Genetic Diversity of Fluorescent Proteins in Caribbean Agariciid Corals

Meredith Meyers, James W. Porter, and John P. Wares

From the Odum School of Ecology, University of Georgia, Athens, GA 30602 (Meyers, Porter, and Wares) and the Department of Genetics, University of Georgia, Athens, GA 30602 (Wares).

Address correspondence to Meredith Meyers at the address above, or e-mail: mkmeyers@uga.edu.

The fluorescent protein (FP) gene family is a highly diverse group of proteins whose expression govern color diversity in corals. Here, we examine the genetic diversity of FPs and the extent to which it can be used to assess phylogenetic relationships within the coral genus Agaricia. Tissue samples were collected throughout the Florida Keys from a wide range of phenotypes within the genus Agaricia (A. agaricites [n = 7], A. fragilis [n = 13], and A. lamarcki [n = 2]), as well as the confluential species Heliothrix cucullata (n = 3). Primers were developed from published cDNA sequences to amplify a region of coding and noncoding sequences of FPs. Cloning reactions were performed to capture the multiple copies of FPs and allele diversity. In the resulting 116 cloned sequences, we identified a 179-bp coding region for phylogenetic analysis. Three distinct clades were found in all 3 species of Agaricia, potentially representing 3 copies of the FP gene. Of the 3 gene copies, 2 contain distinct subclades that display reciprocal monophyly between A. agaricites and A. fragilis, whereas A. lamarcki is polyphyletic. Further resolution of the species phylogeny is necessary to fully understand how genetic diversity within this gene family is distributed among taxa and habitats.

Key words: Agaricia, Agaricidae, fluorescent proteins, gene duplication, nucleotide diversity, scleractinia

Introduction

In corals, the fluorescent protein (FP)-like family is a diverse group of proteins that provide much of the color diversity that is observed on coral reefs (Matz et al. 1999; Lukyanov et al. 2000; Dove et al. 2001). They are unique in that they synthesize a chromophore in 2 or 3 autcatalytic reactions; therefore, variation in protein structure and color is determined by sequence variation in a single expressed gene (Matz et al. 1999). Four classes of FP colors are known for corals (fluorescent green, cyan, red, and nonfluorescent purple-blue) (Labas et al. 2002; Shagin et al. 2004; Alieva et al. 2008). The functional role of the color diversity and the FPs themselves are still not fully understood although they likely play multiple roles including photoprotection for endosymbiotic dinoflagellates, photosynthetic light harvesting, and visual signalling to other reef organisms (Salih et al. 2000; Matz et al. 2006).

Studies examining FP sequence variation within a single coral species, Montastraea cavernosa, determined that expression occurs from at least 4–7 distinct loci within the genome (Kelmanson and Matz 2003) with nucleotide divergence among transcripts ranging from 2% to 17%, not counting indels. Furthermore, phenotypic diversity in M. cavernosa is thought to be determined by differential expression of color-coding genes, rather than sequence divergence, as evidenced by distinct color morphs harboring the same suite of FP genes (Kelmanson and Matz 2003). Further insight into the extent of genetic diversity of FPs in corals is clearly warranted in order to better understand how this component of the photic environment interacts with the endosymbiotic dinoflagellates and microbes that comprise the complex coral “holobiome.”

FP sequences in corals are highly diverse (Matz et al. 1999) and potentially useful for studying variation among and within species, systematic scales that are still difficult to resolve with currently available “universal” genetic markers in corals (Shearer et al. 2002; Vollmer and Palumbi 2004; Hellberg 2006; Shearer and Coffroth 2008; Forsman et al. 2009). Often, researchers need to focus their efforts on development of taxon-specific single locus nuclear markers that contain sufficient variation for species delimitation (Concepcion et al. 2008). As such, it is of use to identify additional, variable nuclear markers that are applicable broadly across the diversity of corals, and FP sequence data may be appropriate. Here, we examine the extent of FP genetic diversity contained within and among species in the phenotypically diverse scleractinian (hard coral) genus Agaricia (lettuce corals) in order to 1) quantify the level and type of diversity found in the Agaricidae, for comparison with the diversity observed in M. cavernosa and 2) investigate whether, given the sequence diversity found in the FP gene family, data from these gene regions are applicable for phylogenetic analysis among species and subspecies. Coral species delineation can be quite difficult given the low levels of diversity found at many of the loci commonly used for this purpose (Shearer et al. 2002; Forsman et al. 2009), so we evaluate whether distinct taxa (particularly A. agaricites and A. fragilis) could be reliably distinguished using these data.
Materials and Methods
Sample Collection and DNA Extraction
Tissue samples of *Agaricia* species (*A. agaricites*, *A. fragilis*, and *A. lamarcki*) and *H. culellata* representing a range of phenotypes were collected from 9 locations in the Florida Keys reef system. Using SCUBA, we took 1–2-cm³ clippings from adult colonies and placed them in ziploc bags. At the surface, tissue clippings were transferred to 15-mL centrifuge tubes with 95% ETOH and stored at −20 °C until extraction. DNA isolations were performed using a modified-cetyltrimethyl ammonium bromide (CTAB) extraction protocol as described in Concepcion et al. (2006). Resulting isolations were then further purified using the Strata-Prep PCR clean-up kit (StrataGene).

Polymerase Chain Reaction and DNA Sequencing
We used published cDNA sequences from *A. agaricites* and *A. fragilis* (GenBank Accession numbers AY037775 and AY647156, respectively (Alieva et al. 2008)) to create primers using Primer3. The resulting primers, F: 5’CTCTGGCGTTCTCTTTGAT and R: 5’CTGCATGACTGGACCATTAG, were used in subsequent amplifications. Polymerase chain reactions (PCRs) were done on 25-µL reaction volumes following PCR protocols of Wares et al. (2009) using 2 units AccuPrime Taq high-fidelity polymerase (Invitrogen) and annealing temperature of 50 °C. All PCR products were cloned using the TOPO-TA cloning kit for sequencing (Invitrogen) according to manufacturers’ instructions. Following the cloning reaction, 6–10 clones per sample were picked and reamplified with M13 primers and 50 °C annealing temperatures. Those PCR products were cleaned up for sequencing using Exonuclease I and Antarctic Phosphatase; cycle sequencing followed previous protocols (Wares et al. 2009). Reaction products were precipitated with isopropanol and resuspended in Hi-Di formamide and read on an ABI 3730XL at the Georgia Genomics Facility. Sequence data were trimmed and aligned using Codon-Code Aligner v.3.5.7. Our sequences included long intron regions that were difficult to unambiguously align across the entire data set and were therefore excluded from analysis. All nucleotides with PHRED scores less than 20 were considered ambiguous. Amino acid sequences were generated using the universal translation in Aligner and compared with reported protein sequences (Alieva et al. 2008).

Phylogenetic Analyses
Maximum parsimony analyses were completed in PAUP* 4.0b10 (Swofford 2003). A heuristic search was applied to the complete nucleotide dataset, and a simple addition sequence was used on starting trees obtained by stepwise addition. Tree-bisection-reconnection was used for branch swapping. Bootstrap scores were calculated with 1000 replicates. Bayesian phylogenetic reconstructions were performed using MrBayes with model parameters as determined in JModeltest. Two independent runs were completed with 4 heated Markov chain Monte Carlo (MCMC) chains, $5 \times 10^6$ generations and a sample frequency of $f = 1000$.

We further analyzed relationships of the gene family using the NeighborNet network analysis in SplitsTree (Huson and Bryant 2006). This analysis provides an alternative visualization of data (to traditional tree-based models) where gene duplication, recombination, or other mechanisms are thought to generate diversity within the system. Overall, our phylogenetic goal was to identify the number of distinct, well-supported clades (likely paralogs) that contained distinct sequences for both focal *Agaricia* species (likely orthologs).

Using DNAsp v.5.10.1 (Librado and Rozas 2009), diversity measures for the coding region were calculated, including the ratio of nucleotide diversity πA/πS between nonsynonymous (A) and synonymous (S) sites, the ratio of replacement substitutions Ka/Ks at nonsynonymous and synonymous sites for comparisons of phylogenetically distinct clades, McDonald–Kreitman tests to identify potential excesses of replacement substitutions, and Tajima’s D within clades to see if patterns of polymorphism deviate from the neutral expectation. Sites with ambiguities were removed from analysis. All of these measures are used to identify the relative pattern of substitution among synonymous and nonsynonymous positions in the coding sequence; each is a weak test for selection (Wayne and Simonsen 1998), but all are used to identify consensus or complementary patterns within and among clades.

Results
After excluding sequences of insufficient length or quality, our sequencing efforts returned 116 haplotypes from 25 individuals (7 *A. agaricites*, 13 *A. fragilis*, 2 *A. lamarcki*, and 3 *H. culellata*; see Supplementary Material online). The number of clones sequenced for *A. agaricites* ($n = 56$) and *A. fragilis* ($n = 45$) is sufficient, given binomial distribution assumptions for recovering gene copies in each species, to ensure data from up to 4 gene copies could be identified with a probability of 99%. We also included 5 cloned sequences recovered from *A. lamarcki* and the 10 unique clones from *H. culellata* in all analyses (see Supplementary Table online). A 179-bp exon region was identified based on alignment with reported cDNA sequences from *A. agaricites* and *A. fragilis* (GenBank Accession numbers AY037775, AY647156, and AY037765) (Alieva et al. 2008). Large regions of poorly aligned intron sequences were trimmed for the purposes of this study.

Within the analyzed 179-bp segment, 102 sites were variable, with 79 sites being parsimony informative. Using jModeltest, we determined that the best-fit model of molecular evolution was K80 (Ti/Tv = 1.504). In Bayesian analysis (Nst = 2, rates = equal), the duplicate runs achieved stationarity and converged after $5 \times 10^6$ generations at an average standard deviation of split frequencies of 0.0049 (Potential Scale Reduction Factor [PSRF] for all parameters = 1.0). The topology of parsimony and Bayesian phylogenies was essentially the same, and support indices (bootstrap and posteriors) are indicated on a single tree (Figure 1a).
The gene family has undergone many apparent duplication events, with several individuals having as many as 9 distinct alleles separated by at least 1 nucleotide polymorphism (8 alleles when only parsimony informative characters are considered), and 7 unique amino acid sequences. Parsimony analysis sorted alleles into 3 distinct clades (presumed paralogs, as each contains both of the focal species) within Agaricia, with Helioseris appearing to be reciprocally monophyletic with Agaricia (clades denoted by Roman numerals, Figure 1a). Additionally, 2–3 subclades within each major clade identify groups of taxa within Agaricia. Tree rooting is uncertain because exact timing of divergences between taxa and divergences between gene copies is unclear with the data given here.

Amino acid variation among clades is substantial (Figure 1b), but some allelic groups are likely non-functional (see stop codon in subclade Ia). Published cDNA sequences of FPs (Alieva et al. 2008) align within subclades IIa, IIb, and IIIa indicating potential functionality of those clades. Network analysis of our sequence data generated a result with a very similar overall arrangement of clades (see Supplementary Material online). We were particularly interested in characterizing the uncertainty associated with clade IIC, given its poor statistical support because the 3 alleles represented came from a unique morph of A. fragilis.

In comparisons among clades, we found elevated Ka/Ks ratios between subclades Ia (which only had haplotypes from A. agaricites) and IIC (with the unique A. fragilis morph) (Table 1; Ka/Ks = 7.62), and between clades I and II (Ka/Ks = 2.33). Additionally, values of Tajima’s D tended to be strongly (and in half of the subclades, significantly) negative, consistent with directional or purifying selection (Table 1).

In general, this shift in the polymorphism frequency was consistent across nonsynonymous and synonymous mutations although in subclades IIIa and IIIb some interesting dynamics were noted: in subclade IIIb, Tajima’s D was more negative at nonsynonymous substitutions than synonymous; and in subclade IIIa, Tajima’s D actually had values of similar magnitude but opposite sign for nonsynonymous and synonymous sites. No McDonald–Kreitman comparisons of clades, as numbered in Table 1, were statistically significant.

**Discussion**

Previous studies identified frequent duplication of green fluorescent protein (GFP)-like gene regions, often among recent lineages on the coral phylogenetic tree, leading to functionally diverse proteins (Labas et al. 2002; Kelmanson and Matz 2003). Gene duplication seems to be frequent across the coral genome (Kenyon 1997; Hislop et al. 2005)

![Figure 1](https://academic.oup.com/jhered/article-abstract/104/4/572/776855)
and a general feature of genome complexity in cnidarians (Technau et al. 2005). Here, we do not attempt to specify the timing of duplication events as there is insufficient information on rates of evolution in this genus; however, we do show that several diversification events have occurred within this taxon. Many of these events predate the radiation of species, as the copies are found in most of the species we explored.

Additional support for duplication events comes from the high GFP diversity within single coral colonies. Although our cloning efforts were extensive, they were not exhaustive—even using high-fidelity polymerase, we cannot discount PCR error or cloning artifacts from generating some of this diversity. However, pairwise comparisons of alleles within individuals suggest only a small minority of alleles that differ by only a single substitution (which does not contribute to the phylogenetic question being resolved here), and recombination analysis indicated that most of the diversity found in our system exhibits little recombinational signal. Our only concern for recombination-based diversity was in clade Iic (see splits tree in Supplementary Material online). However, gene diversity in corals is often created through hybridization of related lineages (Vollmer and Palumbi 2002; Frade et al. 2010) and so ambiguity in this lineage could be either biological or statistical. Overall, diversity of GFP-like alleles is extremely high in agariciid corals. Sequence divergence ranges up to 13.2% average between the Heliozoid and Agaricia lineages, and up to 11.4% divergence among clades within Agaricia. These divergences are similar to reported nucleotide differences among alleles recovered from M. cavernosa (Kelmanson and Matz 2003). Additionally, the number of distinct alleles from a single individual were similar to those found in M. cavernosa. This allelic variation further appears to support the findings of Kelmanson and Matz (2003), where color variation within species was attributed to variations in the relative expression of multiple copies of the FPs within an individual.

The question is whether the variable and potentially functionally important characters found here may be useful for systematic analysis in Agaricia. This genus has a complex taxonomic history and still lacks a resolved phylogeny of all species (Stake, 2007). All 3 Agaricia species collected are found in each of the 3 major GFP lineages shown here, but our cloning efforts only returned 2 individuals that had alleles from all 3 major clades. Thus, additional information will be useful to understand how this diversity is partitioned among taxa, color variants, and habitats. Within 2 of the major clades (I and III), we identify 2 subclades with complete lineage sorting between A. fragilis and A. agaricites, although A. lamarecki is polyphyletic in clade I and is not sampled enough to be certain about representation in remaining clades. It is also of interest to know how these alleles might reflect population structure and gene flow. Given the geographic range of collection sites throughout the Florida Keys, we evaluated population structure among sites using Hudson's Snn (given the small sample sizes and high nucleotide diversity). However, no significant structure was detected in any of the clades based on location, with the exception of clade Iib (calculated in DNAsp; results not shown). This structure was considered to be associated instead with taxonomic differences among collection sites for lineages in that subclade.

Whether all of this allelic diversity is ecologically relevant remains to be determined. Alleles in subclade Ia include a stop codon in most copies, assuming a standard genetic code as with other clades. This suggests that some of the duplicated diversity is non-functional. Many amino acid substitutions exist even within the subclades identified here (Figure 1). Values of \( \pi_a/\pi_s \) ratios are high for some clades (Table 1), suggesting the possibility of diversifying selection. Although the shifts in values of Tajima’s D may be informative with regard to admixture of paralog populations within clades, depending on the sample size from each type (Exwes and Wares 2012), our focus here is on differences among clades in the relative contribution of synonymous or non-synonymous diversity to this statistic (Table 1). In particular, clades IIIa and IIIb suggest different evolutionary dynamics for synonymous and non-synonymous mutations.

Here, we demonstrate as many as 7 distinct amino acid sequences within an individual of A. agaricites. Furthermore, we found up to 40 unique amino acid sequences across all specimens sampled, but we have not shown that all are functional, and have not established the relationship of this diversity to known color classes. However, a reported GFP-like sequence in the green color class from Agaricia (GenBank Accession number AY647156; Alieva et al. (2008)) aligns

<table>
<thead>
<tr>
<th>Comparison</th>
<th>( \pi_a/\pi_s )</th>
<th>Ka/Ks</th>
<th>( -\pi_s )</th>
<th>TajD</th>
<th>TajD N/S ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ia vs. Ib</td>
<td>1.036</td>
<td>0.523</td>
<td>-1.84*, -2.06*</td>
<td>-1.84*, -2.16*, -1.63</td>
<td>0.705</td>
</tr>
<tr>
<td>Ia vs. IIb</td>
<td>0.824</td>
<td>1.501</td>
<td>-2.16*, -1.63</td>
<td>-2.16*, n/a</td>
<td>0.705</td>
</tr>
<tr>
<td>Ia vs. IIC</td>
<td>( \pi_S = 0 )</td>
<td>0.705</td>
<td>-1.63, n/a</td>
<td>-1.63, n/a</td>
<td>0.705</td>
</tr>
<tr>
<td>IIB vs. IIC</td>
<td>0.151</td>
<td>7.624</td>
<td>-0.05, -0.35</td>
<td>-0.05, -0.35</td>
<td>0.151</td>
</tr>
<tr>
<td>IIIa vs. IIIb</td>
<td>1.007</td>
<td>2.328</td>
<td>-0.05, -0.35</td>
<td>-0.05, -0.35</td>
<td>1.007</td>
</tr>
<tr>
<td>I vs. II</td>
<td>0.728</td>
<td>0.473</td>
<td>-0.05, -0.35</td>
<td>-0.05, -0.35</td>
<td>0.728</td>
</tr>
<tr>
<td>I vs. III</td>
<td>0.479</td>
<td>0.473</td>
<td>-0.05, -0.35</td>
<td>-0.05, -0.35</td>
<td>0.479</td>
</tr>
<tr>
<td>II vs. III</td>
<td>0.479</td>
<td>0.473</td>
<td>-0.05, -0.35</td>
<td>-0.05, -0.35</td>
<td>0.479</td>
</tr>
</tbody>
</table>

Groups being compared are numbered as in Figure 1. The individual group values are reported in order with the groups being compared for Tajima’s D (*\( P < 0.05 \)), followed by the ratio of D for nonsynonymous and synonymous substitutions.
with 97% identity to alleles in clades IIIa (E-score = 4e−76). Other color classes Agaricid corals are reported to display based on visual observations include at least orange and yellow fluorescence (Van Moorsel 1983; Mazel et al. 2003). The functionality and/or color variation of haplotypes with unique amino acid substitutions found in our study will need to be examined through experimental approaches.

Further development of the species phylogeny is required before we can fully understand the functional implications of this diversity. Although approaches like reconciled tree analysis (Chen et al. 2000) can improve identification of paralogs in gene families, the phylogeny of Agaricia is still very much in question (Stake 2007; Meyers, unpublished). Here, we show that the diversity characterized for the FP gene family in other scleractinian corals has continued to develop even within recent species radiations, and may provide important clues into the maintenance of coral diversity in varying photic environments. However, as with the difficulty of exploring phylogenetic analysis: with a limited sample of the congeners A. lamarcki, our results show that very recent speciation events are not resolved into monophyletic groups, and the polyphylly of the major lineages A. agaricites and A. fragilis (particularly clade II) suggest a complexity of evolutionary dynamics that limit the utility of these data for phylogenetic analysis.

Supplementary Material
Supplementary material can be found at http://www.jhered.oxfordjournals.org/.

Funding
Environmental Protection Agency (EPA X7-97469002-8); the National Ocean Survey (NOAA/NOS MOA-2001-683); the National Science Foundation (Ecology of Infectious Disease, NSF-1015342); and the University of Georgia Research Foundation (2008–2010).

Acknowledgments
We thank K. Dyer, S. Pankey, V. Kosminin, and the CREMP coral monitoring team for logistical support. Three anonymous reviewers greatly improved the article. Research in the Florida Keys National Marine Sanctuary was conducted under Permit No. FKNMS 2006-058.

References


Wei NWV, Wallace CC, Dai CF, Moothien Pillay KR, Chen CA. 2006. Analyses of the ribosomal internal transcribed spacers (ITS) and the 5.8S gene indicate that extremely high rDNA heterogeneity is a unique feature in the scleractinian coral genus Acropora (Scleractinia; Acroporidae). Zool Stud. 45:404–418.

Received August 10, 2012; First decision October 18, 2012; Accepted March 25, 2013

Corresponding Editor: Jose Lopez