GFP-like Proteins as Ubiquitous Metazoan Superfamily: Evolution of Functional Features and Structural Complexity

Dmitry A. Shagin,[*][1] *Ekaterina V. Barsova,*[*][1] *Yurii G. Yanushevich,*[*]
Arkady F. Fradkov,[*] *Konstantin A. Lukyanov,*[*] *Yulii A. Labas,*[†] *Tatiana N. Semenova,*[‡]
Juan A. Ugalde,[§][||] *Ann Meyers,*[§] *Jose M. Nunez,*[§] *Edith A. Widder,*[¶]
Sergey A. Lukyanov,[*] *and Mikhail V. Matz*[§][#]

*Institute of Bioorganic Chemistry RAS, Moscow, Russia; †Institute of Biochemistry RAS, Moscow, Russia;
‡Shirshov Institute of Oceanology RAS, Moscow, Russia; §Whitney laboratory, University of Florida, St. Augustine;
||Laboratory of Bioinformatics and Gene Expression, INTA–University of Chile, Santiago, Chile; ¶Harbor Branch Oceanographic
Institution, Fort Pierce, Florida; and #Department of Molecular Genetics and Microbiology, University of Florida, Gainesville

Homologs of the green fluorescent protein (GFP), including the recently described GFP-like domains of certain extracellular matrix proteins in Bilaterian organisms, are remarkably similar at the protein structure level, yet they often perform totally unrelated functions, thereby warranting recognition as a superfamily. Here we describe diverse GFP-like proteins from previously undersampled and completely new sources, including hydromedusae and planktonic Copepoda. In hydromedusae, yellow and nonfluorescent purple proteins were found in addition to greens. Notably, the new yellow protein seems to follow exactly the same structural solution to achieving the yellow color of fluorescence as YFP, an engineered yellow-emitting mutant variant of GFP. The addition of these new sequences made it possible to resolve deep-level phylogenetic relationships within the superfamily. Fluorescence (most likely green) must have already existed in the common ancestor of Cnidaria and Bilateria, and therefore GFP-like proteins may be responsible for fluorescence and/or coloration in virtually any animal. At least 15 color diversification events can be inferred following the maximum parsimony principle in Cnidaria. Origination of red fluorescence and nonfluorescent purple-blue colors on several independent occasions provides a remarkable example of convergent evolution of complex features at the molecular level.

Introduction

Green fluorescent protein (GFP) along with its mutants and homologs are widely known today because of their extensive use as in vivo fluorescent markers facilitating biomedical studies (Tsien 1998; Lippincott-Schwartz and Patterson 2003). Investigation of this protein family began more than 40 years ago when GFP from hydromedusa *Aequorea victoria* (synonym *A. aequorea*) was discovered (Johnson et al. 1962; Shimomura, Johnson, and Saiga 1962). Before long, similar green proteins were detected in many bioluminescent coelenterates including various medusae, apparently all luminescent hydroid polyps, and a few others (Herring 1978; Chalfie 1995). Within bioluminescent systems, fluorescent GFP-like proteins function as secondary emitters, making the luminescence spectrum a sharp green peak instead of a wide blue one (e.g., Herring 1978; Ward and Cormier 1978; Gorokhovatsky et al. 2003). This effect is ostensibly an adaptation to better suit the visual systems of potential observers in green coastal waters (Partridge and Cummings 1999). In addition to altering the emission color, in *Renilla* GFP actually improves the bioluminescence quantum yield (Ward and Cormier 1978). Despite the apparently wide distribution of GFP-like proteins in bioluminescent organisms, only a few of them were cloned: GFP from *Aequorea victoria* (Prasher et al. 1992), two more proteins (Xia et al. 2002; Gurskaya et al. 2003) from medusae of the same genus (one of these

proteins, surprisingly, was neither fluorescent nor colored until modified by site-specific mutagenesis [Gurskaya et al. 2003]), and three proteins from Anthozoans of the order Pennatulacea (Szent-Gyorgyi, Bryan, and Szczepaniak 2001). Several GFP-like proteins of apparent bacterial origin can be found by searching the GenBank database (for example, in Azotobacter and Azomonas, accession numbers AF324408.1 and AF324405.1); however, because these sequences are identical to the GFP from *Aequorea victoria*, these entries are more than likely the result of contamination by GFP-containing bacterial cloning vectors.

In 1999 GFP homologs were cloned from non-bioluminescent Anthozoa species (Matz et al. 1999). These proteins exhibited an unexpected color diversity that, in addition to green, also included yellow and red fluorescent and purple-blue nonfluorescent colors (Labas et al. 2002). Recent analysis of GFP-like proteins in the great star coral *Montastraea cavernosa* (Kelmanson and Matz 2003) concluded that proteins of a cyan subtype of the green class have a separate evolutionary history, suggesting that the cyan coloration may serve a specific function. As for the function of fluorescence, the hypothesis originally proposed by Kawaguti (1944), who suggested that fluorescent proteins may be photoprotective, has received the most experimental support thus far with respect to photoprotection of endosymbiotic algae (Salih et al. 2000). Still, the function remains controversial. In many cases the efficacy of intrinsic photoprotection mechanisms of the algae (Gorbunov et al. 2001) surpasses the expected effects offered by fluorescent proteins (Mazel et al. 2003). Moreover, under the photoprotection hypothesis, the reason behind the existing diversity of colors remains elusive. It is possible that photoprotection is one, but not the only, function of fluorescent proteins in Anthozoa. Among

possible alternatives are photosynthesis aid (Salih et al. 2000) and photoreception functions, although no evidence for either has yet been found in related studies (Gorbunov and Falkowski 2002; Gilmore et al. 2003). Still another possibility is a straightforward pigment function: the generation of color effects aimed at the outside observer (Ward 2002). Fluorescent GFP-like proteins seem to be quite suitable for such a task (Mazel and Fuchs 2003); however, it is difficult to imagine where the selective advantage of appearing colorful might lie in the case of Cnidaria.

To date the number of GFP-related, Anthozoa-derived sequences in Genbank is about 100—a striking contrast to just three sequences that were cloned from class Hydrozoa. We recently described the diversity and phylogenetic relationships among anthozoan GFP-like proteins (Labas et al. 2002), but it was impossible to relate this information to GFPs from other sources or even to root the Anthozoan tree, because of the lack of data from other systematic groups.

Another unexpected finding came in 2001 (Hopf et al. 2001), when it was shown that bilaterian animals also harbor an unmistakable GFP-like domain that comprises part of the so-called G2F fragment ("globular fragment two") of extracellular matrix proteins called nidogens (entactins) and fibulins. This GFP-like domain of the G2F fragment (which we later call G2FP domain) has a clearly homologous fold, but is less than 10% identical to GFPs by amino acid sequence. G2FP domains are found in all bilaterian genomes sequenced thus far and are neither colored nor fluorescent, but instead they serve as protein-binding modules that participate in control of the extracellular matrix formation during development (Willem et al. 2002; Tunggal et al. 2003). Although G2FP domains can be successfully aligned with fluorescent GFP-like proteins by structure tracing methods (see online Supplementary Material), it was previously impossible to determine the phylogenetic relationships between these two major groups with suitable confidence, because of the great sequence divergence between them and the lack of other comparably disparate groups in the dataset.

In the present work we describe four new sequences of GFP-like proteins from Hydrozoa, including the yellow fluorescent and the nonfluorescent purple ones. Six green-fluorescent GFP homologs from planktonic Copepoda of the Pontellidae family (phylum Arthropoda, class Crustacea) are also described. The expansion of the Hydrozoa clade and the addition of an entirely new clade of bilaterian fluorescent proteins into the data set provided an opportunity to resolve deep-level phylogenetic relationships within the superfamily for the first time.

The following are the GenBank accession numbers:

Hydrozoa FPs: AY485333 (phiYFP); AY485334 (anm1GFP1); AY485335 (anm1GFP2); AY485336 (anm2CP).
Copepoda FPs: AY268071 (ppluGFP1); AY268072 (ppluGFP2); AY268073 (laesGFP); AY268074 (pmeaGFP1); AY268075 (pmeaGFP2); AY268076 (pdae1GFP).

Materials and Methods
Collection of Samples

A fluorescent stereomicroscope Leica MZ FL-III was used to detect fluorescent specimens in plankton samples. Several bright green-fluorescent Copepoda specimens (phylum Arthropoda; subphylum Crustacea; class Maxillopoda; subclass Copepoda; order Calanoida; family Pontellidae) were found in samples collected in the Gulf Stream, 120 miles east of Charleston, S.C., August 15–17, 2002, 2 h after sunset, at a depth of 0–10 m. Three small hydroid medusae were also selected, one from the Copepoda sample ("anthomedusa 1") and two others from surface plankton samples collected in the Intracoastal Waterway near Whitney laboratory in October–December 2002. One of the medusae was identified to the genus level *Phialidium sp.* (class Hydrozoa, order Hydroida, suborder Leptomeduzae, family Campanulariidae); two others belonged to suborder Anthomedusae of the order Hydroida, so they were denoted as "anthomedusa 1" and "anthomedusa 2."

cDNA Cloning and Protein Analysis

The organisms were fixed in RNAlater solution (Ambion) and stored at −20°C for 1–2 months before processing. Total RNA was isolated from single copepod specimens using a NucleoSpin RNA II kit (Clontech). cDNA was synthesized and amplified with a SMART PCR cDNA Synthesis kit (Clontech) and cloned into PCR-Script vector (Stratagene). About 5×10^4 recombinant clones for each library were visually screened using a fluorescent stereomicroscope. See Supplementary Material online for details of protein purification and spectroscopic and biochemical analyses.

Phylogenetic Analysis

The alignment of cDNA coding regions was analyzed using the Bayesian maximum likelihood method implemented in MrBayes 3.0 (Huelsenbeck and Ronquist 2001) and also by nonparametric bootstrap that used maximum likelihood, minimum evolution, and maximum parsimony criteria in PAUP* version beta 4.10 (Swofford 2002). Protein alignment was analyzed under maximum likelihood criterion using TreePuzzle (Schmidt et al. 2002). See online Supplementary Material for the list of accession numbers of the aligned sequences, details of alignment construction, and phylogenetic methods.

Results and Discussion
Features of the Novel Proteins

Spectral and biochemical features of the novel proteins are summarized in figure 1 and table 1. For spectral data on all the Copepoda proteins, see online Supplementary figure 1; for examples of gel filtration analysis, see online Supplementary figure 2. It is important to note that, in contrast to coral GFP-like proteins, the new Hydrozoa fluorescent proteins are dimeric rather than tetrameric, whereas the new chromoprotein appears to be a monomer. The only Copepoda protein analyzed for oligomerization is also a monomer. This low degree of oligomerization makes

the new proteins promising as candidates for the development of biotechnology tools.

Yellow phiYFP

The primary structure of this protein bears striking similarity in some key positions to engineered variants of *A. victoria* GFP. First, phiYFP has leucine in position 64 instead of the characteristic phenylalanine of GFP and many other fluorescent proteins (numbering corresponds to GFP sequence). Substitution of leucine in this site was found to greatly improve GFP folding, so this mutation was introduced in the commercially available "enhanced" GFP mutant, EGFP (Cormack, Valdivia, and Falkow 1996; Yang et al. 1998). Second, position 65 in phiYFP is occupied by threonine. In GFP, substitution of the wild-type serine for threonine at this position results in a dramatic change in the shape of absorption spectrum, and it is also included in the EGFP mutant. Finally, phiYFP contains tyrosine in the position 203, the key mutation imperative to yellow emission in the commercially available GFP mutant, YFP (Ormo et al. 1996). In contrast to the only other known yellow fluorescent protein zoanYFP (Matz et al. 1999), the absorption spectrum of phiYFP in denaturing conditions perfectly matches the spectrum of denatured GFP (fig. 1*F* and *G*), indicative of identical chromophores. Therefore, phiYFP is the only natural GFP-like protein found to date that utilizes the same structural solutions to adjust its spectral properties as was found for GFP in protein engineering studies.

Chromoprotein anm2CP

The anm2CP absorption spectrum was very similar to that of Anthozoa chromoproteins (Gurskaya et al. 2001*a*). It was suggested (Chudakov et al. 2003) and recently confirmed during analysis of the first crystal structure of the blue chromoprotein Rtms5 from the coral *Montipora efflorescens* (Prescott et al. 2003) that the chromophore in chromoproteins may be a DsRed-type in a nonplanar trans conformation. Another chromophore was reported in chromoprotein asulCP from the sea anemone *Anemonia sulcata* (Martynov et al. 2001). Although the exact structure still awaits confirmation by crystallography, distinction from the DsRed-type chromophore is evident when absorption spectra of denatured asulCP and DsRed are compared (Gross et al. 2000; Martynov et al. 2001) (see fig. 1*H*–*K*). In contrast, spectral properties of anm2CP in a denatured state clearly indicated that it possesses a DsRed-type chromophore (fig. 1*H*–*K*), which then must be isomerized similarly to Rtms5.

Earlier we showed that amino acids at positions 148, 165, and 203 are the key determinants of spectral differences between fluorescent proteins and chromoproteins (Lukyanov et al. 2000; Gurskaya et al. 2001*a*; Bulina et al. 2002; Chudakov et al. 2003). In anm2CP, positions 148, 165, and 203 are occupied by Glu, Cys, and Val, respectively. Although such a combination was never observed in GFP-like proteins, it is more similar to a chromoprotein-like than to a fluorescent protein-like arrangement. One can speculate that bulky Glu148 and Val203 should cause steric hindrance of the chromophore
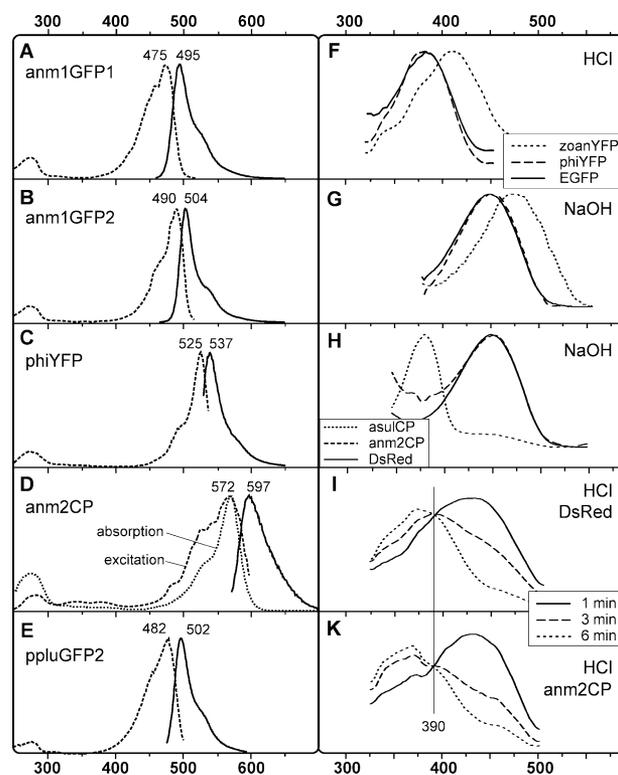

Fig. 1.—Spectroscopic characteristics of the novel proteins. Horizontal axis: wavelength in nanometers. Vertical axis: normalized emission or absorption amplitude. (*A–E*): Excitation (dashed lines) and emission (solid lines) spectra for the novel proteins. For the chromoprotein anm2CP (panel *D*), the absorption curve is shown as a dotted line. Copepoda proteins are spectroscopically similar to each other and therefore represented by a single protein, ppluGFP2 (panel *E*). (*F, G*): Absorption spectra of acid (*F*) and alkali (*G*) denatured yellow proteins and EGFP. (*H*): Absorption spectra of alkali-denatured asulCP, DsRed and amn2CP. (*I, K*): Changes in the absorption spectra of DsRed (*H*) and anm2CP (*I*) in acid during a 6 minute interval reflecting the hydrolysis of a DsRed-type chromophore into a GFP-like form. This behavior is never observed in asulCP (not shown).

cis conformation while small Cys165 should favor its *trans* conformation.

Copepoda GFPs

We identified six closely related (more than 60% identity) GFP-like proteins in copepods: two from *Pontellina plumata*, one from *Labidocera aestiva*, two from cf. *Pontella meadi*, and one from an unidentified species. All proteins demonstrated green fluorescence and possessed similar, but not identical, spectra (fig. 1*E*). Copepoda GFPs represent the most remote clade of fluorescent GFP-like proteins (see below for more discussion on phylogeny); however, the chromophore-forming Tyr66 and Gly67, as well as Arg96 and Glu222 that are presumably involved in autocatalysis, are conserved. A characteristic feature of Copepoda GFP is the absence of tryptophane residues (only laesGFP contains Trp179), which explains their weak excitation at 280 nm.

Absorption spectra of the acid- and alkali-denatured ppluGFP2 exactly matched the denatured GFP spectra (not shown). pH titration of the ppluGFP2 solution in the acidic

Table 1
Properties of Novel GFP-like Proteins from Hydrozoa and Copepoda

Protein Name	Species	Absorption max, nm	Emission max, nm	Molar Extinction Coefficient $M^{-1}cm^{-1}$	Fluorescence Quantum Yield	Oligomeric State	Probable Chromophore Structure[a]
anm1GFP1	Unidentified	475	495	75,000	0.65	Dimer	GFP
anm1GFP2	Unidentified	490	504	ND	ND	Dimer	GFP
phiYFP	*Phialidium sp.*	525	537	115,000	0.60	Dimer	GFP
anm2CP	Unidentified	572	597	120,000	<0.001	Monomer	DsRed (Rtms5?)
ppluGFP1	*Pontellina plumata*	480	500	65,000	0.60	ND	GFP
ppluGFP2	*Pontellina plumata*	482	502	70,000	0.60	Monomer	GFP
laesGFP	*Labidocera aestiva*	491	506	ND	ND	ND	GFP
pmeaGFP1	*Pontella meadi*	489	504	99,000	0.74	ND	GFP
pmeaGFP2	*Pontella meadi*	487	502	98,000	0.72	ND	GFP
pdae1GFP	Unidentified	491	511	105,000	0.68	ND	GFP

[a] The denoted chromophore structures are described in: GFP (Cody et al. 1993); DsRed (Gross et al. 2000); Rtms5 (Prescott et al. 2003).

region (pH 5.0 and below) demonstrated gradual transformation of the 482 nm absorption peak into the peak at 376 nm with an isosbestic point at 415 nm, a behavior characteristic of GFP and many of its mutants. Therefore, ppluGFP2 most likely possesses an imidazolidinone chromophore identical to that of *A. victoria* GFP in the deprotonated state.

Possible Functions of Novel GFP-like Proteins

GFPs of anthomedusae may be linked to bioluminescence, similarly to GFP from *Aequorea victoria*. However, none of our subjects was obviously luminous and the localization of fluorescence in these medusae—extendable parts, such as tentacles and/or manubrium—did not resemble known luminosity patterns, which are confined to the bell (Herring 1978). It is possible that fluorescence in these medusae represents a daylight functional analog of bioluminescence (similar to "blanching" in comb jellies [Mackie 1995]). It may be speculated that upon mechanical stimulation medusa would retract its previously extended tentacles and/or manubrium, thereby rapidly concentrating the fluorescent substance in a small area, which might produce a flash-like visual effect. The function of the yellow fluorescent protein, phiYFP, from *Phiallidium* is more difficult to explain. Although the protein seems to colocalize with luminous regions, it is unlikely to be involved in bioluminescence, because the emission color of this medusa is green (Herring 1978), suggesting the involvement of GFP (that apparently escaped our cloning efforts) rather than YFP.

In the case of Copepoda it would also be tempting to suggest that their GFPs are involved in bioluminescence, because many copepods possess this capability, if not for the fact that no luminous species are known from the Pontellidae family (Herring 1988). Moreover, the fluorescent areas in our specimens are not associated with any visible glands and are certainly quite unlike the luminous gland regions in other copepods (Herring 1988). The Pontellidae family includes many species that are colored blue or purplish by carotenoproteins. Among our specimens, *Labidocera aestiva* exhibited this kind of blue coloration, which did not show any obvious spatial correlation to the fluorescent

regions. It is possible that this blue as well as green fluorescent coloration may serve as counter-shading (blue for open ocean, green fluorescence for near shore waters) or as disruptive coloration directed at near-surface predators that use color vision. It can also be reasoned, given the notable dissimilarity of fluorescent patterns in analyzed species, that these patterns may be used by copepods themselves to recognize conspecific individuals.

The role of nonfluorescent proteins in general is the most mysterious. In Anthozoa they are usually confined to extremities of the organism, such as branch or tentacle tips, and are not correlated to any of the tested physiology parameters (Takabayashi and Hoegh-Guldberg 1995). Finding of such a protein in a hydromedusa only adds more intrigue to the story. Among other possibilities, the protein may serve a shadowing role for the jellyfish's photoreceptors, which would enable the animal to sense the direction of light. Although this function in most animals is served by melanins, there are variations, including a peculiar case of hemoglobin recruited for this role (Burr et al. 2000). The nonfluorescent protein from Hydrozoa represents a third case of independent evolution of a nonfluorescent color (see discussion below).

Evolution of the Superfamily
GFP Homologs Comprise a Superfamily

The group of structural homologs of GFP that all share the GFP-like "beta-can" fold should be regarded as a superfamily following the criteria proposed by the Protein Information Resource (http://pir.georgetown.edu/pirwww/otherinfo/sfdef.html). The main reason for such a classification is that this group unites at least two clearly definable protein families. The first one consists of G2FP domains, which are incapable of autocatalytic chromophore synthesis and are found within multi-domain proteins of the extracellular matrix. The second one includes fluorescent and/or colored proteins capable of synthesizing the chromophore autocatalytically and which are not found in a multi-domain context. We compiled an HMM profile that recognizes all of the currently known members of the superfamily (see online Supplementary Material) and has the potential to identify novel ones.
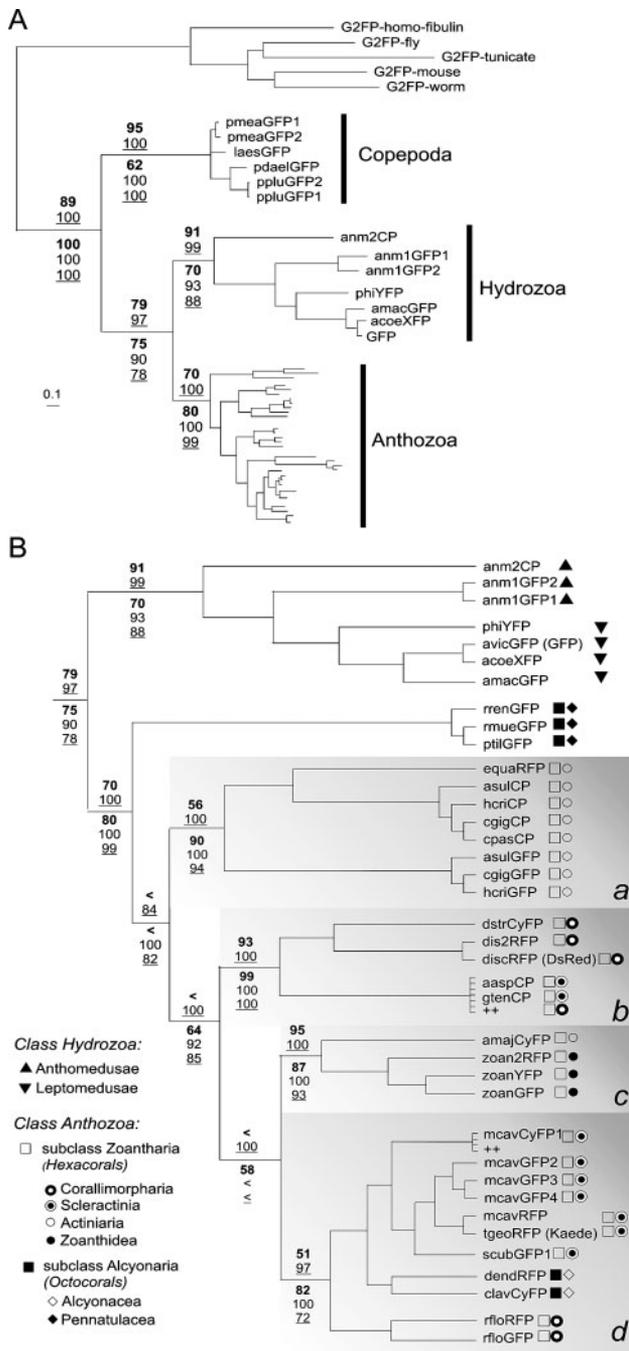
Fig. 2.—(A) Phylogenetic tree of GFP-like proteins. The clade of Anthozoa proteins is represented schematically, for the purpose of illustrating the characteristic branch lengths. Alignment of very remotely related sequences, such as of G2FP domains and fluorescent proteins, was done following the superimposition of known 3-D protein structures (see online Supplementary Material). The values above the branches are Tree-Puzzle support indexes obtained after analysis of protein sequences (bold) and the Bayesian analysis of cDNA (underlined). Below the branches are support values obtained by non-parametric bootstrap analysis of cDNA under the following criteria: maximum likelihood (bold), minimum evolution (normal), and maximum parsimony (underlined). Scale bar: 0.1 replacements/nucleotide. (B) Phylogenetic tree topology of Cnidarian GFP-like proteins. Support values of the four major clades (A, B, C, and D) of Anthozoa proteins and problematic deep internal branches follow the legend to panel A, "<" indicating less than 50% support. All other branches are supported at better than 50% level by all DNA-based methods. Groups of very similar proteins are depicted as terminal

Deep-level Phylogeny

The evolutionary route of the GFP superfamily could be imagined as two separate lineages that diverged by descent, whereby one, in Cnidaria, gained fluorescent properties, while another, in Bilateria, specialized in protein binding and became part of extracellular matrix proteins. It is also possible, however, that these two lineages originated as a result of an ancient gene duplication preceding the separation of Cnidaria and Bilateria, so that both G2FPs and fluorescent GFPs could be found within a single genome, both in Cnidaria and Bilateria. The position of Copepoda fluorescent proteins within the tree helps to discriminate between these two possibilities. Under the first scenario ("G2FPs for Bilateria, GFPs for Cnidaria"), Copepoda proteins can either be G2FPs that evolved fluorescence independently of GFPs, or products of horizontal gene transfer from one of the Cnidarian lineages. Under the second scenario ("ancient duplication"), Copepoda should be seen as an outgroup with respect to Cnidaria, but still much more related to them than to bilaterian G2FPs. Results of phylogenetic analysis (fig. 2A) clearly confirm the latter variant. First, there is very strong statistical support for monophyly of the clade uniting deuterostome and protostome G2FPs ($P < 0.0001$ in AU, KH, and SH tests; Kishino and Hasegawa 1989; Shimodaira and Hasegawa 1999; Shimodaira 2002), which indicates that Copepoda GFPs were already separate at the time of the common ancestor of deuterostomes and protostomes. Second, Copepoda GFPs did not arise within any of the known Cnidaria lineages either, ruling out the possibility of horizontal gene transfer from there. Third, Copepoda sequences are significantly more similar to cnidarian GFPs than to G2FPs; the average maximum likelihood distance between Copepoda and Cnidaria cDNA sequences is 2.2 replacements/nucleotide, which is about four times less than the average distance between Copepoda GFPs and bilaterian G2FPs (8.7 replacements/nucleotide).

Accepting the "ancient duplication" scenario implies that the common ancestor of Cnidaria and Bilateria had both G2FP and GFP proteins. Moreover, by that time the GFP lineage may have already evolved fluorescence; Copepoda GFPs follow the same structural solution to fluorescence as Cnidaria proteins, which is most parsimoniously explained by the appearance of the solution only once before Cnidaria and Bilateria separated 700–900 MYA (Benton and Ayala 2003). If the green fluorescence appeared so early, it must not have been related to bioluminescence or coloration with the purpose of producing visual effects, because the visual systems were (most probably) not yet developed to the necessary extent (Knoll and Carroll 1999; Erwin and Davidson 2002). Notably, this early origination of fluorescence implies that proteins that are members of the fluorescent GFP lineage are likely to be found in virtually any animal.

←

polytomies, with only one or two representative protein names shown. Systematic position of the host is denoted. The color class of a protein is encoded within its name: GFP = green, YFP = yellow, RFP = red, CyFP = cyan, CP = chromoprotein. Alternative protein names, if in use, are given in parentheses.
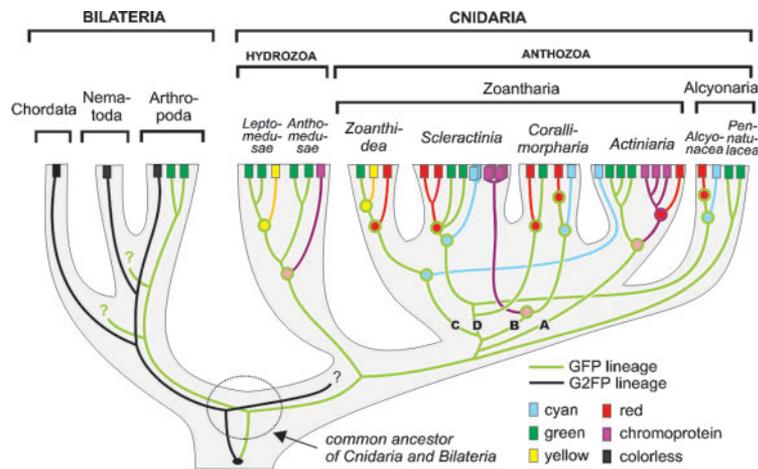
Fig. 3.—Phylogeny of GFP-like proteins (black and colored lines) overlaid upon the general taxonomy of the organisms to illustrate the extent of gene lineage sorting and inferred color diversification events. The latter are shown as circles, with the border corresponding to the parent color and the center to the originating color. See text for discussion on some alternatives in the reconstruction of color diversification events. Four major Anthozoan clades (A, B, C and D) are denoted. Gene lineages ending with question marks indicate that more extensive sampling of the genes in these taxa is required to draw a conclusion.

The only remaining alternative explanation of the presence of fluorescent GFPs in copepods is horizontal gene transfer from some yet unknown non-cnidarian source. Although current data do not warrant such an assumption, it is interesting to note that horizontal gene transfer may indeed have contributed to the distribution of GFP-like proteins, at least among Anthozoa; it seems to be the only possible way to explain the striking sequence similarity of some 60 recently described chromoproteins from Indo-Pacific corals and corallimorphs (Karan et al. 2002).

Color Evolution in Anthozoa

Several phylogeny reconstruction methods—including parsimony, neighbor-joining, and Bayesian maximum likelihood inference—suggested a position of the root for Anthozoa proteins within the branch leading to the pennatulacean GFPs, although support for this topology is not very high, except in the neighbor-joining bootstrap (fig. 2B). We tend to accept this topology as the best guess, especially because it is in agreement with the phylogeny of Anthozoa obtained using some other methods (Won, Rho, and Song 2001). The four major gene lineages of Zoantharia proteins identified previously, clades A, B, C, and D (Labas et al. 2002), were also observed in the new analysis. However, the deep relationships between these lineages remain poorly resolved (fig. 2B).

Color diversity originated independently within different lineages. We inferred the color origination events from the phylogenetic tree topology following the maximum parsimony principle. Green was assumed to be the basal ("background") color because of two considerations: First, green is the only color known in the Copepoda outgrop. Second, according to the current state of knowledge, green proteins have the simplest chromophore structure and represent a kind of "default" state; proteins of all other colors (including the cyan subtype) tend to become green as a result of random mutagenesis (Baird, Zacharias,

and Tsien 2000; Terskikh et al. 2000; Wiehler, von Hummel, and Steipe 2001; K. Lukyanov, unpublished data).

From the current data a total of 15 color diversification events can be inferred (fig. 3). Of these, there are six independent events leading to the appearance of red color, four of cyan color, three of nonfluorescent color, and two of yellow color. In most cases the parsimony principle dictates that a nongreen color evolved from green along a terminal branch of the tree. This fact may indicate the recent origin of color diversity. However, we believe that in this particular case such a pattern is mostly the result of biased sampling of taxa; to date, sampled species from each of the major clades were rarely related more closely than at the order level. Still, there are few cases when deep nongreen nodes can be inferred. Thus, within clade A, the node joining the group of chromoproteins and the red fluorescent protein equaRFP is likely to be either a chromoprotein or a red fluorescent protein (fig. 3) rather than green, because such a scenario would require only one unlikely conversion from a green to nongreen color (Labas et al. 2002) instead of two. Within clade D, as an alternative to the scenario depicted on figure 3, it is equally parsimonious to assume three ancestral nodes along the stem of the clade to be cyan instead of green, which would lead to the inference of some "backward" cyan to green color conversions in their descendants. Theoretically, an origination of green from nongreen is possible and can even be anticipated (Labas et al. 2002), because a nongreen protein may become green simply as a result of the accumulation of random mutations, requiring no color-related selection pressure to achieve the color change.

Our analysis indicates that chromoproteins may be the most ancient evolutionary invention, both in absolute time and relative to other nongreen colors. There are three known cases of independent evolution of a nonfluorescent color: in Hydrozoa and within anthozoan clades A and B (fig. 3). In Hydrozoa and clade B, separation of the

chromoprotein lineage precedes the separation of sub-orders, and in both these cases (and probably also in clade A, as discussed in the previous paragraph) chromoproteins branch off before other nongreen colors.

It would be very interesting to study the pathways of color diversity evolution by means of the reconstruction and recreation of ancestral proteins, which correspond to the nodes within the tree (Chang and Donoghue 2000). It is tempting to speculate that independent origination of color diversity of the reef Anthozoa was the result of some environmental and/or ecological factors that were experienced by all these organisms. Therefore, there is an exciting possibility that identifying such factors would provide a key to the biological significance of the colorfulness of coral reefs in general.

Nongreen Proteins: Convergent Molecular Evolution of Complexity

Red fluorescent proteins perform three consecutive autocatalytic reactions during chromophore synthesis, in contrast to just the first two required for cyans and greens, resulting in an extended chromophore structure (Gross et al. 2000; Wall, Socolich, and Ranganathan 2000; Yarbrough et al. 2001) and therefore representing a higher-complexity level of organization. The fact that red fluorescence evolved from green on several independent occasions must be considered an example of convergent evolution of complexity at the molecular level. Different structural solutions evolved in different clades, substantiating a case for true convergent evolution rather than parallelism. In clade B, the red-emitters DsRed and dis2RFP possess wide and skewed emission spectra and show barely detectable green fluorescence during maturation. Zoan2RFP from clade C and equaRFP from clade A show emission spectra similar to DsRed, but they proceed through a brightly fluorescent green stage during maturation ("fluorescent timer" pheno-type). Finally, red-emitters from clade D (tgeoRFP, mcavRFP1, rfloRFP and dendRFP [the latter was originally described as green by Labas et al. 2002]) have a peculiar narrow emission spectrum, exhibit "timer" phenotype, and require long-wave UV-A light to complete maturation (Ando et al. 2002; M. Matz, unpublished data). It has recently been demonstrated that the chromophore of one of their representatives—tgeoRFP, or Kaede—is indeed structurally different from the one in DsRed (Mizuno et al. 2003). Thus far, there are only a few documented cases of convergent molecular evolution (Zakon 2002), and to our knowledge, none of these cases can be claimed as convergent evolution of complexity. We believe that red-emitting proteins can become an excellent model for studying basic principles of the evolution of complexity at the molecular level, especially since the complex functional feature in question (red color of fluorescence) is so easily tractable in mutagenesis studies.

Yellow fluorescent proteins, of which only two are known at present, clearly represent another case of convergent evolution, because they attained the yellow fluorescent feature by different structural means: zoanYFP apparently has unique chromophore structure, whereas phiYFP relies on modification of the molecular environ-ment of a green-type chromophore. Chromoproteins, which have three independent origins, may represent another case of functional convergence, because different chromophore structures have been reported in different lineages (Marty-nov et al. 2001; Prescott et al. 2003). Another indication of convergent evolution of chromoproteins is the fact that the Hydrozoa protein anm2CP has a unique arrangement of the spectroscopically relevant residues. In wild-type cyan fluorescent proteins (in contrast to the genetically engi-neered cyan variant of GFP from *Aequorea victoria*), the chromophore is the same as in greens, so the phenotype is dependent on the chromophore's molecular environment (Gurskaya et al. 2001b). This fact makes it difficult at present to evaluate whether the structural features that are responsible for cyan color are the same or different in independent cyan lineages and, therefore, whether these lineages represent cases of parallel or convergent evolution of the cyan color.

Implications for Biotechnology

The fact that different organisms yield GFP-like proteins that may be of the same color, but based upon different structural principles, makes the study of natural GFP-like proteins of great value for biomedical imaging technology. First, there is always hope that a new natural protein would possess some very useful feature not seen before in other proteins of the same color, such as the natural timer phenotype (Labas et al. 2002) or photo-activation (Ando et al. 2002) in red fluorescent proteins originating from lineages other than the first-found DsRed (Matz et al. 1999). The second advantage lies in creating a basis for future protein engineering. Sophisticated in vivo imaging applications relying on GFP and its homologs often demand specific properties from a reporter protein of a certain color, which are not inherent to the original protein and need to be engineered, such as sensitivity to external factors, absence of oligomerization, or specific photochem-ical behavior. Proteins demonstrating different structural solutions to achieving the same color may become independent starting points in such engineering efforts, which would significantly raise the chances of success.

Conclusions

The gene tree of the GFP superfamily consists of two major lineages, one uniting all fluorescent proteins and another leading to colorless GFP-like protein-binding domains of the nidogens and related Bilaterian proteins of the extracellular matrix. Representatives of the fluorescent lineage were found not only in Cnidaria, but also in Bilateria (Arthropoda, Crustacea, Copepoda), suggesting that fluorescent GFP-like proteins evolved before separa-tion of Bilateria and Cnidaria and therefore may be more widely utilized as pigments in the animal world than previously thought. In Cnidaria, descendants of the fluorescent lineage formed a multigene family of paralogs very early in evolution, preceding the separation of Anthozoa subclasses. These paralogous lineages underwent extensive duplications and losses throughout the evolution of Anthozoa, leading to intricate sorting of the gene

lineages among extant taxonomic groups. Color diversity and the corresponding complexity of the protein organization evolved independently in several Anthozoa lineages, as well as in Hydrozoa. Apparently, the first color to separate from the ancestral green was non-fluorescent purple-blue, followed by cyans, reds and yellows. Apart from being a key to the origins of the color diversity of contemporary coral reefs, phylogeny of fluorescent GFP-like proteins may represent an excellent model for studying such theoretical issues of molecular evolution such as the origin of complex features, roles of natural selection in evolution of paralogous gene families and evolution of functions in gene families addressed by means of recreating ancestral genes.

Supplementary Material

Supplementary figure 1: Normalized excitation and emission spectra of the new GFP-like proteins from Copepoda.

Supplementary figure 2: Example of a gel filtration assay of the novel proteins.

Detailed description of methods: Cloning, biochemical and spectroscopic characterization, alignment construction and phylogenetic analysis [suppl_methods.pdf].

Alignment of cDNA sequences in NEXUS format used for phylogenetic analysis, containing a PAUP* command block [suppl_align.txt].

Hidden Markov model (HMM) profile for searching protein sequence databases for members of the GFP superfamily using HMMER software (Eddy 1998) [suppl_hmm.txt].

Acknowledgments

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.

Baird, G. S., D. A. Zacharias, and R. Y. Tsien. 2000. Biochemistry, mutagenesis, and oligomerization of DsRed, a red fluorescent protein from coral. Proc. Natl. Acad. Sci. USA **97**:11984–11989.

Benton, M. J., and F. J. Ayala. 2003. Dating the tree of life. Science **300**:1698–1700.

Bulina, M. E., D. M. Chudakov, N. N. Mudrik, and K. A. Lukyanov. 2002. Interconversion of Anthozoa GFP-like fluorescent and non-fluorescent proteins by mutagenesis. BMC Biochem **3**:7.

Burr, A. H. J., P. Hunt, D. R. Wagar, S. Dewilde, M. L. Blaxter, J. R. Vanfleteren, and L. Moens. 2000. A hemoglobin with an optical function. J. Biol. Chem. **275**:4810–4815.

Chalfie, M. 1995. Green fluorescent protein. Photochem. Photobiol. **62**:651–656.

Chang, B. S. W., and M. J. Donoghue. 2000. Recreating ancestral proteins. Trends Ecol. Evol. **15**:109–114.

Chudakov, D. M., V. V. Belousov, A. G. Zaraisky, V. V. Novoselov, D. B. Staroverov, D. B. Zorov, S. Lukyanov, and K. A. Lukyanov. 2003. Kindling fluorescent proteins for precise in vivo photolabeling. Nat. Biotechnol. **21**:452–452.

Cody, C. W., D. C. Prasher, W. M. Westler, F. G. Prendergast, and W. W. Ward. 1993. Chemical structure of the hexapeptide chromophore of the Aequorea green-fluorescent protein. Biochemistry **32**:1212–1218.

Cormack, B. P., R. H. Valdivia, and S. Falkow. 1996. FACS-optimized mutants of the green fluorescent protein (GFP). Gene **173**:33–38.

Eddy, S. R. 1998. Profile hidden Markov models. Bioinformatics **14**:755–763.

Erwin, D. H., and E. H. Davidson. 2002. The last common bilaterian ancestor. Development **129**:3021–3032.

Gilmore, A. M., A. Larkum, A. Salih, S. Itoh, Y. Shibata, C. Bena, H. Yamasaki, M. Papina, and R. Van Woesik. 2003. Simultaneous time resolution of the emission spectra of fluorescent proteins and Zooxanthellar chlorophyll in reef-building corals. Photochem. Photobiol. **77**:515–523.

Gorbunov, M. Y., and P. G. Falkowski. 2002. Photoreceptors in the cnidarian hosts allow symbiotic corals to sense blue moonlight. Limnol. Oceanogr. **47**:309–315.

Gorbunov, M. Y., Z. S. Kolber, M. P. Lesser, and P. G. Falkowski. 2001. Photosynthesis and photoprotection in symbiotic corals. Limnol. Oceanogr. **46**:75–85.

Gorokhovatsky, A. Y., N. V. Rudenko, V. V. Marchenkov, V. S. Skosyrev, M. A. Arzhanov, N. Burkhardt, M. V. Zakharov, G. V. Semisotnov, L. M. Vinokurov, and Y. B. Alakhov. 2003. Homogeneous assay for biotin based on *Aequorea victoria* bioluminescence resonance energy transfer system. Anal. Bioch. **313**:68–75.

Gross, L. A., G. S. Baird, R. C. Hoffman, K. K. Baldridge, 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, N. I. Pounkova, D. B. Staroverov, M. E. Bulina, Y. G. Yanushevich, Y. A. Labas, S. Lukyanov, and K. A. Lukyanov. 2003. Colourless green fluorescent protein homologue from the non-fluorescent hydromedusa *Aequorea coerulescens* and its fluorescent mutants. Biochem. J. **373**:403–408.

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.

Herring, P. J. 1978. Bioluminescence of invertebrates other than insects. Pp. 199–240 *in* P. J. Herring ed. Bioluminescence in action. Academic Press, London.

———. 1988. Copepod luminescence. Hydrobiologia **167/168**: 183–195.

Hopf, M., W. Gohring, A. Ries, R. Timpl, and E. Hohenester. 2001. Crystal structure and mutational analysis of a perlecan-binding fragment of nidogen-1. Nat. Struct. Biol. **8**:634–640.

Huelsenbeck, J. P., and F. Ronquist. 2001. MrBayes: Bayesian inference of phylogenetic trees. Bioinformatics **17**:754–755.

Johnson, F. H., L. C. Gershman, J. R. Waters, G. T. Reynolds, Y. Saiga, and O. Shimomura. 1962. Quantum efficiency of Cypridina luminescence, with a note that of Aequorea. J. Cell. Comp. Physiol. **60**:85–104.

Karan, M., F. Brugliera, J. Mason, E. L. Jones, S. G. Dove, O. Hoegh-Guldberg, and M. Prescott. 2002. Cell visual characteristic-modifying sequences. In patent: WO 02070703-A 179. Nufarm Australia Limited, The University of Queensland.

Kawaguti, S. 1944. On the physiology of reef corals. VI. Study of the pigments. Palao. Trop. Biol. Stn. Stud. **2**:617–674.

Kelmanson, I., and M. Matz. 2003. Molecular basis and evolutionary origins of color diversity in great star coral *Montastraea cavernosa* (Scleractinia: Faviida). Mol. Biol. Evol. **20**:1125–1133.

Kishino, H., and M. Hasegawa. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA-sequence data, and the branching order in Hominoidea. J. Mol. Evol. **29**:170–179.

Knoll, A. H., and S. B. Carroll. 1999. Early animal evolution: emerging views from comparative biology and geology. Science **284**:2129–2137.

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.

Lippincott-Schwartz, J., and G. H. Patterson. 2003. Development and use of fluorescent protein markers in living cells. Science **300**:87–91.

Lukyanov, K. A., A. F. Fradkov, N. G. Gurskaya et al. (12 co-authors). 2000. Natural animal coloration can be determined by a nonfluorescent green fluorescent protein homolog. J. Biol. Chem. **275**:25879–25882.

Mackie, G. O. 1995. Defensive strategies in planktonic coelenterates. Mar. Fresh Behav. Physiol. **26**:119–129.

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., A. F. Fradkov, Y. A. Labas, A. P. Savitsky, A. G. Zaraisky, M. L. Markelov, and S. A. Lukyanov. 1999. Fluorescent proteins from nonbioluminescent Anthozoa species. Nat. Biotechnol. **17**:969–973.

Mazel, C. H., and E. Fuchs. 2003. Contribution of fluorescence to the spectral signature and perceived color of corals. Limnol. Oceanogr. **48**:390–401.

Mazel, C. H., M. P. Lesser, M. Y. Gorbunov, T. M. Barry, J. H. Farrell, K. D. Wyman, and P. G. Falkowski. 2003. Green-fluorescent proteins in Caribbean corals. Limnol. Oceanogr. **48**:402–411.

Mizuno, H., T. K. Mal, K. I. Tong, R. Ando, T. Furuta, M. Ikura, and A. Miyawaki. 2003. Photo-induced peptide cleavage in the green-to-red conversion of a fluorescent protein. Mol. Cell **12**:1051–1058.

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.

Partridge, J. C., and M. E. Cummings. 1999. Adaptation of visual pigments to the aquatic environment. Pp. 251–283 in S. N. Archer et al., eds. Adaptive mechanisms in the ecology of vision. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Prasher, D. C., V. K. Eckenrode, W. W. Ward, F. G. Prendergast, and M. J. Cormier. 1992. Primary structure of the *Aequorea victoria* green-fluorescent protein. Gene **111**:229–233.

Prescott, M., M. Ling, T. Beddoe, A. J. Oakley, S. Dove, O. Hoegh-Guldberg, R. J. Devenish, and J. Rossjohn. 2003. The 2.2 A crystal structure of a pocilloporin pigment reveals a nonplanar chromophore conformation. Structure **11**:275–284.

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

Schmidt, H. A., K. Strimmer, M. Vingron, and A. von Haeseler. 2002. Tree-Puzzle: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics **18**:502–504.

Shimodaira, H. 2002. An approximately unbiased test of phylogenetic tree selection. Syst. Biol. **51**:492–508.

Shimodaira, H., and M. Hasegawa. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. **16**:1114–1116.

Shimomura, O., F. H. Johnson, and Y. Saiga. 1962. Extraction, purification, and properties of Aequorin, a bioluminescent protein from luminous Hydromedusan, Aequorea. J. Cell Comp. Physiol. **59**:223–239.

Swofford, D. L. 2002. PAUP. Phylogenetic Analysis Using Parismony (* and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Szent-Gyorgyi, C. S., B. J. Bryan, and W. Szczepaniak. 2001. Patent US 6232107-B.

Takabayashi, M., and O. Hoegh-Guldberg. 1995. Ecological and physiological differences between two color morphs of the coral Pocillopora-Damicornis. Marine Biol. **123**:705–714.

Terskikh, A., A. Fradkov, G. Ermakova et al. (12 co-authors). 2000. "Fluorescent timer": protein that changes color with time. Science **290**:1585–1588.

Tsien, R. Y. 1998. The green fluorescent protein. Annu. Rev. Biochem. **67**:509–544.

Tunggal, J., M. Wartenberg, M. Paulsson, and N. Smyth. 2003. Expression of the nidogen-binding site of the laminin gamma 1 chain disturbs basement membrane formation and maintenance in F9 embryoid bodies. J. Cell Sci. **116**:803–812.

Wall, M. A., M. Socolich, and R. Ranganathan. 2000. The structural basis for red fluorescence in the tetrameric GFP homolog DsRed. Nat. Struct. Biol. **7**:1133–1138.

Ward, W. W. 2002. Fluorescent proteins: who's got 'em and why? Pp. 123–126 in P. E. Stanley and L. J. Kricka, eds. Bioluminescence and chemiluminescence. World Scientific, Cambridge, UK.

Ward, W. W., and M. J. Cormier. 1978. Energy-transfer via protein-protein interaction in Renilla bioluminescence. Photochem. Photobiol. **27**:389–396.

Wiehler, J., J. von Hummel, and B. Steipe. 2001. Mutants of Discosoma red fluorescent protein with a GFP-like chromophore. FEBS Lett. **487**:384–389.

Willem, M., N. Miosge, W. Halfter, N. Smyth, I. Jannetti, E. Burghart, R. Timpl, and U. Mayer. 2002. Specific ablation of the nidogen-binding site in the laminin gamma 1 chain interferes with kidney and lung development. Development **129**:2711–2722.

Won, J. H., B. J. Rho, and J. I. Song. 2001. A phylogenetic study of the Anthozoa (phylum Cnidaria) based on morphological and molecular characters. Coral Reefs **20**:39–50.

Xia, N. S., W. X. Luo, J. Zhang, X. Y. Xie, H. J. Yang, S. W. Li, M. Chen, and M. H. Ng. 2002. Bioluminescence of *Aequorea macrodactyla*, a common jellyfish species in the East China Sea. Marine Biotechnol. **4**:155–162.

Yang, T. T., P. Sinai, G. Green, P. A. Kitts, Y. T. Chen, L. Lybarger, R. Chervenak, G. H. Patterson, D. W. Piston, and

S. R. Kain. 1998. Improved fluorescence and dual color detection with enhanced blue and green variants of the green fluorescent protein. J. Biol. Chem. **273**:8212–8216.

Yarbrough, D., R. M. Wachter, K. Kallio, M. V. Matz, and S. J. Remington. 2001. Refined crystal structure of DsRed, a red fluorescent protein from coral, at 2.0-A resolution. Proc. Natl. Acad. Sci. USA **98**:462–467.

Zakon, H. H. 2002. Convergent evolution on the molecular level. Brain Behav. Evol. **59**:250–261.