Fluorescent Protein Candidate Genes in the Coral Acropora digitifera Genome

Authors: Chuya Shinzato, Eiichi Shoguchi, Makiko Tanaka, and Nori Satoh
Source: Zoological Science, 29(4) : 260-264
Published By: Zoological Society of Japan
URL: https://doi.org/10.2108/zsj.29.260
Fluorescent Protein Candidate Genes in the Coral Acropora digitifera Genome

Chuya Shinzato*, Eiichi Shoguchi, Makiko Tanaka, and Nori Satoh

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

The vivid coloration of corals depends on fluorescent proteins that include cyan (CFP), green (GFP) and red (RFP) fluorescent proteins, and a non-fluorescent blue/purple chromoprotein. We examined how many genes encoding fluorescent proteins are present in the recently sequenced genome of the coral Acropora digitifera. Based on molecular phylogenetic analysis, we found one, five, one, and three candidate genes for CFP, GFP, RFP, and chromoprotein, respectively. The CFP and GFP genes are clustered in a ~80-kb-long genomic region, suggesting that they originated from an ancestral gene by tandem duplication. Since CFP and GFP possess the same chromophore, the gene clustering may provide the first genomic evidence for a common origin of the two proteins. Comparison between the fluorescent protein genes of closely related coral species suggests an expansion of chromoprotein genes in the A. digitifera genome, and of RFP genes in the A. millepora genome. The A. digitifera fluorescent protein genes are expressed during embryonic and larval developmental stages and in adults, suggesting that the genes play a variety of roles in coral physiology.

Key words: corals, Acropora digitifera, fluorescent protein genes, chromoprotein genes, gene clustering

INTRODUCTION

Corals exhibit a wide range of patterns of coloration (Dove et al., 2001; Mazel et al., 2003; Matz et al., 2006), which largely depends on fluorescent proteins (Matz et al., 1999). Four basic colors of fluorescent proteins are present in corals: three fluorescent proteins including cyan (CFP), green (GFP), and red (RFP), and a non-fluorescent blue/purple chromoprotein (Kelmanson and Matz, 2003; Field et al., 2006). Fluorescent proteins are of comparable size, usually ~230 amino acid residues. However, during evolution, corals acquired the ability to synthesize several distinct types of fluorescent or colored moiety—the chromophore—from the amino acid residues within fluorescent proteins, via two or three consecutive autocatalytic reactions. Individual chromophores can differ dramatically in spectroscopic characteristics (Lukyanov et al., 2006). Among the four basic colors, CFP and GFP possess the same chromophore (Henderson and Remington, 2005).

As with other natural pigments, the variation of fluorescent proteins within a coral colony suggests that these proteins play multiple specific roles in corals (Dove et al., 2001; Matz et al., 2002; Kelmanson and Matz, 2003; Salih et al., 2000). One prominent function is associated with the maintenance of obligate symbiosis with dianoflagellates (Kamaguchi, 1969; Salih et al., 1998). Fluorescent proteins are able to convert shorter wavelengths of light to longer wavelengths, which protects the coral from harmful wavelengths and enhances the available useful light for symbiotic brown algae. In addition, recent studies suggest other roles for fluorescent proteins in corals, serving as visual triggers for other organisms (Wachter, 2006) and as oxygen radical quenchers (Mazel et al., 2003; Bou-Abdallah et al., 2006).

We have sequenced the genome of the coral Acropora digitifera (Shinzato et al., 2011). The ~420 Mbp genome of this coral is estimated to contain 23,668 protein-coding genes. Since the function of coral fluorescent proteins within the holobiont remains undetermined and controversial (Aleeva et al., 2008), the annotation of A. digitifera genes provides the first catalogue of the coral fluorescent protein repertoire. Our genome-wide analysis has revealed that the coral genome contains ten candidate fluorescent protein genes.

MATERIALS AND METHODS

Gene searching

We used the two methods to annotate fluorescent protein genes. The first and most convenient method was a BLAST search using other anthozoan fluorescent protein genes against the A. digitifera gene models (BLASTP) or the assembly (TBLASTN). We used 24 fluorescent proteins from Acropora corals as the search queries (Supplementary Table S1). The second method involved the characterization of a GFP domain. To screen and identify the domain in the gene model, we used the Pfam database (Pfam-A. hmm, release 24.0; http://pfam.sanger.ac.uk) (Finn et al., 2010), which contains 11,912 conserved domains; protein entries matching the conserved domain were identified using HMMER searches (hmm) (Eddy, 1998). Genes exhibiting significant similarity (Blast E-value < 1e-10) with known Acropora fluorescent proteins and having
Fluorescent Protein in *Acropora digitifera*

a GFP domain (Pfam accession: PF01353) were analyzed further.

**Phylogenetic analysis**

Fluorescent protein candidates were aligned using ClustalW (Larkin et al., 2007) with default parameters. Gaps and ambiguous areas were excluded using Gblocks 0.91b (Castresana, 2000) either manually or using default parameters. The accuracies of the multiple alignments were also checked manually. Based on the alignment datasets (104 amino acid residues), phylogenetic trees were constructed by Neighbor-Joining (NJ) as in the study of Shinzato et al. (2011). The bootstrap analysis was replicated 1,000 times. The calculation and tree construction were performed by SeaView (Gouy et al., 2010).

**Transcriptome analysis**

For gene expression analysis, Illumina RNA-seq data for *A. digitifera* embryonic and larval developmental stages (a mixture of RNA samples from eggs, blastulae, gastrulae, swimming larvae, and metamorphosing larvae) and adults (Shinzato et al., 2011) were mapped to *A. digitifera* fluorescent protein genes using the Bowtie software (Langmead et al., 2009), and unique mapped reads were used for RPKM (reads per kilobase per million reads) calculations (Mortazavi et al., 2008). Due to the limitation of the sequencing capacity, RNA samples from embryonic and larval stages were mixed and sequenced (Shinzato et al., 2011).

**RESULTS AND DISCUSSION**

The *A. digitifera* genome contains ten candidate genes for fluorescent proteins

The fluorescent protein complement of *A. digitifera* was surveyed using a combination of BLAST and Pfam domain searching. We found ten genes for distinct fluorescent proteins in this coral genome, tentatively named *Adi-Fluorescent protein-1* to *Adi-Fluorescent protein-10* (Fig. 1, Table 1). Using both *Nematostella* and arthropod fluorescent proteins in BLAST searches did not detect novel fluorescent gene candidates in the genome. A tripeptide –X–Y–G–, where X is highly variable, forms the precursor to the chromophore of fluorescent proteins. The tripeptides were found in most *Adi-Fluorescent proteins*, except for *Adi-Fluorescent protein-6* and *-9* (Fig. 1). The lack of the tripeptide in these two genes is likely due to gene prediction errors in the *A. digitifera* genome project. Recently, Alieva et al. (2008) conducted a broad survey of the diversity and evolution of coral fluorescent proteins. That study indicated 40 novel proteins; these, along with the previously known fluorescent proteins, represented all six suborders of Scleractinia. We carried out molecular phylogenetic analysis of the ten fluorescent proteins from *A. digitifera*. As shown in Fig. 2, this analysis suggested that one protein (*Adi-Fluorescent protein-1*) was from the CFP clade, five (*Adi-Fluorescent protein-2* to *Adi-Fluorescent protein-6*) from the GFP clade, one (*Adi-Fluorescent protein-10*) from the RFP clade, and three (*Adi-Fluorescent protein-7* to *Adi-Fluorescent protein-9*) from the Chromoprotein clade. Due to the limitation of the genomic project, RNA samples from embryonic and larval stages were mixed and sequenced (Shinzato et al., 2011).

**Fig. 1.** Boxshade alignment of putative *Acropora digitifera* fluorescent proteins. Gene names are abbreviated as AdiFP. Chromophore-forming tripeptides are boxed. Note that AdiFP-6 and -9 lack the tripeptides, possibly due to gene prediction errors in the *A. digitifera* genome project (Shinzato et al., 2011). These sequences have been deposited with DDBJ/EMBL/GenBank under accession numbers BR000962–BR000970 and AB698751.

**Table 1.** The expression of fluorescent protein genes in *Acropora digitifera*. Presumed color types of each protein are shown. Numbers represent RPKM (read per kilobase per million reads) for each gene in embryonic and larval developmental stages and adulthood.

<table>
<thead>
<tr>
<th>Gene name</th>
<th>Presumed color type</th>
<th>RPKM at developmental stages</th>
<th>RPKM in adult</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Adi-Fluorescent protein-1</em></td>
<td>cyan</td>
<td>1.72</td>
<td>6.75</td>
</tr>
<tr>
<td><em>Adi-Fluorescent protein-2</em></td>
<td>green</td>
<td>0.37</td>
<td>2.31</td>
</tr>
<tr>
<td><em>Adi-Fluorescent protein-3</em></td>
<td>green</td>
<td>50.61</td>
<td>0.24</td>
</tr>
<tr>
<td><em>Adi-Fluorescent protein-4</em></td>
<td>green</td>
<td>1.50</td>
<td>0.18</td>
</tr>
<tr>
<td><em>Adi-Fluorescent protein-5</em></td>
<td>green</td>
<td>0.17</td>
<td>ND</td>
</tr>
<tr>
<td><em>Adi-Fluorescent protein-6</em></td>
<td>green</td>
<td>0.06</td>
<td>0.12</td>
</tr>
<tr>
<td><em>Adi-Fluorescent protein-7</em></td>
<td>chromoprotein</td>
<td>10.31</td>
<td>1.21</td>
</tr>
<tr>
<td><em>Adi-Fluorescent protein-8</em></td>
<td>chromoprotein</td>
<td>180.32</td>
<td>8.58</td>
</tr>
<tr>
<td><em>Adi-Fluorescent protein-9</em></td>
<td>chromoprotein</td>
<td>237.27</td>
<td>1.56</td>
</tr>
<tr>
<td><em>Adi-Fluorescent protein-10</em></td>
<td>red</td>
<td>641.55</td>
<td>0.14</td>
</tr>
</tbody>
</table>
Fluorescent protein-9) from the chromoprotein clade. Each of the four clades was supported by a nearly 100% bootstrap value (Fig. 2). Therefore, it is likely that Adi-Fluorescent protein-1 encodes CFP, Adi-Fluorescent protein-2 to Adi-Fluorescent protein-6 are GFPs, Adi-Fluorescent protein-10 are RFP, and Adi-Fluorescent protein-7 to Adi-Fluorescent protein-9 are chromoproteins (Table 1). Measuring of the fluorescent spectra of A. digitifera fluorescent proteins will be required to confirm the fluorescent colors of those genes.

Based on the non-redundant protein database (NCBI), another coral, Acropora millepora, contains at least 12 fluorescent protein genes (Fig. 2). As this study reveals the presence of ten fluorescent proteins genes in the A. digitifera genome, we conclude that both Acropora species have a comparable numbers of fluorescent protein genes.

### Expression of fluorescent protein genes

In the A. digitifera genome project, we carried out expressed sequence tag analyses of genes that are expressed during embryonic and larval developmental stages and adulthood, using the Illumina GAIIx sequencer (Shinzato et al., 2011). We exami...
ined mRNA levels of the 10 putative fluorescent protein genes in order to determine whether the corresponding mRNAs are actually expressed in the coral. This analysis showed that all 10 genes are expressed during embryonic and larval developmental stages or adulthood (Table 1); the expression of only one gene, Adi-Fluorescent protein-5, was not detected in adults (Table 1). The expression level differed from gene to gene and at the two different ages. For example, the expression levels of Adi-Fluorescent protein-3 (GFP), -8 (chromoprotein), -9 (chromoprotein) and -10 (RFP) were much higher during developmental stages than adulthood (Table 1), suggesting that these genes play specific functions in the coral embryogenesis. Further studies should be carried out to explore the gene expression profiles in more detail, both temporally and spatially, in association with their functions.

A putative CFP gene and five GFP genes are clustered in the genome

The assembly of the A. digitifera genome sequence reached to contig N50 = 10.7 kbp (the longest contig was 98.2 kbp) and scaffold N50 = 191.5 kbp (six scaffolds exceeded 1 Mbp) (Shinzato et al., 2011). The assembly statistics are at least equal to those of other basal animal genome assemblies (e.g., Srivastava et al., 2010 for a sponge genome). We were therefore able to examine the localization of the fluorescent protein genes in the A. digitifera genome. We found that six genes are tandem clustered in a ~80-kbp-long region of Scaffold-84, which expanded to 463-kbp long (Fig. 3A): Adi-Fluorescent protein-1, Adi-Fluorescent protein-4, Adi-Fluorescent protein-2, Adi-Fluorescent protein-3, Adi-Fluorescent protein-5, and Adi-Fluorescent protein-6, in that order. The former four genes are oriented in the same reading direction, while the latter two are in the opposite direction (Fig. 3A).

As described above, it is presumed that Adi-Fluorescent protein-1 encodes a CFP and Adi-Fluorescent protein-2 to Adi-Fluorescent protein-6 encode GFPs. Of the four basic colors, CFP and GFP possess the same chromophore (Henderson and Remington, 2005), suggesting a close evolutionary relationship between them. Since the six fluorescent protein genes are arranged in tandem in the coral genome, it is highly likely that these six genes originated from an ancestral gene by tandem duplication. Intron positions in ORFs of Adi-Fluorescent protein gene 1-6 are well conserved (Supplementary Fig. S1). In contrast to the case of gene conversion between green and red proteins of Montipora efflorescens (Alieva et al., 2008), no obvious gene conversion is observed in A. digitifera fluorescent proteins (Fig. 1, Supplementary Fig. S1).

Inspection of Figs. 2 and 3 suggests the following evolutionary scenario for these six genes. Adi-Fluorescent protein-6 occupies the most basal phylogenetic position within the clustered genes, which suggests that it may be most closely related to the ancestral gene (Fig. 2). This ancestral gene might have been tandem duplicated to form its sister Adi-Fluorescent protein-5, due to its shared orientation with Adi-Fluorescent protein-6. Next, since Adi-Fluorescent protein-3 and -4 are phylogenetically close to Adi-Fluorescent protein-6 and -5 (Fig. 2), Adi-Fluorescent protein-6 and -5 might have been duplicated simultaneously to form Adi-Fluorescent protein-3 and -4. On this occasion, the orientations may have been reversed (Fig. 3A). Finally, tandem duplication of Adi-Fluorescent protein-3 and -4 occurred independently to form Adi-Fluorescent protein-2 and Adi-Fluorescent protein-1, respectively. Adi-Fluorescent protein-2 retained GFP character, while Adi-Fluorescent protein-1 evolved to encode a protein with CFP character. Although this remains purely a speculative scenario, the data taken together provide the first genomic evidence that CFP and GFP genes can originate by duplication of a common ancestral gene.

Expansion of chromoprotein genes in the A. digitifera genome

Three A. digitifera genes, Adi-Fluorescent protein-7, -8, and -9 formed a well-supported clade of chromoproteins along with those of Goniopora djiboutiensis and A. millepora (Fig. 2). In addition, Adi-Fluorescent protein-7 and -9 are arranged in tandem in Scaffold-947 with the same orientation (Fig. 3B). This indicates that A. digitifera has a more complex chromoprotein repertoire, due to expansion of this family, than these two corals. Adi-Fluorescent protein-8, which is located on Scaffold-528 (data not shown), may be located on the same chromosome as -7 and -9, although we have not found any clues to this in the genome assembly. As described above, the chromoproteins are characterized by higher absorption and lower emission properties (Mazel et al., 2003; Bou-Abdallah et al., 2006) and may have higher 

**ACKNOWLEDGMENTS**

This study was supported in part by KAKENHI (21121505,
REFERENCES


(Received November 22, 2011 / Accepted December 2, 2011)