
- Shimomura O 1979.** Structure of the chromophore of Aequorea green fluorescent protein. *Federation of European Biochemical Societies* 104: 220–222.
- Shinzato C, et al. 2011.** Using the *Acropora digitifera* genome to understand coral responses to environmental change. *Nature* 476: 320–323. doi: 10.1038/nature10249
- Shinzato C, Shoguchi E, Tanaka M, Satoh N 2012.** Fluorescent protein candidate genes in the coral *Acropora digitifera* genome. *Zoolog Sci* 29: 260–264. doi: 10.2108/zsj.29.260

**Suzuki G, Hayashibara T, Shirayama Y, Fukami H 2008.** Evidence of species-specific habitat selectivity of *Acropora* corals based on identification of new recruits by two molecular markers. Marine Ecology Progress Series 355: 149-159. doi: 10.3354/meps07253

**Tamura K, Stecher G, Peterson D, Filipski A, Kumar S 2013.** MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol. 30: 2725-2729. doi: 2710.1093/molbev/mst2197. Epub 2013 Oct 2716.

**Veron JEN. 2000.** Corals of the World. Townsville, Australia: Australian Institute of Marine Science.

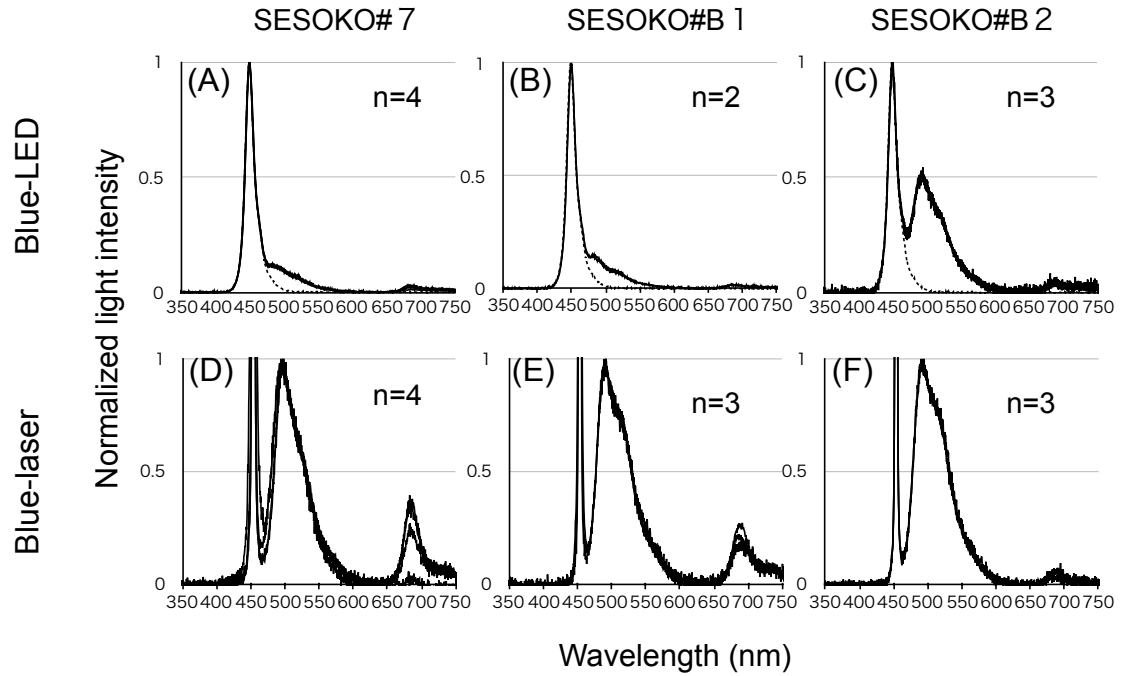
**Walsh JB, and Stephan, Wolfgang 2008.** Multigene Families: Evolution. eLS. doi: [doi: 10.1002/9780470015902.a0001702.pub2]

**Weider LJ, et al. 2005.** The Functional Significance of Ribosomal (r)DNA Variation: Impacts on the Evolutionary Ecology of Organisms. Annual review of ecology, evolution, and systematics 36: 219-242. doi: 10.1146/annurev.ecolsys.36.102003.152620

**Yue JX, Holland ND, Holland LZ, Deheyn DD 2016.** The evolution of genes encoding for green fluorescent proteins: insights from cephalochordates (amphioxus). Sci Rep 6: 28350. doi: 10.1038/srep28350

## Appendix

**Appendix Figure S2.1.** Positions and sequences of primers. **(A)** Positions of primers are indicated by arrows, below the schematic representations of *FP* sequences (*AdiFP2*, *AdiFP10L*, and *AdiFP8L*). Directions of the arrows indicate directionality of primers (5' to 3'). Primer names are given under each arrow. The initiation and terminal codons are indicated by vertical arrows. **(B)** Primer names and sequences.



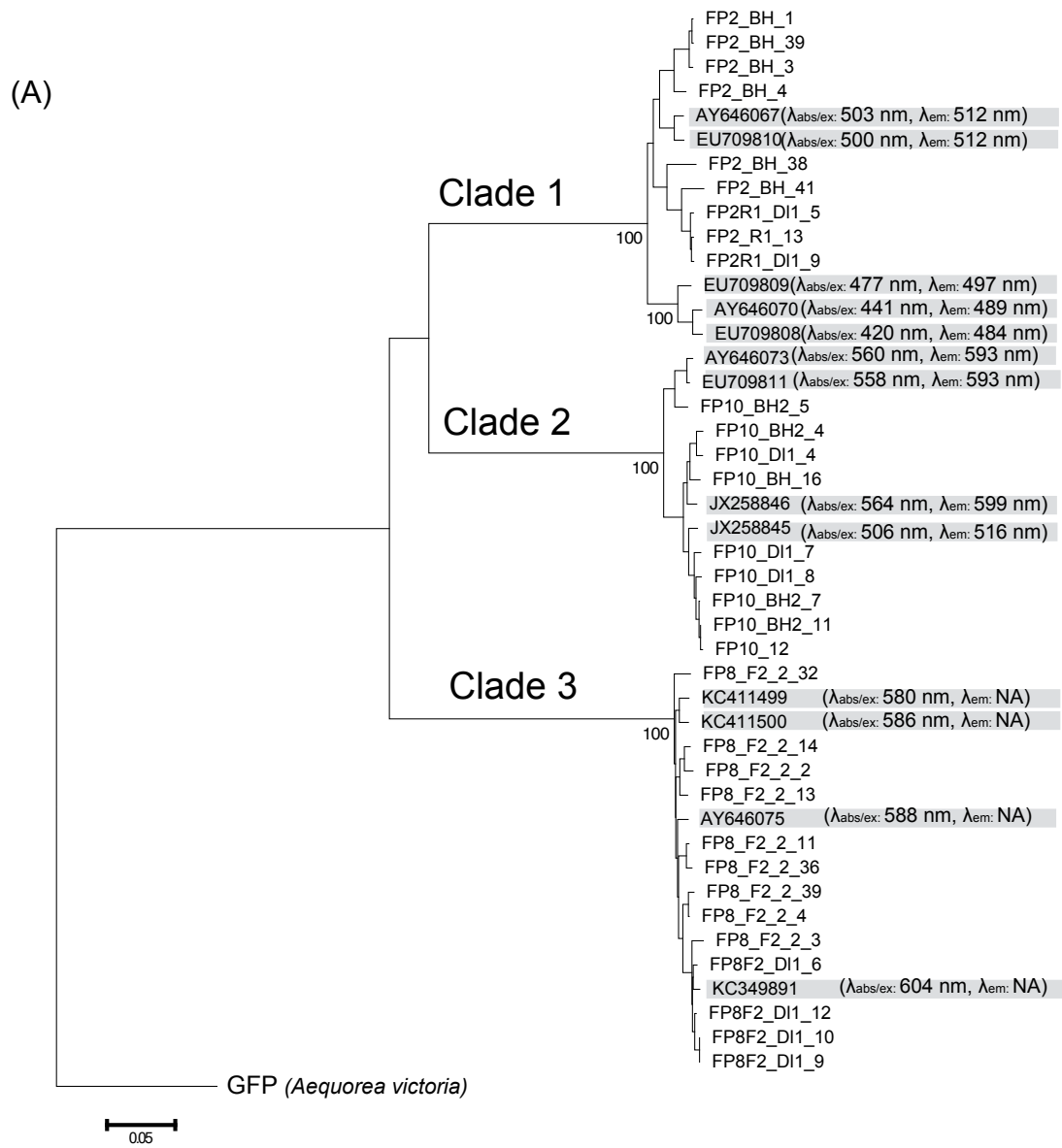
**Appendix Figure S2.2.** Live coral fluorescence measurements in the sea. Emission spectra (solid line) of three colonies of *A. digitifera* (A–F) and LED excitation spectra (dotted line, A–C) are shown. “n” indicates the number of measurements. Measurements from a single colony using the same excitation light are drawn in the same panel. Horizontal and vertical axes indicate wavelengths (nanometers) and normalized light intensities, respectively. Coral colony ID numbers are shown at the top.

```

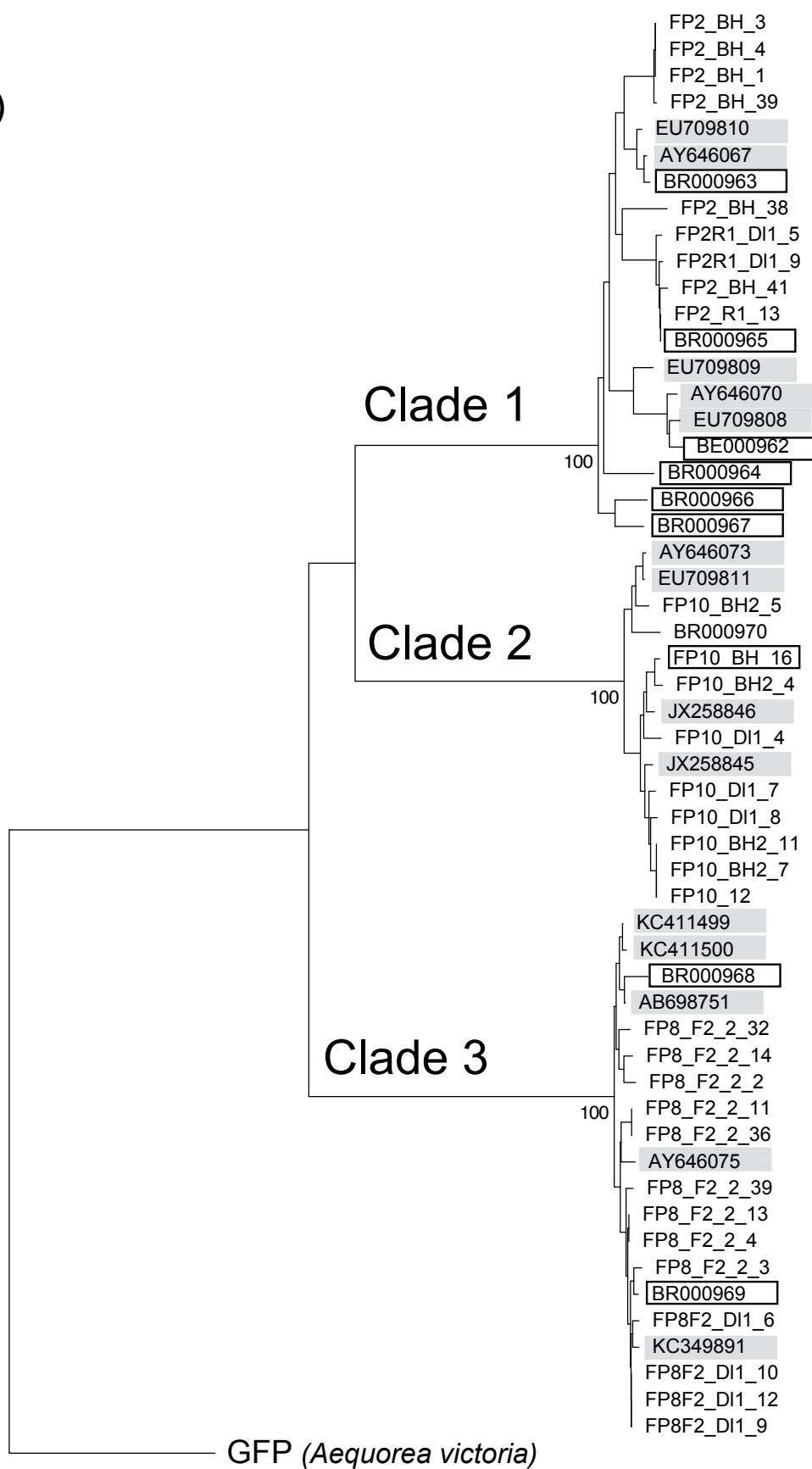
amilGFP(AY646067)  ATGTCCTTATTCAAAGCAAGGCATCGTACAAGAAATGAAGACGAAATACCATATGGAAGGC
AdiFP2 (BR000963)  ATGTCCTTATTCAAAGCAAGGCATCGTACAAGAAATGAAGACGAAATACCATATGGAAGGC
amilRFP(AY646073)  ATGGCTCTGTCAAAGCACGGTTTAACAAAGGACATGACGATGAAATACCACATGGAAGGG
AdiFP10(BR000970)  -----ATGGAAGGG
AdiFP10L          ATGGCTCTGTCAAAGCACGGTCTAACAAGGACATGACGATGAAATACCGGATGGAAGGG
amilCP (AY646075)  -----ATGAGTGTGATCGCTAAACAAATGACCTACAAGGTTTATATGTCAGGC
AdiFP8 (AB698751)  -----ATGACCTACAAGGTTTATATGTCAGGC
AdiFP8L          -----ATGAGTGTGATCGCTAAACAAATGACCTACAAGGTTTATATGTCAGGC

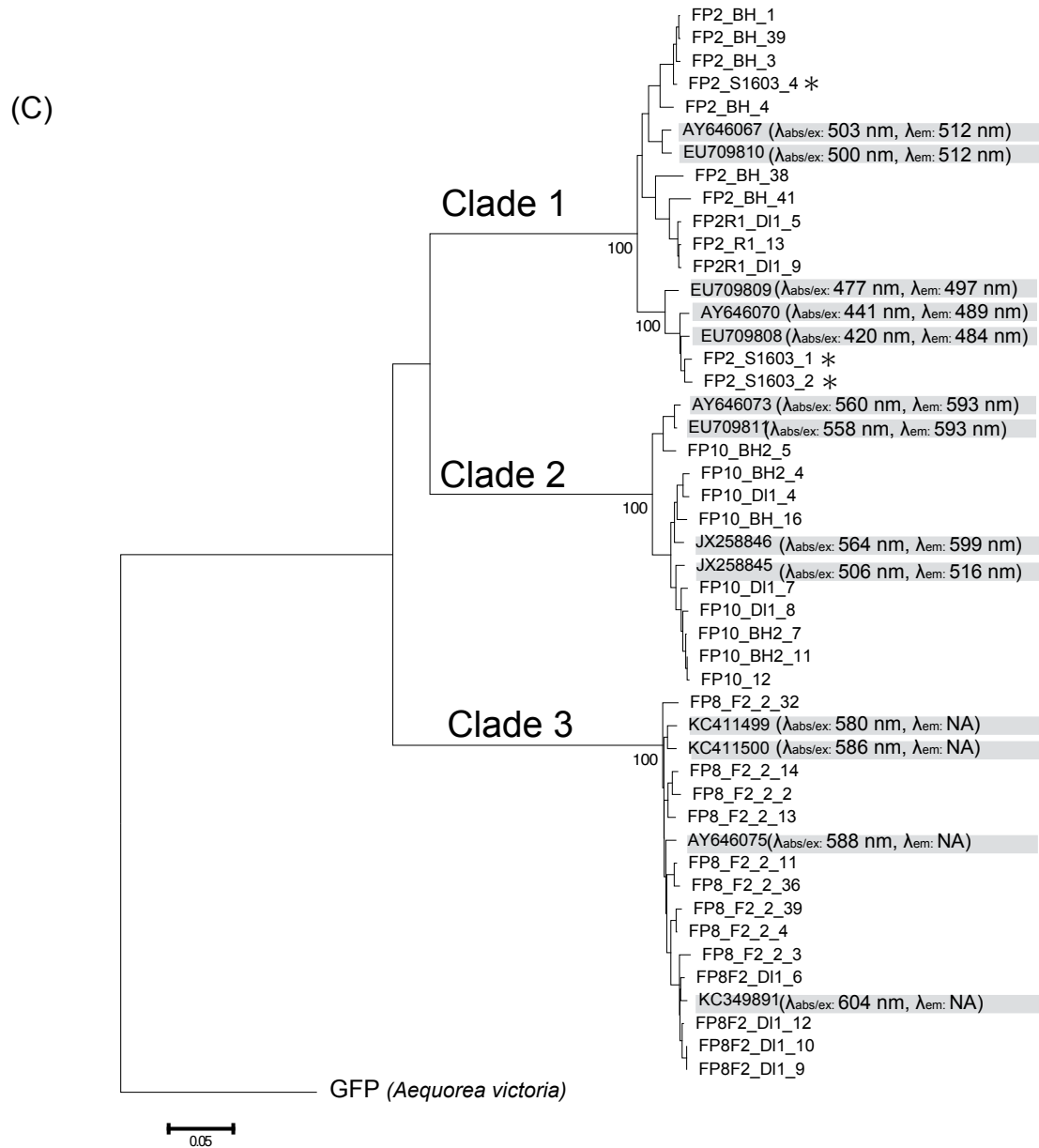
```

**Appendix Figure S2.3.** Alignment of extended 5' *FP* sequence termini. Alignment of 5' termini (60 bp) of *FP*-like sequences (*AdiFP2*, *AdiFP8*, and *AdiFP10*; Shinzato *et al.* 2012), known *FP* sequences with measured fluorescence (amilGFP, amilRFP, and amilCP: shown in gray), and extended 5' termini sequences (*AdiFP10L* and *AdiFP8L*). Extended sequence fragments are underlined. The initiation codon is highlighted in black.



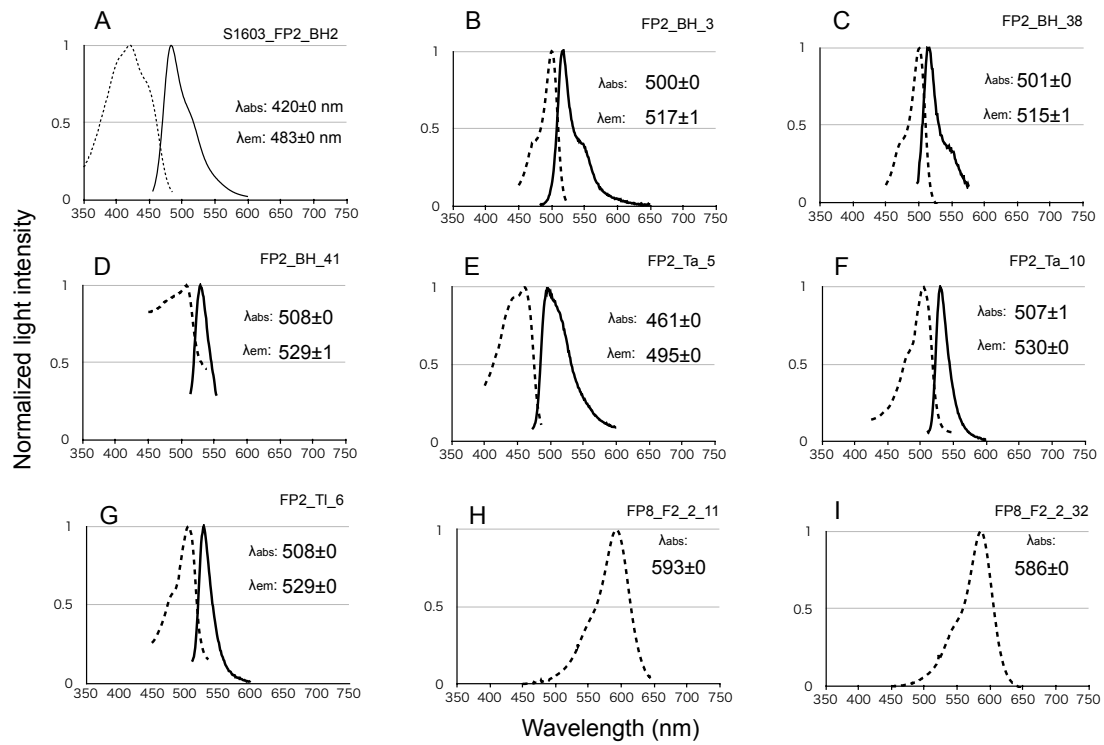
(B)





**Appendix Figure S2.4.** Phylogenetic relationships between newly identified *A. digitifera* FP sequences from an adult and larva with *A. millepora* FP sequences (A), with *A. digitifera* genome sequence data (B), and with three *A. digitifera* FP sequences from an additional adult colony (ID: S1603) (C). The tree was constructed using the neighbor-joining method with the p-distance method. Bootstrap probability for each node was obtained by 1,000 replicates and is shown next to each node. *A. victoria* GFP was used as an outgroup. The scale bar represents 0.05 substitutions per site. (A) *A. millepora* FP sequences from GenBank are shown in gray, with  $\lambda_{\text{abs/ex}}$  and  $\lambda_{\text{em}}$  values in brackets. The accession numbers and reference  $\lambda_{\text{abs/ex}}$  and  $\lambda_{\text{em}}$  values

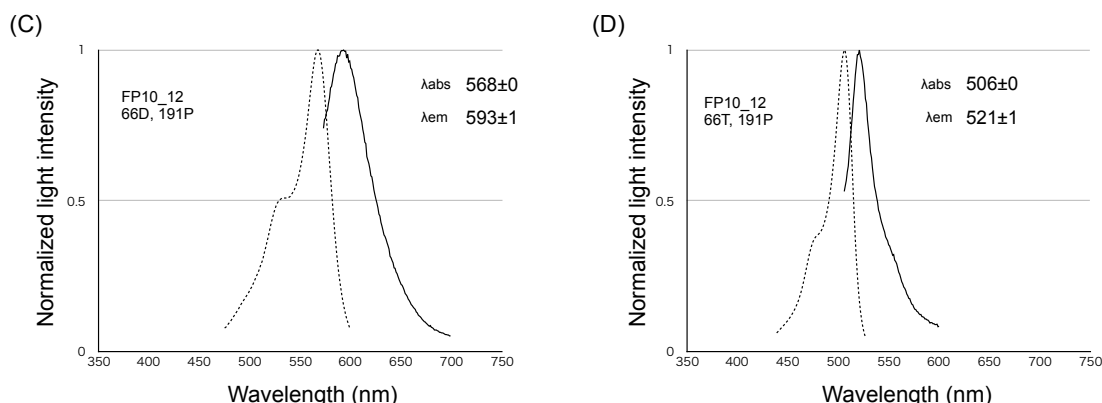
are as follows: AY646070, AY646067, AY646073, and AY646075 (Alieva, et al. 2008); EU709808, EU709809, EU709810, EU709811, KC411499, KC411500, and KC349891 (D'Angelo, et al. 2008); JX258846 (<http://www.ncbi.nlm.nih.gov/nuccore/475653237/>); and JX258845 (<http://www.ncbi.nlm.nih.gov/nuccore/JX258845>). (B) *FP* sequences from *A. digitifera* genomic sequence data (Shinzato, et al. 2012) are enclosed in rectangles. (C) The *FP* sequences from an additional adult colony (ID: S1603) were marked by asterisks.



**Appendix Figure S2.5.** Emission and absorption spectra of recombinant FP proteins. Absorption (dotted line) and emission (solid line) spectra of recombinant FP proteins encoded by S/MWE clade *A. digitifera* (A–D) and *A. tenuis* (E–G) sequences are presented. Absorption spectra (dotted line) of recombinant proteins from CP clade are also shown (H, I). FP sequence names are shown in the upper-right corner of each panel. Horizontal and vertical axes indicate wavelengths (nanometers) and normalized light intensities, respectively.

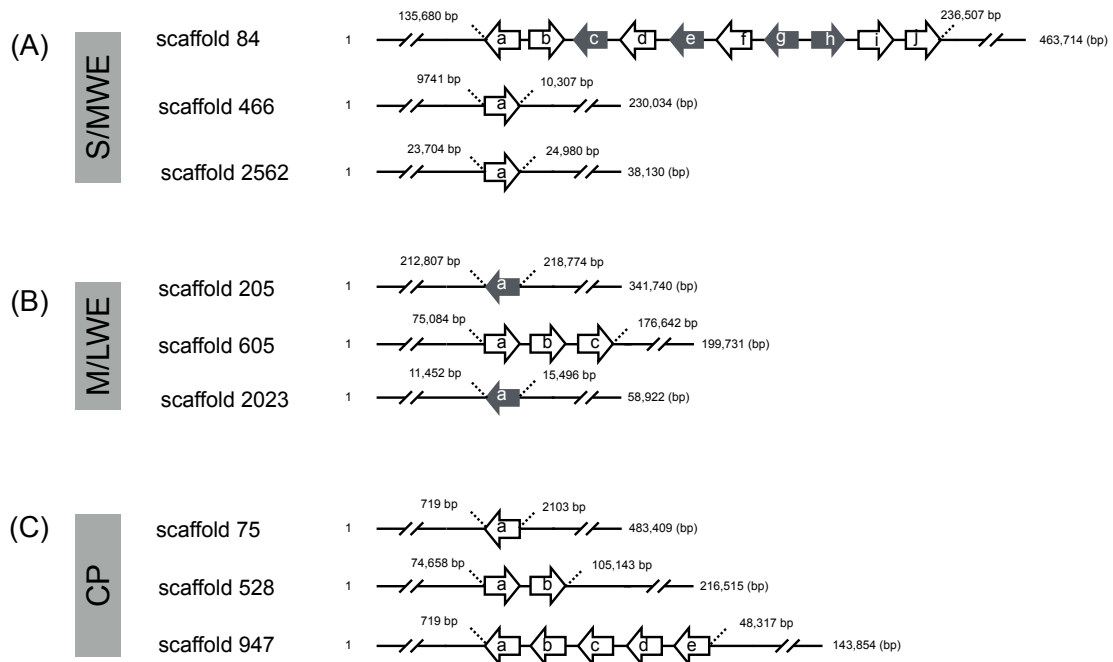
(B)

Nucleotide	1	111111112	222222223	333333334	444444445	555555556	666666667	777777778	888888889	999999999	000000000	111111111	111111111
	123456789	123456789	123456789	123456789	123456789	123456789	123456789	123456789	123456789	123456789	123456789	123456789	123456789
ami1RFP	MA	LKHLGTLK	DMTKMGHG	SVDGHKFVT	GHGNGPFEG	KOTNNLCVE	GGPLPFSEDI	LSAAFYDGNR	VTFTEYQGMV	DFPKNSCPAG	YTHWRSLEIF	DGAVCTSVLE	ITVSVEENCF
FP10_BH2_4		.....R.....	.....S.....	.....I.....	.....V.....	.....D.....	.....Q.....	.....A.....	.....A.....	.....A.....	.....A.....	.....A.....	.....A.....
FP10_12		.....R.....	.....C.....	.....S.....	.....I.....	.....V.....	.....Q.....	.....A.....	.....A.....	.....A.....	.....A.....	.....A.....	.....A.....
Nucleotide	111111111	111111111	111111111	111111111	111111111	111111111	111111111	111111111	111111112	222222222	222222222	222222222	222
	123456789	123456789	123456789	123456789	123456789	123456789	123456789	123456789	123456789	123456789	123456789	123456789	123
ami1RFP	TH	SKFGVGN	FPADGPMVK	MTTNWPEFS	KITLPVPRQI	LKGQVAMYL	LKDGGYRCQ	FSVTITAKSD	BSMPPEWHF	QHKLTREDS	DAKNQKQWL	ENAVASRAL	PG*
FP10_BH2_4		.....Y.....	.....C.....	.....V.....	.....V.....	.....V.....	.....V.....	.....V.....	.....S.....	.....S.....	.....A.....	.....A.....	.....A.....
FP10_12		.....Y.....	.....C.....	.....V.....	.....V.....	.....V.....	.....V.....	.....V.....	.....S.....	.....S.....	.....A.....	.....A.....	.....A.....

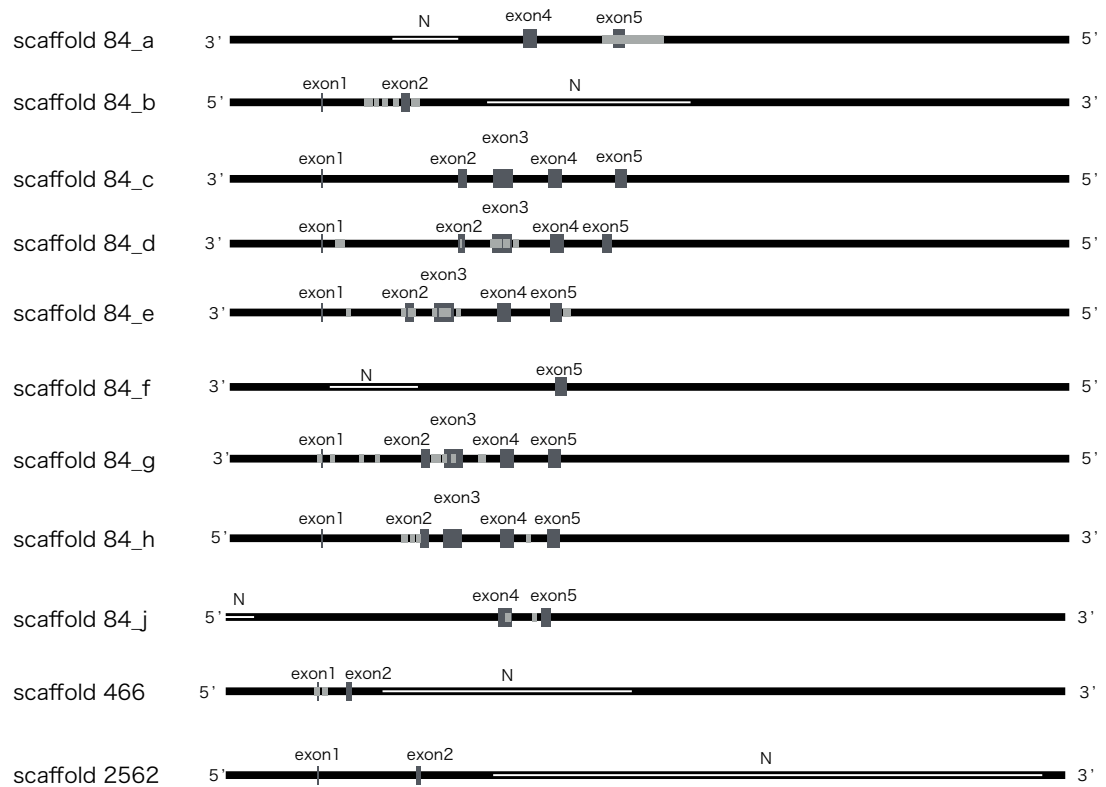


(A) Amino acid alignments of N-terminal sequence fragments from S/MWE (top), M/LWE (middle), and CP (bottom) clades. Amino acid sequences encoded by the newly identified *FP* genes are aligned with ancestral sequences of MWE1 clade (Ancestor-MWE), S/MWE clade

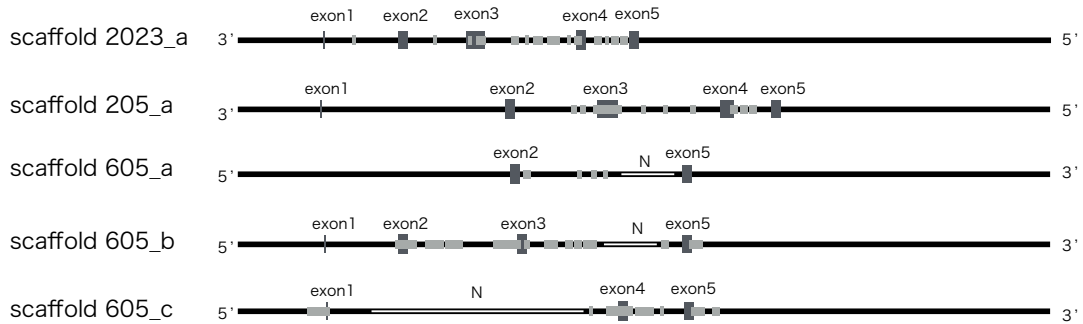
(Ancestor-S/MWE), two subclades of M/LWE (Ancestor-M/LWE), CP clade (Ancestor-CP), and the known *FP* genes. The three chromophore amino acids are colored. Asterisks indicate conserved amino acids. (B) Amino acid alignment of amilRFP, FP10\_BH2\_4, and FP10\_12. Mutated amino acids are shown inside gray boxes. (C) Emission and absorption spectra of recombinant FP proteins. Excitation (dotted line) and emission (solid line) spectra of recombinant proteins obtained after mutagenesis of FP10\_12: double (T66D, S191P; left) and single (S191P; right) mutants.



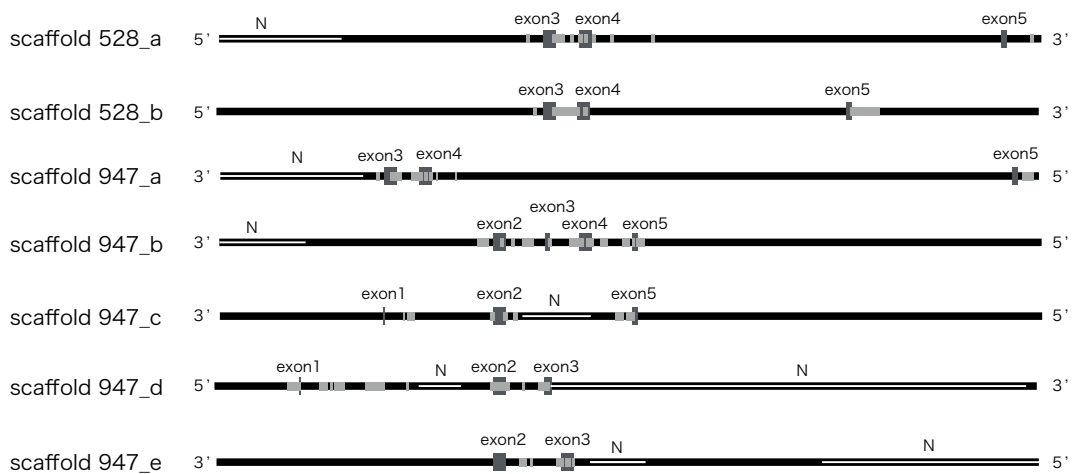
## (D) S/MWE



## (E) M/LWE



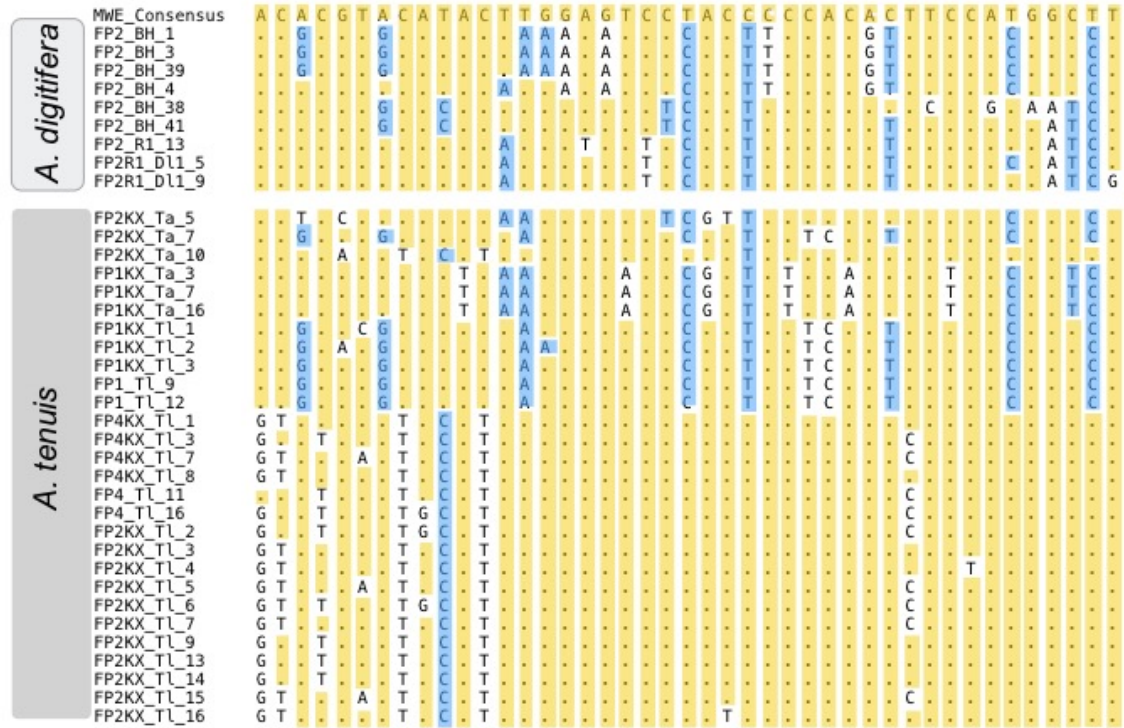
## (F) CP



**Appendix Figure S2.7.** Positions of *FP* genes and high coverage regions in *A. digitifera* genome. *FP* genes from S/MWE (A), M/LWE (B), and CP (C) clades are indicated by arrows. The arrowheads indicate *FP* gene directionality. Gray and open arrows indicate genes that possess full exon sets and genes with un-assembled regions, respectively. Predicted *FP* genes are named in alphabetical order in each scaffold. Letters within the arrows correspond to genomic regions D–F. *FP* gene cluster region positions are shown at both ends. Schematic representations of *FP* genes in S/MWE (D), M/LWE (E), and CP (F) clades are also shown. Exons and high coverage regions are indicated by dark and light gray boxes, respectively. Open boxes designated by “N” indicate un-assembled regions of the genome.

(A)

MWE1



(B)

M/LWE

M/LWE		1	1	1	1	1	1	1	1	1	2	3	4	5	5	5	6	6	6	6	6		
		8	9	1	1	2	2	2	4	4	5	6	6	5	7	2	3	9	1	1	2	4	5
		4	0	1	4	0	3	9	1	7	3	5	7	1	7	5	4	1	2	8	1	2	1
M/LWE_Consensus		G	G	T	C	G	A	T	T	T	G	A	T	G	G	T	C	G	G	C	G	G	A
A. digitifera	FP10_12	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.
	FP10_BH_16	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	A	.
	FP10_BH2_11	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	FP10_BH2_4	.	.	.	T	.	.	.	.	.	C	.	.	.	.	G	.	A	.	.	.	A	.
	FP10_BH2_5	.	A	.	.	.	.	.	.	.	.	A	.	.	.	.	T	.	.	.	.	.	.
	FP10_BH2_7	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	FP10_DL1_4	.	.	.	T	.	.	.	.	.	C	.	.	.	.	G	.	.	.	.	.	.	G
	FP10_DL1_7	T	.	.	.	.	.	.	.	.	.	A	.	.	A	.	A	.	.	.	.	.	.
FP10_DL1_8	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	G	
A. tenuis	FP10KX_TL_1	C	.	C	.	A	.	.	.	.	.	.	.	A	A	.	.	.	.	.	G	T	.
	FP10KX_TL_2	C	.	.	.	G	.	.	.	.	C	.	.	C	.	.	.	A	.	.	T	.	.
	FP10KX_TL_3	.	.	.	.	A	.	.	.	.	C	.	.	C	.	.	.	.	.	.	T	.	.
	FP10KX_TL_6	.	.	.	T	.	G	.	.	.	C	.	.	C	A	.	.	.	.	G	T	.	.
	FP10KX_TL_8	C	.	C	.	G	.	.	.	.	C	.	.	A	A	.	.	.	.	G	T	.	.
	FP10KX_TL_9	C	.	.	.	.	.	.	.	.	C	.	.	C	A	.	.	.	.	G	T	.	.
	FP10KX_TL_12	T	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.
	FP10KX_TL_13	.	.	.	T	.	G	.	.	.	C	.	.	C	.	.	.	.	.	G	T	.	.
	FP10KX_TL_15	.	.	.	T	.	G	.	.	.	C	.	.	C	A	.	.	.	.	G	T	.	.
	FP10KX_TL_16	C	.	C	.	G	.	.	.	.	.	.	.	A	A	.	.	.	.	G	T	.	.

(C)

CP

CP		1	1	1	1	1	1	1	2	2	2	2	3	3	3	3	3	4	4	4	4	4	5	5	6
		1	2	3	4	5	5	9	4	5	6	6	3	6	9	9	0	1	3	6	9	9	0	1	5
		1	9	8	7	0	3	4	3	5	1	4	0	9	0	3	5	7	5	8	2	3	1	0	1
A. digitifera	CP_Consensus	T	G	C	C	G	C	G	C	G	C	G	T	G	T	C	T	G	C	G	T	C	A	T	C
	FP8_F2_2_11	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	C	.
	FP8_F2_2_13	.	.	G	A	.	.	A	.	.	.	.	.	.	.	.	G	.	G	.	.	T	.	C	.
	FP8_F2_2_14	.	.	G	.	.	.	A	.	.	.	.	.	.	G	.	G	.	G	.	A	C	.	C	.
	FP8_F2_2_2	.	.	G	A	.	.	A	.	.	.	.	.	G	.	.	G	.	.	A	C	.	.	C	A
	FP8_F2_2_3	.	A	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	C	A
	FP8_F2_2_32	.	.	.	.	A	.	.	T	.	.	.	.	.	.	.	.	.	.	.	C	.	.	C	.
	FP8_F2_2_36	.	.	G	A	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.
	FP8_F2_2_39	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	C	.
	FP8_F2_2_4	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	C	.
	FP8F2_Dl1_10	.	A	.	.	.	.	.	T	.	.	.	.	.	.	.	G	.	.	.	.	.	.	C	.
	FP8F2_Dl1_12	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	C	.
FP8F2_Dl1_6	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	G	A	.	.	.	.	.	C	.	
FP8F2_Dl1_9	.	A	.	.	.	.	.	T	.	.	.	.	.	.	.	G	.	.	.	.	.	.	C	.	
A. tenuis	FP8KX_Ta_1	.	.	.	.	.	.	C	.	.	.	T	T	.	.	.	.	.	.	.	.	.	G	.	.
	FP8KX_Ta_2	C	.	.	.	A	T	T	.	A	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.
	FP8KX_Ta_4	.	.	.	.	.	.	C	.	.	T	T	.	.	A	.	.	.	.	.	.	.	G	.	.
	FP8KX_Ta_11	.	.	.	.	.	.	C	.	.	T	T	C	.	.	.	.	.	.	.	.	.	G	.	.
	FP8KX_Tl_4	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.
	FP8KX_Tl_17	C	.	.	.	A	T	T	.	A	.	.	.	A	.	.	.	.	.	.	.	.	G	.	.

**Appendix Figure S2.8.** Variable synonymous sites in *FP* sequences. Variable synonymous sites in *FP* sequences from MWE1 (A), M/LWE (B), and CP (C) clades are presented. Nucleotides identical with the consensus (top) are indicated by dots and are colored in yellow. Variable nucleotides shared by *A. digitifera* and *A. tenuis* sequences are colored in blue. Variable nucleotides in the sequences of only one of the two species are uncolored. Nucleotide positions are shown at the top.

(A) *A. digitifera*: MWE1

```

111111|1111122222|2233333334|4444445555|55666666
233600023|5689911125|6901237792|3457990446|78001257
7678256822|5203745658|3164703701|1832342567|39092668

7 0 *** * * * * * * * * * * * * * * * * * * * *
26 0 * * * * * * * * * * * * * * * * * * * *
37 0 * * * * * * * * * * * * * * * * * * * *
38 0 * * * * * * * * * * * * * * * * * * * *
62 0 * * * * * * * * * * * * * * * * * * * *
105 0 * * * * * * * * * * * * * * * * * * * *
106 0 * * * * * * * * * * * * * * * * * * * *
108 0 * * * * * * * * * * * * * * * * * * * *
122 0 * * * * * * * * * * * * * * * * * * * *
132 0 * * * * * * * * * * * * * * * * * * * *
155 0 * * * * * * * * * * * * * * * * * * * *
162 0 * * * * * * * * * * * * * * * * * * * *
180 0 * * * * * * * * * * * * * * * * * * * *
193 0 * * * * * * * * * * * * * * * * * * * *
197 0 * * * * * * * * * * * * * * * * * * * *
214 0 * * * * * * * * * * * * * * * * * * * *
215 0 * * * * * * * * * * * * * * * * * * * *
216 0 * * * * * * * * * * * * * * * * * * * *
225 0 * * * * * * * * * * * * * * * * * * * *
258 0 * * * * * * * * * * * * * * * * * * * *
263 0 * * * * * * * * * * * * * * * * * * * *
291 0 * * * * * * * * * * * * * * * * * * * *
306 0 * * * * * * * * * * * * * * * * * * * *
314 0 * * * * * * * * * * * * * * * * * * * *
327 0 * * * * * * * * * * * * * * * * * * * *
330 0 * * * * * * * * * * * * * * * * * * * *
373 0 * * * * * * * * * * * * * * * * * * * *
377 0 * * * * * * * * * * * * * * * * * * * *
390 0 * * * * * * * * * * * * * * * * * * * *
421 0 * * * * * * * * * * * * * * * * * * * *
431 0 * * * * * * * * * * * * * * * * * * * *
448 0 * * * * * * * * * * * * * * * * * * * *
453 0 * * * * * * * * * * * * * * * * * * * *
472 0 * * * * * * * * * * * * * * * * * * * *
493 0 * * * * * * * * * * * * * * * * * * * *
494 0 * * * * * * * * * * * * * * * * * * * *
502 0 * * * * * * * * * * * * * * * * * * * *
545 0 * * * * * * * * * * * * * * * * * * * *
546 0 * * * * * * * * * * * * * * * * * * * *
567 0 * * * * * * * * * * * * * * * * * * * *
573 0 * * * * * * * * * * * * * * * * * * * *
589 0 * * * * * * * * * * * * * * * * * * * *
600 0 * * * * * * * * * * * * * * * * * * * *
609 0 * * * * * * * * * * * * * * * * * * * *
612 0 * * * * * * * * * * * * * * * * * * * *
626 0 * * * * * * * * * * * * * * * * * * * *
656 0 * * * * * * * * * * * * * * * * * * * *
678 0 * * * * * * * * * * * * * * * * * * * *

# of nontrivial sites = 48

```

(B) *A. digitifera*: M/LWE

```

111112444|5555666
6145992679|2579459
3473672879|5041214

63 0 * * * * *
114 0 * * * * *
147 0 * * * * *
152 0 * * * * *
196 0 * * * * *
197 0 * * * * *
222 0 * * * * *
468 0 * * * * *
477 0 * * * * *
499 0 * * * * *
525 0 * * * * *
550 0 * * * * *
574 0 * * * * *
591 0 * * * * *
642 0 * * * * *
651 0 * * * * *
694 0 * * * * *

# of nontrivial sites = 17

```

(C) *A. digitifera*: CP

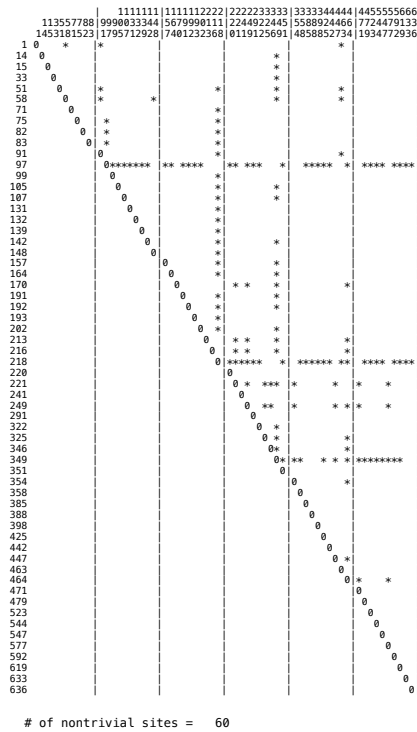
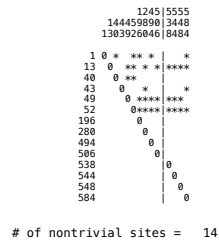
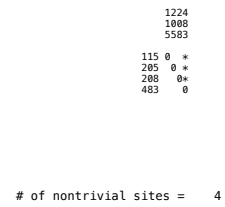
```

111111111|1222223344|4455666
2311222335|8136791922|6925255
7579026587|2910091347|0831835

27 0 * * * * *
35 0 * * * * *
117 0 * * * * *
119 0 * * * * *
120 0 * * * * *
122 0 * * * * *
126 0 * * * * *
135 0 * * * * *
138 0 * * * * *
157 0 * * * * *
182 0 * * * * *
219 0 * * * * *
231 0 * * * * *
260 0 * * * * *
270 0 * * * * *
299 0 * * * * *
311 0 * * * * *
393 0 * * * * *
424 0 * * * * *
427 0 * * * * *
460 0 * * * * *
498 0 * * * * *
523 0 * * * * *
551 0 * * * * *
628 0 * * * * *
653 0 * * * * *
655 0 * * * * *

# of nontrivial sites = 27

```

(D) *A. tenuis*: MWE1(E) *A. tenuis*: M/LWE(F) *A. tenuis*: CP

**Appendix Figure S2.9.** *FP* sequence four-gamete tests. Results of four-gamete tests with *FP* sequences from MWE1 (A), M/LWE (B), and CP (C) clades of *A. digitifera*, and MWE1 (D), M/LWE (E), and CP (F) clades of *A. tenuis*, are shown. Variable site positions are shown at the top and in the left columns. When a recombination event was detected between top and left column sites, the crossing point of the two sites was marked by an asterisk.

Appendix Table S2.1. The specimen information

Sample name	Species	Life stage	Experiment
SESOKO#B1	<i>A. digitifera</i>	adult	fluorescence measurements
SESOKO#B2	<i>A. digitifera</i>	adult	fluorescence measurements
SESOKO#3	<i>A. digitifera</i>	adult	fluorescence measurements
SESOKO#4	<i>A. digitifera</i>	adult	fluorescence measurements
SESOKO#7	<i>A. digitifera</i>	adult	fluorescence measurements
Adigi1	<i>A. digitifera</i>	adult	RNA-seq analysis and FP sequence isolation from cDNA
Adigi2	<i>A. digitifera</i>	adult	FP gene copy number estimation from genomic DNA
Adigi3	<i>A. digitifera</i>	adult	FP gene copy number estimation from genomic DNA
Adigi4	<i>A. digitifera</i>	adult	FP gene copy number estimation from genomic DNA
Adigi5	<i>A. digitifera</i>	adult	FP gene copy number estimation from genomic DNA
S1603	<i>A. digitifera</i>	adult	FP sequence isolation from cDNA
DI1	<i>A. digitifera</i>	larva	FP sequence isolation from cDNA
DI	<i>A. digitifera</i>	larvae	RNA-seq analysis
Ta	<i>A. tenuis</i>	adult	RNA-seq analysis and FP sequence isolation from cDNA
TI	<i>A. tenuis</i>	larvae	RNA-seq analysis and FP sequence isolation from cDNA

Appendix Table S2.2. The average coverage of read mapping.

Total number of reads	Mapped sequence	Length	Average coverage
<i>A. digitifera</i> larvae	<i>AdiFP10L</i>	699	93973
80,164,640	<i>AdiFP8L</i>	666	397662
	<i>AdiFP2</i>	696	18032
<i>A. digitifera</i> adult	<i>AdiFP10L</i>	699	37
94,059,356	<i>AdiFP8L</i>	666	194
	<i>AdiFP2</i>	696	37
<i>A. tenuis</i> larvae	<i>AdiFP10L</i>	699	50167
84,018,960	<i>AdiFP8L</i>	666	313
	<i>AdiFP2</i>	696	50805
<i>A. tenuis</i> adult	<i>AdiFP10L</i>	699	113
88,821,670	<i>AdiFP8L</i>	666	1071
	<i>AdiFP2</i>	696	203

## Appendix text

&gt;CP\_clade\_ancestor

ATGAGTGTGATCGCTAAACAAATGACCTACAAGGTTTATATGTCAGGCACGGT  
CAATGGACACTACTTTGAGGTCGAAGGCGATGGAAAAGGAAAGCCTTACGA  
GGGGGAGCAGACGGTAAAGCTCACTGTCACCAAAGGCGGACCTCTGCCATT  
TGCTTGGGATATTTTATCACCACAGTGTGAGTACGGAAGCATAACCATTACCA  
AGTACCCTGAAGACATCCCTGATTATGTAAAGCAGTCATTCCCGGAGGGATAT  
ACATGGGAGAGGATCATGAACTTTGAAGATGGTGCAGTGTGTACTGTCAGCA  
ATGATTCCAGCATCCAAGGCAACTGTTTCATCTACCATGTCAAGTTCTCTGGT  
TTGAACTTTCCTCCCAATGGACCTGTTATGCAGAAGAAGACACAGGGCTGGG  
AACCCAACACTGAGCGTCTCTTTGCACGAGATGGAATGCTGATAGGAAACAA  
CTTTATGGCCCTGAAGTTAGAAGGAGGCGGTCACTATTTGTGTGAATTCAAAT  
CTACTTACAAGGCAAAGAAGCCTGTGAAGATGCCAGGGTATCACTATGTTGA  
CCGCAAACCTGGATGTAACCAATCACAACAAGGATTACACTTCCGTTGAGCAG  
TGTGAAATTTCCATTGCACGCAAACCTTTGGTCGCC

&gt;MWE\_subclade\_ancestor

ATGTCTTATTCAAAGCAAGGCATCGCACAAGAAATGCGGACGAAATACCATA  
TGGAAGGCAGTGTCAATGGCCATGAATTCACGATCGAAGGTGTAGGAACTGG  
AAACCCTTACGAAGGGAAACAGATGTCCGAATTAGTGATCATCAAGCCTGCG  
GGAAAACCCCTTCCATTCTCCTTTGACATACTGTCAACAGCCTTTCAATATGG  
AAACTTATGCTTCACTAAGTACCCTGCAGACATGCCTGACTATTTCAAGCAAG  
GATTCCCAGATGGAATGTCATATGAAAGGTCATTTCTATTTGAGGATGGAGGA

GTTGCTACAGCTAGCTGGAACATTCGTCTCGAAGGAAATTGCTTCATCCACAA  
TTCCATCTTTCATGGCGTAAACTTTCCCGCTGATGGACCCGTAATGAAAAAGA  
AGACAATTGGCTGGGATAAGTCCTTCGAAAAAATGACTGTGTCTAAAGAGGT  
GTTAAGAGGTGATGTGACTCAGTTTCTTATGCTCGAAGGAGGTGGTTACCAC  
AGATGCCAGTTTCACTCCACCTACAAAACAGAGAAGCCGGTTACACTGCCCC  
CGAATCATGTCGTAGAACATCACATTGTGAGGACTGACCTTGGCCAAACTGC  
AAAAGGCTTCACAGTCAAGCTGGAAGAACATGCTGCGGCTCATGTTAATCCT  
TTGAAGGTAAA

>S/MWE\_clade\_ancestor

ATGTCTTATTCAAAGCAAGGCATCGCACAAGAAATGCGGACGAAATACCATA  
TGGAAGGCAGTGTCAATGGCCATGAATTCACGATCGAAGGTGTAGGAACTGG  
AAACCCTTACGAAGGGAAACAGATGTCCGAATTAGTGATCATCAAGCCTGCA  
GGAAAACCCCTTCCATTCTCCTTTGACATTCTGTCAACAGTCTTTCAATATGG  
AAACAGGTGCTTCACAAAGTACCCTGAAGGAATGCCTGACTATTTCAAGCAA  
GCATTCCCAGATGGAATGTCATATGAAAGGTCATTTCTATTTGAGGATGGAGG  
AGTTGCTACAGCTAGCTGGAACATTCGTCTCGAAGGAAATTGCTTCATCCAC  
AATTCCATCTTTCATGGCGTAAACTTTCCCGCTGATGGACCCGTAATGAAAAA  
GAAGACAATTGGCTGGGATAAGTCATTTCGAAAAAATGACCGTGTCCAAGGA  
GGTGTTAAGAGGTGATGTGACTATGTTTCTTATGCTCGAAGGAGGTGGTTACC  
ACAGATGCCAGTTTCACTCCACTTACAAAACAGAAAAGCCGGTTACACTGCC  
CCCGAATCATGTCGTAGAACATCACATTGTGAGGACTGACCTTGGCCAAAGT

GCAAAAGGCTTCACAGTCAAGCTGGAAGAACATGCTGCGGCTCATGTTAAC  
CCTTTGAAGGTAAATAA

>M/LWE\_subclade\_ancestor

ATGGCTCTGTCAAAGCACGGTCTAACAAAGGACATGACGATGAAATACCGGA  
TGGAAGGGTGTGTTCGATGGGCATAAATTTGTGATCACGGGCCACGGCAATGG  
AAGTCCTTTTGAAGGGAAACAGACTATCAATCTGTGTGTGGTTGAAGGGGG  
ACCCCTGCCATTCTCCGAAGACATTTTGTCTGCTGTGTTTGACTACGGAAACA  
GGGTCTTCACTGAATATCCTCAAGGCATGGTTGACTTTTTCAAGAATTCATGT  
CCAGCTGGATACACATGGCAAAGGTCTTTACTCTTTGAAGATGGAGCAGTTT  
GCACAGCCAGTGCAGATATAACAGTGAGTGTTGAGGAGAACTGCTTTTATCA  
CGAGTCCAAGTTTCATGGAGTGAACTTTCCTGCTGATGGACCTGTGATGAAA  
AAGATGACAACATAATTGGGAGCCATGCTGCGAGAAAATCATACCAGTACCTA  
GACAGGGGATATTGAAAGGGGATGTCGCCATGTACCTCCTTCTGAAGGATGG  
TGGGCGTTACCGGTGCCAGTTCGACACAGTTTACAAAGCAAAGACTGACCC  
GAAAAAGATGCCGGAGTGGCACTTCATCCAACATAAGCTCACCCGGGAAGA  
CCGCAGCGATGCTAAGAACCAGAAATGGCAACTGGCAGAACATTCTGTTGCT  
TCCCGATCCGCATTGCCCTGATAA

# Chapter 3

## **Presence-absence polymorphisms of highly expressed *FP* sequences**

## Abstract

Despite many hypotheses regarding the roles of fluorescent proteins (FPs), their biological roles and the genetic basis of FP-mediated color polymorphisms in *Acropora* remain unclear. In this chapter, I determined the genetic mechanism underlying fluorescent polymorphisms in *A. digitifera*. Using a high-throughput sequencing approach, I found that *FP* gene sequences in *FP* multi-gene family exhibit presence–absence polymorphism among individuals. A few particular sequences in short-to-middle wavelength emission (S/MWE) and middle-to-long wavelength emission (M/LWE) clades were highly expressed in adults, and different sequences were highly expressed in larvae. These highly expressed sequences were absent in the genomes of individuals with low total *FP* gene expression. In adults, presence–absence differences of the highly expressed *FP* sequences were consistent with measurements of emission spectra of corals, suggesting that presence–absence polymorphisms of these *FP* sequences contributed to the fluorescent polymorphisms. The functions of recombinant FPs encoded by highly expressed sequences in adult and larval stages were different, suggesting that expression of *FP* sequences with different functions may depend on the life-stage of *A. digitifera*. Highly expressed *FP* sequences exhibited presence–absence polymorphisms in subpopulations of *A. digitifera*, suggesting that presence–absence status is maintained during the evolution of *A. digitifera*.

subpopulations. The difference in FPs between adults and larvae and the polymorphisms of highly expressed *FP* genes may provide key insight into the biological roles of FPs in corals.

## Introduction

Reef-building corals (Scleractinia) show diverse coloration (Marshall, et al. 2003). It has been proposed that fluorescence is a major determinant of color in corals (Dove, et al. 2001; Kelmanson and Matz 2003; Oswald, et al. 2007). The sources of fluorescence in corals are fluorescent proteins (FPs), which emit visible light, and chlorophyll-a from symbiotic dinoflagellate algae, which emit near infrared light (Mazel 2003; Moisan and Mitchell 2001). FPs are excited by environmental light and emit a longer wavelength than the excitation spectrum (Johnsen 2012). The color of an FP is determined by its amino acid sequence (Field, et al. 2006). The center of the region determining light emission in FPs comprises the tripeptide “–X–Y–G–” (termed a chromophore), where the first peptide X varies among FPs (Henderson and Remington 2005). Based on the emission spectra, FPs are classified into four groups: cyan (CFP), green (GFP), yellow (YFP), and red (RFP) (Alieva, et al. 2008; Labas, et al. 2002). Non-fluorescent chromoprotein (CP) is also classified as an *FP* gene family member based on its amino acid sequence similarity (Labas, et al. 2002).

*FP* genes show signatures of diversifying selection, suggesting that FPs with different spectral ranges may have distinct roles in corals (Field, et al. 2006). Although many biological roles of FPs have been predicted, the biological roles of FPs with different spectral

properties remain unclear. Originally, a photo-protective role of FPs was proposed (Salih, et al. 2000), and roles in photo-protection in algal symbionts have subsequently been proposed for RFP (Gittins, et al. 2015) and CP (Smith, et al. 2013). An antioxidant role has been suggested for CFP, GFP, RFP, and CP; however, FP mutants lacking a chromophore show greater antioxidant activity than that of wild-type FP (Palmer, et al. 2009).

*Acropora* is one of the most abundant coral genera in coral reefs of the Indo-Pacific region (Veron 2000) and the whole genome of *Acropora digitifera* has been decoded (Shinzato, et al. 2011). The natural variation of FPs provides an opportunity to analyze the biological role of FPs. FP-mediated color polymorphisms within *Acropora* species have been reported. For example, the photo-inhibitory effect is thought to be substantially weaker in *Acropora palifera* individuals with the fluorescent phenotype than in those with the non-fluorescent phenotype (Salih, et al. 2000). *A. aspera* with the blue phenotype is less sensitive to changes in light than *A. aspera* with the light blue or cream phenotype at normal temperatures, but *A. aspera* with the blue phenotype is more sensitive to increases in temperature (Dove 2004). In *A. nobilis*, individuals with the blue phenotype show greater zooxanthellae and photosynthetic pigment concentrations than those of individuals with the brown phenotype (Smith, et al. 2013). In *A. valida*, individuals with the purple phenotype show a weaker photo-inhibitory effect than

individuals with the brown phenotype (Smith, et al. 2013). In vitro analyses have demonstrated that CP affects the algal chlorophyll excitation spectrum, supporting a role for CP in the photo-protection of zooxanthellae (Smith, et al. 2013). However, CP expression is not vital for shallow-water corals, as evidenced by the adjacent growth of individuals with different color phenotypes in a shallow-water reef (Smith, et al. 2013). In *A. millepora*, a fluorescence polymorphism has been observed in both larval and adult stages (Gittins, et al. 2015; Kenkel, et al. 2011; Strader, et al. 2016). Redder *A. millepora* larvae exhibit diapause-like physiological characteristics and are less responsive to settlement cues than greener larvae, suggesting that red fluorescence might be a marker of long-range larval dispersal (Kenkel, et al. 2011; Strader, et al. 2016). Adult *A. millepora* exhibit high, medium, and low redness levels under the same light levels in shallow water, and the expression levels of RFP are correlated with reduced photo-damage of zooxanthellae under light stress (Gittins, et al. 2015). These expression levels are correlated with the number of particular *RFP* sequences with a particular promoter type (Gittins, et al. 2015). Although color polymorphisms mediated by different FPs have been reported in many *Acropora* species, the genetic differences have only been analyzed for *RFP* sequences in *A. millepora*.

Several *FP* sequences have been reported for a single species in the genus *Acropora*

(Alieva, et al. 2008; D'Angelo, et al. 2008; Gittins, et al. 2015). In *A. digitifera*, the *FP* multigene family has been comprehensively described. More than 30 *FP* gene copies per individual have been reported in *A. digitifera*, and these copies can be separated into short-/middle-wavelength emission (S/MWE including CFP and GFP), middle-/long-wavelength emission (M/LWE including GFP and RFP), and chromoprotein (CP) clades (Takahashi-Kariyazono, et al. 2016). All known FPs from *Acropora* species belong to one of these three clades in the *FP* multigene family (Takahashi-Kariyazono, et al. 2016). Although *FP* gene copy numbers and several *FP* sequences have been reported in *A. digitifera*, the sequences in this family have not been comprehensively identified. It is necessary to determine all *FP* gene sequences in each individual to understand the genetic basis of fluorescence polymorphisms and the biological roles of FPs.

In this chapter, I focused on exon 3 of the *FP* gene, because it is the longest exon and contains nucleotides coding for a chromophore. I examined all exon 3 sequences from *FP* genes and their expression patterns in adult and larval *A. digitifera*. The presence and absence of highly expressed *FP* sequences contributed to the total expression differences among individuals.

## **Materials and Methods**

### **Specimen collection and species identification**

A branch fragment was collected from each of 11 *A. digitifera* colonies in Sesoko, Okinawa, Japan. In total, 11 coral fragments (sample ID: S1601–08 and S1610–12) were preserved in RNAlater (Thermo Scientific, Waltham, MA, USA). In addition, five different colonies of *A. digitifera* (sample ID: S1401–05) were collected from the field and subsequently maintained at the Sesoko Station aquarium (Tropical Biosphere Research Center, University of the Ryukyus). Bundles of gametes from each of five colonies of *A. digitifera* were mixed to allow fertilization. Larvae were reared by daily transfer to fresh seawater and maintained at about 26°C. A single larva was preserved in RNAlater (Thermo Scientific). Species were identified based on morphology. The collection of samples in this chapter was approved by the Aquaculture Agency of Okinawa Prefecture (permit numbers 26-9 and 28-31).

### **DNA extraction and preparation of DNA libraries for hybridization**

Genomic DNAs were extracted from 12 coral fragments of *A. digitifera* (sample ID: S1403, S1601–08, and S1610–12) using DNeasy Blood & Tissue Kits (QIAGEN, Hilden, Germany). Genomic DNAs were fragmented to 500 bp using the Covaris M220 (Covaris, Inc., Woburn,

MA, USA). Following the manufacturer's instructions, DNA libraries of 11 *A. digitifera* samples (sample ID: S1601–08 and S1610–12, Appendix Table S3.1) were constructed using the TruSeq Nano DNA Library Preparation Kit (Illumina, Inc., San Diego, CA, USA).

### **Preparation of hybridization probes for *FP* gene capture**

The *FP* gene family in *A. digitifera* can be divided into three clades (S/MWE, M/LWE, and CP) (Takahashi-Kariyazono, et al. 2016). *FP* gene sequences were amplified using TaKaRa Ex Taq (Takara, Shiga, Japan) and the following primer sets: MiA\_MWEe2\_F1 and AdiFP2XbaI\_R\_L for the *FP* gene in the S/MWE clade, MiA\_MLWEe2\_F3 and MiA\_MLWe5\_R2 for the *FP* gene in the M/LWE clade, and AdiFP8L\_KpnIF\_L and AdiFP8XbaI\_R2\_L for the *FP* gene in the CP clade. All primer sequences are shown in Appendix Table S3.2. Genomic DNAs of *A. digitifera* (sample ID: S1403) and *Acropora* sp. were used as templates. PCR was performed using the GeneAmp PCR System 9700 (Applied Biosystems, Foster City, CA, USA). PCR conditions for the amplification of *FP* gene sequences were as follows: denaturation for 3 min at 94°C, followed by 35 cycles of denaturation for 1 min at 94°C, annealing for 1 min at 50°C, and extension for 3 min at 72°C.

Following the manufacturer's instructions, *FP* gene sequences in all three clades

were labeled with biotin using the *Label IT* Biotin Labeling Kit (Takara). Biotin-labeled DNA fragments were used as hybridization probes.

### ***FP* gene capture**

*FP* gene capture was performed using the SeqCap EZ Hybridization and Wash Kit (Roche, Basel, Switzerland), SeqCap EZ Accessory Kit v2 (Roche), SeqCap HE-Oligo Kit (Roche), and SeqCap EZ Pure Capture Bead Kit (Roche) following the manufacturer's instructions for SeqCap EZ Library SR (Roche), with minor modifications. Briefly, Biotin-labeled *FP*-gene fragments were used as hybridization probes, instead of the SeqCap EZ library (Roche). Equal amounts of all libraries (in total, 1 µg) were mixed with 135 ng of Biotin-labeled *FP*-gene fragments and were hybridized at 46°C for 72 h. Other procedures were performed in accordance with the manufacturer's instructions.

### **Sequencing of captured DNA libraries and extraction of exon 3 sequences**

Short DNA sequences (paired-end, 250 bp) were determined from the captured libraries using the Illumina HiSeq2500 platform. The nucleotide sequences were deposited in the DDBJ Sequenced Read Archive under accession number DRR120550–DRR120560. Since the average

insert size of the libraries was 500 bp, the overlapping paired reads were merged into one sequence read using CLC Genomics Workbench (<https://www.qiagenbioinformatics.com/>), after the removal of the adaptor sequences and low-quality reads (quality score less than 20). When the paired reads did not overlap, both reads were used, without merging. Merged and unmerged reads were used for sequence identification.

The longest exon containing nucleotides encoding a chromophore, i.e., exon 3 (e3), was used for subsequent analyses. To identify e3 sequences in the *FP* gene, the ancestral sequences of *FP* genes (explained below) were used as a reference for mapping. To construct the ancestral sequences, full-length coding regions of *FP* genes of *Acropora* species were collected from the NCBI database. The accession numbers of the collected sequences are shown in Table S3. Using these *FP* sequences, the ancestral sequences of two sub-clades in the S/MWE, M/LWE, and CP clades (Takahashi-Kariyazono, et al. 2016) were estimated by the maximum likelihood method with a pre-set tree topology using MEGA ver. 7 (Kumar, et al. 2016). The ancestral sequences are listed in the Appendix text.

Short DNA reads from each sample were mapped to the reference sequences composed of the e3 ancestral sequences of two sub-clades in the S/MWE, M/LWE, and CP clades using CLC Genomics Workbench. Reads showing similarity (>80%) to the e3 sequence

(S/MWE and M/LWE clades: length, 225 bp [accessions: BR000963 and XM\_015914911]; CP clade: length, 219 bp [Appendix Figure S3.1]) were mapped to reference sequences. Reads mapped to the ancestral sequences of each of the three clades were separately extracted and assembled using ATGC (GENETYX CORPORATION, Tokyo, Japan) (matching percentage = 100%), and unassembled reads were removed. Assembled contigs for each of the three clades were aligned, and contigs covering the full length of e3 were selected. Sequences in intronic regions were removed from contigs. For each sample, DNA reads (merged reads and other reads) were mapped (complete match) to these assembled e3 contigs using CLC Genomics Workbench. The e3 contigs in which at least two reads covered the full length were identified as e3 sequences, those containing insertions or deletions that cause frameshifts and/or premature stop codons were regarded as pseudo-gene (pseudo-e3) sequences. Among all e3 sequences isolated from 11 individuals, identical sequences were removed. The phylogenetic tree of e3 sequences was constructed using the maximum likelihood method based on the Kimura 2-parameter mode with gamma distributed rates and invariant sites using MEGA ver. 7 (Kumar, et al. 2016). Bootstrap support for each clade was obtained based on 1,000 replicates. E3 sequences were deposited in GenBank under accession numbers LC349488– LC349725.

**Validation of the absence of highly expressed exon 3 sequences**

Following the manufacturer's instructions, DNA libraries of three *A. digitifera* samples (sample ID: S1601, S1603, and S1606) were constructed using the TruSeq DNA PCR-Free Library Preparation Kit (Illumina, Inc.). Short DNA sequences (paired-end, 125 bp) were determined from these three and from the DNA libraries of eight *A. digitifera* samples (sample ID: S1602, S1604–5, S1607–8, and S1610–S1612) described under METHODS (DNA extraction and preparation of DNA libraries for hybridization) by the Illumina HiSeq2500 platform. The nucleotide sequences were deposited in the DDBJ Sequenced Read Archive under accession numbers DRR108003–DRR108012 and DRR108024. To verify for the absence of the highly expressed sequences in the genomes of individuals in which these sequences were not isolated, short reads from the genomes of each individual were mapped to all isolated e3 sequences. Short DNA reads showing similarity (100%) with 80% read lengths were mapped to e3 sequences using CLC Genomics Workbench. When the full length of e3 was covered, with at least one coverage without any mismatches, the e3 sequence was regarded as positive in the genome of the individual. When any reads mapped to the e3 without any mismatches, the e3 sequence was regarded as negative in the genome of the individual. If a few or several reads were mapped to e3, but the full length of e3 was not covered, the e3 sequences were classified

by the second mapping condition. The second mapping was performed under the same criteria as the moderate mapping condition (similarity 100% with 40 % of read length) using the same short DNA reads and e3 sequences. When the full length of e3 was covered without any mismatches under the moderate mapping condition, the e3 sequence was regarded as positive. The remaining e3 sequences were categorized as unclassified.

#### **Estimation of *A. digitifera* *FP* gene copy numbers**

*FP* gene copy numbers in the *A. digitifera* genome were estimated by quantitative PCR (qPCR) using 11 *A. digitifera* specimens. *FP* gene copy number estimation was performed using previously described methods (Takahashi-Kariyazono, et al. 2016) with minor modifications. Briefly, two new reverse primers (MLWE\_qPCR\_R1-3 for M/LWE and MiA\_CP\_e3\_R1-3 for CP) were used for qPCR. Primer sequences are shown in Appendix Table S3.2. qPCR was performed using the Thermal Cycler Dice TP800 (Takara). Genomic DNAs of 11 samples (S1601–8, S1610–12) were used as templates.

#### **RNA extraction and sequencing**

Total RNAs were extracted from 11 adult coral fragments of *A. digitifera*, the same individuals

used for DNA extraction, and a single *A. digitifera* larva using TRIzol reagent (Thermo Fisher Scientific). cDNAs were synthesized from total RNA using the PrimeScript II 1st Strand cDNA Synthesis Kit (Takara). RNAs from 11 adults were used to construct libraries for high-throughput RNA sequencing using the NEBNext Ultra RNA Library Prep Kit for Illumina (New England Bio Labs, Ipswich, MA, USA). Short DNA sequences (paired-end, 125 bp) were determined from the libraries using the Illumina HiSeq2500 platform. The nucleotide sequences were deposited in the DDBJ Sequenced Read Archive under accession numbers DRR108013-DRR108023.

### **Expression levels of exon 3 sequences of *FP* genes**

After the removal of the adaptor sequences and low-quality reads (quality score less than 20), RNA-seq reads from 11 individuals were mapped to *A. digitifera* genome assembly ver.1.1. In addition to the 11 adult individuals, RNA-seq reads from *A. digitifera* larvae (12.1 Gbp, accession: SRX1534820) were downloaded from a public database (DDBJ) and mapped to *A. digitifera* genome assembly ver.1.1. Reads showing high sequence similarity (>90%, >112 bp) were mapped to query sequences and expression values were calculated using CLC genomic workbench. RPKM (Reads Per Kilobase of exon model per Million mapped reads) were used to

normalize expression values to identify candidate housekeeping genes. First, house-keeping genes that showed stable expression among individuals and developmental stages were selected as internal controls for normalization of *FP* gene expression levels. In particular, the variance indexes for RPKM values (standard-deviation/average RPKM) were calculated for all genes among 11 adults and larvae. Thirty-four genes with average RPKM values among 11 adults and larvae of  $\geq 50$  and with a variance index of RPKM among 11 adults and larvae of less than 1.0 were selected as house-keeping genes (Appendix Table S3.4).

To calculate expression values for e3 sequences, RNA-seq reads from 11 adults and larvae (accession: SRX1534820 and DRR054773) were mapped to all isolated e3 sequences using CLC Genomics Workbench. Reads showing high similarity to corresponding e3 sequences (100%, >119 bp) were mapped to those sequences. To normalize mapped reads, the total numbers of reads mapped to the 34 house-keeping genes were used as internal controls. The normalized expression values for *FP* sequences were calculated as follows: (Mapping coverage of the *FP* e3 sequence/Mapped reads on the internal control)  $\times$  100,000, where the mapping coverage of the *FP* e3 sequence was estimated as the reads mapped to the e3 sequence divided by the e3 length (bp) and mapped reads on the internal control indicates reads mapped to the 34 house-keeping genes. The e3 sequences of FPs in the S/MWE, M/LWE, and CP clades

were 225, 225, and 219 bp in length, respectively.

### **Identification of *Symbiodinium* clades in each sample**

To identify the clades of *Symbiodinium* in each sample, RNA-seq reads of 11 individuals were mapped to the nucleotide sequences of nuclear 28S (nr28S) and chloroplast 23S (cp23S) ribosomal DNA from eight clades of *Symbiodinium* (Pochon, et al. 2012) using CLC Genomics Workbench. The accession numbers for each sequence are shown in Appendix Table S3.5.

### **Expression analyses of *Symbiodinium* genes**

To analyze gene expression in symbiotic algae living in all coral individuals, RNA-seq reads of 11 coral individuals showing high similarity (>80%) and lengths of >100 bp were mapped to the transcriptome of *Symbiodinium goreau* (clade C) (Davies, et al. 2018) using CLC Genomics Workbench. The following comparisons were evaluated according to gene expression and gene presence-absence patterns: (1) high expression of *S/Me3\_a\_hil* and the absence of this sequence in the genome; (2) presence and absence of *S/Me3\_a\_hil* in the genome; (3) expression of *M/L-De3\_a* and the absence of this sequence in the genome; (4) expression and no expression of M/LWE clade TYG-type sequences; (5) high (expression values  $\geq 10$ ) and low

expression (expression values  $\leq 5$ ) of total CP sequences. The contigs with significant expression differences ( $p < 0.05$  with FDR correction) between two groups were selected.

### **Identification and cloning of *FP* cDNA sequences, purification, and spectroscopic analyses of recombinant FP proteins**

Full-length cDNAs of *FP* genes were amplified by PCR. Using the cDNA of an *A. digitifera* larva as a template, S/MWE-*FP* full-length cDNAs were amplified using PrimeSTAR GXL DNA Polymerase (Takara). M/LWE-*FP* full-length cDNAs were amplified using PrimeSTAR GXL DNA Polymerase (Takara) from the cDNAs of an *A. digitifera* larva and an adult (sample ID: S1603). CP full-length cDNAs were amplified using TaKaRa Ex Taq from the cDNAs of an *A. digitifera* larva and an adult (sample ID: S1606). PCR conditions for the amplification of full-length cDNAs were as follows: denaturation for 3 min at 94 °C, followed by 30 cycles of denaturation for 1 min at 94 °C, annealing for 1 min at 55 °C, and extension for 1 min at 72 °C. All primer sets and primer sequences are shown in Appendix Table S3.2 and S3.6. PCR products were cloned into the T-Vector pMD20 vector (Takara), and the sequences were verified using the Applied Biosystems Automated 3130xl Sequencer (Foster City, CA, USA).

To construct recombinant FP proteins, vectors with cloned *FP* sequences were used

as templates for subcloning into expression vectors. Full-length *FP* cDNAs were subcloned into the pCold I expression vector (Takara) and then used to transform BL21 *Escherichia coli* cells (Takara). Each clone was grown in 20 mL of LB medium supplemented with ampicillin and IPTG overnight, and the recombinant proteins were extracted by sonication and purified using TALON beads with poly-histidine tags (Takara).

Emission and excitation spectra of purified recombinant FP proteins in 50 mmol/L phosphate buffer solution with 500 mmol/L imidazole, pH 7.0, were measured using the RF-6000 spectro-fluorophotometer (SHIMAZU, Kyoto, Japan). Absorption spectra of purified recombinant CP proteins were measured using a UV-1800 spectrophotometer (SHIMAZU). Measurements were obtained at least three times for each FP.

### **Identification of *FP* genes in the three *A. digitifera* subpopulations**

The publicly available genomic DNA reads of *A. digitifera* collected from the southern Ryukyu Archipelago located in southwestern Japan were used (Appendix Figure S3.2). These individuals have been separated into four subpopulations; Okinawa, Kerama, Yaeyama-South and Yaeyama-North (Shinzato, et al. 2015). Individuals with mapping read coverage  $\geq 6$  were selected for my analysis. The mapping read coverages were based on those in a previous report

(Shinzato, et al. 2015). A total of 91 sets of genomic DNA reads from 40 individuals belonging to Okinawa, 39 individuals belonging to Kerama, and 12 individuals belonging to Yaeyama-North were downloaded. A subpopulation (Yaeyama-South) was excluded from my analysis, because there were only two individuals with mapping read coverage  $\geq 6$  in this subpopulation. The accession numbers of downloaded genomic DNA reads are shown in Table S7. Validation of the existence of highly expressed e3 sequences in each individual from the three subpopulations was performed using the method explained under METHODS (Validation of the absence of highly expressed exon 3 sequences).

### **Live coral fluorescence measurements**

Light emission, including reflectance and fluorescence spectra, was measured from each of 11 *A. digitifera* colonies that were used for DNA and RNA extraction. Visual observation did not reveal clear color differences among 11 *A. digitifera* colonies. An LED source with a peak wavelength of 448 nm was used as the excitation light source. The distances from the excitation light and measurement probe to objects were 6 cm, and spectra were recorded using the Jaz Spectrometer (Ocean Optics, Dunedin, FL, USA) in the dark. For all measurements, light emitted at longer than 660 nm was not attributed to coral fluorescence because chlorophyll *a*

from the symbiotic dinoflagellate algae living within the coral tissues emit light with a primary peak wavelength around 685 nm and a secondary peak at 730 nm (Mazel 2003; Moisan and Mitchell 2001). Fluorescence was estimated by excitation (LED)-emission spectral subtraction.

## Results

### Copy number variation and sequence diversity in the *FP* gene family

I quantified *FP* gene copy numbers in 11 *A. digitifera* genomes by qPCR. I detected 14–24 *FP* gene copies in the S/MWE clade, 5–13 in the M/LWE clade, and 9–15 in the CP clade (Figure 3.1), indicating copy number variation among individuals.

Despite the high estimated copy numbers, only ten *FP* genes (including nine partial genes) were identified in the assembled genome sequence of *A. digitifera* (Shinzato, et al. 2012). This inconsistency may reflect the difficulty of multi-gene family assembly (Mariano, et al. 2015). To identify sequence diversity in *FP* genes, I determined the exon 3 (e3) sequences of *FP* genes without de novo assembly. I chose e3 because it has the longest exon among *FP* genes (those of the S/MWE, M/LWE, and CP clades were 225, 225, and 219 bp in length, respectively) and contains the region that encodes a chromophore. DNA fragments containing *FP* genes in libraries were condensed by *FP* gene capture, and short DNA sequences (paired-end, 250 bp) were determined (2,141,708–2,850,976 reads) from 11 samples (Appendix Table S3.1). These reads from each library were mapped to e3 of S/MWE *FP* (1,069–1,532 reads), M/LWE *FP* (215–655 reads), and CP (628–1,698 reads) genes. Using mapped reads, e3 sequences of *FP* genes were determined from 11 *A. digitifera* genomes (Figure 3.1). I detected

different e3 nucleotide sequences from 11 individuals, including 111 intact and 19 pseudo-sequences in the S/MWE clade, 29 intact and one pseudo-sequence in the M/LWE clade, and 49 intact and 29 pseudo-sequences in the CP clade (Table 3.1). The intact e3 sequences were translated into amino acid sequences. The number of amino acid sequences were 67, 17, and 38 in the S/MWE, M/LWE, and CP clades, respectively. Based on the first peptide of the chromophore, I divided the e3 sequences in the M/LWE clade into two types, TYG-type and DYG-type. The phylogenetic trees based on the nucleotide sequences of e3 are shown in Appendix Figure S3. 3A–C. Among the three clades, I detected the most e3 sequences at both the nucleotide and amino acid sequence levels in the S/MWE clade.

Table 3.1. Number of exon 3 sequences isolated in three clades

Clade	Intact exon3		Pseudo-exon3
	Nucleotide	Amino acid	Nucleotide
S/MWE	111	67	19
M/LWE	29	17	1
CP	49	38	29

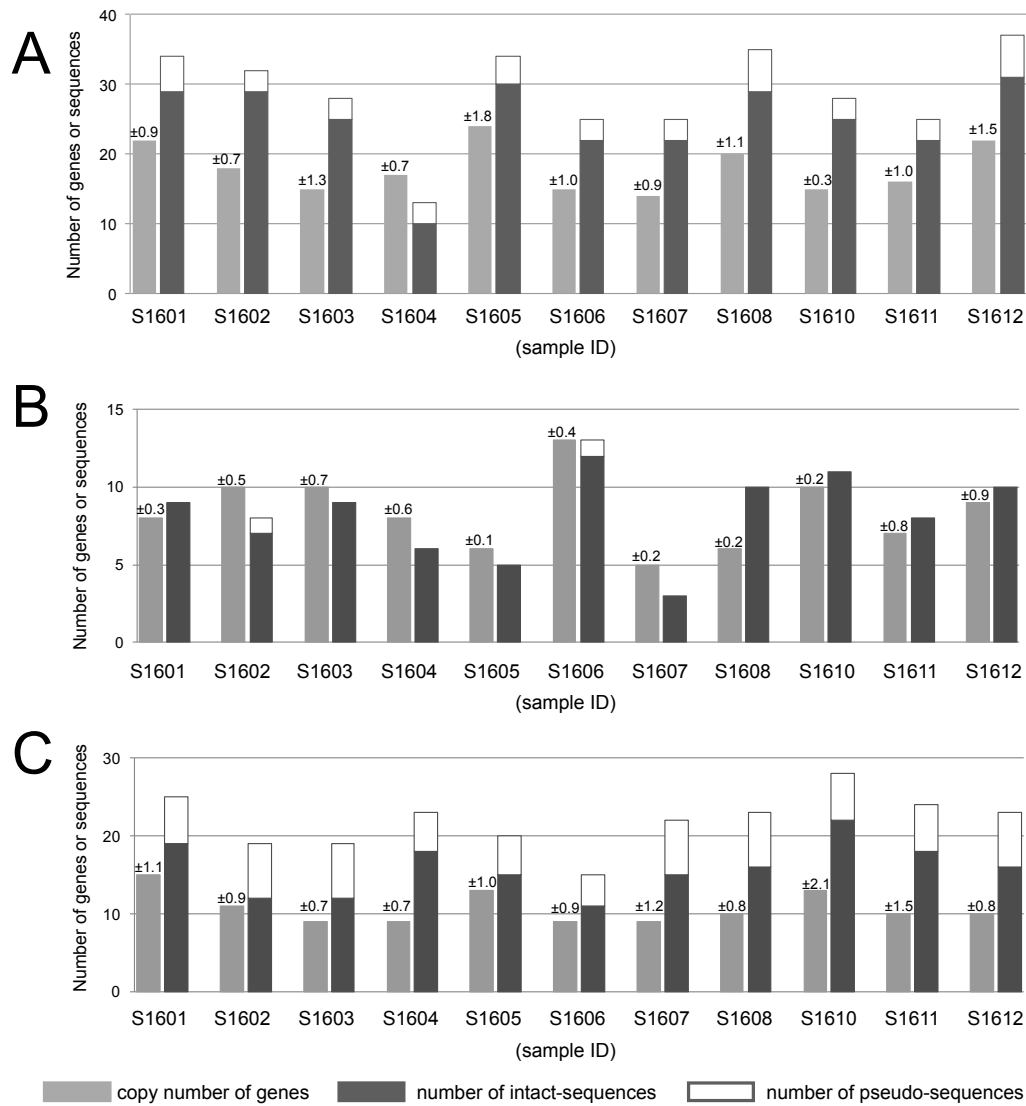


Figure 3.1. *FP* gene copy numbers and total numbers of sequences obtained in each clade *FP* gene copy numbers and exon 3 sequence numbers in the MWE (A), M/LWE (B), and CP (C) clades from 11 adult colonies of *A. digitifera* are given. Copy numbers are shown in light gray with standard deviations obtained from three replicates. Numbers of intact exon 3 sequences are shown in dark gray and pseudo-exon 3 sequences are shown in white. S1601–S1612 indicate sample IDs.

### Expression differences for exon 3 sequences of *FP* genes among adult individuals

I examined the expression levels of each e3 sequence based on RNA-seq reads from 11 adult individuals (41,007,102–109,525,174 reads). In RNA-seq read mapping, 36–50% of RNA-seq reads were mapped to the *A. digitifera* genome ver. 1.1 (Appendix Table S3.9). Expression levels of e3 sequences were normalized by the expression levels of 34 house-keeping genes and categorized into six levels (Figure 3.2). The expressed e3 sequences in larval (see below) and adult stages are summarized in Table 3. 2.

I detected variation in total expression levels in each clade or type of M/LWE among 11 individuals (Table 3.3). In the S/MWE clade, among the individuals with high overall *FP* gene expression (Table 3.3, total expression of the S/MWE clade sequences in S1603, S1605, and S1611), one e3 sequence (SMe3\_010) was highly expressed (Figure 3.2A). I refer to this e3 sequence as “*S/Me3\_a\_hil*,” it was also detected in four other individuals (Figure 3.2A: S1602, S1606–7, and S1610), but at lower levels than those in the three high-expression individuals (Figure 3.3). In the other four individuals (Figure 3.2A: S1601, S1604, S1608, and S1612), I did not detect *S/Me3\_a\_hil* in genomic DNA (250-bp reads) or RNAseq reads. In addition to *S/Me3\_a\_hil*, the other e3 sequence (e3 ID: SMe3\_079) in a single individual (Figure 3.2A:

S1603) showed high sequence similarity with *S/Me3\_a\_hi1* and high expression. I refer to this e3 sequence as “*S/Me3\_a\_hi2*.” Among three individuals with high overall *FP* gene expression in the S/MWE clade, the expression of *S/Me3\_a\_hi1* and *S/Me3\_a\_hi2* sequences accounted for greater than 90% of the total expression (Figure 3.3A). The absence of *S/Me3\_a\_hi1* and *S/Me3\_a\_hi2* in the genomes of individuals shown by the 250-bp reads from genomic DNA was verified using short reads (paired-end, 125 bp). These results indicate that each of the *S/Me3\_a\_hi1* and *S/Me3\_a\_hi2* sequences were present in one individual and absent from another individual, thus, there were presence–absence polymorphisms in these 11 individuals.

In the M/LWE clade, I observed that 4 of 29 e3 nucleotide sequences were expressed in at least one adult. In four adult individuals (Figure 3.2B: S1602, S1604, S1605, and S1607), I did not detect the expression of any e3 sequence. In adults, only one DYG-type sequence (e3 ID: MLe3\_009) was expressed in three samples (Figure 3.2B: S1603, S1606, and S1610), and little expression of TYG-type sequences was observed. I refer to this expressed DYG-type e3 sequence as “*M/L-De3\_a*,” I identified this sequence in four individuals and verified these findings using short reads (paired-end, 125 bp) of genomic DNA from these individuals (Appendix Table S3.7). Four individuals had *M/L-De3\_a*, among which one individual (Figure 3.3B ID: S1602) did not express this sequence. The other six individuals did not possess this

sequence in their genome (Appendix Table S3.7), indicating a presence–absence polymorphism of *M/L-De3\_a*.

In the CP clade, I detected the expression of 20 out of 49 e3 nucleotide sequences in at least one adult. In one sample (ID: S1606), total expression in the CP clade was relatively high (Table 3.3), and this expression was attributed to five e3 sequences (Figure 3.2C). I did not identify an e3 sequence with high expression in the CP clade (Figure 3.2C).

Table3.2. Number of expressed sequences in the larval and adult stages

Clade	Adults and larvae	Only adults	Only larvae
S/MWE	1	34	3
M/LWE_TYG	2	1	7
M/LWE_DYG	0	1	8
CP	4	16	8

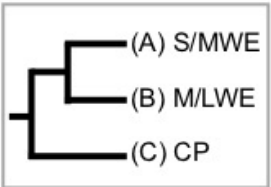


Figure 3.2. Expression of each *FP* sequence in adults and larvae.

Rows indicate the exon 3 sequences belonging to S/MWE clade (A), M/LWE clade (B), and CP clade (C). Columns indicate the normalized expression values for *FP* sequences calculated by the expression of the internal control genes (see METHODS) in 11 adults and larvae. Sample IDs S1601–S1612 are indicated as 01–12. L indicates larvae. A grey box indicates the absence of the sequence in a genome of a sample and a slash mark indicates that the sequence data were not available. Exon 3 sequences were aligned based on the phylogenetic trees in Fig. S2 (UPGMA based on amino acid differences) to show amino acid differences. Black arrow indicates a branch supported by the ML tree in Figure S3. The first peptide of the chromophore for each exon 3 sequences is shown in a rectangle. An asterisk indicates the exon 3 sequence used for the recombinant protein analysis.

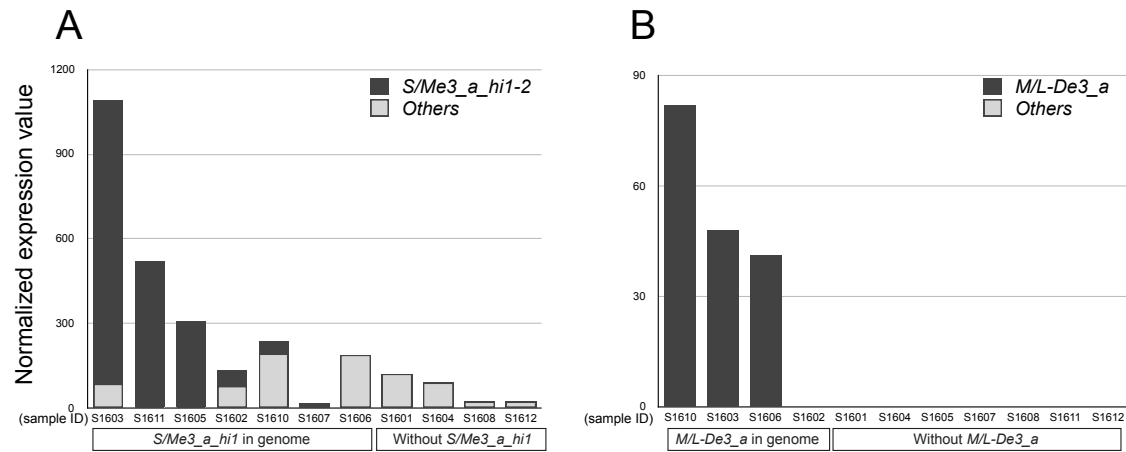


Figure 3.3. Proportions of highly expressed exon 3 sequences in the S/MWE and M/LWE clades

The normalized expression values for *FP* sequences calculated by the expression of the internal control genes (see METHODS) in the (A) S/MWE clade (*S/Me3\_a\_hi1* and *S/Me3\_a\_hi2*) and (B) M/LWE clade (*M/L-De3\_a*) and the expression of all *FP* sequences in each clade, except highly expressed sequences shown in black and gray, respectively. The y-axis indicates the normalized expression value. S1601–S1612 indicate sample IDs.

### Expression differences in exon 3 sequences of *FP* genes between adults and larvae

I compared the expression of each e3 sequence between planula larvae and adults using publicly available RNA-seq reads of larvae (accession: SRX1534820) and RNA-seq reads of 11 adults, respectively. In the S/MWE clade, I observed larval expression of four closely related e3 sequences, and I did not detect the expression of these four e3 sequences in adults, except for low expression in one individual (Figure 3.2A), indicating differences in the regulation of *FP* gene expression between adults and larvae. The total expression of sequences in this clade was lower in larvae than in the adults with the three highest expression levels (Table 3.3).

In the M/LWE clade, I observed higher total expression of both TYG-type and DYG-type sequences in larvae than in adults (Table 3.3). Although I observed lower expression of TYG-type than DYG-type sequences in 11 adults, I detected similar total expression values for TYG-type and DYG-type sequences in larvae. I observed high expression of two (Figure 3.2B: MLe3\_004, and MLe3\_008) and one (Figure 3.2B: MLe3\_012) TYG-type and DYG-type e3 sequences, respectively. Hereafter, I refer to these e3 sequences as “*M/L-Te3\_l\_hi1*,” “*M/L-Te3\_l\_hi2*,” and “*M/L-De3\_l\_hi*.” Among 11 adult individuals, I isolated *M/L-Te3\_l\_hi1*, *M/L-Te3\_l\_hi2*, and *M/L-De3\_l\_hi* from six, four, and two individuals, respectively (Figure 3.2B), and the presence and absence of these sequences in the genome were verified by short

reads (paired-end, 125 bp) of genomic DNA (Appendix Table S3.7). These results indicate that highly expressed sequences in larvae showed presence–absence polymorphisms in adult individuals.

In the CP clade, I observed higher total expression of CP clade sequences in larvae than in adults (Table 3.3). I detected the expression of 12 e3 sequences in larvae, and no or low expression of these e3 sequences in 11 adults (Figure 3.2C). These results indicate differences in *CP* sequences expression between adults and larvae.

The genetic differences in *FP* genes (copy number variation and presence–absence polymorphisms) in individuals used as “adult samples” and “larval samples” might affect the expression patterns of *FP* genes at different life stages. To examine this possibility, I compared two independent RNA-seq data sets of *A. digitifera* larvae (accessions: SRX1534820 and DRR054773). The results showed that two RNA-seq data sets from larvae had similar expression patterns, and that the patterns were different from those of adults (Appendix Figure S3.5). This expression pattern difference was most likely the result of a difference in expression between adults and larvae rather than the result of a difference in genetic background.

In all three clades, some e3 sequences exhibited no expression in both adults and larvae (Figure 3.2A–C). It is possible that these unexpressed sequences are parts of pseudogenes

and/or are expressed at different life stages (e.g., developmental stages) or in specific conditions.

Table3.3. Expression values for specific exon 3 sequences

Sample	S/MWE			M/LWE (DYG)			M/LWE (TYG)	CP
	<i>S/Me3_a_hil-2</i>	Other than <i>S/Me3_a_hil-2</i>	Total	<i>M/L-De3_a</i>	Other than <i>M/L-De3_a</i>	Total	Total	Total
S1601	0	120	120	0	0	0	1	13
S1602	56	77	133	0	0	0	0	8
S1603	1,002	86	1,088	48	0	48	0	6
S1604	0	89	89	0	0	0	0	3
S1605	304	5	309	0	0	0	1	4
S1606	1	188	189	41	0	41	1	43
S1607	16	2	18	0	0	0	0	5
S1608	0	24	24	0	0	0	1	5
S1610	46	191	237	83	0	83	0	10
S1611	514	5	519	0	0	0	1	8
S1612	0	21	21	0	0	0	2	7
Larvae	0	239	239	0	1,236	1,236	1,186	830

### Functional differences among FPs

To evaluate functional differences among *FP* e3 sequences, I cloned full-length *FP* coding regions, constructed recombinant FPs, and measured emission and excitation or absorption spectra. In the S/MWE clade, I cloned four different full-length cDNAs of *FP* genes, i.e., *S/Mcd\_l* expressed in larvae, *S/Mcd\_a* expressed in adults, and *S/Mcd\_a\_hi1* and *S/Mcd\_a\_hi2*, which are highly expressed in adults. The e3 sequences of full-length cDNAs of *FP* genes were identical to the e3 sequences I determined, and the correspondence between full-length cDNA names and e3 sequence IDs is shown in Appendix Table S3.6. The emission peak of recombinant *S/Mcd\_l* expressed in larvae was  $519 \pm 0$  nm (Figure 3.4A), whereas the emission peaks of two recombinant proteins expressed in adults (*S/Mcd\_a* and *S/Mcd\_a\_hi1*) were  $488 \pm 0$  nm and  $483 \pm 0$  nm, respectively (Figure 3.4B). Thus, recombinant FPs encoded by *FP* sequences highly expressed in larvae and adults exhibited different emission peaks.

Furthermore, I detected similar emission peaks for recombinant *S/Mcd\_a\_hi1* (highly expressed) and *S/Mcd\_a* (Figure 3.4B). Recombinant *S/Mcd\_a\_hi1* was excited by light under 380 nm, whereas *S/Mcd\_a* was not highly excited by light in this range. The excitation and emission spectra of recombinant *S/Mcd\_a\_hi1* (shown in black in Figure 3.4B) and *S/Mcd\_a\_hi2* (Figure 3.4C) were identical, indicating that these two highly expressed e3

sequences in adults share the same function in the wavelength of excitation and emission spectra of FP.

In the M/LWE clade, I cloned three different full-length cDNAs of *FP* genes, i.e., *M/L-Tcd\_l\_hi2* (TYG-type expressed in larvae), *M/L-Dcd\_l2* (DYG-type with identical amino acid sequences expressed in larvae), and *M/L-Dcd\_a* (DYG-type expressed in adults). Emission peaks for the recombinant *M/L-Tcd\_l\_hi2*, *M/L-Dcd\_l2*, and *M/L-Dcd\_a* were  $519 \pm 0$  nm,  $613 \pm 2$  nm, and  $606 \pm 1$  nm, respectively (Figure 3.4D–F).

In the CP clade, I cloned three different full-length cDNAs of *FP* genes, i.e., *CPcd\_l1* expressed in larvae, *CPcd\_l2* expressed in larvae, and *CPcd\_a* expressed in adults. Absorption peak values for recombinant *CPcd\_l1*, *CPcd\_l2*, and *CPcd\_a* were  $585 \pm 0$  nm,  $603 \pm 0$  nm, and  $579 \pm 0$  nm, respectively (Figure 3.4G–I). These CPs absorbed long wavelength light.

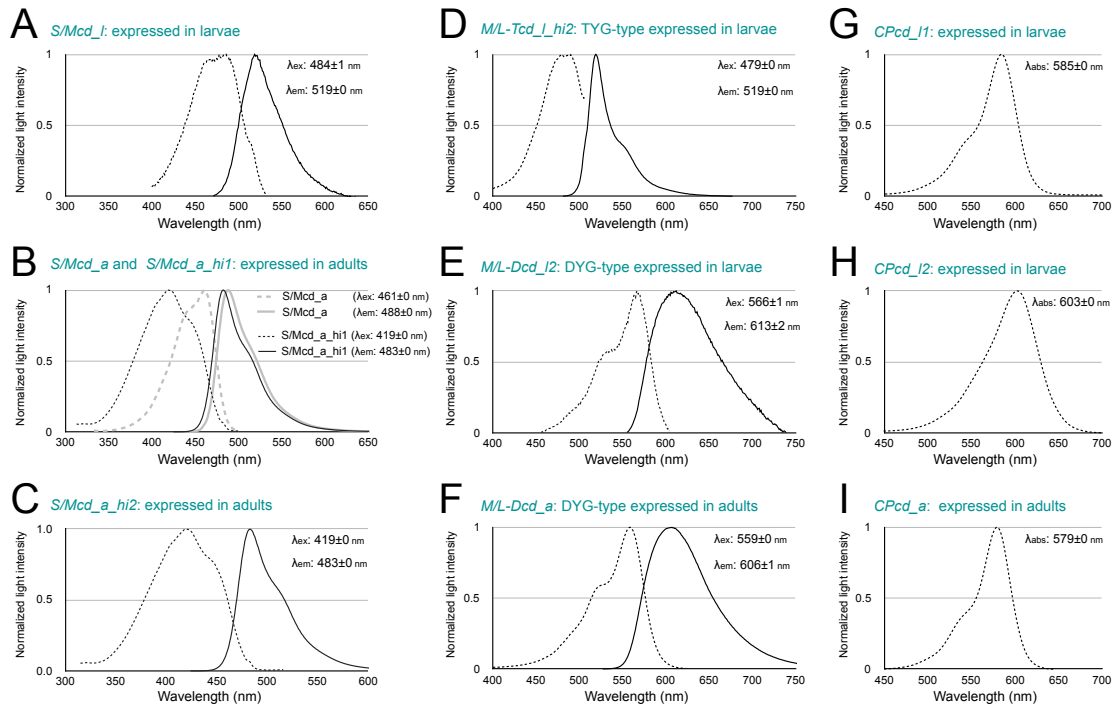


Figure 3.4. Emission and absorption spectra of FP recombinant proteins.

Absorption (dotted line) and emission (solid line) spectra of recombinant proteins belonging to the S/MWE clade (A, *S/Mcd\_l*; B, *S/Mcd\_a* and *S/Mcd\_a\_hi1*; and C, *S/Mcd\_a\_hi2*), M/LWE clade (D, *M/L-Tcd\_l\_hi2*; E, *M/L-Dcd\_l2*; and F, *M/L-Dcd\_a*), and CP clade (G, *CPcd\_a*; H, *CPcd\_l1*; I, *CPcd\_l2*) are shown. Horizontal and vertical axes indicate wavelengths (nanometers) and normalized light intensities, respectively.

### Identification of highly expressed *FP* sequences in three *A. digitifera* subpopulations

In S/MWE, I detected high *S/Me3\_a\_hi1* and *S/Me3\_a\_hi2* expression in three adult individuals with high total *FP* gene expression. Additionally, in the M/LWE clade DYG-type, expression differences among 11 adult individuals were caused by differences in *M/L-De3\_a* expression. In larvae, *M/L-Te3\_l\_hi1*, *M/L-Te3\_l\_hi2*, and *M/L-De3\_l\_hi* in the M/LWE clade were highly expressed.

To confirm whether these sequences existed in *A. digitifera* subpopulations in the southern Ryukyu Archipelago, I examined 102 individuals from three subpopulations (Okinawa, Kerama, and Yaeyama-North) (Shinzato, et al. 2015). I detected the presence of each of *S/Me3\_a\_hi1*, *M/L-De3\_a*, *M/L-Te3\_l\_hi1*, *M/L-Te3\_l\_hi2*, and *M/L-De3\_l\_hi* in the genomes of individuals from the three subpopulations (Table 3.4 and Appendix Table S3.7). I detected the presence and absence of each of *S/Me3\_a\_hi1*, *M/L-Te3\_l\_hi2*, and *M/L-De3\_l\_hi* in the genomes of individuals from the three subpopulations (Table 3.4 and Appendix Table S3.7), indicating that the presence–absence polymorphic state of these sequences was shared by the three subpopulations. I found *S/Me3\_a\_hi2* in two subpopulations, but not in the Yaeyama-North subpopulation containing six individuals without this sequence or with partially

covered (categorized as unclassified, see METHOD). Also, I found only individuals with *M/L-Te3\_l\_hi1* or unclassified individuals in Kerama and Yaeyama-North. In these cases, I was unable to determine whether or not *S/Me3\_a\_hi2* and *M/L-Te3\_l\_hi1* were presence–absence polymorphic in these subpopulations. I found individuals with *M/L-De3\_a*, but no individual without *M/L-De3\_a* or unclassified individuals in Yaeyama-North, suggesting that this sequence is fixed or exists at high frequencies in this subpopulation. However, since the number of individuals in Yaeyama-North used in this analysis was smaller than in the other two subpopulations, it could have occurred by chance.

Table 3.4. Number of individuals that possess or did not possess specific exon 3 sequences in their genome

FP e3 sequence	Subpopulation	Number of individuals			Polymorphic state
		Positive	Negative	Unclassified	
<i>S/Me3_a_hi1</i>	Okinawa (n=51)	28	7	16	Polymorphic
	Kerama (n=39)	27	4	8	Polymorphic
	Yaeyama_North (n=12)	3	5	4	Polymorphic
<i>S/Me3_a_hi2</i>	Okinawa (n=51)	1	11	39	Polymorphic
	Kerama (n=39)	3	10	26	Polymorphic
	Yaeyama_North (n=12)	0	6	6	Unknown
<i>M/L-De3_a</i>	Okinawa (n=51)	13	27	11	Polymorphic
	Kerama (n=39)	9	23	7	Polymorphic
	Yaeyama_North (n=12)	12	0	0	(Presence-fixed)
<i>M/L-Te3_l_hi1</i>	Okinawa (n=51)	31	5	15	Polymorphic
	Kerama (n=39)	32	0	7	Unknown
	Yaeyama_North (n=12)	4	0	8	Unknown
<i>M/L-Te3_l_hi2</i>	Okinawa (n=51)	13	3	35	Polymorphic
	Kerama (n=39)	11	2	26	Polymorphic
	Yaeyama_North (n=12)	2	3	7	Polymorphic
<i>M/L-De3_l_hi</i>	Okinawa (n=51)	13	14	24	Polymorphic
	Kerama (n=39)	19	6	14	Polymorphic
	Yaeyama_North (n=12)	4	4	4	Polymorphic

**Differences of *FP* gene expression were consistent with differences in fluorescence emitted from coral individuals**

The emission spectra (including reflectance and fluorescence) from 11 *A. digitifera* colonies excited by LED (448 nm spectrum peak) were measured in the sea. In individuals with high expression of *S/Me3\_a\_hil*, fluorescence spanned approximately 475–570 nm (Appendix Figure S3.4A–C).

I did not observe fluorescence from four individuals without *S/Me3\_a\_hil* in their genomes (Appendix Figure S3.4D–G). Two of the four individuals with low expression of *S/Me3\_a\_hil* (ID: S1607, 1610) did not show fluorescence (Appendix Figure S3.4J and K). The remaining two individuals (ID: S1602, 1606) with low expression of *S/Me3\_a\_hil* showed fluorescence with a peak at 516 nm (Appendix Figure S3.4H and I). The fluorescence spanning approximately 500–590 nm in one individual (ID: 1606) (Appendix Figure S3.4I) may have been emitted by different FPs, because the emission peak of the FP encoded by *S/Me3\_a\_hil* was 483 nm (Figure 3.4B). The fluorescence from one individual (ID: 1602) (Appendix Figure S3.4H) may have been emitted by a mixture of FPs encoded by *S/Me3\_a\_hil* and other sequences. These results show that fluorescence polymorphisms existed among 11 *A. digitifera*;

individuals with fluorescence spanning approximately 475–570 nm, one individual with fluorescence spanning approximately 500–590 nm, and individuals with no fluorescence. Coral fluorescence spanning approximately 475–570 nm was consistent with the high expression of *S/Me3\_a\_hi1* and absence of *S/Me3\_a\_hi1* in the genome.

### ***FP* expression does not affect gene expression in symbiotic algae**

Photo-protective effects of FPs on photosynthetic algal symbionts have been proposed (Gittins, et al. 2015; Salih, et al. 2000; Smith, et al. 2013). However, I observed differences in *FP* expression among individuals collected from the same shallow water light environment. To compare the conditions of symbiotic algae living in each coral individual, I evaluated the levels of all expressed genes in symbiotic algae by RNA-seq.

First, I identified the clade of symbiotic algae in the 11 individuals that I collected using the mapping results for RNA-seq reads to the nucleotide sequence list of nr28S and cp23S ribosomal DNA from eight clades of algae. In all individuals, RNA-seq reads were mapped to nr28S and cp23S of the clade C, and few reads were mapped to sequences of other clades (Appendix Table S3.5). This result indicates that most symbiotic algae living in the 11 individuals of *A. digitifera* belonged to clade C.

To compare gene expression, RNA-seq reads from 11 adult individuals were mapped to the transcriptome of *S. goreau* (clade C) (Davies, et al. 2018). Approximately 20 to 34 % of RNA-seq reads were mapped to the contigs (Appendix Table S3.9). For all comparisons, the expression levels of algal genes were strongly correlated between groups ( $r = 0.98\text{--}1.00$ , Appendix Table S3.8). In the comparisons based on presence and absence of *S/Me3\_a\_hil* (comparisons #1 and 2) and the expression level of *M/L-De3\_a* (comparison #3), no gene showed significant differences in expression. In the comparisons based on the expression level of M/LWE TYG-type sequences (comparison #4), I detected significant differences ( $p < 0.05$  with FDR correction) in the expression levels of one contig annotated as cytochrome b6-f complex subunit 4 between groups (Appendix Table S3.10). In the comparisons based on the expression level of CP sequences (comparison #5), I detected significant differences ( $p < 0.05$  with FDR correction) in the expression levels of one unannotated contig between groups (Appendix Table S3.10).

## Discussion

### **Genetic mechanism underlying the fluorescence polymorphism among individuals of *A. digitifera***

*Acropora* FPs are thought to be involved in the protection of photosynthetic algal symbionts against sunlight (Gittins, et al. 2015; Salih, et al. 2000; Smith, et al. 2013). However, intraspecific FP-mediated color polymorphisms have been found within the same light environment (Gittins, et al. 2015; Smith, et al. 2013). This natural variation raises questions about the importance of FPs for photo-protection. To address this question, it is necessary to determine the mechanism that generates the intraspecific fluorescence polymorphism.

The expression of *FP* genes in adult *Acropora* individuals is influenced by external stimuli, such as light, heat, and injury. CFP, GFP, RFP, and CP levels increase according to the light intensity (D'Angelo, et al. 2008; Roth, et al. 2010), and CFP is down-regulated in dark stress (DeSalvo, et al. 2011). Heat stress down-regulates GFP (Roth and Deheyn 2013). The expression of CFP increases in response to injury (D'Angelo, et al. 2012). However, these factors cannot explain variation in fluorescence in corals living in the same environment, without injury.

Previously, I reported variation in the sequences and total copy number of particular

*FP* genes in *A. digitifera* (Takahashi-Kariyazono, et al. 2016); therefore, I analyzed copy number variation in 11 adult individuals of *A. digitifera*, living in the same light environment in shallow water. I detected copy number variation for *FP* genes from each of three clades (S/MWE, M/LWE, and CP). The numbers of e3 sequences and the e3 sequence repertoires in each of three clades also differed among individuals. The numbers of e3 sequences were generally higher than the estimated gene copy numbers in the S/MWE and CP clades, suggesting that *FP* genes in these clades exhibit high allelic diversity. Hence, there were large genetic differences in *FP* genes in each of two clades among individuals of *A. digitifera*.

In each of three clades, I detected variation in the total expression levels of *FP* genes among 11 adult individuals. In the S/MWE clade, one sequence (*S/Me3\_a\_hi1*) showed a presence-absence polymorphism, and high expression of this sequence contributed to the total expression levels in the individuals with the top three overall expression levels of *FP* sequences in the S/MWE clade. In this clade, four individuals possessed *S/Me3\_a\_hi1*, but the expression levels were lower than those of the other three individuals with high expression. This lower expression may be due to differences in gene regulation or promotor sequence differences. In the M/LWE clade DYG-type, expression differences among 11 adult individuals were explained by the expression of one sequence (*M/L-De3\_a*) in three individuals that possessed this

sequence. These results suggest that particular *FP* sequences contribute to differences in total *FP* gene expression among individuals. In the CP clade, I did not observe a highly expressed sequence in adult samples. A comparative analysis of the emission spectra and expression of *FP* sequences indicated that fluorescence emission and non-emission of corals were consistent with high expression of *S/Me3\_a\_hil* and the absence of *S/Me3\_a\_hil* in the genome, respectively. According to these results, I conclude that the presence–absence differences of particular *FP* sequences with high expression contribute to the fluorescence differences among individuals in *A. digitifera*.

#### ***FP* gene expression differences between adults and larvae**

The overall expression of *FP* genes in each clade differed substantially between adults and larvae, especially for both DYD- and TYG-type in the M/LWE clade and in the CP clade. High expression of the M/LWE clade in larvae was consistent with the high proportion of RFP in the larval transcriptome of *A. millepora* (Meyer, et al. 2011). These results suggest that FPs in the M/LWE and CP clades have important roles in larvae. However, several adult individuals did not possess the *FP* sequences with high expression in larvae in the M/LWE and CP clades. This result suggests that the individuals without highly expressed sequences in larvae did not express

these sequences when they were larvae. In other words, the intensity of fluorescence and the expression of *FP* genes by highly expressed sequences might be different among larvae. Indeed, a difference in fluorescence has been reported among larvae of *A. millepora* (Kenkel, et al. 2011; Strader, et al. 2016).

In the S/MWE clade, the expressed *FP* sequences were different between adults and larvae. In the S/MWE clade, the *FP* sequence with high expression in larval stage was absent in several adult individuals, suggesting that the expression of this *FP* gene might differ among larval individuals. Hence, the expression of FPs was different between adults and larvae.

### **Hypothetical biological roles of FPs in *A. digitifera***

The biological roles of FPs have been proposed based on natural variation in fluorescence in corals (Gittins, et al. 2015; Salih, et al. 2000; Smith, et al. 2013). However, observed fluorescence variation can be explained by genetic differences or by differences in the regulation of gene expression among individuals. In this chapter, I demonstrated *FP* genetic variation among individuals of *A. digitifera*. Based on genetic differences, it was possible to separate individuals into different groups. Therefore, I examined two proposed biological roles of FPs using genetic differences and RNA-seq data containing reads from algal symbionts.

FPs in adult *Acropora* species have a proposed role in the photo-protection of algal symbionts (Gittins, et al. 2015; Salih, et al. 2000; Smith, et al. 2013). A supportive role in photosynthesis by algal symbionts has also been proposed (Salih, et al. 2000). These two biological roles predict that algal photosynthesis for individuals with high *FP* gene expression is more active than that for individuals with very low expression. In cultured symbiotic algae (genus *Symbiodinium*), up-regulation of a transcript encoding intrinsic membrane-bound light harvesting proteins is caused by the high intensity of light illumination, although most of transcripts encoding genes related to photosynthesis did not show light intensity-regulated expression (Xiang, et al. 2015).

I compared the expression of symbiotic algae within a coral host (*in hospite*) with respect to the presence and absence of highly expressed genes in the S/MWE and M/LWE (TYG- and DYG-type) clades in their genomes. In the case of CP, the groups were categorized by their expression values. In comparisons of S/MWE, M/LWE DYG-type, and CP clades, no gene related to algal photosynthesis showed differential expression, suggesting that *FP* gene expression in the S/MWE, M/LWE DYG-type, and CP clades observed in this chapter did not have strong effects on transcriptional regulatory processes of genes related to algal photosynthesis. A contig annotated as cytochrome b6-f complex subunit 4 was significantly

more highly expressed ( $p < 0.05$  with FDR correction) in individuals without M/LWE TYG-type expression than in individuals with M/LWE TYG-type expression. However, genes encoding the other subunits of cytochrome b6-f complex did not show differential expression (Appendix Table S3.11). Therefore, I could not conclude the effects of M/LWE TYG-type on algal photosynthesis based on the differential expression of only one subunit of the cytochrome b6-f complex. In this chapter, I only studied 11 individuals; thus, it is possible that RNA-seq analysis with larger data sets might detect some more subtle differences.

One finding in this chapter was that there was a difference in FP function between adults and larvae. Adult individuals with fluorescence emission mainly expressed *FP* sequences of the S/MWE clade excited by short wavelength light, including UV light, and emitted fluorescence with a peak under 490 nm. Larvae mainly expressed *FP* genes in the S/MWE clade excited by short wavelength light (visible light, no UV) and emitted fluorescence with a peak at a longer wavelength (519 nm). Among the FP recombinants that I analyzed, both the excitation and emission spectra of larval FPs had longer wavelengths than those of adult FPs. The same shift in fluorescence has been observed in *Seriatopora hystrix* (Roth, et al. 2013), suggesting that FPs may have similar roles in adults and larvae of distantly related coral species. In addition to the S/MWE clade, larvae mainly expressed *FP* genes in the M/LWE and CP clades.

The excitation and emission wavelengths were middle (TYG-type) and long (DYG-type) in the M/LWE clade. CPs absorbed long wavelength light. In summary, it is possible to hypothesized that adults utilized UV to short wavelength light and larvae utilized middle to long wavelength light via FPs, based on the functions of the highly expressed sequences in adults and larvae. These different light usages may provide insight into the biological roles of FPs.

The other key feature of *FP* genes in *A. digitifera* was the pattern of polymorphisms. The highly expressed sequences (*S/Me3\_a\_hi1* and *M/L-De3\_a*) in adults from the S/MWE and M/LWE clades were presence–absence polymorphisms in three *A. digitifera* subpopulations in the southern Ryukyu Archipelago with one exception, *M/L-De3\_a*, in Yeyama-North. In addition, two highly expressed sequences in larvae from the M/LWE clade were presence–absence polymorphisms in all subpopulations, and polymorphisms in fluorescence in larvae were predicted. These polymorphisms suggest that the presence–absence status of highly expressed *FP* sequences has been maintained in *A. digitifera* subpopulations, at least in the southern Ryukyu Archipelago. In other words, the presence and absence of fluorescence emission in adults (short wavelengths) and in larvae (middle to long wavelengths) has persisted throughout the evolution of *A. digitifera* subpopulations. This polymorphic status may be beneficial for the survival of immobile adult corals in various environments, e.g., different light

conditions, depths, currents, and attached substrates. In larvae, the polymorphic status may be beneficial for different dispersal distances, as suggested in *A. millepora* (Kenkel, et al. 2011; Strader, et al. 2016). In summary, the presence–absence polymorphisms of highly expressed *FP* sequences were a notable feature and will be key to unraveling the biological roles of FPs in corals.

In this chapter, I demonstrate the genetic basis of differences in fluorescence among individuals in *A. digitifera*. Comparisons between *FP* genotypes and the habitats or external environments of corals will provide a basis for understanding the biological roles of FPs in corals in the near future.

This chapter is based on Takahashi-Kariyazono *et al.* (2018) published in *Genome Biology and Evolution*.

## References

- Alieva NO, et al. 2008.** Diversity and evolution of coral fluorescent proteins. *PLoS One* 3: e2680. doi: 10.1371/journal.pone.0002680
- D'Angelo C, et al. 2008.** Blue light regulation of host pigment in reef-building corals. *Marine Ecology Progress Series* 364: 97-106. doi: 10.3354/meps07588
- D'Angelo C, et al. 2012.** Locally accelerated growth is part of the innate immune response and repair mechanisms in reef-building corals as detected by green fluorescent protein (GFP)-like pigments. *Coral Reefs* 31: 1045-1056. doi: 10.1007/s00338-012-0926-8
- Davies SW, Ries JB, Marchetti A, Castillo KD 2018.** *Symbiodinium* Functional Diversity in the Coral *Siderastrea siderea* Is Influenced by Thermal Stress and Reef Environment, but Not Ocean Acidification. *Frontiers in Marine Science* 5. doi: 10.3389/fmars.2018.00150
- DeSalvo MK, Estrada A, Sunagawa S, Medina M 2011.** Transcriptomic responses to darkness stress point to common coral bleaching mechanisms. *Coral Reefs* 31: 215-228. doi: 10.1007/s00338-011-0833-4
- Dove S 2004.** Scleractinian corals with photoprotective host pigments are hypersensitive to thermal bleaching. *Marine Ecology Progress Series* 272: 99-116.
- Dove SG, Hoegh-Guldberg O, Ranganathan S 2001.** Major colour patterns of reef-building corals are due to a family of GFP-like proteins. *Coral Reefs* 19: 197-204.
- Field SF, Bulina MY, Kelmanson IV, Bielawski JP, Matz MV 2006.** Adaptive evolution of multicolored fluorescent proteins in reef-building corals. *J Mol Evol* 62: 332-339. doi: 10.1007/s00239-005-0129-9
- Gittins JR, D'Angelo C, Oswald F, Edwards RJ, Wiedenmann J 2015.** Fluorescent protein-mediated colour polymorphism in reef corals: multicopy genes extend the

adaptation/acclimatization potential to variable light environments. *Mol Ecol* 24: 453-465. doi: 10.1111/mec.13041

**Henderson JN, Remington SJ 2005.** Crystal structures and mutational analysis of amFP486, a cyan fluorescent protein from *Anemonia majano*. *Proc Natl Acad Sci U S A* 102: 12712-12717. doi: 10.1073/pnas.0502250102

**Johnsen S. 2012.** The optics of life. New Jersey: Princeton University Press.

**Kelmanson IV, Matz MV 2003.** Molecular basis and evolutionary origins of color diversity in great star coral *Montastraea cavernosa* (Scleractinia: Faviida). *Mol Biol Evol* 20: 1125-1133. doi: 10.1093/molbev/msg130

**Kenkel CD, Traylor MR, Wiedenmann J, Salih A, Matz MV 2011.** Fluorescence of coral larvae predicts their settlement response to crustose coralline algae and reflects stress. *Proceedings of the Royal Society B: Biological Sciences* 278: 2691-2697. doi: 10.1098/rspb.2010.2344

**Kumar S, Stecher G, Tamura K 2016.** MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution* 33: 1870-1874.

**Labas YA, et al. 2002.** Diversity and evolution of the green fluorescent protein family. *Proc Natl Acad Sci U S A* 99: 4256-4261. doi: 10.1073/pnas.062552299

**Mariano DC, et al. 2015.** MapRepeat: an approach for effective assembly of repetitive regions in prokaryotic genomes. *bioinformatics* 11: 276-279.

**Marshall NJ, Jennings K, McFarland WN, Loew ER, Losey GS 2003.** Visual Biology of Hawaiian Coral Reef Fishes. III. Environmental Light and an Integrated Approach to the Ecology of Reef Fish Vision. *Copeia* 3: 467–480.

**Mazel CH 2003.** Contribution of fluorescence to the spectral signature and perceived color of corals. *Limnology and Oceanography* 48: 390-401.

**Meyer E, Aglyamova GV, Matz MV 2011.** Profiling gene expression responses of coral larvae (*Acropora millepora*) to elevated temperature and settlement inducers using a novel RNA-Seq procedure. *Mol Ecol* 20: 3599-3616. doi: 10.1111/j.1365-294X.2011.05205.x

**Moisan TA, Mitchell BG 2001.** UV absorption by mycosporine-like amino acids in *Phaeocystis antarctica* Karsten induced by photosynthetically available radiation. *Marine Biology* 138: 217–227.

**Oswald F, et al. 2007.** Contributions of host and symbiont pigments to the coloration of reef corals. *FEBS J* 274: 1102-1109. doi: 10.1111/j.1742-4658.2007.05661.x

**Palmer CV, Modi CK, Mydlarz LD 2009.** Coral fluorescent proteins as antioxidants. *PLoS One* 4: e7298. doi: 10.1371/journal.pone.0007298

**Pochon X, Putnam HM, Burki F, Gates RD 2012.** Identifying and characterizing alternative molecular markers for the symbiotic and free-living dinoflagellate genus *Symbiodinium*. *PLoS One* 7: e29816. doi: 10.1371/journal.pone.0029816

**Roth MS, Deheyn DD 2013.** Effects of cold stress and heat stress on coral fluorescence in reef-building corals. *Scientific Reports* 3. doi: 10.1038/srep01421

**Roth MS, Fan TY, Deheyn DD 2013.** Life history changes in coral fluorescence and the effects of light intensity on larval physiology and settlement in *Seriatopora hystrix*. *PLoS One* 8: e59476. doi: 10.1371/journal.pone.0059476

**Roth MS, Latz MI, Goericke R, Deheyn DD 2010.** Green fluorescent protein regulation in the coral *Acropora yongei* during photoacclimation. *Journal of Experimental Biology* 213: 3644-3655.

**Salih A, Larkum A, Cox G, Kühl M, Hoegh-Guldberg O 2000.** Fluorescent pigments in corals are photoprotective. *Nature* 408: 850-853.

**Shinzato C, Mungpakdee S, Arakaki N, Satoh N 2015.** Genome-wide SNP analysis explains coral diversity and recovery in the Ryukyu Archipelago. *Sci Rep* 5: 18211. doi: 10.1038/srep18211

**Shinzato C, et al. 2011.** Using the *Acropora digitifera* genome to understand coral responses to environmental change. *Nature* 476: 320-323. doi: 10.1038/nature10249

**Shinzato C, Shoguchi E, Tanaka M, Satoh N 2012.** Fluorescent protein candidate genes in the coral *Acropora digitifera* genome. *Zoolog Sci* 29: 260-264. doi: 10.2108/zsj.29.260

**Smith EG, D'Angelo C, Salih A, Wiedenmann J 2013.** Screening by coral green fluorescent protein (GFP)-like chromoproteins supports a role in photoprotection of zooxanthellae. *Coral Reefs* 32: 463-474. doi: 10.1007/s00338-012-0994-9

**Strader ME, Aglyamova GV, Matz MV 2016.** Red fluorescence in coral larvae is associated with a diapause-like state. *Molecular Ecology* 25: 559-569. doi: 10.1111/mec.13488

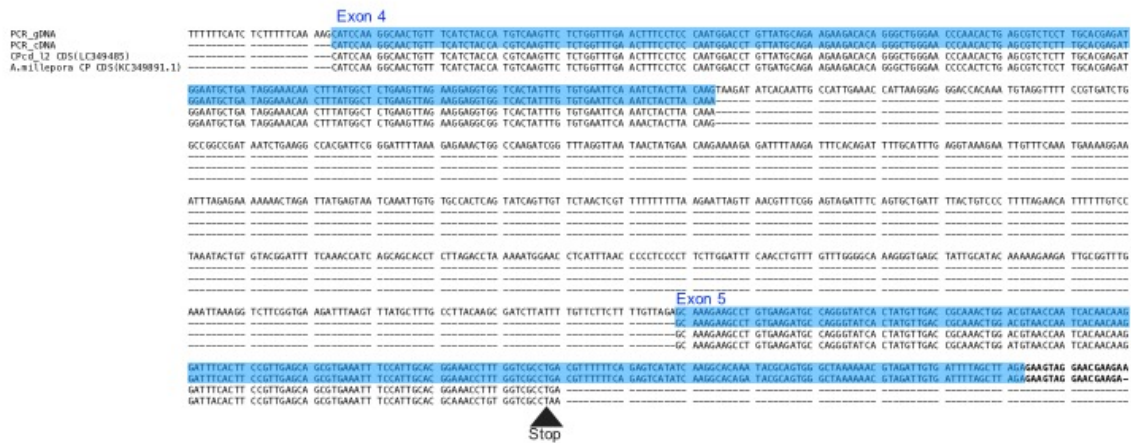
**Takahashi-Kariyazono S, Gojobori J, Satta Y, Sakai K, Terai Y 2016.** *Acropora digitifera* encodes the largest known family of fluorescent proteins that has persisted during the evolution of *Acropora* species. *Genome Biology and Evolution* 8: 3271-3283.

**Veron JEN. 2000.** *Corals of the World*. Townsville, Australia: Australian Institute of Marine Science.

**Xiang T, Nelson W, Rodriguez J, Tolleter D, Grossman AR 2015.** *Symbiodinium* transcriptome and global responses of cells to immediate changes in light intensity when grown under autotrophic or mixotrophic conditions. *Plant J* 82: 67-80. doi: 10.1111/tpj.12789

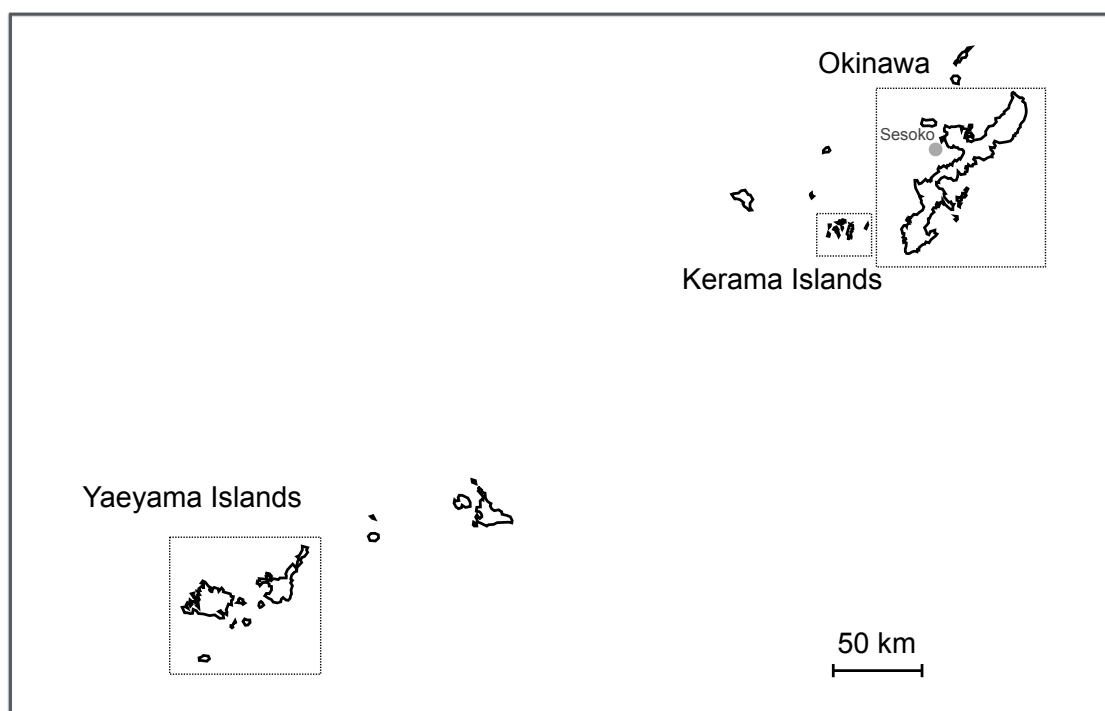
## Appendix

[illegible]

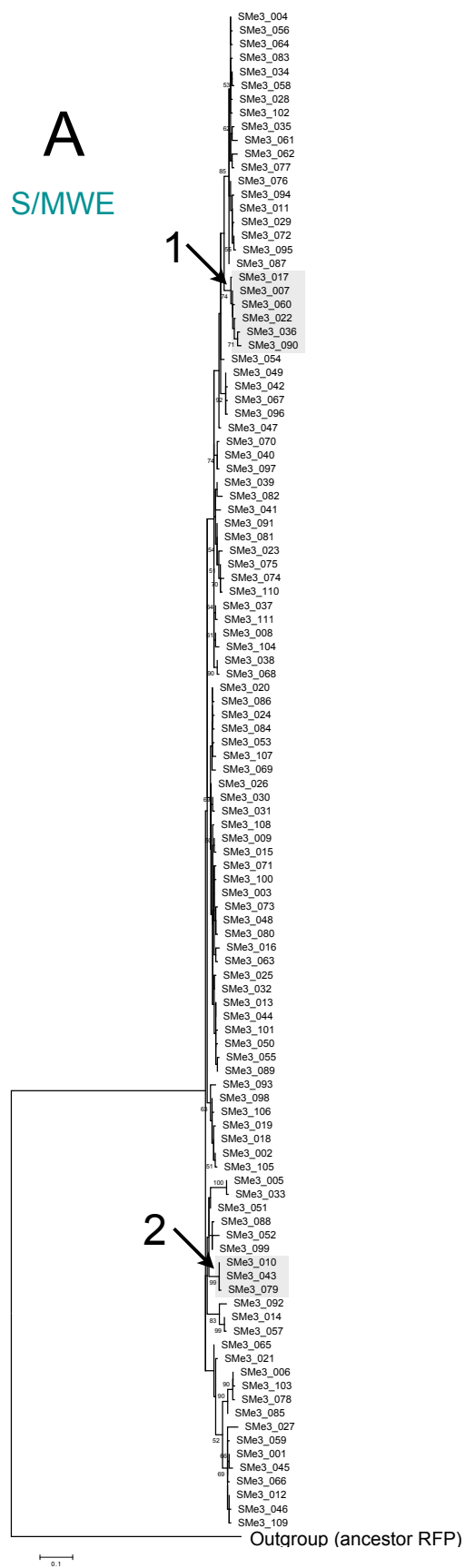


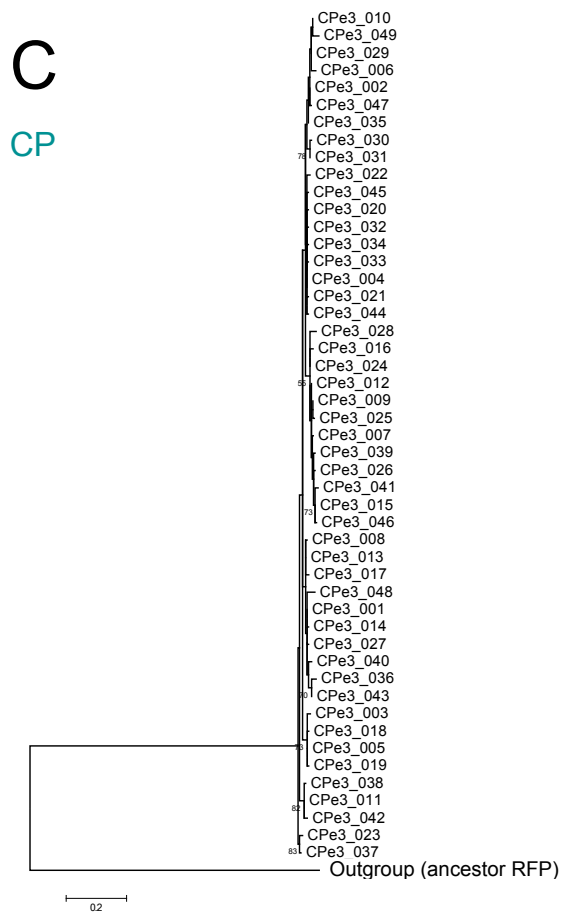
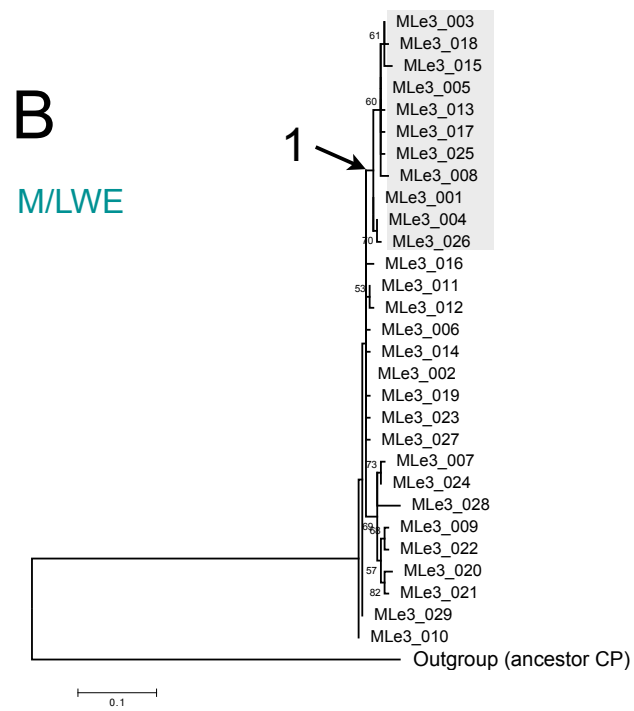
Appendix Figure S3.1. The exon-intron structure of a CP gene

A *CP* gene sequence was amplified from genomic DNA (details are provided below), and was aligned with three sequences; a cDNA sequence of a *CP* gene amplified from *A. digitifera* genomic DNA, coding region sequences of CP from *A. digitifera* (accessions: LC349485), and *A. millepora* (accession: KC349891.1). The exons are shown in blue, and the sequences of primers that were used for PCR amplification are shown by characters in bold. The triangle indicates the positions of start and stop codons. *CP* gene sequences were amplified using TaKaRa Ex Taq (Takara) and the following primer set: FP8\_5U\_F2 and FP8\_5U\_R1. Genomic DNA of *A. digitifera* (sample ID: S1601) was used as a template. PCR conditions for the amplification of *CP* gene sequences were as follows: denaturation for 3 min at 94°C, followed by 35 cycles of denaturation for 1 min at 94°C, annealing for 1 min at 55°C, and extension for 3 min at 72°C. The PCR product was cloned into the T-Vector pMD20 vector (Takara), and the sequences were verified using the Applied Biosystems Automated 3130xl Sequencer (Foster City, CA, USA) and the following primers: M13primerRV, M13primerM4, MiA\_CP\_e3\_R1-3, 180426CPi\_F1, 180426CPe3\_R1, 180427CPi\_F2, and 180427CPi\_R1.



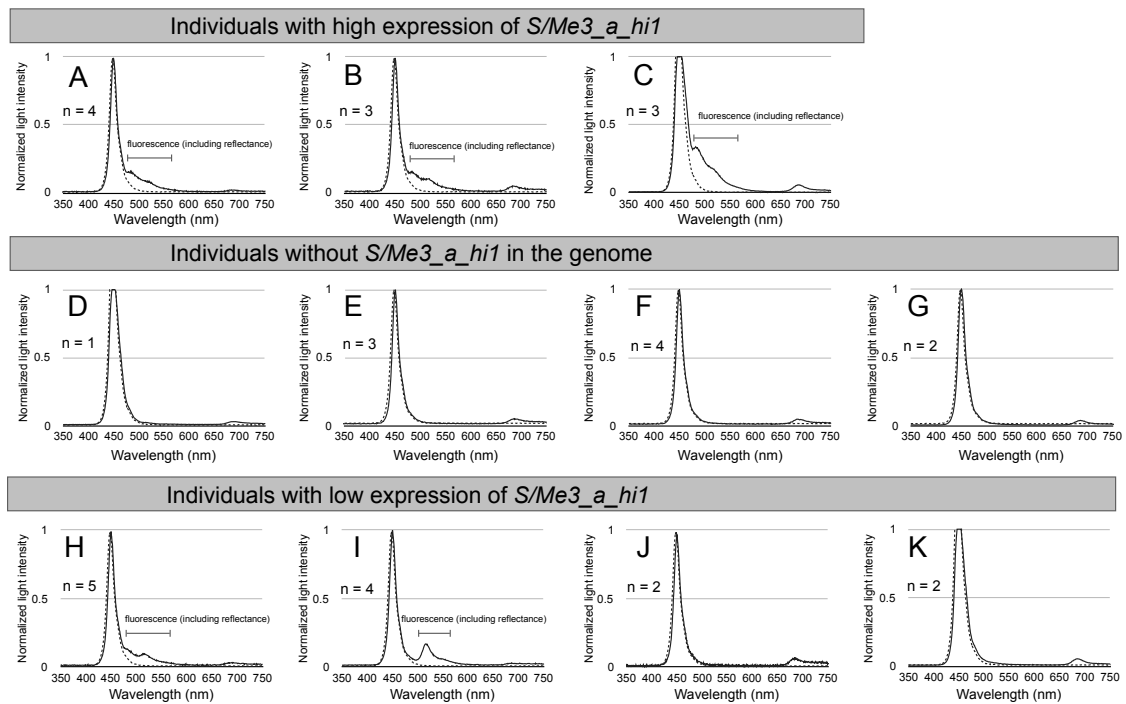
Appendix Figure S3.2. Localities of *A. digitifera* populations





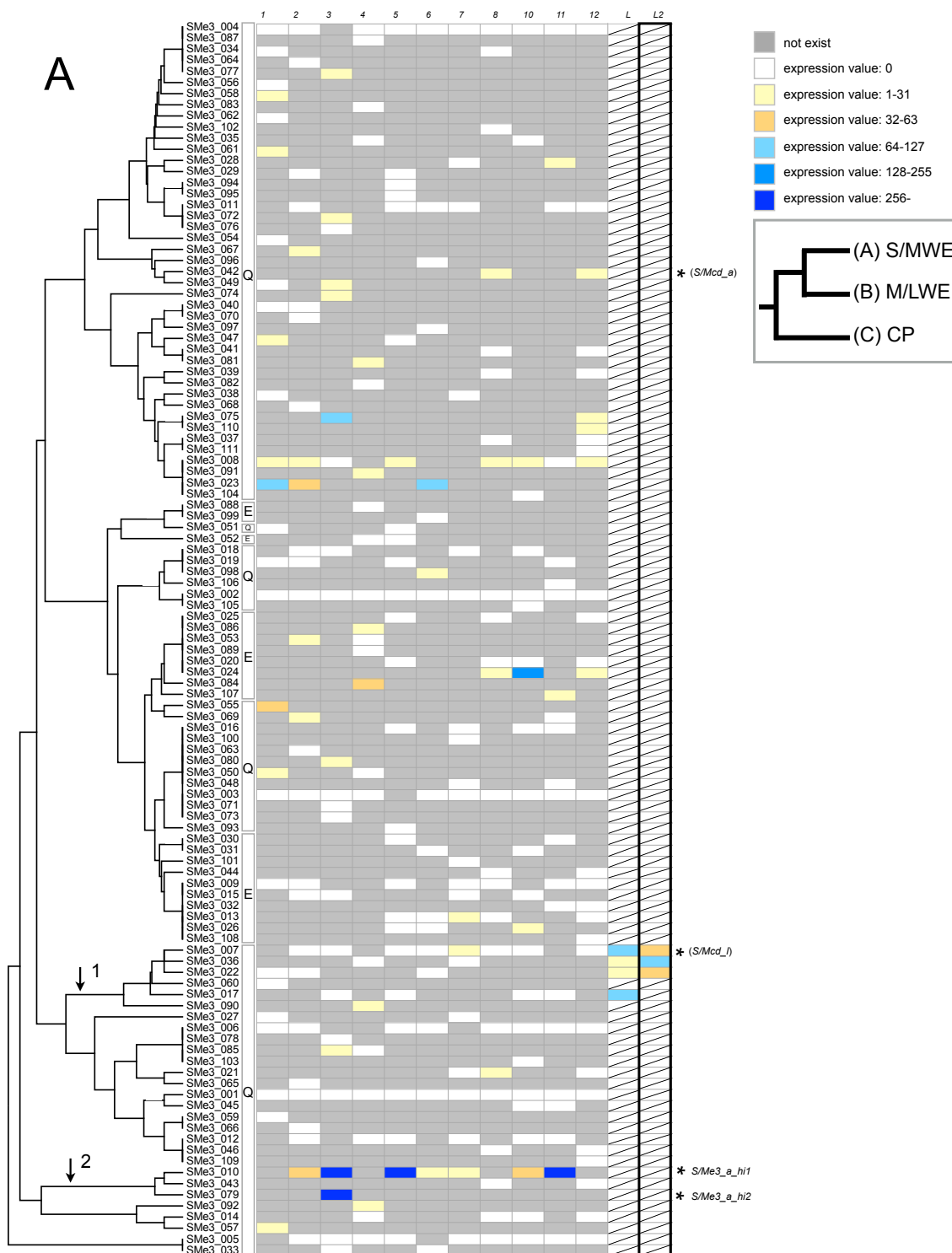
**Appendix Figure S3.3. Phylogenetic relationships based on exon 3 sequences**

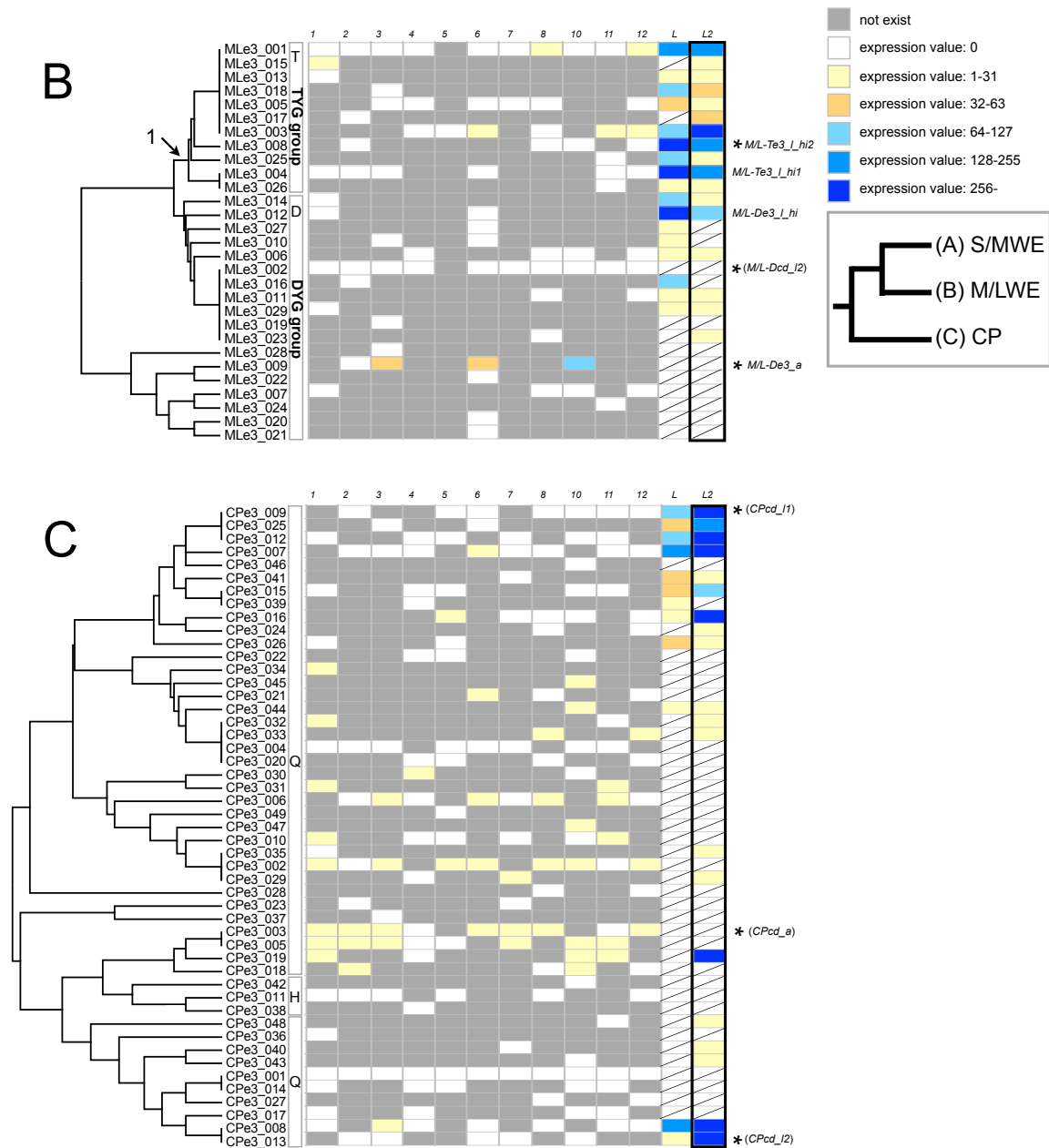
Phylogenetic relationships within each clade were analyzed using the maximum likelihood method with the Kimura 2-parameter model (+G, +I). Bootstrap support for each clade was obtained by 1,000 replicates and probabilities (>50) are shown next to each node. Estimated ancestral RFP was used as an outgroup for the S/MWE clade and the scale bar represents 0.1 substitutions per site (A); estimated ancestral CP was used as an outgroup in the M/LWE clade and the scale bar represents 0.2 substitutions per site (B); and estimated ancestral RFP was used as an outgroup in the CP clade and the scale bar represents 0.1 substitutions per site (C). Black arrow indicates the branch support of the phylogenetic relationships shown in Figure 2.



Appendix Figure S3.4. Live coral fluorescence measurements in the ocean

Horizontal and vertical axes indicate wavelengths (nanometers) and normalized light intensities, respectively. Emission spectra of *A. digitifera* with high expression of *S/Me3\_a\_hi1* (A, S1603; B, S1605; and C, S1611), without *S/Me3\_a\_hi1* in the genome (D, S1601; E, S1604; F, S1608; and G, S1612), and low expression of *S/Me3\_a\_hi1* (H, S1602; I, S1606; J, S1607; and K, S1610) are shown. “n” indicates the number of measurements. A representative emission spectrum is shown by a solid line. The LED excitation spectrum is shown by a dotted line.





### Appendix Figure S3.5. Expression of each *FP* sequence in the additional larval dataset

Rows indicate the exon 3 sequences belonging to clade S/MWE (A), M/LWE clade (B), and CP clade (C). Columns indicate the normalized expression values for *FP* sequences calculated by the expression of the internal control genes (see METHODS) in 11 adults and two independent larval specimens. Sample IDs S1601–S1612 are indicated as 01–12. L indicates the larval specimen (accession: SRX1534820) that is also shown in Figure 2. L2 indicates the larval specimen (accession: DRR054773) added in this figure (the part encircled by thick black line). A grey box indicates the absence of the sequence in a genome of a sample and a slash mark indicates that the sequence data were not available. Exon 3 sequences were aligned based on the phylogenetic trees in Fig. S2 (UPGMA based on amino acid differences) to show amino acid differences. Black arrow indicates a branch supported by the ML tree in Figure S3. The first peptide of the chromophore for each exon 3 sequences is shown in a rectangle. An asterisk indicates the exon 3 sequence used for recombinant protein analysis.

Appendix Table S3.1. The overview of high-throughput sequencing of FP gene captured sequences

Sample ID	Accession number	Total reads	Merge overlapping paires		Mapped reads on exon3 (merged)		Mapped reads on exon3 (not merged)	
			Merged (%)	Not merged (%)	S/MWE	M/LWE	S/MWE	M/LWE
S1601	DRR120550	2,198,726	34.4	65.6	267	69	893	205
S1602	DRR120551	2,503,772	32.7	67.3	295	96	845	287
S1603	DRR120552	2,850,976	33.8	66.2	319	121	834	402
S1604	DRR120553	2,219,502	34.5	65.6	282	64	787	178
S1605	DRR120554	2,361,828	33.4	66.7	405	46	1,127	169
S1606	DRR120555	2,141,708	33.0	67.0	266	119	746	360
S1607	DRR120556	2,625,478	32.1	68.0	316	76	1,001	267
S1608	DRR120557	2,345,116	32.7	67.3	334	89	984	324
S1610	DRR120558	2,592,244	31.5	68.5	351	152	1,097	503
S1611	DRR120559	2,651,848	32.8	67.2	344	118	1,058	335
S1612	DRR120560	2,612,768	33.3	66.7	369	91	1,052	294
								654

Appendix Table S3. 2. Primer sequences

Name of primers	Sequence of primers
FP2_5UHind_F1	5'-CACACAAAGCTTGACRGAGTGAACRCCGTRAACAG-3'
FP2_3UBam_R1	5'- CACACAGGATCCGTAAGTATTGAAAGTGCTTCCATC -3'
FP10_5UHind_F1	5'- CACACAAAGCTTGCARGAGAGTTACTHGASTTTTGTATCAC -3'
FP10_3UBam_R1	5'- CACACAGGATCCCAGAGWAATAAGCATGCGCATGTAAC -3'
FP8_5U_F2	5'- TCTCTTTCACGGAAGGAGCTACTTAA -3'
FP8_5U_R1	5'- GTYTACACTTCTTCGTTCTCTACTTC -3'
AdiFP2KpnI_F2_L	5'-CACACAGGTACCATGTCTCATTCAAAGCAAGGCAT -3'
AdiFP2XbaI_R_L	5'- TGTGTGTCTAGATTATTTAACCTTCAAAGGGTTA -3'
AdiFP10L_KpnIF_L	5'- CACACAGGTACCATGGCTCTGTCAAAGCACGGTCTA -3'
AdiFP10L_KpnIF2_L	5'- CACACAGGTACCATGGCTTTGTCAAAGCACGGTCTA -3'
AdiFP10L_KpnIF3_L	5'- CACAGGTACCATGGCTCTGCCAAAGCACGG -3'
AdiFP10L_XbaIR5_L	5'- TGTGTGTCTAGATTATCAGGCCAATGCGGA -3'
AdiFP10L_XbaIR3_L	5'- TGTGTCTAGATTATCCGGGCAATGCGGATC -3'
AdiFP10L_XbaIR4_L	5'- TGTGTCTAGATTATCAGGGCAATGCGGATC -3'
AdiFP8L_KpnIF_L	5'- CACACAGGTACCATGAGTGTGATCGCTAA -3'
AdiFP8XbaI_R1_L	5'- TGTGTGTCTAGATCAGGCGACCACAGGTTTGC -3'
AdiFP8XbaI_R3_L	5'- TGTGTCTAGATCAGGCGACCAAAGGTTTGC -3'
MiA_MWEe2_F1	5'- AATGAAGACGAAATACCAYATGGAA -3'
MiA_MLWEe2_F3	5'- CCACGGTAATGGAAATCCTTT -3'
MiA_MLWE5_R2	5'- GGTCTTCCCGAATGAGCTT -3'
AdiFP8XbaI_R2_L	5'- TGTGTGTCTAGATCAGGCGACCACAAGTTTGC -3'
MLWE_qPCR_R1-3	5'- CTTTCAATATCCCCTGTCTAGGTACTGGTA -3'
MiA_CP_e3_R1-3	5'- TCTTCAAAGTTCATGATCCTCTCCCATGT -3'
180426CPi_F1	5'- GATTATTCTTAGTGGTCTCTAGAAC -3'
180426CPe3_R1	5'- ATCGCCTTCGACCTCAAAGTAGT -3'
180427CPi_F2	5'- CGGAAGTATAGTTTTTTCACGTTTCAC -3'
180427CPi_R1	5'- GACAGGTTTAGTCTAATCGCATCA -3'

Appendix Table S3. 3. Accession numbers of FP sequences used for estimation of ancestral sequences

S/MWE		M/LWE		CP	
accession number	species	accession number	species	accession number	species
AY646066	<i>Acropora aculeus</i>	LC125047	<i>Acropora digitifera</i>	AY646077	<i>Acropora aculeus</i>
AY646069	<i>Acropora aculeus</i>	LC125048	<i>Acropora digitifera</i>	LC125058	<i>Acropora digitifera</i>
LC125053	<i>Acropora digitifera</i>	LC125049	<i>Acropora digitifera</i>	LC125059	<i>Acropora digitifera</i>
LC125054	<i>Acropora digitifera</i>	LC125050	<i>Acropora digitifera</i>	LC125060	<i>Acropora digitifera</i>
LC125088	<i>Acropora digitifera</i>	LC125051	<i>Acropora digitifera</i>	LC125061	<i>Acropora digitifera</i>
LC125055	<i>Acropora digitifera</i>	LC125052	<i>Acropora digitifera</i>	LC125062	<i>Acropora digitifera</i>
LC125056	<i>Acropora digitifera</i>	LC125067	<i>Acropora digitifera</i>	LC125063	<i>Acropora digitifera</i>
LC125089	<i>Acropora digitifera</i>	LC125068	<i>Acropora digitifera</i>	LC125064	<i>Acropora digitifera</i>
LC125057	<i>Acropora digitifera</i>	LC125069	<i>Acropora digitifera</i>	LC125065	<i>Acropora digitifera</i>
LC125070	<i>Acropora digitifera</i>	JN631072	<i>Acropora millepora</i>	LC125066	<i>Acropora digitifera</i>
LC125071	<i>Acropora digitifera</i>	JX258846	<i>Acropora millepora</i>	LC125072	<i>Acropora digitifera</i>
AY646067	<i>Acropora millepora</i>	JX258845	<i>Acropora millepora</i>	LC125073	<i>Acropora digitifera</i>
EU709810	<i>Acropora millepora</i>	JX258844	<i>Acropora millepora</i>	LC125074	<i>Acropora digitifera</i>
EU709809	<i>Acropora millepora</i>	AY646073	<i>Acropora millepora</i>	LC125075	<i>Acropora digitifera</i>
EU709808	<i>Acropora millepora</i>	EU709811	<i>Acropora millepora</i>	AY646076	<i>Acropora hyacinthus</i>
AY646070	<i>Acropora millepora</i>	AB626607	<i>Acropora tenuis</i>	KC349891	<i>Acropora millepora</i>
AY646068	<i>Acropora nobilis</i>			AY646075	<i>Acropora millepora</i>
AY646072	<i>Acropora nobilis</i>			KC411499	<i>Acropora millepora</i>
AY646071	<i>Acropora nobilis</i>			KC411500	<i>Acropora millepora</i>
EU709806	<i>Acropora pulchra</i>			EU709807	<i>Acropora pulchra</i>
				AY646074	<i>Acropora tenuis</i>

Appendix Table S3.4. The list of house-keeping genes

Gene ID	Product	Length (bp)	Average (RPKM)
LOC107346156	cathepsin D-like (LOC107346156)	5126	230
LOC107328417	transgelin-3-like (LOC107328417)	9812	227
LOC107346828	syntenin-1-like (LOC107346828)	5094	216
LOC107356370	myosin regulatory light chain 12B-like (LOC107356370)	2162	215
LOC107338479	calsyntenin-1-like (LOC107338479)	11466	207
LOC107358260	apoptosis regulator R1-like (LOC107358260)	10628	196
LOC107332968	flotillin-1-like (LOC107332968)	21999	162
LOC107350085	protein lifeguard 4-like (LOC107350085)	5638	150
LOC107352730	ubiquitin-conjugating enzyme E2 L3-like (LOC107352730)	6490	133
LOC107344579	ras GTPase-activating protein-binding protein 2-like (LOC107344579)	3266	128
LOC107338886	calcium-binding protein P-like (LOC107338886)	4308	121
LOC107332675	ras-related C3 botulinum toxin substrate 1-like (LOC107332675)	5764	117
LOC107348027	silk gland factor 3-like (LOC107348027)	1911	101
LOC107337763	CDC42 small effector protein 2-like (LOC107337763)	7812	96
LOC107356181	serine/arginine-rich splicing factor 6-like (LOC107356181)	10956	95
LOC107345356	tetraspanin-3-like (LOC107345356)	11927	90
LOC107338193	ras-related protein Rab-1A-like (LOC107338193)	3841	89
LOC107331231	eukaryotic initiation factor 4A-III (LOC107331231)	7703	85
LOC107335383	14-3-3 protein 1-like (LOC107335383)	8083	79
LOC107339955	plasma membrane calcium-transporting ATPase 1-like (LOC107339955)	6188	74
LOC107357730	secretory carrier-associated membrane protein 1-like (LOC107357730)	13574	72
LOC107332903	far upstream element-binding protein 2-like (LOC107332903)	12684	67
LOC107327056	charged multivesicular body protein 1a-like (LOC107327056)	5193	64
LOC107327350	ras-related protein Rab-5C (LOC107327350)	6422	63
LOC107355951	SUMO-conjugating enzyme UBC9-like (LOC107355951)	4834	62
LOC107351701	nuclear transcription factor Y subunit A-9-like (LOC107351701)	4008	62
LOC107339508	26S protease regulatory subunit 8-like (LOC107339508)	10189	62
LOC107344110	ubiquitin-associated protein 2-like (LOC107344110)	26152	56
LOC107348394	arsenite methyltransferase-like (LOC107348394), partial mRNA	10447	55
LOC107338504	segment polarity protein dishevelled homolog DVL-3-like (LOC107338504)	14665	53
LOC107350555	ras-related protein Rab-18-B-like (LOC107350555)	4831	53
LOC107347947	eukaryotic translation initiation factor 1A-like (LOC107347947)	5723	51
LOC107332574	myosin-2 essential light chain-like (LOC107332574)	971	50
LOC107348743	ubiquitin-conjugating enzyme E2 K-like (LOC107348743)	6541	50

Appendix Table S3.5. The Number of RNA-seq read mapped to nr28S and cp23S sequences of the eight *Symbiodinium* clades

Query information			Sample ID / Number of reads mapped in pairs										
Clade	Region	Accession number	S1601	S1602	S1603	S1604	S1605	S1606	S1607	S1608	S1610	S1611	S1612
A13	cp23S	JN558027	0	0	0	0	0	0	0	0	0	0	0
	nr28S	JN558094	0	0	0	0	0	0	0	0	0	0	0
B1	cp23S	JN557991	0	4	0	0	0	2	0	0	0	0	0
	nr28S	JN558057	0	0	0	0	0	0	0	0	0	0	0
C1	cp23S	JN557969	2,002	3,634	1,166	656	1,072	1,220	920	780	740	1,570	1,142
	nr28S	JN558040	4,334	4,424	2,086	724	1,260	3,204	952	934	1,202	2,428	1,664
D1	cp23S	JN558007	0	0	0	0	0	0	0	0	0	0	0
	nr28S	JN558075	0	0	0	0	0	0	0	0	0	0	0
E1	cp23S	JN558015	0	0	0	0	0	0	0	0	0	0	0
	nr28S	JN558084	0	0	0	0	0	0	0	0	0	0	0
F1	cp23S	JN557997	0	0	0	0	0	0	0	0	0	0	0
	nr28S	JN558066	0	0	0	0	0	0	0	0	2	0	0
G2	cp23S	JN558019	0	0	0	0	0	0	0	0	0	0	0
	nr28S	JN558089	0	0	0	0	0	0	0	0	0	0	0
H1	cp23S	JN557981	2	22	0	0	6	0	6	6	0	2	6
	nr28S	JN558051	0	2	0	0	0	0	2	0	0	0	0

Appendix Table S3.6. Sequence IDs and primer sets

Clade	Exon 3 sequence ID	Additional exon 3 sequence name	CDS clone ID	Primers for full-length cDNA		Primers for subcloning	
				forward primer	reverse primer	forward primer	reverse primer
S/MWE	SMe3_007	-	S/Mcd_I	FP2_5UHind_F1	FP2_3UBam_R1	AdiFP2KpnI_F2_L	AdiFP2XbaI_R_L
	SMe3_042	-	S/Mcd_a	FP2_5UHind_F1	FP2_3UBam_R1	AdiFP2KpnI_F2_L	AdiFP2XbaI_R_L
	SMe3_010	S/Me3_a_hi1	S/Mcd_a_hi1(FP2_S1603_2)		(Takahashi-Kariyazono et al. 2016)		
	SMe3_079	S/Me3_a_hi2	S/Mcd_a_hi2(FP2_S1603_1)		(Takahashi-Kariyazono et al. 2016)		
M/LWE	MLe3_009	M/L-De3_a	M/L-Dcd_a	FP10_5UHind_F1	FP10_3UBam_R1	AdiFP10L_KpnIF3_L	AdiFP10L_XbaIR3_L
	MLe3_012	M/L-De3_I_hi	-	-	-	-	-
	MLe3_002	-	M/L-Dcd_I2	FP10_5UHind_F1	FP10_3UBam_R1	AdiFP10L_KpnIF2_L	AdiFP10L_XbaIR4_L
	MLe3_004	M/L-Te3_I_hi1	-	-	-	-	-
CP	MLe3_008	M/L-Te3_I_hi2	M/L-Tcd_I_hi2	FP10_5UHind_F1	FP10_3UBam_R1	AdiFP10L_KpnIF_L	AdiFP10L_XbaIR5_L
	CPe3_003	-	CPcd_a	FP8_5U_F2	FP8_5U_R1	AdiFP8L_KpnIF_L	AdiFP8XbaI_R1_L
	CPe3_009	-	CPcd_I1	FP8_5U_F2	FP8_5U_R1	AdiFP8L_KpnIF_L	AdiFP8XbaI_R3_L
	CPe3_013	-	CPcd_I2	FP8_5U_F2	FP8_5U_R1	AdiFP8L_KpnIF_L	AdiFP8XbaI_R3_L

Appendix Table S3.7. The existence of highly expressed e3 sequence in each individual

Subpopulation	Sample name	Accession number	FP e3 sequence name					
			S/Me3_a_hi1	S/Me3_a_hi2	M/L-De3_a	M/L-Te3_l_hi1	M/L-Te3_l_hi2	M/L-De3_l_hi
Okinawa	S1601	DRR108003	N	N	U	P	N	P
	S1602	DRR108004	P	U	P	P	U	N
	S1603	DRR108005	P	P	P	P	U	U
	S1604	DRR108006	N	N	N	N	U	U
	S1605	DRR108007	P	U	N	U	N	N
	S1606	DRR108008	P	U	P	P	U	P
	S1607	DRR108009	P	U	N	U	U	N
	S1608	DRR108010	N	N	N	U	U	N
	S1610	DRR108024	U	U	P	N	P	N
	S1611	DRR108011	P	U	N	P	U	N
	S1612	DRR108012	U	N	N	U	P	U
	Okinawa_Sesoko_sesoko1	DRR099286	P	U	P	U	U	U
	Okinawa_Sesoko_sesoko2	DRR099287	U	U	U	P	U	U
	Okinawa_Sesoko_sesoko3	DRR099288	U	U	P	P	U	U
	Okinawa_Sesoko_sesoko4	DRR099289	P	U	P	P	U	P
	Okinawa_Sesoko_Ss5	DRR099282	U	U	N	U	U	U
	Okinawa_Sesoko_Ss6	DRR099283	U	U	N	P	U	U
	Okinawa_Sesoko_Ss7	DRR099284	P	U	N	P	P	U
	Okinawa_Sesoko_Ss8	DRR099285	N	U	P	U	U	U
	Okinawa_Hedo_Hd1	DRR099273	U	U	N	U	U	U
	Okinawa_Hedo_Hd5	DRR099277	N	U	U	U	U	U
	Okinawa_Hedo_Hd8	DRR099280	U	N	N	P	U	U
	Okinawa_Manza_Mz1	DRR099303	P	U	N	P	P	P
	Okinawa_Manza_Mz2	DRR099304	U	N	N	U	P	P
	Okinawa_Manza_Mz3	DRR099291	U	N	U	P	U	P
	Okinawa_Manza_Mz4	DRR099305	P	U	U	N	P	U
	Okinawa_Manza_Mz5	DRR099292	P	N	P	P	U	N
	Okinawa_Manza_Mz6	DRR099293	U	U	U	U	U	N
	Okinawa_Manza_Mz7	DRR099294	P	U	N	P	U	U
	Okinawa_Manza_Mz9	DRR099296	U	U	P	U	P	N
	Okinawa_Manza_MzC1	DRR099298	P	U	N	N	N	N
	Okinawa_Manza_MzC2	DRR099300	U	U	N	U	U	U
	Okinawa_Manza_MzC4	DRR099302	N	N	P	N	U	U
	Okinawa_Ohdo_ohdo1	DRR099334	P	U	P	P	U	N
	Okinawa_Ohdo_ohdo3	DRR099335	U	U	U	P	U	U
	Okinawa_Ohdo_ohdo5	DRR099336	P	U	N	P	U	U
	Okinawa_Ohdo_ohdo7	DRR099337	P	U	N	P	U	U
	Okinawa_Ohdo_Od10	DRR099315	P	U	N	P	P	U
	Okinawa_Ohdo_Od11	DRR099316	P	U	N	P	U	P
	Okinawa_Ohdo_Od12	DRR099317	P	U	N	P	U	U
	Okinawa_Ohdo_Od13	DRR099318	P	U	N	U	U	N
	Okinawa_Ohdo_Od14	DRR099319	U	N	P	P	U	P
	Okinawa_Ohdo_Od15	DRR099320	P	U	U	P	U	P
	Okinawa_Ohdo_Od18	DRR099321	P	U	N	P	P	U
	Okinawa_Ohdo_Od19	DRR099322	P	U	N	P	U	P
	Okinawa_Ohdo_Od22	DRR099323	P	U	U	P	P	P
	Okinawa_Ohdo_Od23	DRR099324	P	U	U	P	P	N
	Okinawa_Ohdo_Od24	DRR099325	U	U	N	P	U	P
	Okinawa_Ohdo_Od25	DRR099326	P	U	U	P	U	U
	Okinawa_Ohdo_Od26	DRR099327	P	U	N	U	P	N
	Okinawa_Ohdo_Od27	DRR099328	N	N	N	P	P	P

Appendix Table S3.7. The existence of highly expressed e3 sequence in each individual

(continued)

Subpopulation	Sample name	Accession number	FP e3 sequence name					
			S/Me3_a_hi1	S/Me3_a_hi2	M/L-De3_a	M/L-Te3_l_hi1	M/L-Te3_l_hi2	M/L-De3_l_hi
Kerama	Kerama_Geruma_KrA1	DRR099380	P	P	P	P	U	U
	Kerama_Geruma_KrA2	DRR099381	P	U	U	P	U	U
	Kerama_Geruma_KrA3	DRR099382	P	U	N	P	P	U
	Kerama_Geruma_KrA4	DRR099383	U	N	N	P	P	N
	Kerama_Geruma_KrA5	DRR099384	U	N	N	U	U	P
	Kerama_Geruma_KrA6	DRR099385	P	U	N	P	P	P
	Kerama_Geruma_KrA7	DRR099386	P	U	P	P	U	U
	Kerama_Geruma_KrA8	DRR099387	U	U	N	U	N	U
	Kerama_Geruma_KrA9	DRR099388	U	P	P	P	U	U
	Kerama_Geruma_KrA10	DRR099374	P	U	N	P	N	U
	Kerama_Geruma_KrA11	DRR099375	P	U	U	P	U	U
	Kerama_Geruma_KrA12	DRR099376	P	U	P	P	P	P
	Kerama_Geruma_KrA13	DRR099377	P	U	N	U	U	P
	Kerama_Geruma_KrA14	DRR099378	N	N	P	P	U	P
	Kerama_Geruma_KrA15	DRR099379	P	U	U	P	U	U
	Kerama_Yakabi_KrC1	DRR099365	N	N	P	P	P	P
	Kerama_Yakabi_KrC2	DRR099366	P	N	N	P	P	N
	Kerama_Yakabi_KrC3	DRR099367	P	U	U	P	U	U
	Kerama_Yakabi_KrC5	DRR099369	P	U	N	P	U	U
	Kerama_Yakabi_KrC6	DRR099370	P	U	N	P	U	P
	Kerama_Yakabi_KrC7	DRR099371	U	N	N	P	U	P
	Kerama_Yakabi_KrC8	DRR099372	P	U	N	P	P	P
	Kerama_Yakabi_KrC9	DRR099373	N	U	P	P	U	N
	Kerama_Yakabi_KrC10	DRR099359	U	N	N	P	U	U
	Kerama_Yakabi_KrC11	DRR099360	P	U	N	P	U	P
	Kerama_Yakabi_KrC12	DRR099361	P	U	N	P	U	P
	Kerama_Yakabi_KrC13	DRR099362	P	U	N	P	U	P
	Kerama_Yakabi_KrC14	DRR099363	U	U	U	P	U	P
	Kerama_Yakabi_KrC15	DRR099364	P	U	U	P	P	P
	Kerama_Zamami_KrE1	DRR099344	P	N	P	P	U	N
	Kerama_Zamami_KrE2	DRR099345	P	U	N	P	U	U
	Kerama_Zamami_KrE3	DRR099346	P	U	N	U	U	P
	Kerama_Zamami_KrE4	DRR099347	P	P	N	U	P	N
	Kerama_Zamami_KrE5	DRR099348	P	U	N	P	U	P
	Kerama_Zamami_KrE6	DRR099349	P	U	N	P	U	P
	Kerama_Zamami_KrE8	DRR099351	P	U	N	P	U	P
	Kerama_Zamami_KrE11	DRR099339	P	U	U	P	P	P
	Kerama_Zamami_KrE12	DRR099340	U	N	P	U	U	N
	Kerama_Zamami_KrE15	DRR099343	N	N	N	U	P	U
Yeahama-North	Yaeyama_Kabira_Isy4	DRR099401	P	U	P	U	U	U
	Yaeyama_Kabira_Isy7	DRR099402	U	N	P	U	N	U
	Yaeyama_Kabira_Isy8	DRR099403	P	U	P	U	P	N
	Yaeyama_Kabira_Isy12	DRR099389	U	N	P	P	U	P
	Yaeyama_Kabira_Isy15	DRR099390	N	U	P	P	N	U
	Yaeyama_Kabira_Isy16	DRR099391	N	U	P	U	N	N
	Yaeyama_Kabira_Isy17	DRR099392	N	N	P	P	U	P
	Yaeyama_Kabira_Isy18	DRR099393	U	U	P	U	U	N
	Yaeyama_Kabira_Isy26	DRR099399	P	U	P	U	P	N
	Yaeyama_Uehara_Irm2	DRR099423	U	N	P	P	U	P
	Yaeyama_Uehara_Irm21	DRR099416	N	N	P	U	U	P
	Yaeyama_Uehara_Irm27	DRR099422	N	N	P	U	U	U

Appendix Table S3.8. The overview of RNA-seq analysis

Comparison#	Clade	Group1	Group2	Group1: sample ID	Group2: sample ID	r	Note
1	S/MWE	High expression of <i>S/Me3_a_hi1</i>	Absence of <i>S/Me3_a_hi1</i> in genome	S1603, 5, 11	S1601, 4, 8, 12	0.99	S1610 (intermediate expression) were not included.
2	S/MWE	Presence of <i>S/Me3_a_hi1</i>	Absence of <i>S/Me3_a_hi1</i> in genome	S1602-3, 5-7, 10-11	S1601, 4, 8, 12	1.00	
3	M/LWE (DYG)	Expression of <i>M/L-De3_a</i>	Absence of <i>M/L-De3_a</i> in genome	S1603, 6, 10	S1601-2, 4-5, 7-8, 11-12	0.99	
4	M/LWE (TYG)	Expression of TYG sequences	No expression of TYG sequences	S1601, 5, 8, 11-12	S1602, 4, 7	0.98	Samples with DYG expression were not included.
5	CP	Total expression $\geq 10$	Total expression $\leq 5$	S1601, 6, 10	S1604-5, 7-8	0.98	

Appendix Table S3.9. The overview of RNA-seq

Sample ID	Accession number	Total Reads	Mapped reads to <i>A. digitifera</i> (%)	Mapped reads to Symbiodinium clade C (%)
<b>S1601</b>	DRR108013	75,591,730	43.5	27.9
<b>S1602</b>	DRR108014	109,525,174	50.1	20.9
<b>S1603</b>	DRR108015	53,761,100	40.1	30.9
<b>S1604</b>	DRR108016	55,333,358	45.5	24.5
<b>S1605</b>	DRR108017	47,341,614	36.3	34.3
<b>S1606</b>	DRR108018	62,627,534	47.7	25.5
<b>S1607</b>	DRR108019	41,007,102	40.0	30.8
<b>S1608</b>	DRR108020	53,338,306	38.9	30.7
<b>S1610</b>	DRR108021	48,461,084	38.0	31.8
<b>S1611</b>	DRR108022	58,739,822	46.1	25.0
<b>S1612</b>	DRR108023	57,109,894	46.0	25.3

Appendix Table S3.10. Differentially expressed genes

Feature ID	Description	Comparison #	FDR p-value correction	Mean RPKM	
				Group 1	Group 2
comp220555_c0_seq1	Cytochrome b6-f complex subunit 4	4	0.00	993.69	1,482.10
comp240712_c1_seq1	(Not annotated)	5	1E-02	65.96	113.28

Appendix Table S3. 11. The expression of genes encoding the cytochrome b6-f complex

Gene	Clade B Gene ID	Clade C transcriptome ID*	FDR p-value	group 1 - Means (RPKM)	group 2 - Means (RPKM)
petA	symbB.v1.2.018517	comp210046_c0_seq1	1.00	0.00	0.00
	symbB.v1.2.018518	comp210046_c0_seq2	1.00	0.00	0.00
	symbB.v1.2.018520	comp248005_c0_seq6	1.00	0.00	0.00
	symbB.v1.2.020193	comp256863_c0_seq1	1.00	26.50	26.57
	symbB.v1.2.020196	comp256863_c0_seq2	1.00	85.39	86.14
		comp256863_c0_seq3	1.00	28.41	28.63
		comp256863_c0_seq4	1.00	0.00	0.00
		comp256863_c0_seq5	1.00	2.02	1.90
		comp256863_c0_seq6	1.00	23.73	23.06
		comp256863_c0_seq7	1.00	0.00	0.00
		comp256863_c0_seq8	1.00	49.59	47.83
		comp256863_c0_seq9	1.00	0.53	0.00
		comp256863_c0_seq10	1.00	34.30	32.06
		comp256863_c0_seq11	1.00	0.00	0.00
		comp256863_c0_seq12	1.00	0.00	0.00
		comp256863_c0_seq13	1.00	24.46	25.83
		comp432697_c0_seq1	1.00	0.00	0.00
petB	JX094309-10	No hits found	-	-	-
petC	symbB.v1.2.006097	comp107053_c0_seq1	1.00	0.00	0.00
	symbB.v1.2.006095	comp115853_c0_seq1	1.00	27.48	26.14
	symbB.v1.2.006093	comp158216_c0_seq1	1.00	6.48	7.55
	symbB.v1.2.006094	comp158216_c0_seq2	1.00	6.64	4.85
		comp158710_c0_seq1	1.00	20.47	21.11
		comp163711_c0_seq1	1.00	43.41	41.87
		comp169889_c0_seq1	1.00	0.00	0.00
		comp169889_c0_seq2	1.00	11.65	13.51
		comp171588_c0_seq1	1.00	9.12	9.30
		comp181060_c0_seq1	1.00	9.29	7.60
		comp181613_c0_seq1	1.00	13.29	13.99
		comp181992_c0_seq1	1.00	0.00	0.00
		comp190291_c0_seq1	1.00	18.82	18.38
		comp190379_c0_seq1	1.00	12.13	10.66
		comp206699_c0_seq1	1.00	26.73	25.41
		comp206699_c0_seq2	1.00	0.19	0.59
		comp208792_c1_seq1	1.00	28.71	26.80
		comp210646_c0_seq1	1.00	3.89	4.19
		comp220141_c0_seq1	1.00	0.23	0.00
		comp220141_c0_seq2	1.00	17.31	17.24
		comp220141_c0_seq3	1.00	26.42	26.33
		comp232436_c0_seq2	1.00	0.00	0.00
		comp232436_c0_seq6	1.00	13.10	12.63
		comp232436_c0_seq10	1.00	0.00	0.00
		comp232436_c0_seq11	1.00	0.01	0.01
		comp232436_c0_seq20	1.00	1.88	1.49
		comp238965_c0_seq1	1.00	0.00	0.00
		comp238965_c0_seq2	1.00	13.80	16.38

Appendix Table S3. 11. The expression of genes encoding the cytochrome b6-f complex  
(continued)

Gene	Clade B Gene ID	Clade C transcriptome ID*	FDR p-value	group 1 - Means (RPKM)	group 2 - Means (RPKM)
		comp238965_c0_seq3	1.00	0.14	0.16
		comp238965_c0_seq4	1.00	5.66	6.38
		comp238965_c0_seq5	1.00	0.00	0.00
		comp238965_c0_seq6	1.00	7.38	7.36
		comp238965_c0_seq7	1.00	0.00	0.00
		comp238965_c0_seq8	1.00	0.00	0.00
		comp238965_c0_seq9	1.00	11.07	10.80
		comp238965_c0_seq10	1.00	0.16	0.08
		comp238965_c0_seq11	1.00	18.63	18.04
		comp238965_c0_seq12	1.00	0.00	0.00
		comp238965_c0_seq13	1.00	23.29	21.75
		comp238965_c0_seq14	1.00	25.50	24.90
		comp238965_c0_seq15	1.00	0.00	0.00
		comp238965_c0_seq16	1.00	0.00	0.10
		comp238965_c0_seq17	1.00	0.00	0.00
		comp238965_c0_seq18	1.00	0.00	0.00
		comp238965_c0_seq19	1.00	0.00	0.00
		comp238965_c0_seq20	1.00	10.92	11.24
		comp238965_c0_seq21	1.00	3.71	3.79
		comp238965_c0_seq22	1.00	8.20	10.54
		comp238965_c0_seq23	1.00	40.55	35.88
		comp238965_c0_seq24	1.00	17.78	16.25
		comp238965_c0_seq25	1.00	4.80	7.99
		comp238965_c0_seq26	1.00	9.24	6.48
		comp264715_c0_seq1	1.00	44.92	45.06
		comp264718_c0_seq1	1.00	54.35	51.84
		comp265278_c0_seq1	1.00	37.25	29.39
		comp265335_c0_seq1	1.00	14.31	14.92
		comp294429_c0_seq1	1.00	0.00	0.00
		comp295943_c0_seq1	1.00	0.00	0.00
		comp296904_c0_seq1	1.00	0.00	0.00
		comp319557_c0_seq1	1.00	0.00	0.00
		comp324745_c0_seq1	1.00	0.00	0.00
		comp343540_c0_seq1	1.00	0.00	0.00
		comp525760_c0_seq1	1.00	0.00	0.00
		comp57399_c0_seq1	1.00	53.57	54.18
		comp94116_c0_seq2	1.00	0.00	0.00
<b>petD</b>	<b>JX094311-14</b>	<b>comp220555_c0_seq1</b>	<b>0.00</b>	<b>993.69</b>	<b>1482.10</b>
petG	symbB.v1.2.003308	comp102784_c0_seq1	1.00	0.00	0.00
	symbB.v1.2.000994	comp113162_c0_seq2	1.00	0.00	0.00
	symbB.v1.2.009131	comp117799_c0_seq1	1.00	9.39	7.07
		comp184508_c0_seq2	1.00	4.00	4.26
		comp184508_c0_seq3	1.00	0.17	0.00
		comp38200_c0_seq2	1.00	0.00	0.00
		comp71070_c0_seq1	1.00	0.00	0.00
petL	-	-	-	-	-
petM	-	-	-	-	-

Appendix Table S3.11. The expression of genes encoding the cytochrome b6-f complex (continued)

Gene	Clade B Gene ID	Clade C transcriptome ID*	FDR p-value	group 1 - Means (RPKM)	group 2 - Means (RPKM)
petN	symbB.v1.2.029346 symbB.v1.2.029345	comp42866_c0_seq1	1.00	65.57	67.95
petO	-	-	-	-	-
ccsA	symbB.v1.2.004806	comp110324_c0_seq1	1.00	0.00	0.00
		comp110324_c0_seq2	1.00	0.00	0.00
		comp281714_c0_seq1	1.00	0.00	0.00
		comp382187_c0_seq1	1.00	0.00	0.00
		comp44007_c0_seq1	1.00	45.91	43.97
ccs1	symbB.v1.2.008118	comp236532_c0_seq1	1.00	41.62	43.90
		comp288527_c0_seq1	1.00	0.00	0.00

\* Appendix Table S3.11

I searched genes that encode the subunits of cytochrome b6-f complex in the assembled transcriptome of *Symbiodinium goreau* (Davies, et al. 2018) by blast search (Altschul, et al. 1990). In *S. minutum*, eight genes encoding the subunits of cytochrome b6-f complex had been reported (Mungpakdee, et al. 2014), and these genes were used as queries of tblastn search (Altschul, et al. 1990) against transcriptome of *S. goreau*. The cut-off e-value of this search was  $1e^{-10}$ .

Reference

1. **Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ 1990.** Basic local alignment search tool. J Mol Biol. 215: 403-410.
2. **Davies SW, Ries JB, Marchetti A, Castillo KD 2018.** *Symbiodinium* Functional Diversity in the Coral *Siderastrea siderea* Is Influenced by Thermal Stress and Reef Environment, but Not Ocean Acidification. Frontiers in Marine Science 5. doi: 10.3389/fmars.2018.00150
3. **Mungpakdee S, et al. 2014.** Massive gene transfer and extensive RNA editing of a symbiotic dinoflagellate plastid genome. Genome Biol Evol 6: 1408-1422. doi: 10.1093/gbe/evu109

## Appendix text

## &gt;Acropora\_CFP

ATGTCTTATTCAAAGCAAGGCATCGTACAAGCAATGAAGACGAAATACCATATGGA  
AGGCAGTGTCAATGGCCATGAATTCACGATCGAAGGTGTAGGAACTGGAAACCCTT  
ACGAAGGCACACAGATGTCCGAATTAGTGATCACCAAGCCTGCAGGAAAACCCCTT  
CCATTCTCCTTTGACATTCTGTCAACAGTCTTTCAATATGGAAACAGGTGCTTCACA  
AAGTACCCTGAAGGAATGACTGACTATTTCAAGCAAGCATTCCCAGATGGAATGTC  
ATATGAAAGGTCATTTCTATTTGAGGATGGAGGAGTTGCTACAGCCAGCTGGAACA  
TTCGTCTCGAAGGAGATTGCTTCATCCACAAATCCATCTATCATGGCGTTAACTTTC  
CCGCTGATGGACCCGTAATGAAAAAGAAGACAATTGGCTGGGATAAGTCCTTCGAA  
AAAATGACTGTGTCCAAAGACGTGTTAAGAGGTGATGTGACTATGTTTCTTATGCTC  
GAAGGAGGTGGTTACCACAGATGCCAGTTTCACTCCACTTACAAACCAGAGAAGCC  
GGTTAACTGCCCCCGAATCATGTCGTGGAACATCACATTGTGAGGACTGACCTTG  
GCCAAAGTGCAAAAGGCTTCACAGTCAAGCTGGAAGAACATGCTGCGGCTCATGTT  
AACCCCTTTGAAGGTTAAATAA

## &gt;Acropora\_GFP

ATGTCTTATTCAAAGCAAGGCATCGTACAAGAAATGAAGACGAAATACCATATGGA  
AGGCAGTGTCAATGGCCATGAATTCACGATCGAAGGTGTAGGAACTGGAAACCCTT  
ACGAAGGGACACAGATGTCCGAATTAGTGATCATCAAGCCTGCGGGAAAACCCCTT  
CCATTCTCCTTTGACATACTGTCATCAGTCTTTCAATATGGAAACAGGTGCTTCACA  
AAGTACCCTGCAGACATGCCTGACTATTTCAAGCAAGCATTCCCAGATGGAATGTC  
ATATGAAAGGTCATTTCTATTTGAGGATGGAGGAGTTGCTACAGCCAGCTGGAACA  
TTCGTCTCGAAGGAAATTGCTTCATCCACAAATCCATCTATCATGGCGTAACTTTC  
CCGCTGATGGACCCGTAATGAAAAAGAAGACAATTGGCTGGGATAAGTCCTTCGAA  
AAAATGACTGTGTCCAAAGAGGTGTTAAGAGGTGATGTGACTATGTTTCTTATGCT  
CGAAGGAGGTGGTTACCACAGATGCCAGTTTCACTCCACTTACAAAACAGAGAAGC  
CGGTTAACTGCCCCCGAATCATGTCGTAGAACATCACATTGTGAGGACTGACCTT  
GGCCAAAGTGCAAAAGGCTTCACAGTCAAGCTGGAAGAACATGCTGCGGCTCATGT  
TAACCCTTTGAAGGTTAAATAA

## &gt;Acropora\_RFP

ATGGCTCTGTCAAAGCACGGTCTAACAAAGGACATGACGATGAAATACCGGATGG

AAGGGTGTGTCGATGGGCATAAATTTGTGATCACGGGCCACGGCAATGGAAATCCT  
TTCGAAGGGGAAACAGACTATCAATCTGTGTGTGGTTGAAGGGGGACCCCTGCCATT  
CTCCGAAGACATTTTGTCTGCTGTGTTTGACTACGGAAACAGGGTCTTCACTGAATA  
TCCTCAAGGCATGGTTGACTTTTTCAAGAATTCATGTCCAGCTGGATACACATGGCA  
AAGGTCTTTACTCTTTGAAGATGGAGCAGTTTGCACAGCCAGTGCAGATATAACAG  
TGAGTGTTGAGGAGAACTGCTTTTATCACGAGTCCAAGTTTCATGGAGTGAACTTTC  
CTGCTGATGGACCTGTGATGAAAAAGATGACAATAATTGGGAGCCATCCTGCGAG  
AAAATCATACCAGTACCTAGACAGGGGATATTGAAAGGGGATGTCGCCATGTACCT  
CCTTCTGAAGGATGGTGGGCGTTACCGGTGCCAGTTCGACACAATTTACAAAGCAA  
AGACTGACCCGAAAAAGATGCCGGAGTGGCACTTCATCCAACATAAGCTCACCCGG  
GAAGACCGCAGCGATGCTAAGAACCAGAAATGGCAACTGGCAGAACATGCTGTTG  
CTTCCCGATCCGCATTGCCCCGATAA

>Acropora\_CP

ATGAGTGTGATCGCTAAACAAATGACCTACAAGGTTTATATGTCAGGCACGGTCAA  
TGGACACTACTTTGAGGTCTGAAGGCGATGGAAAAGGAAAGCCTTACGAGGGGGAG  
CAGACGGTAAAGCTCACTGTCCACCAAGGGCGGACCTCTGCCATTTGCTTGGGATAT  
TTTATCACCACAGTTTCAGTACGGAAGCATAACCATTCACCAAGTACCCTGAAGACA  
TCCCTGACTATGTAAAGCAGTCATTCCCAGAGGGATATACATGGGAGAGGATCATG  
AACTTTGAAGATGGTGCAGTGTGTACTGTCAGCAATGATTCCAGCATCCAAGGCAA  
CTGTTTCATCTACCATGTCAAGTTCTCTGGTGTGAACTTTCCTCCCAATGGACCTGTT  
ATGCAGAAGAAGACACAGGGCTGGGAACCCAACACTGAGCGTCTCTTTGCACGAG  
ATGGAATGCTGATAGGAAACAACCTTTATGGCTCTGAAGTTGGAAGGAGGTGGTCAC  
TATTTGTGTGAATTCAAATCTACTTACAAGGCAAAGAAGCCTGTGAGGATGCCAGG  
GTATCACTATGTTGACCGCAAACCTGGATGTAACCAATCACAACAAGGATTACACTT  
CCGTTGAGCAGTGTGAAATTTCCATTGCACGCAAACCTCTGGTCGCCTGA

# **Chapter 4**

## **Presence–absence polymorphisms of single-copy genes**

**Abstract**

Despite the importance of characterizing genetic variation among coral individuals for the conservation of coral reefs, the correlation between genetic and functional variation is still poorly understood. In this chapter, I detected a high frequency of genes showing presence–absence polymorphisms (PAPs) for single-copy genes in *Acropora digitifera*. Among single-copy genes, some proportion exhibited PAPs, including transposable element (TE)-related genes. Among genes showing PAPs (PAP genes), roughly half were expressed in adults and/or larvae, and the PAP status was associated with differential expression among individuals. Although most of PAP genes were uncharacterized or had ambiguous annotations, these genes were specifically distributed in cnidarian lineages, suggesting that these genes have functional roles related to traits specific to cnidarians or the family Acroporidae. Some proportion of *A. digitifera* PAP genes were also PAPs in *A. tenuis*, the basal lineage in the genus *Acropora*, indicating that PAPs were shared among species in *Acropora*. Expression differences caused by a high frequency of PAP genes may be a novel genomic feature in the genus *Acropora*.

Details of this chapter (pp. 163-217) is omitted in this version.

# **Chapter 5**

## **Perspectives**

The stony corals are the basis of coral reefs that harbor the most biologically diverse ecosystems in warm shallow water. The fluorescence of corals and its genetic determinants, fluorescent protein (*FP*) genes have been one of the enigmatic problems in corals. Many biological roles of *FPS* had been proposed, but are still inconclusive. One of the main factors making the difficulty to analyze the role of *FPS* was missing information of the genetic basis of *FPS* in stony corals.

In my doctoral thesis, I documented the genetic basis of fluorescence in one of the stony corals, *Acropora digitifera*. In the genome of this species, *FP* genes formed a multi-gene family composed of each of three monophyletic clades with an unexpectedly large number of genes (chapter 2). In this multi-gene family, the copy number variation and differences in sequence repertoires of *FP* genes among individuals were observed. Especially, highly expressed *FP* sequences were presence-absence polymorphic status (chapter 3). The multi-*FP* gene family with presence-absence polymorphisms is the genetic basis of fluorescence and fluorescent polymorphism in *A. digitifera*.

During the study of fluorescence in *A. digitifera*, I found a novel feature of the genome of *A. digitifera*. Presence-absence polymorphisms (PAPs) were frequent in single-copy genes of *A. digitifera*. Half of the genes showing PAP were expressed and contributed to differences in gene expression among individuals of *A. digitifera*. Over half of genes showing PAP were specifically distributed in cnidarian lineages, suggesting that they have functional roles associated with traits specific to cnidarians or the family Acroporidae (chapter 4).

My doctoral thesis raises new questions to be solved. As a next step, I focus on two main questions. I isolated over 180 different *FP* sequences from *A. digitifera* individuals, though only several *FP* sequences were highly expressed in adults. Therefore, first question is

when and where does each *FP* gene in the multi-gene family express? Detail expression analyses using individuals from different developmental stages, seasons, environments will answer this question. Observations of temporal and spatial expression change of each *FP* gene will help to understand the biological roles of FPs.

Second question is the functions and biological roles of genes showing PAP. Although the majority of PAP genes were uncharacterized or had ambiguous annotations, over half of these genes were distributed in other cnidarians. These facts raise the following question; what are the functions of genes showing PAP in corals? Functional analyses of genes showing PAP using model animals in the phylum Cnidaria will answer this question. For example, the genome and transcriptomes of a starlet sea anemone *Nematostella vectensis* are available, and many techniques such as gene suppression, genome editing, mRNA misexpression, and inducible promoters have been developed for this animal. Understanding of the functions of genes showing PAP will provide insight into genetic variation among individuals in corals.

# Acknowledgements

I deeply appreciate Dr. Masayuki Hatta (Ochanomizu University, Japan) for introducing me to the world of coral reefs and his help with sampling. I am thankful to Dr. Kazuhiko Sakai (University of the Ryukyus, Japan) for supporting species identification and measurements of fluorescence in the field. I am thankful to Dr. Mutsumi Nishida (University of the Ryukyus, Japan) for arranging the fluorescence measurements from live corals. I thank Drs. Yinqiang Zheng and Imari Sato (National Institute of Informatics, Japan) for helpful discussions of fluorescence measurements.

I greatly appreciate Dr. Yoko Satta (SOKENDAI, Japan) for her support throughout my doctoral program. I also would like to show my appreciation to Dr. Yohey Terai (SOKENDAI, Japan). He encouraged me to study fluorescence of corals. I am thankful to Dr. Jun Gojobori (SOKENDAI, Japan), and the members of Satta laboratory and the Department of Evolutionary Studies of Biosystems, SOKENDAI for their helpful discussions. Finally, I would like to thank my family.

This work was supported by Grant-in-Aid for JSPS Fellows to S. T. K..