

Molecular Basis and Evolutionary Origins of Color Diversity in Great Star Coral *Montastraea cavernosa* (Scleractinia: Faviida)

Ilya V. Kelmanson and Mikhail V. Matz

Whitney Laboratory, University of Florida

Natural pigments are normally products of complex biosynthesis pathways where many different enzymes are involved. Corals and related organisms of class Anthozoa represent the only known exception: in these organisms, each of the host-tissue colors is essentially determined by a sequence of a single protein, homologous to the green fluorescent protein (GFP) from *Aequorea victoria*. This direct sequence-color linkage provides unique opportunity for color evolution studies. We previously reported the general phylogenetic analysis of GFP-like proteins, which suggested that the present-day diversity of reef colors originated relatively recently and independently within several lineages. The present work was done to get insight into the mechanisms that gave rise to this diversity. Three colonies of the great star coral *Montastraea cavernosa* (Scleractinia, Faviida) were studied, representing distinct color morphs. Unexpectedly, these specimens were found to express the same collection of GFP-like proteins, produced by at least four, and possibly up to seven, different genetic loci. These genes code for three basic colors—cyan, green, and red—and are expressed differently relative to one another in different morphs. Phylogenetic analysis of the new sequences indicated that the three major gene lineages diverged before separation of some coral families. Our results suggest that color variation in *M. cavernosa* is not a true polymorphism, but rather a manifestation of phenotypic plasticity (polyphenism). The family level depth of its evolutionary roots indicates that the color diversity is adaptively significant. Relative roles of gene duplication, gene conversion, and point mutations in its evolution are discussed.

Introduction

Coral reef ecosystems are characterized by an amazing variety of colors, but the evolutionary roots and functional role of this diversity remain largely unknown. It is currently believed that the major color determinants of the reef-building corals are proteins homologous to green fluorescent protein (GFP), which are responsible for the majority of colors superimposed upon the overall brownish hue provided by endosymbiotic zooxanthellae (Matz et al. 1999; Lukyanov et al. 2000; Dove, Hoegh-Guldberg, and Ranganathan 2001). These proteins possess a remarkable ability to synthesize the chromophore inside their own globules using the residues of their own polypeptide chain as substrates (Heim, Prasher, and Tsien 1994; Ormo et al. 1996; Gross et al. 2000; Martynov et al. 2001). The structure and molecular environment of the mature chromophore, and therefore the resulting color, are determined solely by the sequence of the protein. This fact provides a unique opportunity to apply the comprehensive suite of methods for molecular sequence analysis to directly address questions related to color evolution (Matz et al. 2002).

The current data set of sequenced and spectroscopically characterized GFP-like proteins of the class Anthozoa includes four color types, possessing different chromophore structures: green, yellow, orange-red, and non-fluorescent purple-blue (Labas et al. 2002). Quite unexpectedly, phylogenetic analysis of the data set showed that proteins of different colors do not constitute separate clades, but are intermixed within the terminal branches of the phylogenetic tree. This fact suggested that the color diversity, instead of being a product of prolonged evolution, originated recently and independently within several

lineages (Labas et al. 2002). Deciphering the evolutionary scenario that resulted in such a phylogenetic pattern is a challenging task, which at the present moment is additionally complicated by insufficient understanding of the adaptive role (or roles) of different colors in Anthozoa.

One of the basic issues regarding the evolution of anthozoan colors that can be addressed at the present state of knowledge is the basis of intraspecific color variation. Many coral species exhibit these variations, ranging from fluorescent blue to fluorescent red and nonfluorescent purple (Mazel 1995, 1997; Veron 2000). Two alternative mechanisms of generating diversity of coloration are possible, which are usually called polymorphism and polyphenism. Polymorphism is the situation when differences in color appearance can be traced to differences in the genome, in which case the color appearance of an organism is largely determined at the moment of zygote formation and there is little possibility for it to change afterwards, except for the intensity. This explanation of color diversity in corals seems particularly attractive, considering the relative ease of certain types of color conversion by random mutations (Lukyanov et al. 2000; Gurskaya et al. 2001b). In contrast, polyphenism is recognized when the same genome gives rise to different phenotypes. In the case of corals, color polyphenism would mean the existence of the same collection of genes coding for GFP-like proteins in all color morphs, the differences in color appearance being due to the changes in relative levels of expression of these genes. In this case, the color can be a much more flexible character during the organism's lifetime than it is allowed under the polymorphism model. The color diversity may also be a result of combination of these two models, in which case both allelic polymorphism and expression level variations would play significant roles.

In the present study, we used the great star coral *Montastraea cavernosa* to gain insight into this problem. This species is one of the most common reef-building corals in the Caribbean, exhibiting several color variants

Key words: green fluorescent protein, anthozoa, color evolution.

E-mail: matz@whitney.ufl.edu.

Mol. Biol. Evol. 20(7):1125–1133, 2003

DOI: 10.1093/molbev/msg130

Molecular Biology and Evolution, Vol. 20, No. 7

© Society for Molecular Biology and Evolution 2003; all rights reserved.

(Mazel 1995, 1997; Veron 2000). Our goals were to evaluate the number of genes coding for GFP-like proteins that are expressed in an individual colony and to determine whether the color differences between morphs are due to variations in expression levels of the color-coding genes or mutations within their coding sequences.

Methods

Sample Collection and cDNA Preparation

Samples of *M. cavernosa* were collected in the Florida Keys National Marine Sanctuary, at the Keys Marine Station, Key Largo (FKNMS authorization permit FKNMS-2000-009). The color morphs were sampled during day dives on the basis of their appearance and also during night dives when the fluorescence was observed using a handheld ultraviolet flashlight or blue dive light (UltraBlue, NightSea). The sampling involved chiseling off small pieces of colonies containing five to seven polyps, which were then brought to the laboratory and kept in a flow-through aquarium. Samples of tissue for cloning were scraped off these pieces when necessary. Emission spectra of the pieces were measured using an Ocean Optics spectrometry setup that included a LS-1 tungsten light source, a USB2000 UV-vis spectrometer, a reflectance probe, and a variable bandpass filter to set the excitation wavelength. Total RNA was isolated and cDNA was prepared according to invertebrate-adapted protocols described earlier (Matz 2002).

Cloning of cDNAs for GFP-like Proteins

The complete cloning procedure included three stages: (1) amplification of initial fragments with 11 different combinations of degenerate primers; (2) obtaining 5' and 3' flanks for all detected sequence variants; and (3) amplification and cloning into bacterial expression vectors of all open reading frames (ORFs) using primers designed to match all determined ORF termini. See online Supplementary Material for oligonucleotides' sequences and protocol details. Once the full-scale cloning was done on one colony of *M. cavernosa*, the last stage (full ORF cloning for bacterial expression) was repeated on two other colonies using the same set of primers designed for ORF termini. The bacterial colonies were screened for fluorescence using a stereomicroscope (Leica MZ FL III), and at least 96 fluorescent clones per cDNA sample were replated as streaks onto new plates, the same clones being also grown as overnight cultures for plasmid isolation and sequencing. The emission spectra of the streaks were measured using the Ocean Optics setup described above.

Contig Assembly

The sequences of the expression constructs were aligned and subjected to phylogenetic analysis using unweighted parsimony within PAUP* package (Swofford 1993) and maximum-likelihood/quartet puzzling approach implemented in TreePuzzle software (Strimmer and von Haeseler 1996). This analysis reliably separated the sequences into groups corresponding to different genes and/or alleles, avoiding possible confusion due to sequencing

errors and occasional point mutations introduced during PCRs. The coding sequences were then separated into contigs as suggested by phylogenetic analysis. Each contig was required to include at least two sequences, which must share at least two nucleotide substitutions that were not observed in other sequences (i.e., represent true synapomorphs). Moreover, each contig was required to contain sequences that encoded spectroscopically identical products. For the colony on which full-scale cloning was done, after subdivision of ORFs into separate contigs, the corresponding 5' and 3' cDNA termini were selected manually among the sequences obtained during cDNA flanks' amplification, requiring an exact match within the overlap of the flank sequence and the inferred coding sequence.

Recombinant Protein Purification and Spectroscopy

Representative clones for each of the detected genes/allelomorphs were grown in a 200 ml volume of LB medium overnight, and the recombinant protein was isolated from them using metalloaffinity chromatography, taking advantage of the 6×His tags attached at the step of coding sequence amplification, using standard protocol (<http://www.qiagen.com/literature/xpreslit.asp#expressionist>). Emission spectra of the products were determined using a USB-2000 spectrometer (Ocean Optics); the excitation spectra were measured using an LS50-B spectrofluorometer (Perkin Elmer).

Data Analysis

For contigs' tree reconstruction, the alignment of cDNA sequences was made by ClustalX (Thompson et al. 1997) and corrected manually according to triplet coding and considerations from protein structure (Matz et al. 1999). The tree of contigs was reconstructed using TreePuzzle software (Strimmer and von Haeseler 1996), with the following parameters: HKY model of DNA evolution (Hasegawa, Kishino, and Yano 1985) was used for maximum-likelihood distance calculation (parameters estimated from the dataset), gamma distribution of sites variability was assumed (alpha parameter estimated from the data set), and 10,000 puzzling steps were done to obtain support values. For gene conversion detection, GENECONV software (version 1.81) was used (Sawyer 1989; Hartl and Sawyer 1991). The tree of "nonconverted" contigs with addition of the rest of the known scleractinian proteins was constructed as a part of a global phylogenetic tree of all known GFP-like proteins (see online Supplementary Material) using MrBayes software (Huelsenbeck and Ronquist 2001). This program performs a Markov chain Monte Carlo simulation to search for the areas of maximum likelihood on a landscape of all possible trees and then builds a consensus from the trees found in these areas, at the same time calculating the posterior (Bayesian) probability for occurrence of each node. The alignment of coding sequences of all known GFP-like proteins was analyzed (see online Supplementary Material), with the exception of the sequences that were identified as possible products of gene conversion by GENECONV. Due to this criterion, in addition to the three sequences described in

this paper (mc6, g6, and r7), amajGFP (GenBank accession number AF168421), dis3GFP (GenBank accession number AF420593), and scubGFP2 (GenBank accession number AY037771) were also excluded. The Markov chain was run with the following settings: mutation rate classes were assigned according to the position within a codon (first, second and third); substitution model was general time reversible (Tavare 1986); number of parallel chains run was four; number of iterations was 200,000; sample frequency was 10; number of initial trees discarded during consensus building (“burn-in” value) was 8,000.

Assessment of Gene Expression Levels in Color Morphs

Pools of fragments of all detected target cDNAs were amplified from each of the morphs using nondegenerate primers that matched any of the cloned cDNA types equally well. The pools were cloned and sequenced, to calculate the number of occurrences of each cDNA type in the mixed populations. Since the primers had the same match quality to all the analyzed genes, and the lengths of the amplified fragments were practically the same, the efficiency of amplification for all the genes should be similar enough to expect that in the resulting product, the relative amounts of different types of GFP-like protein-coding sequences would reflect their original proportion in the mRNA sample. This is the same principle that was utilized in coamplification methods of quantitative PCR, where a control template of known amount was amplified along with an assayed template to provide reference at the amplification endpoint (see Raeymaekers 2000 for review). The sequencing sample size (77 for each morph) guaranteed with 95% confidence that any target mRNA comprising 4% or more of the original fragment pool would be detected.

Results

Cloning of GFP-like Proteins from *Montastraea cavernosa*

cDNA for initial analysis was obtained from a single colony of *M. cavernosa* that exhibited both green and red fluorescent colors (the colony was not characterized spectroscopically in vivo). The cloning was a PCR-based, three-stage procedure, designed to detect as many variants of the target cDNAs as possible. The last stage consisted in preparation of a small-scale bacterial expression library of the target cDNAs, so that the obtained clones could be both sequenced and characterized spectroscopically. The phylogenetic analysis of the sequences of the expression clones subdivided the pool into six clusters that were then used to build contigs, designated mc1 to mc6. The contigs were separated by at least 17, and up to 103, single-base changes within the ORF, corresponding to 2.7% to 16.7% overall nucleotide difference. Some contig-specific insertions/deletions (indels) were observed: mc2, mc3, and mc4 had a nine-base insertion, plus there was an additional 21-base insertion immediately after the initiation codon in cluster mc3. Sequences of cluster mc1 carried another characteristic three-base insertion, and clusters mc5 and mc6 shared another three-base insertion near the end of the

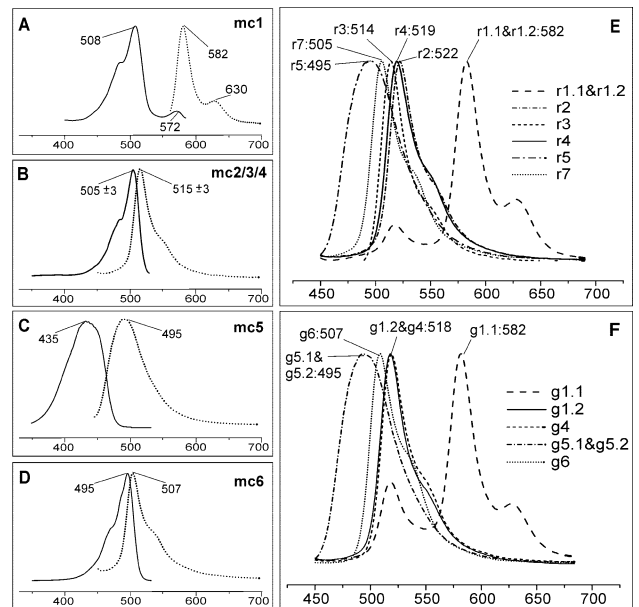


FIG. 1.—Fluorescent properties of GFP-like proteins found in three colonies of *M. cavernosa*. Horizontal axis shows wavelength in nanometers; vertical axis shows fluorescence intensity. (A–D) Proteins identified during initial full-length cDNA cloning. Solid lines indicate excitation; dotted lines indicate emission. Proteins mc2, mc3, and mc4 exhibited very similar characteristics and are represented by a consensus graph (B). Excitation or emission maxima are indicated for each curve. (E and F) Emission spectra of proteins found in two color morphs: red (E) and green (F). Names of proteins and emission maxima are indicated for each curve.

ORF. The flanks of the inferred cDNAs (the 24-base 5'-most and 20-base 3'-most parts of the coding sequences, as well as untranslated regions) were selected manually from previously obtained RACE results.

The emission spectra of the obtained collection can be roughly subdivided into three color bands: cyan, green, and red. It must be noted that in this case “cyan” does not correspond to a separate color class of GFP-like proteins, but rather is a subtype of a green class, as defined by Labas et al. (2002). Cyan and red bands were represented each by a single contig (mc5 and mc1 [fig. 1C and A], respectively), and the rest of the clones were emitting in the green range. Of these, representatives of the contig mc6 were clearly distinguishable from the rest of the green clones by positions of excitation and emission maxima (fig. 1D). The contigs mc2, mc3, and mc4 contained clones with very similar excitation and emission curves (fig. 1B).

In some cases, significant similarity (although never a complete identity) was found between a contig’s sequence and existing GenBank entries, suggesting that the product of the same gene was analyzed. Thus, mc1 was different by only three nucleotide sites from mcavRFP cDNA, mc3 differed by two sites from mcavGFP2 (Labas et al. 2002), and mc5 differed by four sites from the sequence of “cyan fluorescent protein mRNA” (Falkowski and Sun 2001). Surprisingly, mc2 was almost identical (two nucleotide difference) to the “green fluorescent protein mRNA” from *Montastraea faveolata*—another species from the same genus (Lesser and Barry 2001).

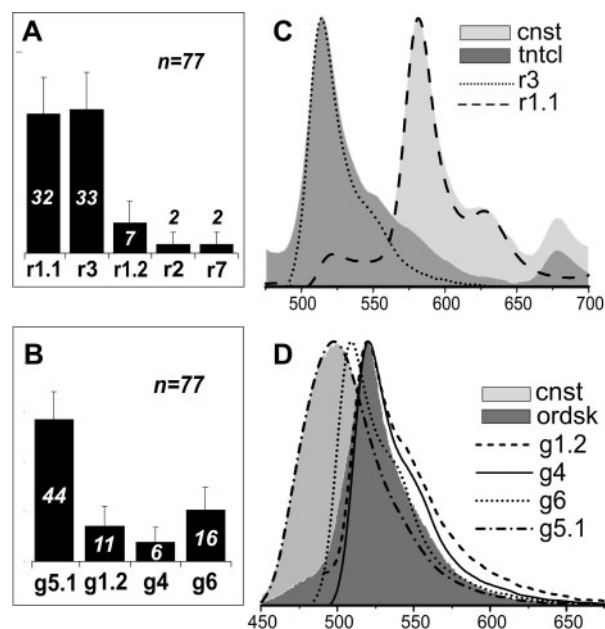


FIG. 2.—(A and B) histograms of relative abundance of transcripts for different GFP-like proteins in red (A) and green (B) morphs. Numbers on bars indicate the number of clones corresponding to a particular protein found within a random sample of 77. Error bars show 95% confidence interval based on the sample size. (C and D) Superimposition of the emission curves of the GFP-like proteins encoded by the most abundant mRNAs upon *in vivo* emission curves of the red (C) and green (D) morphs. Horizontal axis shows wavelength in nanometers; vertical axis shows fluorescence intensity. *In vivo* curves are filled with shades of gray; “cnst” = coenosteum, “tntcl” = tentacles, “ordsk” = oral disk.

cDNAs of GFP-like Proteins in Color Morphs

Two color morphs (red and green) were compared with regard to the spectral properties and expression levels of the encoded GFP-like proteins. The red morph exhibited orange-red emission in coenosteum (tissue between polyps) and oral disk, whereas the tentacles were weakly green-fluorescent. The green morph had a cyan-colored coenosteum, green oral disk, and green or nonfluorescent tentacles (see fig. 2C and D for *in vivo* emission curves).

The red coral colony yielded seven GFP-like contigs, designated r1.1, r1.2, r2, r3, r4, r5, and r7. Four of them (r1.1, r3, r4, and r5) were very similar to the previously detected sequences (mc1, mc3, mc4, and mc5, respectively), and contig r2 fell within the mc2/3/4 clade (greens), although not being clearly related to any of the sequences there. Contig r7 was not closely related to any of the previously detected ones, although phylogenetic grouping, indel pattern, and emission maximum pointed to mc6 as its closest relative. Finally, contig r1.2 was similar to mc1 but differed from it by 17 nucleotide substitutions (2.7% overall difference). Its product possessed an identical emission curve to those of mc1 and r1.1.

The green colony yielded six contigs (fig. 3A), designated g1.1, g1.2, g4, g5.1, g5.2, and g6. Two of them (g5.1 and g5.2) were similar to the previously identified mc5 (cyan) and exhibited identical emission spectra of the protein products. There were two green-coding cDNAs similar to others identified previously:

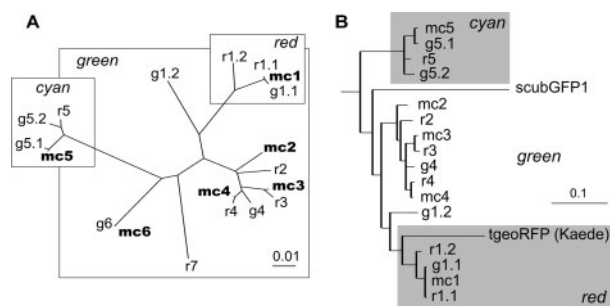


FIG. 3.—(A) Unrooted phylogenetic tree for all *M. cavernosa* GFP-like proteins described here. The tree was constructed using TreePuzzle software (Strimmer and von Haeseler 1996); all the nodes show puzzling support values exceeding 80 (10,000 puzzling attempts). (B) Phylogeny of currently known GFP-like proteins from order Scleractinia, with exception of the sequences that presumably resulted from gene conversion (mc6, g6, r7, and scubGFP2) and very distantly related genCP from *Goniopora tenuidens*. The tree is a part of the bigger phylogenetic tree for all known GFP-like proteins that was constructed using MrBayes (Huelsenbeck and Ronquist 2001) and is found in the online Supplementary Material. Only nodes with posterior probability more than 0.97 are shown. On both panels, general emission waveband of the proteins (cyan, green, or red) is indicated.

contig g6 was identical to mc6, and contig g4 clustered with mc4. The fifth contig was closer to mc1 (red) than to any other types by sequence similarity and indel pattern. However, it contained 58 nucleotide substitutions (9.3% overall difference) in comparison with mc1, and moreover, the corresponding protein product emitted in the green range with no trace of a red peak. Surprisingly, one of the contigs was nearly identical to the red-coding mc1 (contig g1.1) and indeed encoded a red fluorescent product. This contig was represented by two identical expression clones, which contained at least four shared nucleotide substitutions in comparison with the previously encountered red-coding cDNAs (mc1, r1.1, and r1.2). After the finding of these clones, the red fluorescence was actually detected in several isolated areas of the green morph.

Gene Conversion Analysis

Possible gene conversion events between the sequences of the contigs were studied using GENECONV (Sawyer 1989; Hartl and Sawyer 1991). Gene conversion is any process that causes a segment of DNA to be copied onto another segment of DNA, which in most cases means recombination. Given an alignment of DNA or protein sequences, GENECONV looks for aligned segments for which a pair of sequences is sufficiently similar to be suggestive of past gene conversion. Pairwise *P*-values are assigned to compare each fragment with the maximum similarity that might have been expected for that sequence pair in the absence of gene conversion. Global *P*-values compare each fragment with all possible fragments for the entire alignment. In the output, GENECONV identifies a pair of sequences from the alignment, between which gene conversion apparently happened, a *P*-value, and the limits of the copied segment. In a recent evaluation study, GENECONV was found to be one of the most powerful among the gene conversion detection methods (Posada 2002). The summary of the GENECONV output for the

current data set is given in table 1. GENECONV suggested several gene conversion pairs with participation of sequences mc6/g6 (these two contigs are identical) and r7, but the other sequence of the pair always comes from the mc2/3/4 clade. Taking into account high similarity within the clade, these results can be explained by only two gene conversion events, whereby an ORF fragment of one of the sequences from mc2/3/4 clade was copied into mc6/g6 and r7. Notably, the predicted copied segments for mc6/g6 and r7 map to the same region within the ORF and differ only in length (about 120 bases for mc6/g6 and 90 bases for r7) (see table 1). Removal of this segment from the alignment did not change the topology of the phylogenetic tree.

Relative Abundances of Transcripts of Color-Coding Genes in Color Morphs

For the red morph, two types of fragments were found to be significantly ($P < 0.05$) more abundant than others. These two fragments corresponded to r1.1 and r3, coding for red and green colors, respectively (see fig. 2A). Remarkably, the emission curves of the encoded products matched almost perfectly the red and green fluorescence observed in the coral in vivo (fig. 2C). In a green colony, only one of the four detected mRNAs was significantly more abundant than others—the cyan-coding g5.1, its protein product possessing exactly the same emission curve as the coenosteum of the coral (fig. 2B). Of the other three detected mRNAs, two (g1.2 and g4) encoded proteins with the same emission maxima as observed in the oral disks of the coral, although none of their emission curves matched in vivo fluorescence exactly (fig. 2D). The possible cause of this lack of match is the presence of the product of the fourth of the detected mRNAs, g6, in the same areas, which may result in a shortwave shoulder observed in the in vivo emission curve.

Discussion

Number of Color-Coding Loci

The unexpected diversity of mRNAs coding for GFP-like proteins found in a single coral species prompts a question as to whether these transcripts represent different loci or alleles of the same locus. Although this question cannot be completely settled without extensive studies of genomic sequences (which is beyond the scope of this work), there is a strong circumstantial evidence in favor of the “different loci” scenario. First, since *M. cavernosa* is a diploid organism, the maximum number of alleles per nuclear locus cannot exceed two, so finding as many as six or even seven different homologous transcripts certainly indicates the multilocus nature of color coding. Second, at least the major clades in the contig’s tree (fig. 3A) are separated by a degree of nucleotide difference that is too great to attribute to allelic polymorphism (9% to 16%, not counting indels); moreover, the 3’-UTRs in mc1, mc5, mc6, and the mc2/3/4 group are even more divergent. Meanwhile, studies of genetic diversity in corals suggest that their levels of allelic variability are much lower, even in noncoding nuclear loci. Thus, in two species of *Acropora* (*A. cervicornis* and

Table 1
Results of Gene Conversion Analysis by GENECONV

Sequence 1	Sequence 2	Converted Segment ^a			Simulated <i>P</i> -Value
		Start	End	Length	
mc6/g6	mc2	363	497	135	0.0051
mc6/g6	mc4	381	497	117	0.0305
r7	mc4	409	499	91	0.0127
r7	r3	412	499	88	0.0133
r7	mc3	414	499	86	0.0139

^a The coordinates correspond to the alignment MCs-cDNA_trim.fas found in Supplementary Material online.

A. palmata), there were at most two nucleotide differences between alleles per 448 positions of PaxC intron (Van Oppen et al. 2000), corresponding to 0.4% overall difference. Similar low allelic variability was found in these two species for minicollagen and calmodulin introns (Vollmer and Palumbi 2002). In a study of genetic diversity within a species from the same genus as *Montastraea cavernosa*, *M. annularis*, among 27 isolates of a rRNA gene fragment (internal transcribed spacers plus 5.8s rRNA gene), the nucleotide difference did not exceed 1.5% (Fukami et al. 2001). A third argument in favor of the multiple loci hypothesis is that the protein products corresponding to mRNAs of different clades tend to be distributed differently within the coral colony, indicative of independent gene expression control. For example, in a green morph, cyan color was expressed in the coenosteum, whereas green (represented by proteins belonging to mc2/3/4 clade, plus g6) was concentrated in the tentacles and oral disk. In a red colony, the tentacles were also green (due to r3, belonging to mc2/3/4 clade), whereas the coenosteum was red (mc1 clade) (see fig. 2C). The final argument is based on the topology of the phylogenetic tree for all known GFP-like proteins, the relevant subset of which is shown on figure 3B (the complete tree can be found in Supplementary Material online). There are four more fluorescent GFP-like proteins from stony corals (order Scleractinia) currently reported: red-emitting Kaede from *Trachyphyllia geoffroyi* (family Trachyphyllidae [Ando et al. 2002]), or tgeoRFP following the nomenclature of Labas et al. (2002); two green-emitting proteins from *Scolymia cubensis* (family Mussidae)—scubGFP1 and scubGFP2 (Labas et al. 2002); and a chromoprotein genCP from *Goniopora tenuidens* (Gurskaya et al. 2001a). genCP and scubGFP2 had to be excluded from the analysis, since genCP belongs to different ancient lineage of GFP-like proteins (Labas et al. 2002), and scubGFP2 apparently resulted from gene conversion, as suggested by GENECONV software (data not shown). In the global phylogenetic tree constructed for all known GFP-like proteins, the remaining two sequences (scubGFP1 and tgeoRFP) do not appear as sister groups to the cluster of proteins from *M. cavernosa* (family Faviidae), but fall within them (fig. 3B). As it is discussed in more detail below, their placement indicates that the separation of the three major sequence/color clades found in the tree preceded separation of at least two of the three represented coral families, which is improbable for alleles of the same locus.

We can therefore conclude that there are at least three separate loci coding for GFP-like proteins in *M. cavernosa*, which correspond to the clades in the contig tree and three major colors: cyan, green, and red (fig. 3A). It must be noted that these three colors represent only two color/structural classes of GFP-like proteins as suggested by Labas et al. (2002), since cyan proteins have the same chromophore as greens (Gurskaya et al. 2001b) and therefore should be considered a subtype of structural green class. Furthermore, the green color, as represented by the clade mc2/3/4, must involve at least two loci, because there are two cases when three different sequences that fell into this clade were obtained from the same specimen (mc2-mc3-mc4 and r2-r3-r4). Most probably, the sequences not included into any of the clades (contigs mc6/g6, r7 and g1.2) also represent separate loci, on the basis of the high nucleotide divergence (difference between each other and mean difference with any of the major clades is 9% to 15%). On the other hand, it is possible that sequences within the major clades represent alleles of the same locus (or two loci in the case of the mc2/3/4 clade), although the within-clade differences are still high (2.7% to 3.7% between sequences from a single specimen). Therefore, the conservative estimate of the number of color-coding loci in *M. cavernosa* is at least four (two corresponding to the cyan and red clades and two corresponding to the mc2/3/4 clade) and can be as high as seven if contigs mc6, r7, and g1.2 represent separate loci.

Polymorphism Versus Polyphenism

Polymorphism of a trait generally means that in natural populations it comes in several states; the states are inheritable and directly or indirectly correspond to allelic variations at certain loci. Color polymorphism is common in all organisms, from flowers (Subramaniam and Rausher 2000) to bears (Lynch and Ritland 1999), and provides an excellent basis for studies in color evolution and ecology. On the other hand, polyphenism means that the trait is not preset genetically and is determined by the environmental cues. In this case, the color of an organism, combined with the knowledge of factors that affect it, may help to draw conclusions about life history of the organism or its current condition, depending on when the color is determined. Thus, in butterflies, the wing color pattern may be an indicator of the diel cycle conditions (Nijhout 1997) or temperature (Roskam and Brakefield 1999) under which the animal has developed (see Beldade and Brakefield 2002 for review). In grasshoppers and locusts, color polyphenism may be dependent on population density and was suggested to be a basis for evolution of aposematic ("warning") coloration (Sword 2002). A remarkable case of polyphenism that may include color is seen in tadpoles of several frog species, which assume a different morphology and behavior depending on the presence of predators (dragonfly larvae) in the pond (McCollum and Van Buskirk 1996; Van Buskirk, McCollum, and Werner 1997).

In the present study, two different color morphs of *Montastraea cavernosa*, green and red, were found to contain, and express, the same functional suite of color-coding genes. We also demonstrated that the differences

in color appearances of these morphs were because of differences in expression levels (specifically, mRNA abundance) of these genes relative to one another, and not because some of the coding sequences in one or another morph were altered by mutations. These data certainly rule out at least the most straightforward polymorphism scenario, whereby the color differences are due to variations in protein sequences encoded by alleles at color-coding loci. However, the relative expression levels may also be genetically predetermined. Thus, for *Pocillopora damicornis*, Takabayashi and Hoegh-Guldberg (1995) demonstrated that the intensity of pink coloration, which is most probably due to a GFP-like protein from a non-fluorescent class (Lukyanov et al. 2000), can be a dynamic trait since it is light-inducible. However, only originally pink colonies showed noticeable induction, which was interpreted as evidence of genetic control. It must be noted, however, that Yu et al. (2000) recently argued that pink and brown morphs of *P. damicornis* show enough ecological and physiological differences to suggest that they may represent different species. In any case, our data for *M. cavernosa* can only be interpreted as polyphenism if factors other than genetic background are found to determine the color at some point of the coral's life cycle, or changes in the hue of coloration are observed during the lifetime of the coral. Todd et al. (2002a, 2002b) documented environment-induced color change in a Faviid coral *Favia speciosa* (same family as *M. cavernosa*): fragments of the green colony changed their color to red after 4 to 7 months after relocation to greater depth. It is tempting to note that our superficial observations from field sites where the samples were collected are in accord with these results. Although both green and red *M. cavernosa* colonies can be found in shallow (3 to 5 m) and deep (12 to 14 m) waters, the red colonies seem to be more abundant at greater depths. Another case of color change was observed in a specimen of mushroom coral *Fungia fungites*, which went from brown to intense purple and then to bright fluorescent green over the period of several months, apparently in response to varying lighting conditions (R. Rowan, personal communication). So, in *Favia speciosa* and *Fungia fungites*, color definitely is a polyphenetic trait. It appears highly probable that the same is also true for *Montastraea cavernosa*, especially taking in account its multilocus principle of color coding that is well suited to produce phenotypic plasticity. This issue can be settled in further experiments to confirm whether or not the color in *M. cavernosa* is inheritable, and if not (as the polyphenism scenario predicts), which factors determine the color and when they act during the life cycle. At the moment, we do not have enough data to discuss possible functions of the three different colors of *M. cavernosa* without this information at hand, not to mention the color ecology and physiology components that are missing for *M. cavernosa* and only beginning to be developed for other coral species.

Evolution of Intraspecific Sequence/Color Diversity

Our data suggest that the color of *Montastraea cavernosa* is the product of three mechanisms that give rise to

sequence diversity of the color-coding genes—gene duplication, gene conversion, and point mutations. Gene duplications played a major role, leading to separation of genes for three main colors, cyan, green, and red. Some of the duplication events happened even before separation of the families Faviidae, Mussidae, and Trachyphyllidae, as suggested by the placement of non-*Montastraea* sequences within the phylogenetic tree. All these three families belong to the suborder Faviina. According to the fossil record, Mussidae may have existed as a separate lineage since the mid-Jurassic (about 170 to 190 MYA), whereas separation of Faviidae and Trachyphyllidae happened much later, in the Eocene (about 50 MYA) (Veron 1995). The non-Faviidae sequences (scubGFP1 representing family Mussidae, and Kaede, or tgeORFP, representing Trachyphyllidae) do not form sister groups to the cluster of *M. cavernosa* proteins, as it would be expected if the families' separation preceded the genes' separation. Instead, Kaede protein falls within the mc1 clade (in accord with its red emission color), whereas scubGFP1 is basal to the group of mc1-mc2/3/4 clades (fig. 3B). This topology indicates that the three color lineages separated before separation of Faviidae and Trachyphyllidae, whereas separation of Mussidae from other represented coral families preceded the divergence of mc1 and mc2/3/4 (red and green) lineages. The split between mc5 clade (cyan) and the rest of colors is the most ancient, preceding the separation of all three coral families. This finding was unexpected, since cyan GFP-like proteins were previously considered merely a variation within a green class (Labas et al. 2002). Apparently, at least in studies devoted to color evolution and ecology, cyan should be regarded as a separate class of GFP-like proteins, most probably having specific function in corals.

Gene conversion is copying of sequence information between two different loci within a genome and can be a prominent evolutionary mechanism for gene families (Graham 1995; Ohta 2000). When happening on the scale of a whole gene, this mechanism homogenizes the sequences within a gene family (Elder and Turner 1995; Graham 1995; Kupriyanova 2000), whereas conversion of gene fragments produces an opposite effect: it generates combinatorial diversity (McCormack, Tjoelker, and Thompson 1991; Knight and Winstead 1997). Among *M. cavernosa* GFP-like proteins, two sequences (mc6/g6 and r7) appear to be resulting from gene conversion events of the latter (diversifying) type. In both cases, a fragment mapping to the same region within the ORF was copied from a sequence belonging to mc2/3/4 clade (table 1). Diversifying gene conversion seems to be a minor factor in evolution of the GFP-like proteins of *M. cavernosa* and Anthozoa in general, since it is detected only within a few sequences. In addition to mc6/g6 and r7, within the collection of 30 currently known GFP-like proteins except the ones from *Montastraea* genus (see online Supplementary Material), GENECONV suggested gene conversion events in only three sequences: amajGFP (GenBank accession number AF168421), dis3GFP (GenBank accession number AF420593), and scubGFP2 (GenBank accession number AY037771) (Labas et al. 2002; Matz, Lukyanov, and Lukyanov 2002), affecting stretches of

ORF 35 to 77 bases long (data not shown). It is tempting to speculate that rare occurrence of gene conversion between the gene family members indicates that they are rarely located close to each other within the genome (Graham 1995; Ohta 2000). However, this conclusion requires estimation of the frequency of “homogenizing” gene conversion on a whole-gene scale, which would require extensive analysis of genome sequences and is therefore beyond the scope of this work.

On the basis of molecular organization of GFP-like proteins and in vitro mutagenesis experiments, we earlier suggested that the third mechanism of diversity generation, point mutations, should play a special role in the evolution of colors within the GFP-like protein family: it should drive the change of red-shifted colors into green. Therefore, the very persistence of red-emitting proteins in nature can be viewed as an argument in favor of adaptive significance of the red color, since natural selection for the red must be involved to counterbalance the random mutation drive (Labas et al. 2002). We further speculated that if both red and green colors are adaptively significant, there may be cases when the color-converted reds that became green were not eliminated from the population, but shifted function according to their new color. The data presented here provide indirect evidence in favor of this hypothesis. First, the persistence of the separate lineage of red-emitting proteins since before the coral families diverged is consistent with the idea that red color is adaptively significant and is maintained by natural selection. Second, the finding of an abundantly expressed green-emitting protein belonging to the red clade by sequence similarity as well as by indel pattern (contig g1.2 from the green morph) may indicate that the predicted shift-function scenario has been realized. However, these facts should be treated with caution. In a multigene family with high frequency of gene duplication (and, most probably, gene loss) there is ample room for alternative explanation for the observed phylogenetic pattern. To settle the question about color evolution pathways, detailed statistical analysis of selection forces exerted upon individual sites within the protein family members (Suzuki and Gojobori 1999; Yang and Bielawski 2000) is currently underway, as well as analysis of ancestral sequences at the nodes of phylogenetic tree (Chang, Kazmi, and Sakmar 2002).

Conclusions

We found that in *Montastraea cavernosa*, the color differences between colonies are explained by varying levels of expression (specifically, transcript abundances) of a set of genes coding for GFP-like proteins. The set includes at least four, and probably up to seven, separate genetic loci and encodes proteins emitting at three general wavebands: cyan (wide emission peak at around 495 nm), green (narrow emission peak at 505 to 520 nm), and red (emission at 580 nm). Proteins grouped by the criterion of emission waveband also compose separate clades in the phylogenetic tree. Gene duplication should be considered a major mechanism that led to the present-day color diversity. Placement of GFP-like proteins from other

representatives of stony corals within the phylogenetic tree suggests that these gene duplications started before separation of the coral families, as early as the Jurassic period. Further studies will be directed at understanding details of color evolution history and pinpointing factors that affect and maintain the existing color diversity.

Supplementary Material Online

Alignment of cDNA sequences coding for all reported GFP-like proteins, trimmed for phylogenetic analysis, in FASTA format—ALLGFPs_cDNA_trim.fas.

Phylogenetic tree constructed by MrBayes using the above alignment—phylogenetic_tree.pdf.

Alignment of full-length cDNA sequences and synthetic expression constructs reported here, in FASTA format—MCs-cDNA.fas.

The same alignment as in MCs-cDNA.fas, trimmed for phylogenetic analysis, in FASTA format (was used to construct the unrooted tree in fig. 3A)—MCs-cDNA_trim.fas.

Acknowledgments

We are grateful to James Netherton III (Whitney Laboratory, University of Florida) for help with coral collection and manuscript preparation and to Nick V. Grishin (University of Texas) for providing access to computing resources. This study was supported by NIH (NIGMS) grant to M.V.M. (R01 GM66243).

Literature Cited

- Ando, R., H. Hama, M. Yamamoto-Hino, H. Mizuno, and A. Miyawaki. 2002. An optical marker based on the UV-induced green-to-red photoconversion of a fluorescent protein. *Proc. Natl. Acad. Sci. USA* **99**:12651–12656.
- Beldade, P., and P. M. Brakefield. 2002. The genetics and evolution of butterfly wing patterns. *Nat. Rev. Genet.* **3**:442–452.
- Chang, B. S. W., M. A. Kazmi, and T. P. Sakmar. 2002. Synthetic gene technology: applications to ancestral gene reconstruction and structure-function studies of receptors. *Meth. Enzymol.* **343**:274–294.
- Dove, S. G., O. Hoegh-Guldberg, and S. Ranganathan. 2001. Major colour patterns of reef-building corals are due to a family of GFP-like proteins. *Coral Reefs* **19**:197–204.
- Elder, J. F., and B. J. Turner. 1995. Concerted evolution of repetitive DNA-sequences in eukaryotes. *Q. Rev. Biol.* **70**:297–320.
- Falkowski, P. G., and Y. Sun. 2001. GenBank accession no. AY056460.
- Fukami, H., A. F. Budd, D. R. Levitan, R. Kersanach, and J. Jara. 2001. Genetic evaluation of species boundaries and hybridization of the *Montastraea annularis* complex using the molecular markers. Published only in GenBank database. Accession numbers AB065307–AB065334.
- Graham, G. J. 1995. Tandem genes and clustered genes. *J. Theor. Biol.* **175**:71–87.
- Gross, L. A., G. S. Baird, R. C. Hoffman, K. K. Baldrige, and R. Y. Tsien. 2000. The structure of the chromophore within DsRed, a red fluorescent protein from coral. *Proc. Natl. Acad. Sci. USA* **97**:11990–11995.
- Gurskaya, N. G., A. F. Fradkov, A. Terskikh, M. V. Matz, Y. A. Labas, V. I. Martynov, Y. G. Yanushevich, K. A. Lukyanov, and S. A. Lukyanov. 2001a. GFP-like chromoproteins as a source of far-red fluorescent proteins. *FEBS Lett.* **507**:16–20.
- Gurskaya, N. G., A. P. Savitsky, Y. G. Yanushevich, S. A. Lukyanov, and K. A. Lukyanov. 2001b. Color transitions in coral's fluorescent proteins by site-directed mutagenesis. *BMC Biochem.* **2**:6.
- Hartl, D. L., and S. A. Sawyer. 1991. Inference of Selection and Recombination from Nucleotide-Sequence Data. *J. Evol. Biol.* **4**:519–532.
- Hasegawa, M., H. Kishino, and K. Yano. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **22**:160–174.
- Heim, R., D. C. Prasher, and R. Y. Tsien. 1994. Wavelength mutations and posttranslational autoxidation of green fluorescent protein. *Proc. Natl. Acad. Sci. USA* **91**:12501–12504.
- Huelsenbeck, J. P., and F. Ronquist. 2001. MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**:754–755.
- Knight, K. L., and C. R. Winstead. 1997. Generation of antibody diversity in rabbits. *Curr. Opin. Immunol.* **9**:228–232.
- Kupriyanova, N. S. 2000. Conservation and variation of ribosomal DNA in eukaryotes. *Mol. Biol.* **34**:637–647.
- Labas, Y. A., N. G. Gurskaya, Y. G. Yanushevich, A. F. Fradkov, K. A. Lukyanov, S. A. Lukyanov, and M. V. Matz. 2002. Diversity and evolution of the green fluorescent protein family. *Proc. Natl. Acad. Sci. USA* **99**:4256–4261.
- Lesser, M. P., and T. M. Barry. 2001. GenBank Accession no. AF406766.
- Lukyanov, K. A., A. F. Fradkov, N. G. Gurskaya et al. (11 co-authors). 2000. Natural animal coloration can be determined by a nonfluorescent green fluorescent protein homolog. *J. Biol. Chem.* **275**:25879–25882.
- Lynch, M., and K. Ritland. 1999. Estimation of pairwise relatedness with molecular markers. *Genetics* **152**:1753–1766.
- Martynov, V. I., A. P. Savitsky, N. Y. Martynova, P. A. Savitsky, K. A. Lukyanov, and S. A. Lukyanov. 2001. Alternative cyclization in GFP-like proteins family: the formation and structure of the chromophore of a purple chromoprotein from *Anemonia sulcata*. *J. Biol. Chem.* **276**:21012–21016.
- Matz, M. V. 2002. Amplification of representative cDNA samples from microscopic amounts of invertebrate tissue to search for new genes. Pp. 3–18 in B. W. Hicks, ed. *Green fluorescent protein applications and protocols*. Humana Press, Totowa, NJ.
- Matz, M. V., A. F. Fradkov, Y. A. Labas, A. P. Savitsky, A. G. Zarskiy, M. L. Markelov, and S. A. Lukyanov. 1999. Fluorescent proteins from nonbioluminescent Anthozoa species. *Nat. Biotechnol.* **17**:969–973.
- Matz, M. V., K. A. Lukyanov, and S. A. Lukyanov. 2002. Family of the green fluorescent protein: journey to the end of the rainbow. *Bioessays* **24**:953–959.
- Mazel, C. H. 1995. Spectral measurements of fluorescence emission in Caribbean Cnidarians. *Mar. Ecol. Prog. Ser.* **120**:185–191.
- . 1997. Coral fluorescence characteristics: excitation—emission spectra, fluorescence efficiencies, and contribution to apparent reflectance. *Ocean Optics XIII, SPIE* **2963**:240–245.
- McCollum, S. A., and J. Van Buskirk. 1996. Costs and benefits of a predator-induced polyphenism in the gray treefrog *Hyla chrysocelis*. *Evolution* **50**:583–593.
- McCormack, W. T., L. W. Tjoelker, and C. B. Thompson. 1991. Avian B-cell development—generation of an immunoglobulin repertoire by gene conversion. *Annu. Rev. Immunol.* **9**:219–241.
- Nijhout, H. F. 1997. Ommochrome pigmentation of the linea and rosa seasonal forms of *Precis coenia* (Lepidoptera: Nymphalidae). *Arch. Insect Biochem. Physiol.* **36**:215–222.

- Ohta, T. 2000. Mechanisms of molecular evolution. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **355**:1623–1626.
- Ormo, M., A. B. Cubitt, K. Kallio, L. A. Gross, R. Y. Tsien, and S. J. Remington. 1996. Crystal structure of the *Aequorea victoria* green fluorescent protein. *Science* **273**:1392–1395.
- Posada, D. 2002. Evaluation of methods for detecting recombination from DNA sequences: empirical data. *Mol. Biol. Evol.* **19**:708–717.
- Raeymaekers, L. 2000. Basic principles of quantitative PCR. *Mol. Biotechnol.* **15**:115–122.
- Roskam, J. C., and P. M. Brakefield. 1999. Seasonal polyphenism in *Bicyclus* (Lepidoptera: Satyridae) butterflies: different climates need different cues. *Biol. J. Linn. Soc.* **66**:345–356.
- Sawyer, S. 1989. Statistical tests for detecting gene conversion. *Mol. Biol. Evol.* **6**:526–538.
- Strimmer, K., and A. von Haeseler. 1996. Quartet puzzling: a quartet maximum likelihood method for reconstructing tree topologies. *Mol. Biol. Evol.* **13**:964–969.
- Subramaniam, B., and M. D. Rausher. 2000. Balancing selection on a floral polymorphism. *Evolution* **54**:691–695.
- Suzuki, Y., and T. Gojobori. 1999. A method for detecting positive selection at single amino acid sites. *Mol. Biol. Evol.* **16**:1315–1328.
- Swofford, D. L. 1993. PAUP—a computer-program for phylogenetic inference using maximum parsimony. *J. Gen. Physiol.* **102**:A9.
- Sword, G. A. 2002. A role for phenotypic plasticity in the evolution of aposematism. *Proc. R. Soc. Lond. B Biol. Sci.* **269**:1639–1644.
- Takabayashi, M., and O. Hoegh-Guldberg. 1995. Ecological and physiological differences between two color morphs of the coral *Pocillopora damicornis*. *Mar. Biol.* **123**:705–714.
- Tavare, L. 1986. Some probabilistic and statistical problems of the analysis of DNA sequences. *Lect. Math. Life Sci.* **17**:57–86.
- Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**:4876–4882.
- Todd, P. A., R. C. Sidle, and L. M. Chou. 2002a. Plastic corals from Singapore: 1. *Coral Reefs* **21**:391–392.
- . 2002b. Plastic corals from Singapore: 2. *Coral Reefs* **21**:407–408.
- Van Buskirk, J., S. A. McCollum, and E. E. Werner. 1997. Natural selection for environmentally induced phenotypes in tadpoles. *Evolution* **51**:1983–1992.
- Van Oppen, M. J. H., B. L. Willis, H. Van Vugt and D. J. Miller. 2000. Examination of species boundaries in the *Acropora cervicornis* group (Scleractinia, Cnidaria) using nuclear DNA sequence analyses. *Mol. Ecol.* **9**:1363–1373.
- Veron, J. E. N. 1995. *Corals in space and time*. University of New South Wales Press, Sydney, Australia.
- . 2000. *Corals of the world*. Australian Institute of Marine Science, Townsville MC.
- Vollmer, S. V., and S. R. Palumbi. 2002. Hybridization and the evolution of reef coral diversity. *Science* **296**:2023–2025.
- Yang, Z. H., and J. P. Bielawski. 2000. Statistical methods for detecting molecular adaptation. *Trends Ecol. Evol.* **15**:496–503.
- Yu, J. K., T. H. Liao, C. A. Chen, L. S. Fang, and W. S. Tsai. 2000. Do color patterns of *Pocillopora damicornis* reflect Zooxanthellae diversity? *Coral Reefs* **19**:98–99.

David Irwin, Associate Editor

Accepted March 24, 2003