

Diversity and evolution of the green fluorescent protein family

Y. A. Labas*, N. G. Gurskaya†, Y. G. Yanushevich†, A. F. Fradkov†, K. A. Lukyanov†, S. A. Lukyanov†, and M. V. Matz^{§¶}

*Institute of Ecology and Evolution, Leninsky Prospekt 33, Moscow 117071, Russia; †Institute of Bioorganic Chemistry RAS, Miklukho-Maklaya 16/10, Moscow 117871, Russia; and ‡Whitney Laboratory, University of Florida, 9505 Ocean Shore Boulevard, St. Augustine, FL 32080

Edited by Roger Y. Tsien, University of California at San Diego, La Jolla, CA, and approved January 24, 2002 (received for review October 17, 2001)

The family of proteins homologous to the green fluorescent protein (GFP) from *Aequorea victoria* exhibits striking diversity of features, including several different types of autocatalytically synthesized chromophores. Here we report 11 new members of the family, among which there are 3 red-emitters possessing unusual features, and discuss the similarity relationships within the family in structural, spectroscopic, and evolutionary terms. Phylogenetic analysis has shown that GFP-like proteins from representatives of subclass Zoantharia fall into at least four distinct clades, each clade containing proteins of more than one emission color. This topology suggests multiple recent events of color conversion. Combining this result with previous mutagenesis and structural data, we propose that (i) different chromophore structures are alternative products synthesized within a similar autocatalytic environment, and (ii) the phylogenetic pattern and color diversity in reef Anthozoa is a result of a balance between selection for GFP-like proteins of particular colors and mutation pressure driving the color conversions.

Green fluorescent protein (GFP) from bioluminescent jellyfish *Aequorea victoria* is widely known for its usefulness in biotechnology. Its unique ability to synthesize chromophore within itself, without any need for external substrates or cofactors except molecular oxygen (1), made it an excellent *in vivo* marker of gene expression and protein localization in various biological systems. However, it was not until the discovery of GFP-like proteins in nonbioluminescent representatives of class Anthozoa (2) that the actual diversity of GFP-like proteins was fully realized. GFP-like proteins are responsible not only for fluorescent colors in Anthozoa, but also for nonfluorescent purple-blue colors (3–5). It was concluded that, apart from brownish pigmentation caused by symbiotic zooxanthellae, most of the colors perceived in reef Anthozoa are the results of GFP-like proteins, and therefore, they should be considered major color determinants in these organisms (6). With these findings, the GFP research takes a new very interesting direction related to evolutionary ecology rather than biotechnology. Anthozoa represent a unique case when each color is essentially determined by a sequence of a single protein molecule, thus offering an opportunity to apply molecular phylogenetic tools directly to the study of color evolution. It can be expected that in this way it will be eventually possible to obtain clear clues as to the origins of the legendary variety of reef colors, the phenomenon that, despite its prominence, still awaits general scientific explanation.

In this article, we present 11 new GFP-like proteins, bringing the total number of cloned and spectroscopically characterized family members to 27. Taking advantage of this large body of sequence-spectral data, we propose classification criteria for the family and perform phylogenetic analysis. Our data provide clues to the structural basis of color diversity of GFP-like proteins, as well as to the evolutionary forces producing and maintaining this diversity in reef Anthozoa.

Materials and Methods

Collection of Samples. Samples (100–500 mg of tissue) of *Montastraea cavernosa*, *Condylactis gigantea*, *Scolymia cubensis*, and

Ricordea florida were collected at the Florida Keys Marine Sanctuary (Long Key) under National Marine Sanctuary authorization FKNMS-2000-009. The samples were collected during night dives, and candidate specimens were selected on the basis of their appearance under a UV flashlight. Other samples (*Dendronephthya* sp., *Heteractis crispera*, *Discosoma* sp. 3, and *Zoanthus* sp. 2) were obtained from private seawater aquariums.

Cloning, Expression, and Spectroscopic Analysis of GFP-Like Proteins.

Total RNA was isolated from the tissue samples following the protocol described in ref. 7. Total cDNA was amplified by using the SMART cDNA amplification kit (CLONTECH). These amplified cDNA samples were used to amplify 3' fragments of cDNAs coding for GFP-like proteins and then to obtain the missing 5' flanks, exactly as described in ref. 2. After determining the complete cDNA sequence, the coding regions were amplified by using the same cDNA samples as were used to clone the 3' and 5' flanks as templates. An upstream (“N-terminal”) primer had a 5' heel (5'-t**TGAtTGAtGAAGGAGA**aatatc), carrying stop codons (bold) in all frames and bacterial ribosome-binding site (underlined), followed by the target cDNA sequence (20–22 bases) starting with the initiation codon of the ORF. The downstream (“C-terminal”) primer was 22–25-bases-long and corresponded to the antisense sequence of cDNA around the stop codon of the ORF. The resulting fragments were cloned by using a pGEM-T vector cloning kit (Promega) following the manufacturer's protocol, using *Escherichia coli* JM109 strain as host. The colonies were grown on LB/agar/carbenicillin plates supplemented with 0.3 mM of IPTG for 16–20 h at 37°C and then incubated for 2 days at 4°C. The fluorescent colonies were selected by using a fluorescent microscope and streaked widely on new plates. The same colonies were used for overnight culture inoculation followed by plasmid isolation and sequencing, to confirm the identity of the clone. The bacteria were harvested from the plates, suspended in 1 ml of PBS, and disrupted by sonication. The lysate was cleared by centrifugation (5,000 × *g*, 10 min at room temperature), and its fluorescent properties were determined by using an LS-50B spectrofluorometer (Perkin-Elmer); emission spectra were corrected for the dependence of photomultiplier sensitivity on the wavelength. For mcavRFP and rfloGFP, the “early” samples were harvested after 24 h at 37°C; the “late” samples were harvested after 24 h at 37°C followed by 4 days at 4°C.

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: GFP, green fluorescent protein; RFP, red fluorescent protein; CP, chromoprotein.

Database deposition: The sequences reported in this paper have been deposited in the GenBank database [accession nos. AY037776 (cgigGFP), AF420592 (hcigGFP), AF420593 (dis3GFP), AF420591 (dendGFP), AY037769 (mcavGFP), AY037772 (rfloGFP), AY037767 (scubGFP1), AY037771 (scubGFP2), AY037770 (mcavRFP), AY037773 (rfloRFP), and AY059642 (zoan2RFP)].

¶To whom reprint request should be addressed. E-mail: matz@whitney.ufl.edu.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Phylogenetic Analysis. The alignment of GFP-like proteins (see supporting information, which is published on the PNAS web site, www.pnas.org) was constructed after ref. 2, taking into account constraints of the protein structure. Subsequently, the DNA alignment was made by following the protein alignment, excluding the poorly aligned N- and C-terminal regions. The phylogenetic tree was constructed by using TREE-PUZZLE software (8) under an HKY model of DNA evolution (9), assuming that the variability of sites follows γ distribution with α parameter estimated from the data set. The tree was confirmed to be the maximum likelihood tree by PAML software (<http://abacus.gene.ucl.ac.uk/software/paml.html>; ref. 10) under an REV model (11). The tree built by TREE-PUZZLE from protein alignment [JTT model (12)] had the same topology but lower support values because of the smaller number of informative sites in the protein alignment.

Results and Discussion

Nomenclature. For the sake of clarity of phylogenetic analysis representation, and in accord with the newly proposed classification principle (see below), in this article we are using new nomenclature for GFP-like proteins. Our protein identification tags include a four-letter leader composed of the first letter of the genus name and three initial letters of the species name, followed by definition of color type: GFP, green; RFP, red; YFP, yellow; and CP, chromoprotein (nonfluorescent). When the species is not defined, the leader has four initial letters of the genus name. In the case of multiple unidentified species of the same genus, a number is added to the leader (such as in dis3GFP or zoan2RFP). In the case of several proteins of the same color type found in the same species, the number is added to the color definition (such as in scubGFP1 and scubGFP2). For *A. victoria* GFP and drFP583 from *Discosoma* sp., widely accepted common names, GFP and DsRed, are kept.

New GFP-Like Proteins. A total of 11 new GFP-like proteins was cloned and spectroscopically characterized. Their spectral features are summarized in Table 1; the excitation/emission spectra for all of them can be found in supporting information.

The new set includes representatives exhibiting features not seen before in Anthozoan GFP-like proteins. Two green proteins from *C. gigantea* (cgigGFP) and *H. crispa* (hcriGFP) possess double-peaked excitation spectra very similar to those of wild-type GFP, suggesting that their chromophores undergo photoconversion between neutral and ionized states (13, 14). The red-emitting protein zoan2RFP, although very similar to DsRed in the shape of excitation/emission curves, behaves as a “timer”—it turns green at first and then matures into red (Fig. 1 *A* and *B*), similarly to one of the mutant variants of DsRed (15). The two new red-emitters from great star coral *Montastraea cavernosa* (mcavRFP) and Florida corallimorph *Ricordea florida* (rflorRFP) also show a “timer” phenotype (Fig. 1 *C–F*). In contrast to zoan2RFP, they failed to mature completely into red in our bacterial expression trials, which resulted in two-peak emission spectra such as those shown in Fig. 1 *D* and *F*. Both these proteins exhibited two features that make them different from the rest of the red-emitters. First, the red emission band in the more mature form had a major excitation peak virtually identical to the one of the immature green form, the yellow-orange excitation peak being significantly smaller (Fig. 2). In other orange-red proteins, the red emission is excited best in the yellow-orange region (Table 1, spectra E). This unusual shape of excitation spectra may be the result of photoconversion of the ionization states of the chromophore [by analogy with green proteins (13, 14)]. The striking similarity of major excitation peaks for mature and immature proteins, however, makes it tempting to suggest that in mcavRFP and rflorRFP, the “built-in” fluorescence reso-

nance energy transfer (FRET) from the immature green form of the protein to the mature red form is the major mechanism giving rise to red emission. The second specific feature of mcavRFP and rflorRFP is the shape of their red emission curve: its major peak is narrow and almost symmetrical in contrast to the wide and skewed emission peak of the other red-emitters. In addition, there is a distinctive shoulder at approximately 630 nm (compare spectra E and F in Table 1). Because a GFP-type photoconversion, if it is present, is unlikely to have so much effect on the shape of emission spectra (16), these differences may indicate alterations in the covalent structure of the chromophore as compared with other red-emitters. At present, it is unclear whether these two features are linked at the structural or photochemical levels. Further research is necessary to gain better insight into this problem.

Structural/Spectral Types of GFP-Like Proteins. We propose to classify GFP-like proteins by their color as it appears to human eye. On the basis of the currently available data set of sequenced and spectroscopically characterized proteins, we discriminate four color types of GFP-like proteins: green, yellow, orange-red, and purple-blue, or chromoproteins (Table 1). The proteins having excitation/emission spectra such as in row A of Table 1 are sometimes called cyan or even blue (6, 17), because of their appearance under certain microscopy optics, but to the human eye, the color of these proteins after purification still appears bright green (2, 18). It must be noted that, in addition to the four colors discussed here, corals exhibit at least two more: truly blue with emission maximum at less than 475 nm and far-red with emission maximum over 600 nm (6, 17). No corresponding GFP-like proteins have been cloned thus far.

All of the GFP-like proteins reported to date share the same fold of polypeptide chain, termed “ β -can” (19, 20). The report of a truncated but functional chromoprotein from *Anemonia sulcata* (4) was later proven to be incorrect (21). However, there are substantial differences between these color types as far as the chromophore structure is concerned (see Table 1). In GFP (green color), the chromophore is formed by residues 65–67 (Ser-Tyr-Gly) as a result of condensation between the carbonyl carbon of Ser-65 and the amino nitrogen of Gly-67 that produces a five-member ring, followed by the dehydrogenation of the Tyr-66 methylene bridge (1, 19, 20). All of the green proteins apparently possess the same chromophore, and the differences in the spectral shapes are explained by modifications of its environment (13, 14, 18). In the red protein DsRed, the chromophore synthesis includes one more stage that extends the conjugated π -system of the chromophore—dehydrogenation of the bond between the α -carbon and amino nitrogen of the first chromophore-forming residue (22–24). Meanwhile, in the chromoproteins representative of asulCP, cyclization leads to the formation of a six-member rather than five-member ring, and the critical step in creating the extended conjugated π -system is breakage of the polypeptide chain immediately before the chromophore (21). Notably, no other chromoprotein contains such a chain break, as demonstrated by denaturing electrophoresis of the bacterial expression products (data not shown), indicating that the chromophore structure of asulCP is the exception rather than the rule within this color type. Biochemical and mutagenesis studies of the yellow zoanYFP (yellow) suggested that this protein has yet another chromophore structure (18, 25).

It must be concluded that pronounced color differences among GFP-like proteins indicate differences in chromophore structures, which justify the use of color for classification. However, different chromophores might be found even in the

Table 1. Summary of spectral features and chromophore structures in the family of GFP-like proteins

| Protein ID (original ID) | GenBank accession # | Taxonomy Genus species (Class, Sub-class, Order) | Excitation maxima, nm | Emission maxima, nm | Representative spectra excitation emission | COLOR | Representative chromophore structure |
|--|--|--|--|--|--|-------------|--------------------------------------|
| amajGFP (amFP486) dstrGFP (dsFP483) clavGFP (cFP484) | AF168421 AF168420 AF168424 | <i>Anemonia majano</i> (Anthozoa, Zoantharia, Actiniaria) <i>Discosoma striata</i> (Anthozoa, Zoantharia, Corallimorpharia) <i>Clavularia sp.</i> (Anthozoa, Alcyonaria, Alcyonacea) | 458 456 443 | 486 484 483 | | GREEN | |
| GFP cgigGFP hcriGFP | M62653 AY037776 AF420592 | <i>Aequorea victoria</i> (Hydrozoa, ..., Hydroida) <i>Condylactis gigantea</i> (Anthozoa, Zoantharia, Actiniaria) <i>Heteractis crispa</i> (Anthozoa, Zoantharia, Actiniaria) | 395, 471 399, 482 405, 481 | 508 496 500 | | | |
| ptilGFP rmueGFP zoanGFP (zFP506) asulGFP (asFP499) dis3GFP dendGFP ncavGFP rflorGFP scubGFP1 scubGFP2 | AY015995 AY015996 AF168422 AF322221 AF420593 AF420591 AY037769 AY037772 AY037767 AY037771 | <i>Ptilosarcus sp.</i> (Anthozoa, Alcyonaria, Pennatulacea) <i>Renilla muelleri</i> (Anthozoa, Alcyonaria, Pennatulacea) <i>Zoanthus sp.</i> (Anthozoa, Zoantharia, Zoanthidea) <i>Anemonia sulcata</i> (Anthozoa, Zoantharia, Actiniaria) <i>Discosoma sp.3</i> (Anthozoa, Zoantharia, Corallimorpharia) <i>Dendronephthya sp.</i> (Anthozoa, Alcyonaria, Alcyonacea) <i>Montastraea cavernosa</i> (Anthozoa, Zoantharia, Scleractinia) <i>Ricordea florida</i> (Anthozoa, Zoantharia, Corallimorpharia) <i>Scolymia cubensis</i> (Anthozoa, Zoantharia, Scleractinia) <i>Scolymia cubensis</i> (Anthozoa, Zoantharia, Scleractinia) | 500 498 496 403, 480 503 494 506 508 497 | 508 510 506 499 512 508 516 518 506 506 | | | |
| zoanYFP (zFP538) | AF168423 | <i>Zoanthus sp.</i> (Anthozoa, Zoantharia, Zoanthidea) | 494, 528 | 538 | | YELLOW | ? |
| DsRed (drFP583) dis2RFP (dsFP593) zoan2RFP | AF168419 AF272711 AY059642 | <i>Discosoma sp.1</i> (Anthozoa, Zoantharia, Corallimorpharia) <i>Discosoma sp.2</i> (Anthozoa, Zoantharia, Corallimorpharia) <i>Zoanthus sp.2</i> (Anthozoa, Zoantharia, Zoanthidea) | 558 573 552 | 583 593 576 | | ORANGE-RED | |
| ncavRFP rflorRFP | AY037770 AY037773 | <i>Montastraea cavernosa</i> (Anthozoa, Zoantharia, Scleractinia) <i>Ricordea florida</i> (Anthozoa, Zoantharia, Corallimorpharia) | 508, 572 506, 566 | 520, 580 517, 574 | | | |
| asulCP (asCP) | AF246709 | <i>Anemonia sulcata</i> (Anthozoa, Zoantharia, Actiniaria) | 568 | none | | PURPLE-BLUE | |
| hcriCP (hcCP) cgigCP (cgCP) cpasCP (cpCP) gfenCP (gCP) | AF363776 AF363775 AF383155 AF383156 | <i>Heteractis crispa</i> (Anthozoa, Zoantharia, Actiniaria) <i>Condylactis gigantea</i> (Anthozoa, Zoantharia, Actiniaria) <i>Condylactis passiflora</i> (Anthozoa, Zoantharia, Actiniaria) <i>Goniopora tenuidens</i> (Anthozoa, Zoantharia, Scleractinia) | 578 571 571 580 | none none none none | | | ? |

Note that this article uses different names for GFP-like proteins than proposed in original publications (the original names, where available, are given in brackets in the first column, see text for nomenclature details).

proteins of the same color, as it happens within the group of chromoproteins and probably within the orange-red group.

Phylogeny. For the phylogenetic tree representation (Fig. 3), GFP-like proteins from the subclass Alcyonaria—ptilGFP, rmueGFP (26), dendGFP, and clavGFP (2)—were treated as outgroups with respect to other proteins that were isolated from organisms of another Anthozoa subclass, Zoantharia. GFP from *A. victoria* (27) was excluded from the analysis, because a very long branch leading to it could not be placed within the tree with sufficient confidence.

Quite unexpectedly, the grouping of proteins in the phylogenetic tree did not follow color classes or chromophore structures, or even the systematic position of the host. Instead, there are at least four separate clades of Zoantharia proteins (Fig. 3, clades A–D) placed between two outgroup branches containing Alcyonaria proteins. This topology suggests that the four Zoantharia clades already existed before separation of the subclasses Alcyonaria and Zoantharia. The evidence that separation of the clades preceded the separation of Zoantharia orders is even more convincing. On one hand, all of the clades except clade A include proteins from more than one order; on the other hand, none of the orders except the under-sampled Zoanthidea has all their GFP-like proteins within one clade (Table 2).

Most importantly, the tree reveals multiple relatively recent

events of color diversification. Clade B unites proteins of three color types, green, red, and nonfluorescent, cloned from animals of orders Scleractinia and Corallimorpharia. Composition of clade A suggests that separation of nonfluorescent and green colors happened independently in Actiniaria. Clade C suggests very recent separation of green, yellow, and red colors. Clade D suggests two independent events of separation of green and orange-red colors, in addition to the two observed in clades B and C.

The frequency of color conversions within separate lineages makes it impossible, at least for the current data set, to determine what was the color of the common ancestor of GFP-like proteins. Still, the available body of information provides grounds for proposing general principles of maintenance of the present-day color diversity, both on molecular and ecological levels.

Molecular Basis of Color Conversion. Because a chromophore synthesis pathway in DsRed is an extended form of the GFP pathway (22–24), it can be easily imagined that any mutation damaging the additional autocatalytic stage in DsRed would convert it into green protein. Indeed, at least seven different mutant variants of DsRed emitting in the green range were found during random and site-specific mutagenesis (15, 28, 29). Similar reasoning should apply to the two new red proteins, because their red emission also arises as a result of further modification of the green-emitting chromophore.

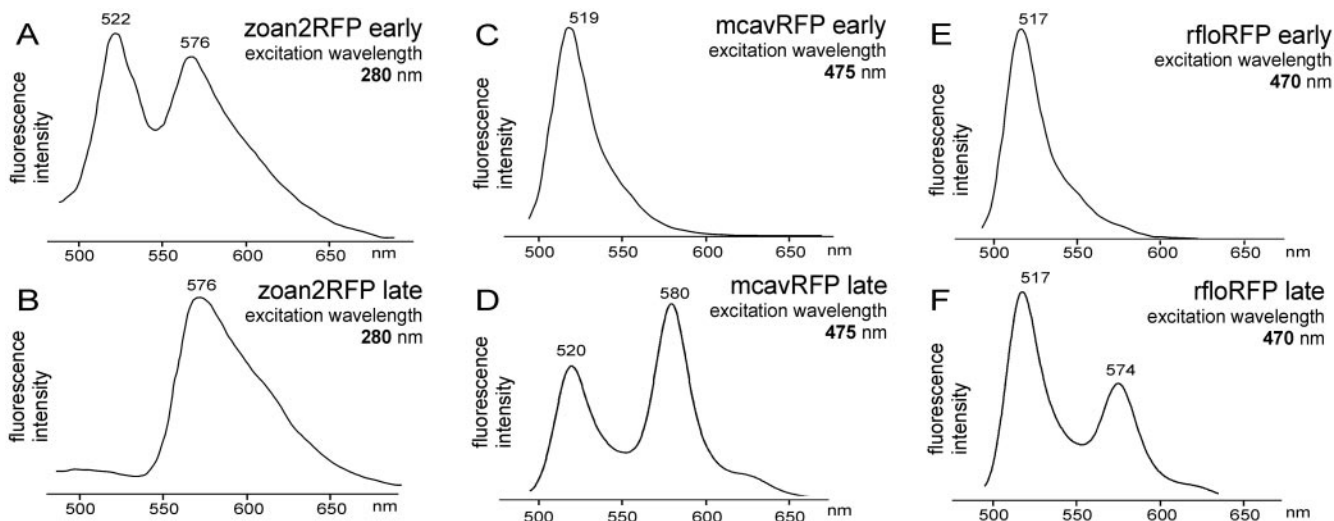


Fig. 1. Changes of emission spectra during maturation of the new red-emitters, zoan2RFP (A and B), mcavRFP (C and D), and rfloRFP (E and F). The excitation wavelength is given within each graph. Horizontal axis is wavelength in nanometers, vertical axis is fluorescence intensity. Maturation stages: A, C, and E, early; B, D, and F, late (see *Materials and Methods* for details). All of the three proteins exhibit “timer” phenotype (green emission at first and red emission arising later). Note that zoan2RFP matures significantly faster than mcavRFP and rfloRFP. Even at the “early” stage, the red emission peak is very pronounced, and by the “late” stage, the protein converts into red completely. In contrast, mcavRFP and rfloRFP fail to undergo such a complete maturation.

It has been shown that a single amino acid replacement can convert a chromoprotein into a DsRed-like red fluorescent protein (3, 5). It is particularly unexpected for asulCP from *A. sulcata*, which has been directly demonstrated to contain a very dissimilar chromophore; and it still seems unlikely that its red fluorescent mutant variant actually switches to synthesizing a DsRed-type chromophore instead of an original one. However, random mutations in this mutant variant resulted in the appearance of green-emitting forms (3). Because no green-emitting intermediate stage was present in the original asulCP autocatalytic pathway (21), formation of a green-emitting structure in these mutants signifies a substantial deviation most probably toward a GFP/DsRed type of chromophore formation sequence, judging by the shape of excitation/emission spectra of the green asulCP mutants (3).

Finally, yellow protein zoanYFP (yellow) also can be converted into a green-emitting state by at least two different amino acid replacements (18).

Taking these data into account, the following explanation of the observed phylogenetic pattern seems plausible. Different chromophore structures, even the most dissimilar ones, are alternative products synthesized with the help of a basically similar autocatalytic environment, rather than outcomes of prolonged evolution of different catalytic mechanisms. Apparently, just a few amino acid changes in the protein may act like a switch between alternative pathways, as exemplified by mutagenesis results on asulCP chromoprotein.

Color Diversity: Random Variations or Functional Differences? The adaptive role of Anthozoa coloration is still unclear. The early idea was that the GFP-like proteins serve as specific UV-blocking agents (30, 31) has been substantially modified in the course of recent research; at present, photoprotection in a more general sense is being discussed as the most likely functional assignment (6, 17). It was suggested that fluorescent pigments of corals might serve as a screen scattering the light that reaches endosymbiotic zooxanthellae, and that this photoprotection function may be turned into the opposite—light collection in low-light conditions—when the fluorescent layer is positioned below the zooxanthellae instead of above them (17). This function fits well the spectral properties of GFPs, especially as far

as the photoprotection part is concerned: in addition to efficient scattering, they would convert the light from the waveband very near the maximum of photosynthesis action spectrum into the waveband where photosynthetic pigments absorb 1.5–2 times less (32, 33). Meanwhile, the absorption/emission bands of most of the other GFP-like proteins do not match the peaks of photosynthesis action spectrum. For them it was further suggested that they could serve as enhancers of the photoprotection function of greens by means of fluorescent coupling: the energy transfer from green to yellow-red pigments would result in further conversion of the incoming light into wavelengths barely affecting photosynthesis (17). The physiological relevance of this mechanism remains to be investigated. Notably, in the two new orange-red proteins, mcavRFP and rfloRFP, a single protein type achieves this extent of conversion by absorbing in the same waveband as greens but emitting in the red region (Table 1, spectra F).

Perhaps surprisingly, only for nonfluorescent purple-blue coloration of Actiniaria, the topology of the phylogenetic tree suggests that its function is different from the function of green. Each of the three species (*A. sulcata*, *C. gigantea*, and *H. crispa*) yielded a nonfluorescent and a green protein, which are grouped according to color (Fig. 3, clade A). This grouping means that these colors existed as separate entities since the time before separation of the three genera. It must be added that green and purple-blue colors have the same pattern of differential localization in these organisms: green uniformly covers the whole tentacle (and in some cases, the rest of the animal body) and is displaced only at the tentacle tip by purple-blue. This fact constitutes an additional argument in favor of functional differences between these colors.

At the same time, no color-related grouping was observed for the orange-red color. Also, high similarity of yellow zoanYFP (yellow) to the green and red proteins within its clade makes it unlikely that yellow proteins would be grouped according to color when more representatives of this color class will be described. This situation might be viewed as an argument in favor of the “random variation” scenario that these colors are products of accumulation of the random mutations under relaxed environmental restraints and have no adaptive significance of their own (34, 35). Although orange-

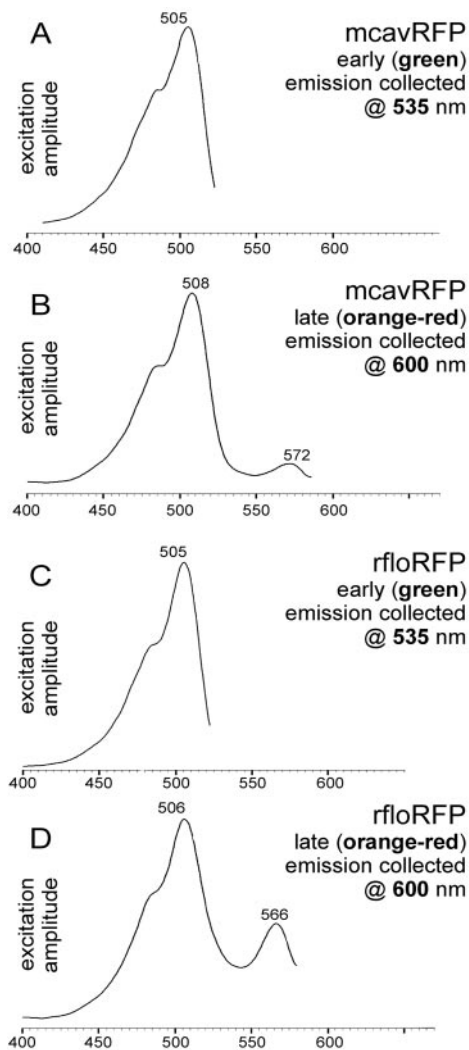


Fig. 2. Details on excitation spectra of mcavRFP (A and B) and rfloRFP (C and D). Wavelengths at which the emission was monitored are given within the graphs. (A and C) Excitation spectra of the green emission band in the immature protein, lacking the red emission. (B and D) Excitation spectra of the red emission band in a more mature form. Horizontal axis is wavelength in nanometers, vertical axis is emission amplitude. Note that in both proteins the major excitation peaks for immature green and mature red forms are virtually identical to each other.

red and yellow proteins can be easily converted into green by random mutations, the relative probability of backward conversion should be also considered. Mutation pressure would work as the sole factor generating the color diversity only in the case of approximately equal probabilities of forward and backward conversions. If, in contrast, the “to green” conversion is much more likely than reversion, selection should be involved to counterbalance the action of mutation pressure; otherwise, nongreen proteins would be eventually driven into a green state. At least in the case of orange-red proteins, where red emission is caused by auxiliary autocatalysis, the probability of damaging the necessary autocatalytic environment must be higher than the probability of its restoration once the damage is done. Indeed, the only successful attempt to engineer a green protein into a red-emitting state required two simultaneous site-directed replacements, which nevertheless resulted in only partial color conversion (18). A good example is also GFP from *A. victoria*: despite years of effort, no

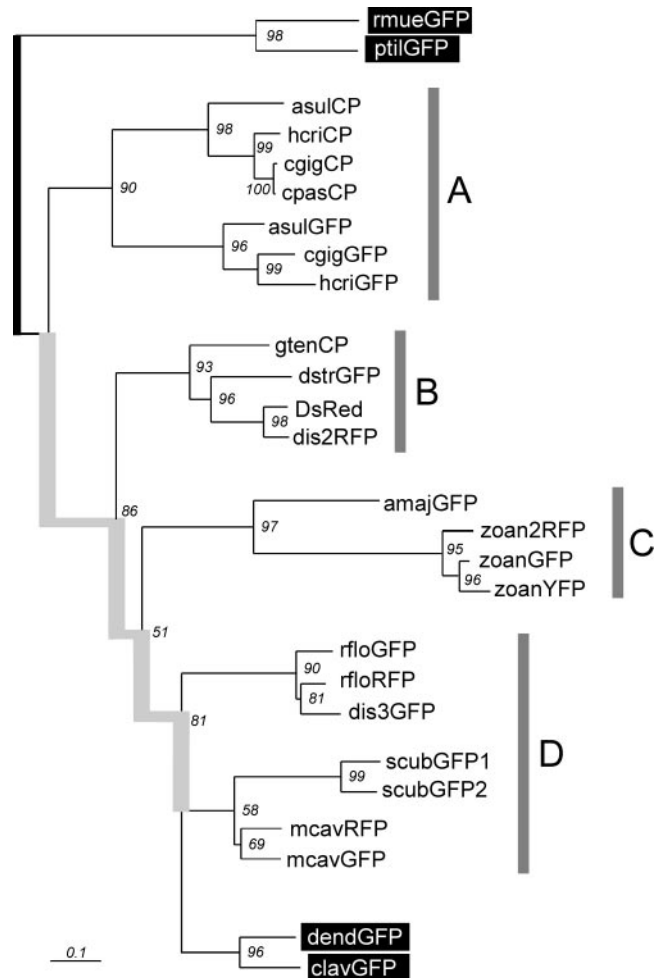


Fig. 3. The maximum-likelihood phylogenetic tree for the current data set of Anthozoan GFP-like proteins. Numbers at nodes denote the quartet-puzzling support values (1,000 puzzling attempts). Proteins from Alcyonaria subclass, which are considered outgroups, are labeled in white on black. The “stem” of the tree (thick gray line), joining two rooting groups, putatively reflects the diversity of GFP-like proteins before the separation of Alcyonaria and Zoantharia subclasses. Gray bars marked A–D denote the four distinct clades of GFP-like proteins found in Zoantharia. [Bar = 0.1 replacements/site.]

significantly red-shifted variants of it were found during random mutagenesis. Similar reasoning probably applies to yellow GFP-like proteins, because no less than three specific simultaneous replacements were necessary to convert green protein zoanGFP into yellow (18).

Thus, in our view, a more likely evolutionary scenario to explain the observed phylogenetic pattern is that at least some of the different colors are responsible for different functions and

Table 2. Summary of the major clades of GFP-like proteins from sub-class Zoantharia

| Clade | Colors | Zoantharia orders |
|-------|--------------------------------|-----------------------------------|
| A | Green, purple-blue | Actiniaria |
| B | Green, orange-red, purple-blue | Corallimorpharia, Scleractinia |
| C | Green, yellow, orange-red | Actiniaria, Zoanthidea |
| D | Green, orange-red | Corallimorpharia, Scleractinia |

are maintained by natural selection. Although color-converted mutant variants do appear because of mutation pressure, they have an option to assume another function according to their new color and thus avoid elimination. To test this hypothesis, two types of further research are essential: ecological and physiological studies on the functions of nongreen colors, including studies of relevance of fluorescent coupling between different

colors for photoprotection; and molecular characterization of GFP-like proteins in morphs of the same species to determine the role of mutation pressure in generating color polymorphism.

This work was supported by Russian Foundation for Fundamental Research Grants 01-04-49037 (to S.A.L.) and 99-04-48873 (to Y.A.L.), and partially by CLONTECH.

1. Heim, R., Prasher, D. C. & Tsien, R. Y. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 12501–12504.
2. Matz, M. V., Fradkov, A. F., Labas, Y. A., Savitsky, A. P., Zharaisky, A. G., Markelov, M. L. & Lukyanov, S. A. (1999) *Nat. Biotechnol.* **17**, 969–973.
3. Lukyanov, K. A., Fradkov, A. F., Gurskaya, N. G., Matz, M. V., Labas, Y. A., Savitsky, A. P., Markelov, M. L., Zharaisky, A. G., Zhao, X., Fang, Y., et al. (2000) *J. Biol. Chem.* **275**, 25879–25882.
4. Wiedenmann, J., Elke, C., Spindler, K. D. & Funke, W. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 14091–14096.
5. Gurskaya, N. G., Fradkov, A. F., Terskikh, A., Matz, M. V., Labas, Y. A., Martynov, V. I., Yanushevich, Y. G., Lukyanov, K. A. & Lukyanov, S. A. (2001) *FEBS Lett.* **507**, 16–20.
6. Dove, S. G., Hoegh-Guldberg, O. & Ranganathan, S. (2001) *Coral Reefs* **19**, 197–204.
7. Matz, M. V. (2002) in *Green Fluorescent Protein: Applications and Protocols*, ed. Hicks, B. W. (Humana, Totowa, NJ), pp. 3–18.
8. Strimmer, K. & von Haeseler, A. (1996) *Mol. Biol. Evol.* **13**, 964–969.
9. Hasegawa, M., Kishino, H. & Yano, K. (1985) *J. Mol. Evol.* **22**, 160–174.
10. Yang, Z. (1997) *Comput. Appl. Biosci.* **13**, 555–556.
11. Yang, Z. H., Goldman, N. & Friday, A. (1994) *Mol. Biol. Evol.* **11**, 316–324.
12. Jones, D. T., Taylor, W. R. & Thornton, J. M. (1992) *Comput. Appl. Biosci.* **8**, 275–282.
13. Brejc, K., Sixma, T. K., Kitts, P. A., Kain, S. R., Tsien, R. Y., Ormo, M. & Remington, S. J. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 2306–2311.
14. Palm, G. J., Zdanov, A., Gaitanaris, G. A., Stauber, R., Pavlakis, G. N. & Wlodawer, A. (1997) *Nat. Struct. Biol.* **4**, 361–365.
15. Terskikh, A., Fradkov, A., Ermakova, G., Zharaisky, A., Tan, P., Kajava, A. V., Zhao, X., Lukyanov, S., Matz, M., Kim, S., Weissman, I. & Siebert, P. (2000) *Science* **290**, 1585–1588.
16. Heim, R., Cubitt, A. B. & Tsien, R. Y. (1995) *Nature (London)* **373**, 663–664.
17. Salih, A., Larkum, A., Cox, G., Kuhl, M. & Hoegh-Guldberg, O. (2000) *Nature (London)* **408**, 850–853.
18. Gurskaya, N. G., Savitsky, A. P., Yanushevich, Y. G., Lukyanov, S. A. & Lukyanov, K. A. (2001) *BMC Biochem.* **2**, 6.
19. Ormo, M., Cubitt, A. B., Kallio, K., Gross, L. A., Tsien, R. Y. & Remington, S. J. (1996) *Science* **273**, 1392–1395.
20. Yang, F., Moss, L. G. & Phillips, G. N., Jr. (1996) *Nat. Biotechnol.* **14**, 1246–1251.
21. Martynov, V. I., Savitsky, A. P., Martynova, N. Y., Savitsky, P. A., Lukyanov, K. A. & Lukyanov, S. A. (2001) *J. Biol. Chem.* **276**, 21012–21016.
22. Gross, L. A., Baird, G. S., Hoffman, R. C., Baldrige, K. K. & Tsien, R. Y. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 11990–11995.
23. Wall, M. A., Socolich, M. & Ranganathan, R. (2000) *Nat. Struct. Biol.* **7**, 1133–1138.
24. Yarbrough, D., Wachter, R. M., Kallio, K., Matz, M. V. & Remington, S. J. (2001) *Proc. Natl. Acad. Sci. USA* **98**, 462–467.
25. Savitsky, A. P., Lukyanov, K. A., Gurskaya, N. G., Savitsky, P. A., Matz, M. V., Fradkov, A. F. & Lukyanov, S. A., 11th International Symposium on Bioluminescence and Chemoluminescence, September 6–10, 2000, Pacific Grove, CA.
26. Szent-Gyorgyi, C. S., Bryan, B. J. & Szczepaniak, W. (2001) U.S. Patent 6232107-B.
27. Prasher, D. C., Eckenrode, V. K., Ward, W. W., Prendergast, F. G. & Cormier, M. J. (1992) *Gene* **111**, 229–233.
28. Baird, G. S., Zacharias, D. A. & Tsien, R. Y. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 11984–11989.
29. Wiehler, J., von Hummel, J. & Steipe, B. (2001) *FEBS Lett.* **487**, 384–389.
30. Kawaguti, S. (1944) *Palao Trop. Biol. Stn. Stud.* **2**, 617–674.
31. Catala, R. (1959) *Nature (London)* **183**, 949.
32. Kinzie, R. A., Jokiel, P. L. & York, R. (1984) *Mar. Biol.* **78**, 239–248.
33. Kuhl, M., Cohen, Y., Dalsgaard, T., Jorgensen, B. B. & Revsbech, N. P. (1995) *Mar. Ecol. Prog. Ser.* **117**, 159–172.
34. Fox, D. L. (1974) in *The Significance of Zoochromes*, ed. Needham, A. E. (Springer, New York).
35. Wicksten, M. K. (1989) *Bull. Mar. Sci.* **45**, 519–530.