2005

Discovery, cloning, and analysis of novel fluorescent proteins from various color morphs of Corynactis californica

Christine E. Schnitzler
California State University, Monterey Bay

DISCOVERY, CLONING, AND ANALYSIS OF NOVEL FLUORESCENT PROTEINS FROM VARIOUS COLOR MORPHS OF *CORYNACTIS CALIFORNICA*

A Thesis Presented to the Faculty
of
California State University Monterey Bay
through
Moss Landing Marine Laboratories

In Partial Fulfillment
of the Requirements for the Degree
Master of Science in Marine Science

By
Christine E. Schnitzler
May 2005

APPROVED FOR THE DEPARTMENT OF
MOSS LANDING MARINE LABORATORIES

Dr. Jonathan Geller
Associate Professor of Biological Sciences
Moss Landing Marine Laboratories

Dr. Steven Haddock
Scientist
Monterey Bay Aquarium Research Institute

Dr. Kenneth Coale
Director
Moss Landing Marine Laboratories

APPROVED FOR THE UNIVERSITY

ABSTRACT

DISCOVERY, CLONING, AND ANALYSIS OF NOVEL FLUORESCENT PROTEINS FROM VARIOUS COLOR MORPHS OF *CORYNACTIS CALIFORNICA*
by Christine E. Schnitzler

Although the number of fluorescent protein (FP) genes cloned from the GFP family continues to increase, few studies of GFP-type pigments in non-bioluminescent, non-symbiotic organisms have been attempted. The first goal of this study was to locate, clone, characterize, and analyze fluorescent proteins from an organism exhibiting these traits in order to better understand their evolution and function. I successfully cloned two full-length GFP homologs by applying a FACS-based screening method to a cDNA library constructed from a temperate corallimorpharian, *Corynactis californica.* The full-length coding regions of each gene were subcloned into an expression vector and bacterial cultures were used to express the proteins. Spectral properties of purified proteins were characterized and chromophore maturation behavior was examined. Phylogenetic methods were also used to analyze the new gene sequences in relation to homologous GFP family members.

After discovering two GFP-like proteins in a single red morph, I investigated six additional morphs of *Corynactis californica,* and found indications of a variety of fluorescent pigments based on fluorescence emission spectra from live specimens. The second goal of this study was to identify and describe the variation in fluorescent pigments among morphs of *C. californica* and to relate the *in vivo* emission patterns and colors to FP genes cloned from and expressed in each morph. Specifically, I found that all morphs express a similar suite of GFP-like proteins, generated by at least three to four

genetic loci, which code for three colors: green, orange, and red. The genes exhibit tissue-specific expression patterns that differ by morph, and two major expression patterns emerged. Sequence and phylogenetic analyses comparing the new FP genes from *C. californica* to one another and to homologous members of the GFP family indicate that FP genes from this species are most closely related to one another, but that FP genes arose in an ancestor to the Anthozoa before speciation events separating anthozoan subclasses, including the Corallimorpharia. Possible ecological roles of variations in fluorescent pigmentation among morphs of *C. californica* are also discussed.

ACKNOWLEDGEMENTS

I would like to express my deep gratitude and appreciation to the faculty, staff and students at Moss Landing Marine Labs, and to the staff of the Monterey Bay Aquarium Research Institute (MBARI) for supporting me throughout my career as a graduate student. This includes committee members Jon Geller, who provided help and guidance at crucial moments, Kenneth Coale, who had a special talent for mediation, and Steve Haddock, who provided daily inspiration, encouragement, instruction, support, and friendship.

This thesis would not have been possible without the collaborative relationship established between MBARI and Maxygen Corporation. Dr. Robert Keenan taught me just about everything I know about proteins, and Rob McCord led the FACS-based screening. Claus Krebber and Mike Lassner provided valuable discussions during the early phases of the project. Members of the Bioluminescence Lab at MBARI, including Christy ("Agent H") Herren, Lynne Christianson, Jody Brewster, and Erica McSwane-Williams helped in countless ways. Karen Osborn was my phylogenetics guru and Josie Hardy was my enthusiastic intern. Members of the Invertebrate Zoology Lab at MLML provided friendship and encouragement. I would like to thank Shane Anderson (UCSB) for collecting the Santa Barbara specimens; and Lonny Lundsten, Jason Felton, John Douglas, Diana Steller, and Kim Reisenbichler for diving support. I would also like to thank the Barry, DeLong, Johnson, Kolber, and Vrijenhoek labs for use of equipment. Funding from MBARI, Maxygen, the Dr. Earl and Ethyl Myers Oceanographic and Marine Biology Trust, the Packard Foundation, and the Martha Johnston Memorial Scholarship made this research possible. Thanks to my family for providing love and support every step of the way, and to my boyfriend, Mike Gallagher, for taking care of me and everything else on a daily basis. I couldn't have done it without you!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Fluorescence is the result of an energy transfer mechanism. It occurs when a substance absorbs light energy of a certain wavelength from an external source and immediately reëmits light of lower energy and a longer wavelength. The energy difference between the excitation and emission wavelength maxima is known as the Stokes shift.

The phenomenon of fluorescence in nature is primarily attributed to the synthesis and exhibition of complex organic molecules known as pigments. A variety of fluorescent pigments are present in a diverse range of marine organisms, from mollusk shell pigments (Woodbridge 1961) to the phycobiliproteins of both red algae and cyanobacteria (Glazer 1989), but the biochemistry of the fluorescent portion of various pigment molecules (the chromophore) remains largely unknown (Tsien 1998). Chlorophyll, another fluorescent pigment common to the marine environment, is produced by photosynthesizing organisms, including marine algae, which are often found as endosymbionts in host organisms. Peak chlorophyll fluorescence occurs in the red region of the spectrum (685 nm) and extends into the infrared region to around 800 nm. The far-red emission of chlorophyll can indicate whether fluorescence originates from the host organism or its endosymbionts. Members of the green fluorescent protein (GFP) family represent yet another type of fluorescent pigment found in many corals and related organisms (Dove et al. 2001). These pigments are unique biochemically and distinct from all other fluorescent pigments; they possess a chromophore that is synthesized autocatalytically without external enzymes or substrates.

Although the first recorded observation of marine fluorescence was over three-quarters of a century ago in a sea anemone (Phillips 1927), only anecdotal information regarding fluorescence has been recorded for most organisms. This is changing, and surveys of coral fluorescence, for example, have become increasingly common (Kawaguti 1944, 1966, Logan et al. 1990, Mazel 1995, 1997, Fuchs and Mazel 1999, Fuchs 2001, Mazel et al. 2003). It is not clear how common GFP-type pigments are in the marine environment, since confirming whether fluorescence is due to a GFP-type pigment or to another pigment source is not obvious.

Studies of the GFP family began with the discovery of a green fluorescent protein from the bioluminescent hydrozoan *Aequorea victoria* (Shimomura et al. 1962). Similar green-emitting proteins were found in related bioluminescent organisms (Hastings and Morin 1969, Cormier et al. 1975), but the most significant breakthrough occurred when the gene encoding *A. victoria*'s green fluorescent protein was cloned (Prasher et al. 1992). A second breakthrough occurred when six genes homologous to GFP, with emissions at a variety of wavelengths, were cloned from five non-bioluminescent anthozoan species (Matz et al. 1999). Nearly 100 fluorescent protein (FP) genes in the GFP family with fluorescence emission maxima spread across the visible spectrum have now been cloned from 34 cnidarian species (Fradkov et al. 2000, Wiedenmann et al. 2000, Labas et al. 2002, Wiedenmann et al. 2002, Shagin et al. 2004), and four copepod species (Shagin et al. 2004). The longest-wavelength native (unmodified) GFP-type protein emits at 611 nm (Wiedenmann et al. 2002).

Our knowledge of many aspects of the GFP family is expanding rapidly. A combination of remarkable features, principally autocatalytic chromophore formation, chromophore stability, and predictable emission color, makes these proteins useful biological markers (Zhang et al. 2002, Verkhusha and Lukyanov 2004). After *A. victoria*'s GFP was cloned, novel genes were discovered, then modified (genetically engineered) to expand their range of use (Campbell et al. 2002). Proteins that become fluorescent upon green light irradiation (Lukyanov et al. 2000), change peak emission with UV irradiation (Ando et al. 2002, Wiedenmann et al. 2004), or change emission color over time (Terskikh et al. 2000) have received recent attention for their unique properties. Beyond applications in biotechnology and biomedicine, GFP-type proteins are ideal subjects for evolutionary biologists interested in color and protein evolution, and ecologists interested in the functional adaptations of fluorescence and the regulation of gene expression in specific environments.

Since a single gene sequence determines the chromophore and resulting emission color of each FP, the evolution of fluorescent pigmentation can be traced via molecular methods (Matz et al. 2002). Phylogenetic analyses have been applied to existing FP gene sequences and current evidence points to multiple origins of color diversity within the Anthozoa, indicating that different colors have evolved several times, possibly in response to external conditions (Shagin et al. 2004). Reconstruction of ancestral fluorescent proteins also suggests multiple origins of different colors (Ugalde 2004).

Anthozoan species with FPs often exhibit numerous color morphs. Two possible evolutionary mechanisms that could introduce this intraspecific variation in pigmentation are polymorphism and phenotypic plasticity (polyphenism). The distinction between

them is significant because one model (polymorphism) treats color as a genetically inherited character, determined at the moment of fertilization, while the other (polyphenism) makes color a more flexible trait, capable of responding to changing external conditions. The genetic basis of within-species pigment variation was addressed in a single study of *Montastraea cavernosa*, a reef-building coral (Kelmanson and Matz 2003). In *M. cavernosa*, it appears that variation in pigmentation among morphs is due to polyphenism. Studies of color variation in other FP-bearing species would provide evidence to support or reject this mechanism as the basis of intraspecific pigment variation. Understanding the underlying genetic basis of variation may help to explain why and how the variation arose.

Although the number of native FP genes cloned continues to increase, the adaptive functions of FPs in marine organisms remain poorly understood. Functions of fluorescent pigmentation are currently subject to much speculation (Salih et al. 2000, Ward 2002, Shagin et al. 2004, Haddock et al. 2005), but direct experimentation has focused exclusively on reef corals that host symbiotic algae (zooxanthellae). In these organisms, the hypothesized function is regulation of the light environment to benefit zooxanthellae.

Few studies of GFP-type pigments in azooxanthellate, non-bioluminescent organisms have been attempted. The goal of this study was to locate, clone, characterize, and analyze fluorescent proteins from an organism exhibiting these traits in order to better understand their evolution and function. Initial screening resulted in multiple candidate species, but ultimately, efforts focused on a temperate corallimorpharian, *Corynactis californica.*

Only two studies of pigmentation in *Corynactis californica* exist. The first distinguished its fluorescence from algal pigmentation (Limbaugh and North 1956); the second eliminated carotenoid pigments as the primary source of color variation among morphs, attributing it instead to, "an unidentified, water-soluble pigment similar to a bile-like pigment, calliactine, isolated from the sea anemone *Sagartia parasitica*" (West 1979). Results from the study detailed here demonstrate the source of fluorescent pigmentation in this species, as well as the genetic origin of variation in fluorescent pigmentation among color morphs of this species. Chapter One details the discovery, cloning, and expression of two novel GFP-type proteins from *C. californica.* Chapter Two focuses on the evolutionary basis of within-species variation in fluorescent pigmentation. This topic was addressed by studying fluorescent pigments in seven color morphs of *C. californica* on whole animal (*in vivo*) and molecular (*in vitro*) levels.

CHAPTER 1

DISCOVERY, CLONING, AND EXPRESSION OF TWO NOVEL FLUORESCENT PROTEINS FROM *CORYNACTIS CALIFORNICA*

INTRODUCTION

The overwhelming majority of organisms from which GFP-type fluorescent proteins have been discovered are cnidarians that are either bioluminescent or zooxanthellate. Species exhibiting these characters have been targeted for isolation of new FP genes because of the hypothesized functions of the fluorescent proteins in the organisms that produce them.

Functions of Fluorescent Proteins

Bioluminescence is the process by which visible light is emitted by an organism as the result of a chemical reaction. The widely accepted function of GFP-type proteins in bioluminescent organisms is that they serve to alter the wavelength of light produced in the bioluminescent reaction. Without FPs, light would emit in the blue range, but with FPs present, the emitted light is of a longer-wavelength, usually in the green range. This is accomplished through a mechanism known as radiationless energy transfer, and GFP-type proteins are often referred to as "secondary emitters" in this context. Blue-green light (440-506 nm) transmits best through seawater, and is the most common color of bioluminescence in the ocean (Haddock and Case 1999). Producing green light instead of blue is thought to be advantageous in a shallow pelagic environment, because green light is brighter (less attenuated) at shallower depths. On the other hand, blue light is transmitted more effectively than green at deeper depths (Haddock and Case 1999). The

observation that deep-water species emit bluish luminescence and shallow species emit greenish luminescence was first made by Nicol (1958) and further investigated by Haddock and Case (1999). It also has been shown that visual pigments in some shallow-living fishes have evolved to detect a broader spectrum of light, and in some deeper-living fishes, a narrower spectrum (470-490 nm) centered on blue light (Douglas et al. 1995). This demonstrates that organisms for which the detection of bioluminescence is potentially important have evolved to maximize detection abilities in specific environments.

Zooxanthellate organisms are thought to regulate the amount and wavelength of sunlight reaching their algal endosymbionts either through photoprotection or photosynthesis enhancement (photoenhancement). In photoprotection, shallow-dwelling organisms use fluorescent pigments to redirect harmful excess sunlight away from zooxanthellae (Kawaguti 1944, Salih et al. 2000, Gilmore et al. 2003). In photoenhancement, light-limited organisms use FPs to absorb short wavelengths of light, transform them to longer wavelengths more suitable for photosynthesis, and emit them to zooxanthellae (Kawaguti 1969, Schlichter et al. 1994, Salih et al. 2000).

All of these theories, however, fail to account for the presence of GFP-type pigments in non-bioluminescent, non-symbiotic organisms, or for the wide range of fluorescent colors found in nature (Shagin et al. 2004). Conspicuous fluorescent coloration displayed for interspecies communication has been discussed as an alternative hypothesis, but remains untested (Ward 2002).

Corynactis californica

After preliminary investigations of fluorescence in several temperate non-symbiotic non-bioluminescent anthozoans, we chose to study the fluorescent pigmentation of *Corynactis californica* (Corallimorpharia: Corallimorphidae). The Corallimorpharia is the smallest extant order within the anthozoan subclass Zoantharia (Hexacorallia), containing approximately 40 species distributed worldwide in tropical and temperate waters (Carlgren 1949, Dunn 1982). Morphologically, corallimorpharians are most similar to stony corals (scleractinians), although, like sea anemones (actiniarians), they lack a calcareous endoskeleton. Recent molecular analysis of mitochondrial gene order demonstrates that corallimorpharians are indeed closer to scleractinians than actiniarians (M. Medina, personal communication). Tropical corallmorpharians usually harbor zooxanthellae, and are suspension feeders that use tentacles to capture prey, though at least three species also capture prey by envelopment (Hamner and Dunn 1980). GFP-type genes have been cloned previously from two corallmorpharian genera (Table 1); both are zooxanthellate and located in tropical waters.

Corynactis californica Carlgren, 1936 (strawberry or club-tipped anemone), exhibits a wide variety of color morphs that include shades of lavender, pink, orange, brown, and red (West 1979). Tentacles are most often unpigmented, but can be white, brown, or orange (personal observation). Tentacle tips, packed with powerful stinging cells (nematocysts) used for prey capture and defense, appear white in color.

The geographic range of *Corynactis californica* extends from Washington, USA to central Baja California, Mexico, and it can be found at depths from the lower intertidal zone to at least 50 meters (Chadwick 1987). It often occurs in areas with little or no direct

sunlight, such as underneath rock ledges or in caves (Hand 1954), or completely covered by macroalgae (personal observation). *C. californica* reproduces sexually by annual spawning and asexually by longitudinal fission (Chadwick and Adams 1991, Holts and Beauchamp 1993), and it commonly occurs as large aggregations consisting of identical asexually-produced clones exhibiting the same pigmentation (Haderlie et al. 1980). Copepods, nauplius larvae, and other small animals comprise the diet of this active suspension feeder (Haderlie et al. 1980).

Discovering Novel FP Genes

Cloning novel GFP-type genes is routinely accomplished through cDNA expression library construction and screening (Ando et al. 2002, Wiedenmann et al. 2002, Shagin et al. 2004). Screening methods vary, but direct or microscopic visual inspection of bacterial colonies under blue or UV excitation is common (Ando et al. 2002, Wiedenmann et al. 2002). Although visual screening methods are effective, they are laborious and restricted by the limits of human vision or the resolution of a microscope. Cormack et al. (1996) showed the utility of fluorescence-activated cell sorting (FACS) library screening for isolating modified fluorescent proteins with improved properties. Bessette and Dougherty (2004) recently used the technique to isolate native FPs. In FACS, automated sorting devices (cytometers) measure and quantify information such as fluorescence emission wavelength and intensity from cells as they continuously pass through a beam of light. Cells exhibiting target characteristics can be isolated and collected for further analysis.

We successfully cloned two full-length GFP homologs by applying a FACS-based screening method to a cDNA library constructed from *C. californica*. The full-length coding regions of each gene were subcloned into an expression vector and bacterial cultures were used to express the proteins. Spectral properties of purified proteins were characterized and chromophore maturation behavior was examined under various pH and salt conditions. Phylogenetic methods were also used to analyze the new FP gene sequences in relation to other GFP family members.

MATERIALS AND METHODS

Collection, Initial Screening and RNA Isolation

Anthozoan specimens were collected with hand tools via SCUBA from either Monterey Bay or Santa Barbara, CA (California Scientific Collecting Permit #803069). All specimens were examined for fluorescence under a Nikon SMZ-1500 stereoscopic dissecting microscope with epifluorescent attachment (Model P-FLA) with long-pass emission filters and excitation at 410 nm, 470 nm, and 500 nm. Undissected tissue from the body wall of one polyp of *Corynactis californica* was used for RNA isolation. Total RNA was isolated using TRIzol Reagent (Invitrogen Corp., Carlsbad, CA) according to manufacturer's instruction and treated with DNase I (Invitrogen Corp.) to remove any contaminating genomic DNA (Simms et al. 1993).

cDNA Library Construction and Screening

A cDNA library was constructed using RNA isolated from *Corynactis californica* using SMART (Switching Mechanism At 5′ end of RNA Transcript) cDNA synthesis technology (Zhu et al. 2001) via a SMART cDNA synthesis kit (BD Biosciences Clontech, Palo Alto, CA) using the "cDNA synthesis by Long-Distance PCR" method according to manufacturer's instruction, including *Sfi* I digestion, and size fractionation (target size was 1 KB). The library was ligated into an expression vector (pTriplEx2, BD Biosciences Clontech) containing a *lac* promoter and ampicillin resistance, transformed in *Escherichia coli* (TG1 strain, Stratagene, La Jolla, CA), plated on agar supplemented with 50 μg/ml carbenicillin (carb) and incubated at 37°C overnight following manufacturer's instructions. A plasmid library of approximately 1 x 10^6 unique clones

was calculated. Agar plates were scraped, and bacteria were combined in a 50 ml tube and grown in LB supplemented with 50 μg/ml carb. Expression was induced by adding isopropyl ß-D-thiogalactoside (IPTG) to the culture at a final concentration of 1.0 mM. At various time points (4, 8, 16 hours), a small aliquot of culture was subjected to FACS screening with a MoFlo cytometer (DakoCytomation, Carpinteria, CA). During the screening, the following excitation and emission wavelengths were monitored: Excitation: 407 nm, Emission: 510-550 nm; Excitation: 488 nm, Emission: 510-550 nm, 560-600 nm, 615-645 nm, and 655-685 nm. A gating strategy isolating positive events from the various emission bandwidths was used to select 400 cells from a total of 8 x 10^6 events. Cells exhibiting detectable fluorescence above background were sorted onto agar plates supplemented with 50 μg/ml carb and 1.0 mM IPTG and visually checked for expression. Plasmid DNA was prepared from a total of 168 clones with a QIAprep Spin Miniprep Kit (Qiagen, Valencia, CA). These samples were then subjected to automated DNA sequencing (ABI 4500, Applied Biosystems, Foster City, CA).

Expression and Purification of Fluorescent Proteins

The sequence encoding each FP gene was amplified from plasmid DNA by polymerase chain reaction (PCR) using specific primers with 6xHis tags at their 5′ ends. Products were cloned into a low copy number, chloramphenicol resistant expression vector (pCK1; provided by Maxygen Corp.) using *Sfi I* restriction enzyme sites, transformed in *E. coli* (TG1 strain; Stratagene), and plated on agar supplemented with 50 μg/ml chloramphenicol (cam) and 1.0 mM IPTG. A strong IPTG-inducible *lac* promoter present in the pCK1 vector drove the expression of the N-terminal 6xHis tagged protein.

After growing overnight at 37°C, individual fluorescent colonies were selected from the plates and inoculated into "starter cultures" consisting of 5 ml of LB supplemented with 50 μg/ml cam and 0.2% (w/v) D-glucose (Sigma-Aldrich Corp., St. Louis, MO) and grown overnight while shaking at 37°C. The next day, LB supplemented with 50 μg/ml cam and 1.0 mM IPTG was inoculated (1:100) with the starter culture and allowed to express for 4 hours while shaking at 37°C. The remaining volume of starter culture was used for plasmid DNA preparation (QIAprep Spin Miniprep Kit; Qiagen) to provide template for DNA sequence confirmation of each construct using vector specific primers. Bacterial cultures were centrifuged at 15,000 g for 5 minutes and protein was harvested from the bacteria by lysis for 30 minutes at room temperature with B-PER Bacterial Protein Extraction Reagent (Pierce Biotechnology, Rockford, IL) supplemented with 0.5 M NaCl, 10 units of Omnicleave Endonuclease (Epicentre, Madison, WI) per OD_{600} of bacterial culture, and 600 units of Ready-Lyse Lysozyme Solution (Epicentre) per OD_{600} of culture. Lysed cells were centrifuged at 15,000 g for 5 minutes and the supernatant was purified by immobilized metal ion affinity chromatography (IMAC) through a Poly-Prep Chromatography Column (Bio-RAD, Hercules, CA) prepared with 0.5 ml bed volume of Ni-NTA resin (Qiagen) that was pre-washed twice with 10 ml wash buffer (20 mM Hepes, pH 7.5, 0.5 M NaCl). Protein was eluted from the column with 3 ml elution buffer (5 mM Hepes, pH 7.5, 0.5 M NaCl, 100 mM L-Histidine). Protein concentration was determined by measuring absorbance at 260 nm using a UV-Vis spectrophotometer (Shimadzu UV-1601).

Determination of Spectral Properties and Chromophore Maturation

To examine fluorescence emission over time, two 150 ml LB cultures of TG1 cells bearing the plasmids supplemented with 50 μg/ml cam were induced with 1.0 mM IPTG and grown in the dark for 4 hours 37°C. Protein was isolated immediately and purified by IMAC as described above. Time zero was considered the time of lysis of the cells. Fluorescence emission spectra of purified proteins were collected using a SpectraPro-150 grating spectrometer (Acton Research Corporation, Acton, MA) coupled to a Princeton Instruments MicroMAX back-illuminated CCD camera (Roper Scientific, Tucson, AZ) every 25 minutes for a period of 16 hours as the proteins matured at room temperature.

To test for the effect of light on chromophore maturation, cultures of both proteins expressed in *E. coli* were induced for six hours and grown in both full light and complete darkness. To measure pH effects on maturation, proteins were lysed and eluted as described previously. Aliquots of each protein were diluted four-fold into 100 mM phosphate solutions buffered at pH 2.5-9.5 in a 96-well format and allowed to mature overnight at room temperature. Fluorescence emission spectra were collected as soon as proteins were diluted and again after overnight maturation. To examine the effect of salt concentration on maturation, proteins were purified from bacterial cultures as before; aliquots of each protein were diluted into phosphate buffer supplemented with 100-800 mM NaCl and allowed to mature overnight at room temperature. Again, fluorescence emission spectra were collected before and after maturation.

Quantum Yield

The quantum yield of a fluorescent protein is a ratio of photons absorbed to photons emitted as fluorescence. To determine the quantum yield, fluorescence excitation and emission spectra were obtained using a Perkin-Elmer LS-50B scanning spectrofluorometer, and absorbance was measured with a Shimadzu UV-1601 UV-Vis spectrophotometer. Fluorescence quantum yields were determined according to the Parker-Reas method (Williams et al. 1983) by comparison with two different standards, rEGFP (BD Biosciences Clontech) and Fluorescein (Sigma-Aldrich Corp.). Briefly, absorbance and fluorescence were measured from five dilutions of each standard, and five dilutions of each purified protein sample. The slope of the relationship between absorbance and fluorescence is proportional to the quantum yield, so each sample was calibrated against two standards, and as a control, the standards were also compared to each other.

Oligomerization State and Purity

An aliquot of each protein (800 μg total) was loaded on a Superose 12HR 10 x 300-mm column (Pharmacia LKB Biotechnology, Uppsala, Sweden) pre-equilibrated with running buffer (20 mM Hepes, pH 7.5, 0.5 M NaCl) and analyzed by gel filtration, or size-exclusion, chromatography. Each protein was compared to rEGFP (monomer), rHcRed1 (dimer), and rDsRed2 (tetramer) protein standards (BD Biosciences Clontech). A 4-12% pseudo-native Bis-Tris SDS-PAGE gel (Invitrogen Corp.) was loaded with 8 μg per lane of each purified protein to check protein purity.

Sequence and Phylogenetic Analyses

Sequences were aligned using T-coffee (Notredame et al. 2000) using previous alignments as a guide (Carter 2004, Shagin et al. 2004) and adjusted manually according to triplet coding in MacClade version 4.07 (Maddison and Maddison 2000). All numbering is based on the original GFP from *A. victoria* (GenBank Accession No. M62653). Percentage similarity was calculated as the percentage fraction identical characters in pairwise aligned sequences, treating gaps as an additional character.

MrModeltest 3.5 (Posada and Crandall 1998) was used to choose the most appropriate molecular evolution model for Bayesian analyses. The relative fit of models was assessed by the Akaike information criterion (AIC). Smaller values of AIC are preferred, and the model was chosen according to this criterion (Akaike 1974, Posada and Crandall 2001).

Bayesian analyses were conducted using MrBayes 3.0b4 (Huelsenbeck and Ronquist 2001). Four chains of the Markov chain Monte Carlo (MCMC) method were run for each analysis. Each Markov chain, three heated and one cold, was started from a random tree and all four chains ran simultaneously for three million generations, with trees being sampled every 300 generations. Tracer v.1.0.1 (Rambaut and Drummond 2003) was used to determine the number of generations until burn-in. These generations were then discarded to ensure the Markov chain had converged on the same topology before inference from the MCMC data set was made. To verify convergence, two searches were run for three million generations and one search was run for 10 million generations. Bayesian analyses were run with an outgroup (*R. reniformis* GFP), and without an outgroup designated and compared.

Maximum parsimony (MP) analyses were conducted with the PAUP* software program version beta 4.10 (Swofford 2000). All characters were equally weighted and gaps were treated as missing data for the MP analysis. Parsimony trees of anthozoan FP genes were reconstructed using the heuristic search option with the tree bisection-reconnection (TBR) branch swapping algorithm and 1,000 random-sequence addition replicates. 1,000 bootstrap replicates were added by using the same criteria as in the original search. All MP searches were unrooted. FP genes from *Renilla muelleri* (rmueGFP; GenBank Accession No. AY015996), *R. reniformis* (rrenGFP; GenBank Accession No. AF372525) and *Ptilosarcus* sp. (ptilGFP; GenBank Accession No. AY015995) were used post-analysis to root the resulting trees.

RESULTS

Collection, Initial Screening and RNA Isolation

Non-bioluminescent, non-symbiotic temperate anthozoan species were targeted for fluorescence. *In vivo* fluorescence emission spectra were recorded from six species (Table 2). RNA was isolated and a cDNA library was constructed from a single red *Corynactis californica* polyp.

cDNA Library Screening

Using a FACS-based library screening approach, two new full-length GFP homologs were identified: one yellow, designated ccalYFP1 (Figures 1a and 1b; GenBank Accession No. AY823227); one red, designated ccalRFP1 (Figures 2a and 2b; GenBank Accession No. AY823226). Both genes are 666 base pairs and 221 amino acids in length. Sequence-based calculations estimated a molecular weight of 28.9 kilodaltons (kDa) for ccalYFP1, and 28.8 kDa for ccalRFP1. The FACS-based screening method proved to be quite effective. If available, this method may be used as an alternative to conventional visual screening methods when screening cDNA libraries for novel FPs.

Determination of Spectral Properties, Chromophore Maturation, and Quantum Yield

The yellow protein, ccalYFP1, has an excitation maximum of 513 nm, emission maximum of 524 nm (Figure 3a), small Stokes shift (11 nm), and a quantum yield of 0.77 (Table 3). The chromophore of ccalYFP1 matures immediately (T< 2 hours) in darkness or full light. Variations in pH and salt concentration had little effect on the emission

maximum of ccalYFP1, beyond extinguishing fluorescence at pH 5.6 (data not shown). On the gel filtration column, the protein migrated as a dimer.

The red protein, ccalRFP1, has an excitation maximum of 560 nm, highly red-shifted emission maximum of 600 nm (Figure 3b), relatively large Stokes shift (40 nm), and quantum yield of 0.46 (Table 3). The chromophore of ccalRFP1 reaches full maturation after approximately 16 hours in darkness or full light at room temperature by passing through a bright green intermediate stage (518 nm; Figure 4). Gel filtration indicates that, like most native red fluorescent proteins, ccalRFP1 is tetrameric.

The emission maximum of ccalRFP1 is affected by variations in pH during chromophore maturation (Figures 5a and 5b). High pH (9.8) results in the brightest emission, but the protein does not fully mature from green to red over time. Slightly lower pH (8.6) gives an intermediate result, with a transition from 518 nm to an emission maximum around 570 nm. Mid-range pH (7.4 to 5.6) results in a transition from 518 nm to 600 nm over time, although there is still some emission at 518 nm. At pH below 5.6, fluorescence is extinguished. Varying salt concentration (100-800 mM NaCl) had no significant effect on the overnight maturation and final emission color of ccalRFP1; most of the protein fully matured to a 600 nm emission maximum by day 2, regardless of salt concentration (Figures 6a and 6b).

Sequence and Phylogenetic Analyses

The chromophore of GFP-type proteins forms from a tripeptide that is encoded by the residues X-Tyr-Gly (positions 65-66-67 according to original GFP numbering), where X varies among FPs. Protein crystal structure (Ormö et al. 1996, Gross et al. 2000, Wall

et al. 2000, Yarbrough et al. 2001, Prescott et al. 2003, Remington et al. 2005, Wilmann et al. 2005) and mutagenesis studies (Li et al. 1997, Gurskaya et al. 2001, Bulina et al. 2002, Terskikh et al. 2002, Yanushevich et al. 2002, Shagin et al. 2004) have confirmed the chromophore sequence and identified additional residues important for fluorescence for many GFP-type proteins.

An amino acid sequence alignment of ccalRFP1 and ccalYFP1 with homologous FPs shows regions of identity among all proteins considered and highlights the variability of the first codon in the inferred chromophore tripeptide (Figure 7). Based on the sequence alignments, the two new FPs possess 86% nucleotide and 80% amino acid similarity to one another (Table 4). They are less than 30% similar to GFP (Prasher et al. 1992) on the amino acid level. Their next closest neighbor among published sequences is *Anemonia majano* GFP (amajGFP; Matz et al. 1999), with which they share 61-64% amino acid similarity (Table 4).

The GTR+I+Γ model, which assumes general time reversibility (GTR), a certain proportion of invariable sites (I), and a gamma approximation of the rate-variation among sites (Γ), represented the optimal evolutionary model for Bayesian analysis. Of the three million generations, 2,000 were considered the "burn-in" period and discarded. All searches converged on similar parameter estimates.

Bayesian analysis reconstructed a gene tree with four distinct clades (Figure 8). Table 5 provides details about ingroup FP genes used in both analyses. Clade 1 is most distant from the *C. californica* FPs and includes one red protein, equaRFP, from a tropical anemone. Clade 2 includes red proteins from a tropical corallimorpharian genus, *Discosoma* (discRFP and dis2RFP). The two FPs from *C. californica* form a well-

supported subclade nested within Clade 3. Their sister group is a subclade of proteins from *Anemonea majano* (amajCyFP and amajGFP) and *Astrangia lajollaensis* (alajGFP). The next closest neighbors to ccalRFP1 and ccalYFP1 in Clade 3 include FPs from *Zoanthus* sp. (zoan2RFP, zoanYFP, zoanGFP). Clade 4 includes red proteins from three tropical scleractinian corals (tgeoRFP, mcavRFP, lhemRFP), and one tropical corallimorpharian (rfloRFP).

MP analysis reconstructed consensus trees with the same topology as the Bayesian tree. Overall, the MP analysis showed low bootstrap support for some nodes, whereas the Bayesian analysis had relatively high posterior probability values for nearly all nodes (Figure 9). However, posterior probability values are known to be considerably less stringent than bootstrap values (Suzuki et al. 2002, Alfaro et al. 2003) and should be considered upper limits of confidence for the relationships depicted at these nodes.

DISCUSSION

The discovery of two novel GFP-type proteins from *Corynactis californica* dramatically updates our understanding of its pigments. In their purified form, the proteins have remarkable features which distinguish them from other native FPs. Further study will shed light on which residues influence emission color, and what factors affect chromophore maturation, especially in the red protein. Although FPs have previously been isolated from tropical corallimorpharian species, this marks the first discovery of fluorescent proteins in a temperate non-symbiotic corallimorpharian. The presence of FPs in non-bioluminescent, non-symbiotic organisms points towards functions, such as interspecies communication, that have not yet been thoroughly investigated.

Protein Purification and Characterization

Protein can easily be purified from any vector that contains a peptide tag comprised of multiple histidine (His) residues by immobilized metal ion affinity chromatography (IMAC). The affinity of histidine's imidazole side chains for metal ions such as nickel is the key to this procedure. Although IMAC is the most common purification method for GFP-type proteins, certain drawbacks exist. The His tag, and its location on the construct (either N- or C-terminal) for example, may have an effect on the maturation and folding of FPs (Yakhnin et al. 1998). Recent data suggests that certain physical characteristics of novel FPs should not be measured from proteins purified in this manner (McRae et al. 2005).

Biochemistry

Before the discovery of ccalYFP1, only two yellow fluorescent proteins had been isolated, one from a zoanthid (zoanYFP; Matz et al. 1999); one from a hydrozoan (phiYFP; Shagin et al. 2004). Each has a unique chromophore structure, indicating an independent, convergent evolution of yellow color (Shagin et al. 2004). The chromophore of ccalYFP1 (positions 65-67) forms from the residues Cys-Tyr-Gly (Figure 7), representing a chromophore found only in two green FPs from a temperate stony coral (*A. lajollaensis*; Bessette and Daugherty 2004). Isoleucine is found in position 203 in ccalYFP1 instead of tyrosine, which holds this position in commercial YFP (Ormö et al. 1996) and phiYFP from a hydrozoan, and is known to be the key residue for producing yellow emission in commercial YFP. The yellow emission of ccalYFP1 must derive from other residues.

The synthesis of a red-emitting chromophore is achieved by the completion of three consecutive autocatalytic reactions, which results in an extended chromophore structure. This can be compared to green- or cyan-emitting chromophores, which only require two consecutive reactions (Gross et al. 2000, Wall et al. 2000, Yarbrough et al. 2001). Red fluorescent proteins have been classified into three separate phenotypes according to their chromophore structure (Shagin et al. 2004). The "DsRed" phenotype has a wide emission spectrum and barely detectable green fluorescence during maturation (e.g. DsRed and dis2RFP, Matz et al. 1999). The "fluorescent timer" phenotype has emission similar to DsRed, but proceeds through a bright green stage during maturation (e.g. zoan2RFP, Labas et al. 2002; equaRFP, Wiedenmann et al. 2002). The "UV-A requiring fluorescent timer" phenotype has a narrow emission spectrum, the timer

phenotype, and a need for long-wave UV-A light to mature (e.g. tgeoRFP, Ando et al. 2002; mcavRFP1, rfloRFP, and dendRFP, Shagin et al. 2004; lhemRFP, Wiedenmann et al. 2004).

The chromophore of ccalRFP1 is formed by Thr-Tyr-Gly (Figure 7), which is also found in phiYFP and EGFP, a modified version of GFP (Cormack et al. 1996, Yang et al. 1998). In GFP, the substitution of threonine for wild-type serine at position 65 results in a dramatic shift in the shape of the absorption spectrum (Shagin et al. 2004). The chromophore of ccalRFP1 clearly exhibits the "fluorescent timer" phenotype, reaching full maturation over a period of 16 hours. This is much faster than the original red fluorescent protein - DsRed, which requires 48 hours to reach 90% maximal fluorescence (Baird et al. 2000), but slightly slower than equaRFP, which reaches 90% maximal fluorescence after only 12 hours (Wiedenmann et al. 2002). Recent DsRed mutants have been engineered for faster maturation, including two variants that fully mature in less than 2 hours (Bevis and Glick 2002, Campbell et al. 2002).

Sequence and Phylogenetic Analyses

In analyses from other studies, fluorescent proteins of various emission colors from the same species do not always occur next to one another on reconstructed gene trees (Shagin et al. 2004). The two fluorescent proteins from *C. californica*, however, form a monophyletic group, indicating that their FPs evolved more recently than the speciation event that separated them from their sister group - proteins from *Anemonea majano*, an Indo-Pacific sea anemone (amajCyFP and amajGFP), and *Astrangia lajollaensis*, a temperate colonial cup coral (alajGFP). The two yellow anthozoan

fluorescent proteins (zoanYFP and ccalYFP1) both fall within Clade 3, indicating that yellow emission may have evolved in their common ancestor. Two proteins that exhibit the "fluorescent timer" chromophore phenotype, ccalRFP1 and zoan2RFP, occur together in Clade 3, though another fluorescent timer (equaRFP) occurs in a separate clade (Clade 1), indicating that this phenotype could have arisen multiple times through evolutionary history resulting in convergent evolution. Four red proteins that exhibit the "UV-A requiring fluorescent timer" phenotype occur together in Clade 4, indicating that a common ancestor to this group may have produced this chromophore phenotype. FP genes from three corallimorpharian genera (*Corynactis, Discosoma,* and *Ricordea*) are found in three different clades, suggesting that FP genes evolved prior to any speciation events separating corallimorpharians from stony corals. Tree topologies reconstructed from MP and Bayesian analyses generally agree with those found by Shagin et al. (2004).

Biological Significance

FPs may be involved in internal physiological processes and fluorescent coloration may be a nonessential byproduct (Wicksten 1989). Some scientists argue, however, that in competitive environments, color patterns are very important for the survival of those organisms that produce them (Watt 1990). Fluorescence may also have evolved sometime in the past for conditions, such as an endosymbiotic relationship, no longer present (Wicksten 1989). It is not clear, however, why an organism would continue to expend energy to synthesize and display complex fluorescent molecules for a condition no longer present. When investigating possible biological functions of

fluorescent pigments, careful consideration of the specific organism, its physiology, its interspecies interactions, and its habitat should be attained before conclusions are drawn.

The discovery of FPs of many colors in non-bioluminescent non-symbiotic anthozoans raises the possibility for a variety of functions for fluorescence in the marine environment. Fluorescence may provide a visual contrast that functions as a signal in interspecies relationships (Ward 2002). Conspicuous coloration as a form of communication has been thoroughly investigated in terrestrial and aquatic environments (Cott 1957). Colors may signal to predators or competitors that an organism is toxic or dangerous, or colors may attract prey or mates. Prey attraction via fluorescent "lures" has been recently suggested for a species of deep-sea siphonophore (*Erenna* sp. A; Haddock et al. 2005). Fluorescent signaling towards potential predators and competing males of the same species has been documented in *Lysiosquillina glabriuscula*, a mantis shrimp (Mazel et al. 2004). In another example, a species of Australian parrot, *Melopsittacus undulates*, apparently uses its fluorescent plumage to attract the attention of the opposite sex (Arnold et al. 2002).

Since *Corynactis californica* lacks endosymbionts and often occurs in light-limited environments, the function of fluorescence is most likely unrelated to light regulation. Although *C. californica* lacks any specialized visual system, it is possible that it uses blue downwelling sunlight to excite its fluorescent pigments as a signal for interspecies communication. Known predators of *C. californica* (*Calliostoma annulatum*, *Latiaxis oldroydi*, *Dermasterias imbricata,* and *Pisaster giganteus*) have limited visual ability and principally rely on chemical cues to search for prey, so they likely would not respond to fluorescent signals. Similarly, known competitors (*Balanophyllia elegans*,

Astrangia lajollaensis, *Paracyathus stearnsii*) do not have the ability to detect visual signals. Other species including crustaceans and fish, however, may possess the visual systems and pigments required to detect fluorescence. Recent work has shown that interspecies variation in visual pigments correlates significantly with microhabitat preference among species of surfperches living at similar depths in Monterey Bay (Cummings and Partridge 2001). Data presented in that study also demonstrate that surfperches possess visual pigments that could detect a wide range of fluorescent signals. An interesting future study could explore correlations between the visual pigments of target species and fluorescent signals from organisms such as *Corynactis californica.*

Conclusions

FACS-based library screening resulted in the cloning of two novel FPs from a single *Corynactis californica* polyp, one yellow (ccalYFP1), and one red (ccalRFP1). The yellow protein is only the third native yellow fluorescent protein known from the GFP family. At 600 nm, the red protein has the second-longest emission maximum of all native FPs. Exhibiting the "fluorescent timer" phenotype by passing through a green intermediate, ccalRFP1 reaches full maturation over a period of 16 hours. The discovery of fluorescent proteins in a non-bioluminescent non-symbiotic anthozoan organism highlights possibilities of adaptive significance of fluorescence unrelated to light regulation. The patterns and colors of fluorescence in *C. californica* and similar species may hold meaning for organisms that possess the visual pigments to distinguish them.

CHAPTER 2

A STUDY OF THE INTRASPECIFIC VARIATION IN FLUORESCENT PIGMENTATION AMONG MORPHS OF *CORYNACTIS CALIFORNICA*

INTRODUCTION

Organisms can respond to changing external variables by altering their morphology or behavior (Meroz-Fine et al. 2003, Relyea 2004). Morphological responses in marine invertebrates include changing the color or pattern of pigmentation or adjusting pigmentation levels (Wicksten 1989, Tupen 1999). Pigment modification can be achieved through evolutionary time by developing inherited genetic differences (polymorphisms) through mutation or natural selection, or within an organism's lifetime by exhibiting phenotypic plasticity in pigment expression in response to environmental factors (Hansson 2004). Differences in pigmentation among morphs of the same species can be attributed to patterns of inheritance and expression encoded by multiple loci or to variations in sequence encoded by different alleles at the same locus (Kelmanson and Matz 2003). As studies of the diversity of color patterns on butterfly wings have shown (Brakefield and French 1999), the origin of differences in pattern or color of pigmentation can be quite complex and may evolve from some combination of both genes and environment. In many cases, factors influencing pigmentation differences among morphs of the same species are unknown, and even basic analyses of the pigments involved have not been attempted (Wicksten 1989).

Recent studies have begun to show that pigmentation in many anthozoan species is due to proteins from the GFP family (Matz et al. 1999, Dove et al. 2001, Labas et al.

2002, Kelmanson and Matz 2003, Bessette and Daugherty 2004, Wiedenmann 2004a). The synthesis of the chromophore of these proteins is unique among pigment molecules because it is achieved autocatalytically without substrates or cofactors. This enables phylogenetic studies of color evolution when sequences are known because each protein sequence encodes a single color, and emission colors of individual proteins can be determined by expression in bacterial systems (Matz et al. 2002).

Two novel FPs, one yellow (ccalYFP1, emission max. 524 nm, GenBank Accession No. AY823227), and one red (ccalRFP1, emission max. 600 nm, Accession No. AY823226) were cloned from a single polyp of a red *Corynactis californica* morph collected from Monterey Bay, CA (see Chapter 1). This species exhibits a wide variety of color morphs, and after these proteins were discovered, six additional color morphs of *C. californica* were collected from various depths, their fluorescence emission spectra were investigated, and evidence of a variety of fluorescent pigments and patterns was found. This raised the question of how genes from the GFP family are involved in the intraspecific variation in pigmentation among color morphs of *C. californica.* The purpose of this study was to identify and describe the variation in fluorescent pigments among morphs of *Corynactis californica* and to relate the *in vivo* emission patterns and colors to FP genes cloned from and expressed in each morph.

A molecular approach was employed to connect FP gene expression with *in vivo* emission spectra recorded from live polyps of *C. californica.* Based on preliminary investigations and a related study of a tropical coral (Kelmanson and Matz 2003), we hypothesized that fluorescent pigmentation is a multi-locus trait in *C. californica*, and that intraspecific variation in pigments among morphs is due to tissue-specific expression

patterns. The first goal of our molecular work was to determine full-length sequences of additional genes coding for FPs in each of seven color morphs. Another was to clone, express, and purify any new proteins and record their emission spectra for the purpose of understanding how individual proteins contribute to the overall coloration patterns of different morphs. The next goal was to determine which FP genes were expressed in each color morph investigated. Sequence and phylogenetic analyses were used to compare the new FP genes from *C. californica* to one another and to other members of the GFP family to better understand their evolution. The ecological role of variations in fluorescent pigmentation among sympatric morphs of *C. californica* is also discussed.

MATERIALS AND METHODS

Collection and *In Vivo* Fluorescence Emission Spectra Measurements

Multiple polyps from individual clones of seven distinct color morphs (designated Lavender, Pink1, Pink2, Red, Orange1, Orange2, and Orange3) of *C. californica* were collected with hand tools, using SCUBA for subtidal specimens, from three locations in Monterey Bay, CA (California Scientific Collecting Permit 803069; Table 6) and examined for fluorescence under a Nikon SMZ-1500 epifluorescence microscope (described in Chapter 1). Specimens were kept alive in a 14.5°C running seawater lab at the Monterey Bay Aquarium Research Institute (MBARI). A Nikon Coolpix 5000 digital camera attached to the microscope was used to document images of whole animal (*in vivo*) fluorescence. All fluorescence images were taken under blue light (470 nm) excitation. Fluorescence emission spectra of live specimens were collected with a USB-2000 fiber-optic spectrometer (Ocean Optics, Dunedin, FL) under blue light excitation. Tissue-specific emission spectra from five tissue types (mouth, oral disc, base of tentacles, tentacles, and body column) were measured by focusing the microscope on a single type of tissue of an individual polyp, then holding the fiber optic up to the eyepiece of the microscope to record the spectrum through the microscope.

RNA Isolation and Degenerate RT-PCR to Determine Additional Novel FP Genes

Total RNA was isolated from single whole polyps of each of the seven color morphs (used for spectral measurement) with a RNeasy Mini Kit (Qiagen) according to manufacturer's instruction, treated with DNase I (Invitrogen), and stored at -80°C until needed. cDNA was synthesized from total RNA via reverse transcription with the

SuperScript III First-Strand Synthesis System for RT-PCR (Invitrogen). All control reactions suggested by the manufacturer were processed with experimental samples. For the cDNA synthesis, first an RNA/primer mixture consisting of 5 μl RNA, 5 μM oligo$(dT)_{20}$, and 1 mM dNTPs was incubated at 65°C for 5 minutes. Then, a cDNA synthesis mix consisting of 1x RT buffer, 5 mM $MgCl_2$, 10 mM dithiothreitol (DTT), 40 units (U) RNase OUT (Invitrogen), and 200 U SuperScript III reverse transcriptase (RT) was added to the RNA/primer mixture and incubated at 50°C for 50 minutes. The cDNA synthesis reaction was terminated by incubating at 85°C for 5 minutes. The RNA template was removed via RNase digestion by adding 2 U of *E. coli* RNase H and incubating at 37°C for 20 minutes. cDNA was stored at -20°C until needed for PCR.

A set of degenerate primers [some positions of the primer have more than one possible nucleotide base (Kwok et al. 1994)] based on the ccalYFP1 and ccalRFP1 sequences (ccal48F/ccal647R; Table 7A) was designed to attempt to amplify additional fragments of genes coding for FPs in this species. Primers were designed to amplify the internal portion of the gene and two of 47 nucleotides in this primer pair were degenerate (Table 7A).

Degenerate PCR reactions for each color morph cDNA consisted of 2 μl cDNA, 2 U Platinum *Taq* polymerase (Invitrogen), 1x Platinum *Taq* buffer, 0.2 μM of each degenerate primer, and 200 μM dNTPs. Cycling parameters were: 94°C for 2 minutes, 35 cycles of 94°C for 15 seconds, T(°C) for 30 seconds, and 72°C for 1 minute, followed by 72°C for 5 minutes. Annealing temperatures [T(°C)] for all PCR reactions are listed throughout Table 7. Resulting PCR products were purified using the QIAQuick Gel Extraction Kit (Qiagen), ligated into a TA cloning vector (pCR2.1-TOPO, Invitrogen),

transformed into chemically competent *Escherichia coli* (Top 10 strain, Invitrogen), and plated on agar supplemented with kanamycin (50 μg/ml) and 5-bromo-4-chloro-3-indolyl-ß-D-galactopyranoside (X-Gal, 40 μg/ml) for blue-white selection following manufacturer's instructions. Ten white colonies were chosen from each of seven plates to inoculate overnight cultures for plasmid DNA purification using a QIAprep Spin Miniprep Kit (Qiagen). Purified plasmid DNA was digested with EcoRI (Promega, Madison, WI) for one hour at 37°C and separated by gel electrophoresis to check for correct-size inserts. DNA was sequenced in both directions using M13 primers directed to the cloning vector (Table 7I) and the BigDye Terminator version 3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA) with an automated sequencer (ABI 3100, Applied Biosystems). Sequences were assembled and edited with Sequencher version 4.2.2 (Gene Codes Corp., Ann Arbor, MI) and aligned by eye in MacClade version 4.07 (Maddison and Maddison 2000). Two unique cDNA fragments (named ccalFP-A and ccalFP-B) with sequence homology to ccalYFP1 and ccalRFP1 resulted from this approach and were selected for further study.

A second set of degenerate primers based on the ccalYFP1, ccalRFP1, ccalFP-A, and ccalFP-B sequences (ccal1F/ccal666R; Table 7A) was designed after the full-length coding regions of the two new genes were determined (see Ends Cloning below). These primers were designed to amplify the internal portion of the gene, and 13 of 48 nucleotides in this primer pair were degenerate (Table 7A). Degenerate PCR with this set of primers was carried out on cDNA previously synthesized from all seven color morphs. PCR reactions and conditions were the same as the first round of degenerate PCR, except 2 U TaqPlus Precision polymerase and 1x TaqPlus Precision buffer (Stratagene) were

used. All cloning, transformation, sequencing, and analysis steps were the same as before. Twenty-four clones were chosen from the seven plates. An additional cDNA fragment (named ccalFP-C) homologous to other FP genes from *Corynactis californica* was cloned from this second set of degenerate primers.

Determining cDNA Ends

To determine the sequence of the 5′ and 3′ ends of the three new gene fragments, two approaches were taken. In the first, rapid amplification of cDNA ends (RACE) was applied (Frohman et al. 1988) using SMART cDNA synthesis technology (Zhu et al. 2001) via a SMART RACE cDNA Amplification Kit (BD Biosciences Clontech) according to manufacturer's instruction. SMART technology provides a mechanism for producing full-length cDNAs. This is based on the ability of PowerScript RT to add a dC-rich tail to the 3′ end of first-strand cDNA, and the ability of the BD SMART oligo's dG-rich residues to anneal to the dC-rich tail and serve as an extended template for reverse transcription.

"RACE-Ready" cDNA was synthesized from the total RNA isolated previously from each color morph via reverse transcription using BD PowerScript RT and RACE primers (Table 7B) provided with the SMART RACE kit. To prepare 5′-RACE-Ready cDNA, 3 μl total RNA was combined with 1 μl 5′-CDS primer and 1 μl BD SMART II A oligo. To prepare 3′-RACE-Ready cDNA, 3 μl total RNA was combined with 1 μl 3′-CDS primer A. Both reactions were incubated at 70°C for 2 minutes and transferred to ice for 2 minutes. A mixture of 1x First-Strand buffer, 2 mM DTT, 1 mM dNTPs, and 1 μl BD PowerScript RT was added to each tube and incubated at 42°C for 1.5 hours. First-

strand reactions were diluted with Tricine-EDTA buffer per manufacturer instruction and incubated at 72°C for 7 minutes to terminate the RT reaction. RACE-Ready cDNA was stored at -20°C until needed.

A 5′ and 3′ gene-specific RACE primer was designed for each novel sequence (Table 7C), and used at a concentration of 0.2 μM in PCR reactions with 1x Universal Primer A Mix, 200 μM dNTPs, 1x BD Advantage 2 PCR buffer, 1x BD Advantage 2 Polymerase Mix and 2.5 μl of either 5′- or 3′-RACE-Ready cDNA. All control reactions recommended by the manufacturer were carried out simultaneously. Cycling parameters were as follows: 5 cycles of 94°C for 30 seconds and 72°C for 2 minutes; followed by 5 cycles of 94°C for 30 seconds, 70°C for 30 seconds, 72°C for 2 minutes; followed by 25 cycles of 94°C for 30 seconds, 68°C for 30 seconds, and 72°C for 2 minutes. Products were visualized by gel electrophoresis, and additional cycles of 94°C for 30 seconds, 68°C for 30 seconds and 72°C for 2 minutes were added, five at a time, until a correct-sized band was visible on the agarose gel. PCR products were purified, cloned into pCR2.1-TOPO, transformed, sequenced, and analyzed as described for the degenerate PCR. RACE-generated sequences were aligned with the novel gene fragments and in cases where overlapping regions matched exactly and new information was produced for either end of the gene fragment, new sequences were assembled in MacClade version 4.07.

Where the RACE approach failed, a non-coding region degenerate PCR strategy was employed. This involved utilizing non-coding region sequence directly adjacent to the coding region of ccalYFP1 and ccalRFP1. The SMART cDNA library generated from *Corynactis californica* (described in Ch. 1) produced large-insert clones whose sequences

extended beyond the coding region of each full-length FP. Some homology was found in these regions between the ccalYFP1 and ccalRFP1 sequences. This "extra" information was utilized to design a 5′- and a 3′-degenerate primer based on these non-coding regions (Table 7D). The 5′-degenerate primer (OUT-5′cory) had six degenerate nucleotides out of 24, and the 3′-degenerate primer (OUT-3′cory) had eight degenerate positions out of 24. Gene-specific primers for each new fragment (Table 7E) were paired with either the 5′- or the 3′-degenerate non-coding region primer in PCR. Reactions consisted of 2 μl cDNA generated with the SuperScript III System (described above), 2 U TaqPlus Precision polymerase (Stratagene), 1x TaqPlus Precision buffer, 0.2 μM of each primer, and 200 μM dNTPs. Cycling parameters were: 94°C for 2 minutes, 35 cycles of 94°C for 15 seconds, T(°C) for 30 seconds, 72°C for 1 minute, followed by 72°C for 5 minutes. Products were purified, cloned into the pCR2.1-TOPO vector, transformed, sequenced, and analyzed as described before. When aligned, overlapping regions of ends sequences and the FP gene fragments had to agree with the new fragments exactly to be considered a match, and full-length sequences were assembled in MacClade v. 4.07.

Full Length Cloning

The complete coding regions of all five FP genes from *Corynactis californica* were amplified via PCR from a single color morph cDNA using gene-specific primers designed for each FP gene (Table 7F) and a high-fidelity, proofreading polymerase. PCR reactions consisted of 2 μl cDNA, 2.5 U PfuUltra Hotstart polymerase (Stratagene), 1x PfuUltra HF buffer, 0.2 μM each gene-specific primer, and 800 μM dNTPs. Cycling parameters were: 95°C for 2 minutes, 35 cycles of 95°C for 30 seconds, T(°C) for 30

seconds, 72°C for 1 minute, followed by 72°C for 10 minutes. PfuUltra Hotstart polymerase does not leave A-overhangs for TA cloning, so PCR products were A-tailed by adding 1x *Taq* PCR Master Mix (Qiagen) and incubating at 72°C for 10 minutes. These products were cloned into the pCR2.1-TOPO vector, and transformed as described for degenerate PCR. Multiple clones of each gene were sequenced with M13 primers, and plasmids whose inserts were exact matches to the assembled full-length genes were used as templates for PCR to amplify full-length products to be ligated into an expression vector.

Gene-specific primers for amplification of full-length products (Table 7G) were designed with restriction enzyme sites engineered on either end to produce products suitable for restriction enzyme cloning into an expression vector, and with appropriate spacing to ensure that the resulting clones would be in-frame with the start codon and a 6xHis tag in the expression vector. Two expression vectors used in this study, pQE31XN-TEV.T3 (pQE31X-TEV) and pQE31XN-DsRed.T3 (pQE31X-non-TEV), modified from commercially available pQE-31 vectors (Qiagen) to contain NcoI and NotI sites in the multiple cloning site, were generously provided by Dr. B. Glick (U. of Chicago; Table 8). PCR consisted of 100 ng full-length plasmid DNA, 2.5 U PfuUltra Hotstart polymerase, 1x PfuUltra HF buffer, 0.2 μM each gene-specific expression primer, and 800 μM dNTPs. Cycling parameters were: 95°C for 2 minutes followed by 35 cycles of 95°C for 30 seconds, T(°C) for 30 seconds, and 72°C for 1 minute, followed by 72°C for 10 minutes.

Full-length PCR products were purified using the QIAQuick Gel Extraction Kit (Qiagen), and double digested with 25 U NcoI [New England Biolabs (NEB), Beverly,

MA] and 25 U NotI (NEB), 1x NEBuffer 3, and 1x bovine serum albumin (BSA) for 16 hours at 37°C. Digested DNA was purified using a GENECLEAN Kit (Qbiogene, Irvine, CA), quantified on an agarose gel, and ligated into an expression vector. Expression vectors were prepared by double digestion with 30 U NcoI and 30 U NotI, 1x NEBuffer 3, and 1x BSA for 16 hours at 37°C, then treated with Antarctic Phosphatase (NEB) for 1 hour at 37°C to remove 5′ phosphate groups from DNA to prevent self-ligation. Vector DNA was then purified and quantified in the same way as insert DNA. Ligation reactions consisted of 400 U T4 DNA ligase (NEB), 1x T4 DNA Ligase Reaction Buffer, and a 1:3 ratio of vector to insert DNA, and were carried out at room temperature for 10 minutes. Ligated DNA was transformed into *E. coli* (XL1-Blue strain, Stratagene) and transformants were plated on agar supplemented with 100 μg/ml ampicillin and 1.0 mM IPTG. Plates grew overnight at 37°C and were transferred to 4°C to develop. Fluorescent colonies were chosen under blue light excitation while visualizing them through the epifluorescence microscope. Individual colonies were inoculated into "starter cultures" of 2.0 ml LB supplemented with 100 μg/ml ampicillin and 0.2% (w/v) D-glucose to repress expression and grown overnight at 37°C. These cultures were used for three different purposes: (1) for plasmid DNA purification (QIAprep Spin Miniprep Kit, Qiagen) and sequence confirmation using pQE vector primers (Table 7I) (2) to inoculate expression cultures (see Expressing FPs in *E. coli* below) and (3) to inoculate overnight growth in 10% glycerol-LB for storage at -80°C.

Expressing FPs in *E. coli* and Protein Purification

Cultures for expression were inoculated 1:100 into LB supplemented with 100 μg/ml ampicillin and 1.0 mM IPTG and grown shaking (250rpm) at 30°C overnight. Bacterial cultures were centrifuged at 18,000 g for 10 minutes at 4°C. Protein was harvested from *E. coli* by lysis with B-PER buffer (Pierce Biotechnology) supplemented with 0.5 M NaCl, endonuclease, and lysozyme (see Chapter 1 for details) for 30 min on ice. Lysed bacteria were centrifuged at 30,000 g for 10 minutes at 4°C. Cleared lysate was purified through a Poly-Prep Chromatography column (Bio-RAD) with a 0.5 ml bed volume of Ni-charged Resin (Bio-RAD) by immobilized metal ion affinity chromatography following the same procedure detailed in Chapter 1. Emission spectra of the purified proteins were determined with the SpectraPro-150 grating spectrometer described in Chapter 1. Emission spectra were taken immediately following purification and after overnight maturation at room temperature.

Determining FP Gene Expression

The degenerate PCR approach used to clone additional novel FP cDNA fragments was also used to determine which FP genes were expressed by each morph. PCR reactions were the same as described for the degenerate PCR. The PCR conditions were also the same, except the annealing temperature was lowered to 52°C for the ccal48F/ccal647R primer set. Five independent PCR reactions were run for each color morph; three utilized the first set of degenerate primers (ccal48F/ccal647R), and two used the second set (ccal1F/ccal666R). PCR products were cloned into the pCR2.1-TOPO vector, and transformed as described before. A total of approximately 50 colonies were

chosen from each color morph for plasmid DNA preparation from all PCR reactions combined. Only plasmids with correct-size inserts were chosen for DNA sequencing with M13 primers.

Some FP genes were not amplified from certain color morphs using the degenerate PCR approach, so gene-specific primers (Table 7H) were designed and used with cDNA from each color morph in PCR. In two cases (ccalYFP1 and ccalRFP1), the full-length primers already designed were specific enough to be used as gene-specific primers. PCR reactions consisted of 2 μl cDNA, 2.5 U PfuUltra Hotstart polymerase, 1x PfuUltra HF buffer, 0.2 μM each gene-specific primer, and 200 μM dNTPs. Cycling parameters were: 94°C for 2 minutes, followed by 35 cycles of 94°C for 30 seconds, T(°C) for 30 seconds, and 72°C for 1 minute, followed by 72°C for 10 minutes. Products were cloned into pCR2.1-TOPO, transformed, sequenced, and analyzed as described before.

In remaining cases where FP genes were not amplified from specific color morph cDNA templates using gene-specific primers, genomic DNA (gDNA) was used as a template to see if the genes were at least present in the genome of each color morph, if not expressed at detectable levels. Positive control reactions with gene-specific primers and gDNA from morphs already found to express those genes were run together with experimental reactions. gDNA was isolated with DNAzol (Molecular Research Center, Inc., Cincinnati, OH) according to manufacturer's instruction from tissue from the same exact color morph clones that had been used for *in vivo* spectral measurements and RNA isolation. The same gene-specific primers used with the cDNA approach (Table 7H) were used with gDNA in PCR reactions consisting of 100 ng gDNA, 1.5 U *Taq* DNA

polymerase (Promega), 1x *Taq* buffer, 800 μM $MgCl_2$, 0.2 μM each gene-specific primer, and 200 μM dNTPs. Cycling parameters were: 94°C for 2 minutes, followed by 35 cycles of 94°C for 30 seconds, T(°C) for 30 seconds, and 72°C for 1 minute, followed by 72°C for 10 minutes. Products were cloned into the pCR2.1-TOPO vector, transformed, sequenced, and analyzed as described before.

Sequence and Phylogenetic Analyses

Sequence and phylogenetic analyses were similar to those detailed in Chapter 1. Sequences were added to the alignment created for Chapter 1, aligned in MacClade v. 4.07, and adjusted manually according to triplet coding. Multiple red proteins from other anthozoan species were included in the alignment for comparison (see Table 5 for information). The only known native orange FP (ceriOFP), from a tube-dwelling anemone (*Cerianthus* sp., GenBank Accession No. AY296063) was also added to the alignment. Percentage similarity was calculated as the percentage fraction identical characters in pairwise aligned sequences, treating gaps as an additional character.

MrModeltest 3.5 (Posada and Crandall 1998) was used to choose the most appropriate molecular evolution model for maximum likelihood (ML) and Bayesian analyses, which were conducted with PAUP* version beta 4.10 (Swofford 2000) and MrBayes 3.0b4 (Huelsenbeck and Ronquist 2001), respectively. ML analysis was completed by heuristic search using 10 random addition sequences using the tree bisection-reconnection (TBR) branch-swapping algorithm. One hundred bootstrap replicates were completed with the same criteria as the original search. For Bayesian analyses, four chains, three heated and one cold, of the Markov chain Monte Carlo

(MCMC) method, were run for each analysis. Three replicate analyses were run to verify convergence, two for 3 million and one for 10 million generations, each with trees sampled every 300 generations. Tracer v.1.0.1 (Rambaut and Drummond 2003) was used to determine the number of generations considered burn-in. These generations were then discarded to ensure the Markov chain had converged on the same topology before inference from the MCMC data set was made. *R. reniformis* GFP was used as the outgroup for the Bayesian analysis.

Maximum parsimony (MP) analysis was conducted with the PAUP* software program version beta 4.10 (Swofford 2000). All characters were equally weighted and gaps were treated as missing data. MP trees of anthozoan FP genes were reconstructed using the heuristic search option with the TBR branch swapping algorithm and 1,000 random-sequence addition replicates. One thousand bootstrap replicates were added by using the same criteria as in the original search. All searches were unrooted. FP genes from *Renilla muelleri* (rmueGFP; GenBank Accession No. AY015996), *R. reniformis* (rrenGFP; GenBank Accession No. AF372525) and *Ptilosarcus* sp. (ptilGFP; GenBank Accession No. AY015995) were used as outgroups to root the resulting MP trees.

RESULTS

In Vivo Fluorescence Emission Spectra Measurements

Fluorescence spectral properties were compared for five tissue types among seven color morphs of *Corynactis californica* (Table 9). Among the samples, fluorescence emission was detected in all tissues of all morphs measured, although some tissues also have nonfluorescent pigments, visible in paired images of morphs under white and blue light (Figure 10a-ab). Slight variation in emission spectra was detected among polyps of individual clones of each color morph, indicated by the standard deviation (Table 9). Most morphs had multiple emission peaks for each tissue type with a single peak dominating, indicating a mixture of pigments co-localized in individual tissues. Overall, emission peaks centered around three wavelengths (518 nm, 563 nm, and 599 nm), though not all peaks were present in all tissue types of all morphs. The tissue-specific mixture of fluorescent pigments resulted in a variety of fluorescent patterns, visible in images of whole animal fluorescence (Figure 10).

Spectra from each tissue type and each morph are also represented graphically (Figures 11-13). One representative spectrum (out of five) was chosen for each tissue type of each morph to create these graphs. All morphs emit at two or three different wavelengths, with one or two emission peaks dominating (Figure 11). Morphs can be grouped according to whether one or two major peaks are emitted. There did not seem to be a pattern within any of the five individual tissue types (Figure 12); each of the three peaks was emitted in each tissue type by at least one morph (Figure 12; Table 9). All morphs emit fluorescent signals with emission maxima centered at two to three common

wavelengths (Figure 13), indicating that similar mixtures of fluorescent pigments are present in all morphs, but that they are often expressed in a tissue-specific manner.

Two general patterns of fluorescence emerged: Pattern A is characterized by a single dominant emission peak emitted in all tissues (Pink2, Orange2, Orange1, and Red morphs; Figure 14; Table 9); Pattern B is characterized by bright fluorescence emission at the base of the tentacles and around the mouth, but little fluorescent emission in other tissues (Lavender, Pink1 and Orange 3 morphs; Figure 15; Table 9). Pattern A was found in morphs collected from all three habitats/depths, while Pattern B was only found in morphs collected from 50 feet (Table 6).

The emission wavelengths were not always the same for all morphs exhibiting pattern A or B. For example, the Orange2 morph had a dominant emission peak around 562 nm, whereas the other morphs exhibiting Pattern A had a single dominant peak in the 599-601 nm range (Figure 14; Table 9). The Lavender morph was unique among morphs displaying Pattern B: tissue at the base of the tentacles emitted around 562 nm, whereas the mouth had emission around 597 nm (Figure 15; Table 9). The Pink1 and Orange3 morphs had similar emission maxima in the base of tentacles (562-563 nm) and around their mouths (567-568 nm; Table 9; Figure 15).

Degenerate RT-PCR to Determine Additional Novel FP Genes

Seventy clones were sequenced from the first degenerate PCR effort. Of these, two represented novel FP gene fragments, homologous to ccalYFP1 and ccalRFP1, and each derived from a different color morph source cDNA (Table 10). Fifteen clones were sequenced from the second degenerate PCR effort. One novel FP gene fragment (ccalFP-

C) resulted from this effort, and the source color morph was different from those that had provided novel sequences before (Table 10).

Determining cDNA Ends

The SMART RACE approach resulted in 3′ sequences for both ccalFP-A and ccalFP-B gene fragments. The template cDNA came from a different morph than the initial gene fragment for ccalFP-A, but the same morph as the initial fragment for ccalFP-B (Table 10). SMART RACE PCR reactions for the ccalFP-A and ccalFP-B 3′ ends each took a total of 45 cycles before a band was visible by gel electrophoresis. RACE reactions designed to produce 5′ products for the ccalFP-A and ccalFP-B fragments were unsuccessful.

The degenerate non-coding region PCR approach was successful in amplifying 5′ cDNA ends for the ccalFP-A and ccalFP-B fragments, and 3′ and 5′ ends for the ccalFP-C fragment. Table 10 specifies which color morphs provided the source material for amplifying these products. In one case (the ccalFP-C 5′ end), two independent PCR reactions generated the same sequence and provided confidence in the resulting 5′ end sequence.

Full-length FP sequences were assembled for each new gene using the 5′ and 3′ fragments. Figures 16-18 provide the full-length nucleotide and inferred amino acid sequences for the three novel FPs, distinct from ccalYFP1 (Figure 1a and b) and ccalRFP1 (Figure 2a and b). The full-length sequences derived from the ccalFP-A and ccalFP-B fragments were each 681 nucleotides and 226 amino acids in length (Figures 16

and 17), and sequence-based calculations indicate that both have a molecular weight of 29.5 kDa.

The ccalFP-C fragment resulted in a full-length sequence 666 nucleotides and 221 amino acids long (Figure 18) – similar in length to both ccalYFP1 and ccalRFP1. Sequence-based calculations indicate a molecular weight of 28.7 kDa.

Full Length Cloning

The source color morph cDNA used to amplify the full-length coding region of each new fluorescent protein is detailed in Table 10. All PCR reactions were successful, although in most cases, even though a high-fidelity proofreading enzyme was used, multiple clones had to be sequenced before full-length clones with no point mutations or frame shifts were produced.

Expressing FPs in *E. coli* and Protein Purification

The full-length ccalFP-A cDNA sequence resulted in a red fluorescent protein that, after purification from either pQE31X expression vector (Table 8), transitions from a bright green immature form (emission max. 512 nm) to a red final state (emission max. 595 nm) after overnight maturation and was renamed ccalRFP2 (Figure 19). The full-length ccalFP-B cDNA sequence resulted in an orange fluorescent protein that, when purified from either pQE31X vector, matures immediately to a 562 nm emission maximum, though it has a green intermediate form (emission max. 515 nm) and was renamed ccalOFP1 (Figure 20). The full-length ccalFP-C cDNA sequence resulted in a green fluorescent protein that, when purified from either the pQE31X-TEV or non-TEV

vector matures immediately to a final emission maximum of 517 nm and was renamed ccalGFP1 (Figure 21). Nomenclature for FPs followed established classification (Labas et al. 2002). Sequences were deposited into the GenBank database under accession numbers DQ065851 (ccalRFP2), DQ065853 (ccalOFP1), and DQ065852 (ccalGFP1).

The original two FPs from *C. californica* were also expressed in and purified from the modified pQE31 vector system. When purified from either pQE31X vector, their maturation behavior and final emission maxima were similar to results from the pCK vector (Ch. 1). ccalYFP1 showed a final emission maximum of 524 nm (data not shown). ccalRFP, however, had a green intermediate of 525 nm (instead of 518 nm) and a final red emission peak at around 608 nm (compared with 600 nm). After 72 hours at room temperature, ccalRFP1 did not fully mature in this system; a relatively small proportion of the protein was still in the green intermediate phase (Figure 22).

Determining FP Gene Expression

A total of 210 clones were sequenced to detect the presence or absence of expression of each FP gene in each color morph (Table 11). Degenerate PCR with color morph cDNA gave mixed results. Out of 35 possible instances, 19 FPs were amplified from the seven color morphs using this approach (Table 11). Relative abundances of each FP were not quantitatively determined, and there are many possible types of bias that can influence the number of clones obtained from each PCR and cloning effort. It is tempting to note, however, that results from the cloning do provide some evidence that individual color morphs may express certain FPs in higher proportion than others (numbers; Table 11).

In some cases, a single FP gene was either amplified or cloned in much higher proportion than others, even from multiple independent PCR reactions. This was true for Pink2, Orange1, and Red color morphs, all of which had ccalRFP1 sequenced much more frequently than any other FP gene. Orange2 had the most even distribution of FP genes, with four of the genes amplified in similar proportions using this approach.

In seven cases, the expression of FP genes was detected only when gene-specific primers were applied in PCR (indicated by "GS"; Table 11). This approach was quite successful in amplifying ccalOFP1. These results demonstrate that ccalOFP1 was expressed by all color morphs at the time of RNA isolation.

In some cases, expression of certain FPs corresponded well with the *in vivo* emission spectra recorded from each color morph. For example, the Pink2, Orange1, and Red morphs had emission spectra centered around 599-601 nm for all tissues (Table 9), and all were found to express a red protein (ccalRFP1, emission max. 600 nm; Table 11). The Lavender morph expressed ccalRFP2 (emission max. 595 nm) and ccalGFP1 (emission max. 517 nm), which corresponded well with the *in vivo* spectra found in multiple tissues, such as the mouth and oral disc (Table 9).

In other cases, the connection was not as straightforward. Expression of ccalOFP1 was detected in the Pink2 and Red morphs (Table 11), although no orange emission peaks were measured from either morph (Table 9). It is possible that ccalRFP1 was expressed at such high levels in these morphs that spectral detection of ccalOFP1 was obscured. Close examination of spectra show slight shoulders of orange emission in some tissues (Figure 14). Expression of ccalRFP1 and ccalRFP2 was detected in the Orange 2 and Orange3 morphs (Table 11), though red *in vivo* emission peaks were not detected in

these morphs (Table 9). A shoulder of red emission can be seen in spectra of multiple Orange2 and Orange3 tissues (Figure 11), however, which may explain the detection of expression of these proteins.

No genes were amplified using the gDNA and specific primer approach. Positive control reactions for the gDNA approach also failed. In nine cases, FP genes were not amplified from any of the PCR-based methods used to detect them.

Sequence and Phylogenetic Analyses

An amino acid alignment (Figure 23), shown relative to ccalRFP1, illustrates regions of similarity and difference among *Corynactis californica* FPs and homologous GFP-type proteins, with an emphasis on yellow, orange, and red-emitting proteins. ccalRFP2 and ccalOFP1 are five amino acids longer than the other ccalFPs. The additional residues occur at the end of the sequence, and not as insertions (Figure 23).

Nucleotide and amino acid sequence homologies between FPs are represented as percentages (Table 12). In some cases, sequence homologies shown in Table 12 differ slightly from those reported in Table 4 due to minor differences in the overall sequence alignment used for the calculations. The three new FP genes possess 86-89% nucleotide and 81-85% amino acid sequence homology to each other and 84-90% nucleotide and 77-87% amino acid homology to ccalYFP1 and ccalRFP1 (Table 12). On the nucleotide level, ccalRFP2 and ccalOFP1 are quite similar to ccalRFP1 (90%), and to one another (89%), while ccalGFP1 is most similar to ccalOFP1 and ccalRFP1 (87%).

The GTR+I+Γ model fit the data best according to the Akaike criterion (Akaike 1974) for ML and Bayesian analyses. Of the three million generations, 2,000 were

considered the "burn-in" period and discarded from the Bayesian analysis. All searches converged on similar parameter estimates.

Bayesian, ML, and MP analyses all reconstructed gene trees with five distinct clades (Figure 24-26), though the topology of the ML tree differed slightly (Figure 26). At some nodes, the ML and MP trees showed relatively low bootstrap support, whereas the Bayesian tree had relatively high posterior probability values for nearly all nodes (Figure 25). The addition of ceriOFP to the alignment added a fifth clade to the trees reconstructed previously (Figures 8 and 9). Clade 1 is most distant from the *C. californica* FPs and includes a single orange protein, ceriOFP. Clade 2 includes one red protein, equaRFP, and is the equivalent of Clade 1 in Figures 8 and 9. Clade 3 includes red proteins from *Discosoma* (discRFP and dis2RFP) and is similar to Clade 2 in Figures 8 and 9. The five FPs from *C. californica* form a well-supported subclade nested within Clade 4 (Clade 3 in Figures 8 and 9). In all of the consensus trees, ccalRFP1 and ccalOFP1 are grouped together, with strong support for ccalRFP2 as sister to them. ccalGFP1 falls between this group and ccalYFP1. Their sister group is a subclade of proteins from *Anemonea majano* (amajCyFP and amajGFP) and *Astrangia lajollaensis* (alajGFP). The next closest neighbors of the ccalFPs within Clade 4 include FPs from *Zoanthus* sp. (zoan2RFP, zoanYFP, zoanGFP). Clade 5 includes four red proteins (tgeoRFP, mcavRFP, lhemRFP, rfloRFP) and is equivalent to Clade 4 from the previous analysis, though there are minor changes in the location of certain branches within this clade (e.g. tgeoRFP, lhemRFP).

The sister group to the *C. californica* FP genes includes amajGFP, which has 67-70% nucleotide and 60-63% amino acid similarity to the group. Another FP in the same

major clade as the *C. californica* FPs, zoanYFP, possesses 61-63% nucleotide and 52-56% amino acid similarity to the group. Other red FPs have 55-60% nucleotide and 43-49% amino acid similarity to the ccalFPs. The FP used to root phylogenetic trees, rrenGFP, displays 47-49% nucleotide and only 32-35% amino acid similarity to the ccalFPs (Table 12).

DISCUSSION

Morphs of *Corynactis californica* show variation in the pattern and emission spectra of their fluorescent pigmentation. In this study, these pigmentation differences were characterized via *in vivo* and *in vitro* molecular methods. Five fluorescent protein genes were isolated from various morphs of this species. Determining which genes were expressed by each morph provided data to address whether intraspecific pigment variation among color morphs of *Corynactis californica* was a multilocus trait and to relate *in vivo* emission spectra to *in vitro* expression of FP genes in each morph. Sequence and phylogenetic analyses provided additional evidence for understanding how the pigments in *C. californica* are related to one another and to homologous members of the GFP family.

Biochemistry

The discovery of three additional fluorescent proteins from the degenerate RT-PCR approach brings the total number of fluorescent protein genes in *C. californica* to five. Each protein has unique biochemical and spectral properties. Emission spectra of these purified proteins occur in regions of the visible spectrum where proteins have not yet been found (ccalYFP1, ccalOFP1, ccalRFP1), or are relatively rare (ccalRFP2; Figure 27) among native FPs.

The chromophores of ccalRFP2, ccalOFP1, and ccalGFP1 are all formed by Thr-Tyr-Gly (Figure 20), the same as ccalRFP1. The new red protein, ccalRFP2, is similar to ccalRFP1 by exhibiting the "fluorescent timer" phenotype, proceeding through a green intermediate before reaching full maturation.

The orange protein, ccalOFP1, is only the second known orange-emitting native fluorescent protein from the GFP-family. The other orange protein, ceriOFP, has a chromophore formed by Gln-Tyr-Gly (Figure 23). Future studies of protein crystal structure, such as those completed for other GFP-like proteins (Remington et al. 2005, Wilmann et al. 2005), and site-directed mutagenesis of specific residues (Terskikh et al. 2002, Shagin et al. 2004, Wiedenmann 2004b) may determine chromophore orientation, specific residues, or other factors that may influence the orange emission color.

The green protein, ccalGFP1, has the same chromophore as EGFP (Cormack et al. 1996), demonstrating that this chromophore has produced green emission in an engineered mutant. Although ccalYFP1 and ccalGFP1 emit at similar wavelengths, their chromophores differ. The chromophore of ccalYFP1 is formed by Cys-Tyr-Gly, which has only been seen in green proteins from *Astrangia lajollaensis* (Bessette and Daugherty 2004). Further study is needed to precisely determine how emission color is obtained in these proteins.

Determining FP Gene Expression

It should be emphasized that the degenerate PCR sampling effort was not exhaustive, and the results of relative abundances could change with increased sampling. Also, PCR results could be biased by the most abundant template in the reaction and the best primer match. Methods such as quantitative PCR or Northern blotting would be useful in determining quantitative levels of relative abundances.

The yellow protein, ccalYFP1, was only detected in the red morph using gene-specific primers. This is interesting because ccalYFP1 was one of the two proteins found

from the same red morph by the FACS-screening method detailed in Chapter 1. This could indicate that ccalYFP1 is expressed in relatively low abundance in the Red morph, but that the FACS-screening method was sensitive enough to detect it.

In cases where amplification from cDNA with gene-specific primers failed to confirm expression, amplification from gDNA was attempted to determine if the genes were present in the genomes of each morph. Although no genes were detected using the gDNA and specific primer approach, it does not necessarily mean that these genes are not present in the genomes of the morphs in question; it could indicate that the extraction of the gDNA template failed, that the gene-specific primers were poorly designed, or that there were introns in the region of the primers.

Overall, in nine cases, FP genes were not amplified from any of the PCR-based methods used to detect them. This does not necessarily mean that these genes are not expressed in the morphs examined. There could be very low levels of expression of these FP genes that are masked by high expression levels of other FPs. Increased sampling effort, or tissue-specific PCR could result in detection of proteins expressed at very low levels, or only in certain tissues. It is also possible that these genes are present in the genome of each morph, but were not being actively expressed at the moment of RNA isolation. Since the gDNA method failed, it is impossible to conclude whether these genes are present in the genomes of these morphs or not.

Multi-Locus Nature of Pigmentation

The data presented here show the same suite of at least three to four homologous FP genes (and possibly five or more) is expressed in seven color morphs of *Corynactis*

californica collected at various depths from Monterey Bay, CA (Table 11), indicating that pigmentation is a multi-locus trait in this species, since the number of unique alleles per locus should not exceed two in a diploid organism. The nucleotide differences among the five FP genes (10-16%; Table 12) are also greater than what would be expected for different alleles of the same locus.

The variation in pigment colors and patterns among morphs results from varying the expression level of each protein in specific tissues at some stage in development. This points to independent control of gene expression by tissue type and also supports the multi-locus nature of pigmentation. This is clear in the Pink2, Orange1, and Red morphs, which all have major emission peaks around 600 nm in all tissues (Table 9), and appear red under blue excitation (Figure 14), but are expressing yellow and orange proteins together with red proteins (Table 11). The yellow and orange proteins are presumably expressed at much lower levels than the red proteins in this morph. It is also possible that the yellow and orange proteins were not detected *in vivo* because they are participating in co-localized energy transfer, with their emission energy channeled into added excitation energy for the red proteins. Similarly, the Orange2 morph has a dominant emission peak of 562 nm in all tissues (Table 9), and appears orange under blue light (Figure 14), but the expression of ccalRFP1 and ccalRFP2 was detected along with ccalOFP1 (Table 11).

Additional evidence derives from Pattern B morphs (Figure 15): all express the same suite of FP genes (Table 11), though spectral differences were detected within tissues among the three morphs (Table 9, Figure 15). The spectral variation is due to tissue-specific differences in expression levels of specific FPs, resulting in different patterns. These results confirm that the evolutionary basis of fluorescence in *C.*

californica is not simple single locus mechanism. The data indicate that it is possible that certain morphs do not contain one or two of the five genes in their genomes, so underlying differences in inheritance may exist among certain morphs, especially between morphs exhibiting either Pattern A or B.

Phenotypic Plasticity vs. Genotypic Inheritance

Although data presented here demonstrate that fluorescent pigmentation is a multi-locus trait in this species, whether the variation in tissue-specific expression among morphs is due to inherited differences in their genomes, environmentally mediated phenotypic plasticity, or a mixture of both mechanisms can only be confirmed with further studies, such as test crosses or transplantation experiments. Inherited differences could arise from allelic variations in the multiple color-coding loci or regulatory loci and be maintained in the population by balancing selection. Phenotypic plasticity, reflected by changes in relative expression levels of color-coding loci, could be controlled by signal transduction pathways that are influenced by external factors.

Experimentally, inherited differences in genetic background could be confirmed with laboratory-based test crosses. Two morphs could be crossed to see how pigmentation segregates in the following generations. To date, *C. californica* has not been induced to spawn in the lab, so this method may prove to be difficult. Alternately, transplantation experiments would show no difference in pigmentation when morphs are transplanted to a new location if the variation in pigmentation is a fixed character under genetic control.

Phenotypic plasticity can be established if specific external factors that influence changes in pigment expression are found (Kelmanson and Matz 2003). It is unknown whether individual specimens of *Corynactis californica* can change their pigmentation throughout their lives, or if expression levels are determined at some specific point in their life cycle, such as upon settlement or during development. Observations in the laboratory indicate that pigment levels in adult *C. californica* can be altered over a period of several months (personal observation), but only a loss of pigmentation was observed, not a change in emission color or pattern of pigmentation. Asexually derived clones of individual morphs exhibit the exact same patterns of pigmentation, indicating that if pigmentation is flexible, expression levels may be fixed at some stage. Similarity among clonemates, however, could merely reflect a shared genotype, or a shared environment. It would be challenging to follow pigmentation through the development of an individual from larvae to adult polyp in the laboratory because *C. californica* reproduces only one time per year (Holts and Beauchamp 1993), and it is not known to spawn under laboratory conditions. Because environmental conditions may change only slightly during a period of years, it also may be impossible to detect differences in pigment expression through short-duration field observations. The alternative to observational studies is experimental manipulation. Since pigmentation may be fixed in adult polyps, newly settled polyps would be ideal for field studies, such as transplantation experiments. Exposing developing *C. californica* to gradients or combinations of environmental factors in the field or lab should result in an alteration of pigmentation if environmentally-induced phenotypic plasticity occurs in this species (Todd et al. 2002a, 2002b, Relyea 2004).

Sequence and Phylogenetic Analyses

Red fluorescent proteins require three consecutive autocatalytic reactions to complete their chromophore synthesis, whereas green proteins only require two reactions, resulting in a extended, more complex chromophore structure in red proteins (Gross et al. 2000, Yarbrough et al. 2001). The evolution of red emission from green has been seen multiple times in the GFP family, and may represent convergent evolution of complexity (Shagin et al. 2004, Ugalde 2004). This also seems to be true for proteins from *C. californica.* The consensus trees reconstructed by ML, Bayesian, and MP analyses indicate that the red and orange proteins in *C. californica* evolved from the yellow and green proteins (Figure 24-26). Similarity calculations and consensus trees show that ccalRFP1, ccalRFP2, and ccalOFP1 are quite similar to one another. This may indicate that residues that contribute to orange or red color of emission may have evolved together from a green ancestor.

It is not clear whether FPs is *C. californica* arose through time via gene duplication and random mutation, becoming fixed in the population because their expression was advantageous, or if they were specifically selected for by their emission color (Matz et al. 2002). The five genes from *C. californica* cluster together, indicating they evolved after the speciation events that separate them from their closest neighbors. FP genes from corallimorpharian genera, on the other hand, can be found in Clades 3 (*Discosoma*), 4 (*Corynactis*), and 5 (*Ricordea*), indicating these FP genes descended from genes that existed before the separation of this anthozoan subclass and that selection may be acting on color or chromophore phenotype rather than taxonomy. More molecular phylogenetic studies of corallimorpharian, scleractinian, and actiniarian species are

needed before further conclusions regarding the evolution of FP genes in these groups can be made (Matz et al. 2002).

Biological Significance

Fluorescent and nonfluorescent pigments exist in *Corynactis californica* and each may serve a different function. In the context of the subtidal environment of *C. californica*, nonfluorescent lavender, orange, or red pigments make the animal appear dark and inconspicuous. Fluorescent pigments, on the other hand, make it more noticeable and prominent. The variation of pigmentation within different tissues of an individual clone of *Corynactis californica* can be observed by eye (Figures 10, 14, and 15), or in spectral measurements (Table 9, Figure 11), and the contrast created by some patterns can be quite striking to the observer. In Pattern B morphs, a bright ring of fluorescence is seen at the base of the tentacles, with another concentrated area of pigmentation around the opening of the mouth (Figure 15). These patterns, and the variations found within each pattern, may hold the key to understanding what advantage fluorescence may convey to *C. californica* and related organisms.

Although the adaptive significance of color in the marine environment has long been studied (Cott 1957), recent attention has been given to the roles of color and visual pigments of coral reef fishes (Marshall 2000, Losey et al. 2003, Marshall et al. 2003a, 2003b). If fluorescence is used for interspecies communication in *Corynactis californica*, as discussed in Chapter 1, the two patterns may serve different functions or may attract particular species. Different color morphs of a coral, *Acropora nasuta,* for example, facilitate the coexistance of competing species of gobies of the genus *Gobiodon*, which

are obligate symbionts of scleractinian corals (Munday et al. 2001). Studies of associations between *C. californica* and organisms that possess color vision may determine specific functions for the different fluorescent patterns seen among morphs.

Corynactis californica does not harbor endosymbiotic algae, but it does exhibit multiple morphs that emit fluorescent signals at wavelengths that span the visible spectrum. Thus, it is unlikely that the function of its fluorescent pigmentation is related to the regulation of light levels. In a distributional study of a closely related species (*Corynactis viridis*), polyps reacted negatively (by retracting) when exposed to increased levels of direct sunlight, and some showed rhythmic activity (expanding by night and retracting by day), indicating that response mechanisms to light levels are already present in these organisms (Muntz et al. 1972). In a distributional study of *C. californica* in relation to three of its competitors, Chadwick (1991) reported that abundance depended on proximity to the top and base of each rock, probably due to vertical gradients in physical and biological factors. Quantitative surveys of subtidal areas examining patterns in the distribution of color morphs may provide insight into which environmental parameters induce differences in pigment expression and into the possible functions of fluorescent pigmentation.

The recent discovery of FP genes in members of the Ceriantharia [a green protein from *Cerianthus membranaceus* (Wiedenmann 2004a), and an orange protein from *Cerianthus* sp., GenBank Accession No. AY296063] is interesting to note because this group is non-symbiotic and believed to be basal within the Zoantharia (France et al. 1996, Berntson et al. 1999). If FP genes derive from an ancestor to this non-symbiotic group, it is possible that FPs originally evolved for reasons other than light regulation, and that

organisms with symbionts evolved this function later. This information, together with data presented in this study, provides strong evidence for multiple functions of the wide range of FPs within and beyond the Anthozoa.

Conclusions

Although one previous study (Wicksten 1989) alluded to the "neon glow" of *Corynactis californica*, the source of fluorescent pigmentation in this species was previously unknown. This study demonstrates that fluorescent proteins from the GFP family are major pigment determinants in this species. Demonstrating that a non-symbiotic anthozoan has evolved multiple FPs of varying emission color, and that intraspecific variation in pigmentation is determined by the tissue-specific expression of these proteins, indicates that organisms may utilize FPs for reasons beyond light regulation, such as interspecies communication. Our data show that fluorescent pigmentation is a multi-locus trait, with three to four loci expressed in each morph. If the fluorescent pigmentation is a plastic character, the expression of individual FPs should vary with specific environments, and future studies may uncover factors influencing patterns and levels of expression in *C. californica.* The multi-locus nature of pigmentation in *C. californica,* and related species with fluorescent proteins, could allow for broad distribution and acclimatization to shifting external conditions, which could prove crucial to their long-term survival in the face of local and global environmental change.

LITERATURE CITED:

Akaike, H. 1974. A new look at the statistical model identification. IEEE Trans. Autom. Contr. **19**:716-723.

Alfaro, M. E., S. Zoller, and F. Lutzoni. 2003. Bayes or bootstrap? A simulation study comparing the performance of Bayesian Markov chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence. Mol. Biol. Evol. **20**:255-266.

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. **99**:12651-12656.

Arnold, K. E., I. P. F. Owens, and N. J. Marshall. 2002. Fluorescent signaling in parrots. Science **295**:92.

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. **97**:11984-11989.

Berntson, E. A., S. C. France, and L. S. Mullineaux. 1999. Phylogenetic relationships within the Class Anthozoa (Phylum Cnidaria) based on nuclear 18S rDNA sequences. Molecular Phylogenetics and Evolution **13**:417-433.

Bessette, P. H., and P. S. Daugherty. 2004. Flow cytometric screening of cDNA expression libraries for fluorescent proteins. Biotechnology Progress **20**:963-967.

Bevis, B. J., and B. S. Glick. 2002. Rapidly maturing variants of the *Discosoma* red fluorescent protein (DsRed). Nature Biotechnology **20**:83-87.

Brakefield, P. M., and V. French. 1999. Butterfly wings: the evolution of development of colour patterns. BioEssays **21**:391-401.

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 Biochemistry **3**:1-8.

Campbell, R. E., O. Tour, A. E. Palmer, P. A. Steinbach, G. S. Baird, D. A. Zacharias, and R. Y. Tsien. 2002. A monomeric red fluorescent protein. Proc. Nat. Acad. Sci. **99**:7877-7882.

Carlgren, O. 1949. A survey of the Ptychodactiaria, Corallimorpharia and Actiniaria. Kungliga Svenska Vetenskapsakademiens Handlingar **1**:1-129.

Carter, R. W., Schmale, M.C., and P.D.L. Gibbs. 2004. Cloning of anthozoan fluorescent proteins. Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology **138**:259-270.

Chadwick, N. E. 1987. Interspecific aggressive behavior of the corallimorpharian *Corynactis californica* (Cnidaria: Anthozoa): effects on sympatric corals and sea anemones. Biol. Bull. **173**:110-125.

Chadwick, N. E. 1991. Spatial distribution and the effects of competition on some temperate Scleractinia and Corallimorpharia. Marine Ecology Progress Series **70**:39-48.

Chadwick, N. E., and C. Adams. 1991. Locomotion, asexual reproduction, and killing of corals by the corallimorpharian *Corynactis californica*. Hydrobiologia **216/217**:263-269.

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

Cormier, M. J., K. Hori, Y. D. Karkhanis, J. M. Anderson, J. E. Wampler, Morin, J.G., and J. W. Hastings. 1975. Evidence for similar biochemical requirements for bioluminescence among the coelenterates. J. Cell. Physiol. **81**:291-298.

Cott, H. B. 1957. Adaptive coloration in animals. Methuen, London, UK.

Cummings, M. E., and J. C. Partridge. 2001. Visual pigments and optical habitats of surfperch (Embiotocidae) in the California kelp forest. Journal of Comparative Physiology Part A **187**:875-889.

Douglas, R. H., J. C. Partridge, and A. J. Hope. 1995. Visual and lenticular pigments in the eyes of demersal deepsea fishes. Journal of Comparative Physiology Part A **177**:111-122.

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.

Dunn, D. F. 1982. Cnidaria. Pages 669-706 *in* S. P. Parker, editor. Synopsis and Classification of Living Organisms, Volume 1. McGraw-Hill Book Company, New York.

Fradkov, A. F., Y. Chen, L. Ding, E. V. Barsova, M. V. Matz, and S. A. Lukyanov. 2000. Novel fluorescent protein from *Discosoma* coral and its mutants possesses a unique far-red fluorescence. FEBS Letters **479**:127-130.

France, S. C., P.E. Rosel, J.E. Agenbroad, L.S. Mullineaux, and T. D. Kocher. 1996. DNA sequence variation of mitochondrial large-subunit rRNA provides support for a two subclass organization of the Anthozoa (Cnidaria). Molecular Marine Biology and Biotechnology **5**:15-28.

Frohman, M. A., M. K. Dush, and G. R. Martin. 1988. Rapid amplification of cDNA ends- protocols developed to facilitate the cloning of the 5' (5'-RACE) or 3' (3'-RACE) ends of a cDNA using information obtained from an incomplete cDNA clone. Proceedings of the National Academy of Sciences USA **85**:8998-9002.

Fuchs, E. 2001. Separating the fluorescence and reflectance components of coral spectra. Applied Optics **40**:3614-3621.

Fuchs, E., and C. Mazel. 1999. Unmixing coral fluorescence emission spectra and predicting new spectra under different excitation conditions. Applied Optics **38**:486-494.

Gilmore, A. M., A. W. D. Larkum, A. Salih, S. Itoh, Y. Shibata, C. Bena, H. Yamasaki, M. Papina, and R. V. Woesik. 2003. Simultaneous time-resolution of the emission spectra of fluorescent proteins and zooxanthellae chlorophyll in reef-building corals. Photochemistry & Photobiology **77**:515-523.

Glazer, A. N. 1989. Light guides. Directional energy transfer in a photosynthetic antenna. J. Biol. Chem. **264**:1-4.

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. **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. 2001. GFP-like chromoproteins as a source of far-red fluorescent proteins. FEBS Lett. **507**:16-20.

Haddock, S. H. D., and J. F. Case. 1999. Bioluminescence spectra of shallow and deep-sea gelatinous zooplankton: ctenophores, medusae and siphonophores. Marine Biology **133**:571-582.

Haddock, S. H. D., C. W. Dunn, P. R. Pugh, and C. E. Schnitzler. 2005. Bioluminescent and red fluorescent lures in a deep-sea siphonophore. Science **In press**.

Haderlie, E. C., C. Hand, and W. B. Gladfelter. 1980. Cnidaria (Coelenterata): the sea anemones and allies. Pages 64-65 *in* R. H. Morris, D. P. Abbott, and E. C. Haderlie, editors. Intertidal Invertebrates of California. Stanford University Press, Stanford, CA.

Hamner, W. M., and D. F. Dunn. 1980. Tropical corallimorpharia (Coelenterata: Anthozoa): feeding by envelopment. Micronesia **16**:37-41.

Hand, C. 1954. The Sea Anemones of Central California. Part I. The Corallimorpharian and Athenarian Anemones. The Wasmann Journal of Biology **12**:345-375.

Hansson, L.-A. 2004. Plasticity in pigmentation induced by conflicting threats from predation and UV radiation. Ecology **85**:1005-1016.

Hastings, J. W., and J. G. Morin. 1969. Calcium-triggered light emission in *Renilla*. A unitary biochemical scheme for coelenterate bioluminescence. Biochemical and Biophysical Research Communications **37**:493-498.

Holts, L. J., and K. A. Beauchamp. 1993. Sexual reproduction in the corallimorpharian sea anemone *Corynactis californica* in a central California kelp forest. Marine Biology **116**:129-136.

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

Kawaguti, S. 1944. On the physiology of reef corals. VI. Study on the pigments. Palau Trop. Biol. Stn. Stud **2**:617-673.

Kawaguti, S. 1966. Electron microscopy on the fluorescent green of reef corals with a note on mucous cells. Biol. J. Okayama University **2**:11-21.

Kawaguti, S. 1969. The effect of green fluorescent pigment on the productivity of the reef corals. Micronesica **5**:313.

Kelmanson, I. V., and M. V. Matz. 2003. Molecular basis and evolutionary origins of color diversity in great star coral *Montastraea cavernosa* (Scleractinia: Faviida). Molecular Biology and Evolution **20**:1125-1133.

Kwok, S., S. Y. Chang, J. J. Sninsky, and A. Wang. 1994. A guide to the design and use of mismatched and degenerate primers. PCR Methods and Applications **3**:S39-S47.

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. Proceedings of the National Academy of Sciences USA **99**:4256-4261.

Li, X., G. Zhang, N. Ngo, X. Zhao, S. R. Kain, and C.-C. Huang. 1997. Deletions of the *Aequorea victoria* green fluorescent protein define the minimal domain required for fluorescence. J. Biol. Chem. **272**:28545-28549.

Limbaugh, C., and W. J. North. 1956. Fluorescent, Benthic, Pacific Coast Coelenterates. Nature **178**:497-498.

Logan, A., K. Halcrow, and T. Tomascik. 1990. UV excitation-fluorescence in polyp tissue of certain scleractinian corals from Barbados and Bermuda. Bull. Mar. Sci. **46**:807-813.

Losey, G. S., W. N. McFarland, E. R. Loew, J. P. Zamzow, P. A. Nelson, and N. J. Marshall. 2003. Visual biology of Hawaiian coral reef fishes. I. Ocular transmission and visual pigments. Copeia **3**:433-454.

Lukyanov, K. A., A. F. Fradkov, N. G. Gurskaya, M. V. Matz, Y. A. Labas, A. P. Savitsky, M. L. Markelov, A. G. Zaraisky, X. Zhao, Y. Fang, W. Tan, and S. A. Lukyanov. 2000. Natural animal coloration can be determined by a nonfluorescent green fluorescent protein homolog. J. Biol. Chem. **275**:25879-25882.

Maddison, D. R., and W. P. Maddison. 2000. MacClade 4: analysis of phylogeny and character evolution. Sinauer Associates, Inc., Sunderland, MA.

Marshall, N. J. 2000. Communication and camouflage with the same "bright" colours in reef fish. Philosophical Transactions of the Royal Society of London, Series B **355**:1243-1248.

Marshall, N. J., K. Jennings, W. N. McFarland, E. R. Loew, and G. S. Losey. 2003a. Visual biology of Hawaiian coral reef fishes. II. Colors of Hawaiian coral reef fish. Copeia **3**:455-466.

Marshall, N. J., K. Jennings, W. N. McFarland, E. R. Loew, and G. S. Losey. 2003b. Visual biology of Hawaiian coral reef fishes. III. Environmental light and an integrated approach to the ecology of reef fish vision. Copeia **3**:467-480.

Matz, M., 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.

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. Nature Biotechnology **17**:969-973.

Mazel, C., T. W. Cronin, R. W. Caldwell, and N. J. Marshall. 2004. Fluorescent enhancement of signaling in a mantis shrimp. Science **303**:51.

Mazel, C. H. 1995. Spectral measurements of fluorescence emission in Caribbean cnidarians. Mar. Ecol. Prog. Ser **120**:185-191.

Mazel, C. H. 1997. Coral fluorescence characteristics: excitation - emission spectra, fluorescence efficiencies, and contribution to apparent reflectance. Ocean Optics XIII, SPIE **2963**:240-245.

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

McRae, S. R., C. L. Brown, and G. R. Bushell. 2005. Rapid purification of EGFP, EYFP, and ECFP with high yield and purity. Protein Expression & Purification **41**:121-127.

Meroz-Fine, E., I. Brickner, Y. Loya, and M. Ilan. 2003. The hydrozoan coral *Millepora dichotoma*: speciation or phenotypic plasticity? Marine Biology **143**:1175-1183.

Munday, P. L., G. P. Jones, and M. J. Caley. 2001. Interspecific competition and coexistance in a guild of coral-dwelling fishes. Ecology **82**:2177-2189.

Muntz, L., T. A. Norton, F. J. Ebling, and J. A. Kitching. 1972. The ecology of Lough Ine. XVIII. Factors controlling the distribution of *Corynactis viridis* Allman. Journal of Animal Ecology **41**:735-750.

Nicol, J. A. C. 1958. Observations on luminescence in pelagic animals. Journal of the Marine Biological Association of the United Kingdom **37**:705-752.

Notredame, C., D. Higgins, and J. Heringa. 2000. T-Coffee: a novel method for multiple sequence alignments. Journal of Molecular Biology **302**:205-217.

Ormö, 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.

Phillips, C. E. S. 1927. Fluorescence of sea anemones. Nature **119**:1-17.

Posada, D., and K. A. Crandall. 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics **14**:917-918.

Posada, D., and K. A. Crandall. 2001. Selecting the best-fit model of nucleotide substitution. Systematic Biology **50**:580-601.

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.

Rambaut, A., and A. Drummond. 2003. Tracer v. 1.0.1. Available from evolve.zoo.ox.ac.uk.

Relyea, R. A. 2004. Fine-tuned phenotypes: tadpole plasticity under 16 combinations of predators and competitors. Ecology **85**:172-179.

Remington, S. J., R. M. Wachter, D. K. Yarbrough, B. Branchand, D. C. Anderson, K. Kallio, and K. A. Lukyanov. 2005. zFP538, a yellow-fluorescent protein from *Zoanthus*, contains a novel three-ring chromophore. Biochemistry **44**:202-212.

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

Schlichter, D., U. Meier, and H. W. Fricke. 1994. Improvement of photosynthesis in zooxanthellate corals by autofluorescent chromatophores. Oecologia **99**:124-131.

Shagin, D. A., E. V. Barsova, Y. G. Yanushevich, A. F. Fradkov, K. A. Lukyanov, Y. A. Labas, T. N. Semenova, J. A. Ugalde, A. Meyers, J. M. Nunez, E. A. Widder, S. A. Lukyanov, and M. V. Matz. 2004. GFP-like proteins as ubiquitous metazoan superfamily: evolution of functional features and structural complexity. Molecular Biology and Evolution **21**:841-850.

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

Simms, D., P. E. Cizdziel, and P. Chomczynski. 1993. TRIzol: a new reagent for optimal single-step isolation of RNA. Focus (Life Technologies) **15**:99-102.

Sun, Y., E. W. J. Castner, C. L. Lawson, P. G. Falkowski, and. 2004. Biophysical characterization of natural and mutant fluorescent proteins cloned from zooxanthellate corals. FEBS Letters **570**:175-183.

Suzuki, Y., G. V. Glazko, and M. Nei. 2002. Overcredibility of molecular phylogenies obtained by Bayesian phylogenetics. Proc. Natl. Acad. Sci. USA **99**:16138-16143.
Swofford, D. L. 2000. PAUP*. Phylogenetic analysis using parsimony (and other methods). Version 4. Sinauer Associates, Inc., Sunderland, MA.
Terskikh, A., A. Fradkov, G. Ermakova, A. Zaraisky, P. Tan, A. V. Kajava, X. Zhao, S. Lukyanov, M. Matz, S. Kim, I. Weissman, and P. Siebert. 2000. "Fluorescent timer": protein that changes color with time. Science **290**:1585-1588.
Terskikh, A. V., A. F. Fradkov, A. G. Zaraisky, A. V. Kajava, and B. Angres. 2002. Analysis of DsRed mutants. The Journal of Biological Chemistry **277**:7633-7636.
Todd, P. A., R. C. Sidle, and L. M. Chou. 2002a. Plastic corals from Singapore: 1. Coral Reefs **21**:391-392.
Todd, P. A., R. C. Sidle, and L. M. Chou. 2002b. Plastic corals from Singapore: 2. Coral Reefs **21**:407-408.
Tsien, R. Y. 1998. The green fluorescent protein. Annu. Rev. Biochem. **67**:509-544.
Tupen, J. W. 1999. Shell form and color variability in *Alia carinata* (Neogastropoda: Columbellidae). The Veliger **42**:249-259.
Ugalde, J. A., Chang, B.S.W., and M.V. Matz. 2004. Evolution of coral pigments recreated. Science **305**:1433.
Verkhusha, V. V., and K. A. Lukyanov. 2004. The molecular properties and applications of Anthozoa fluorescent proteins and chromoproteins. Nature Biotechnology **22**:289-296.
Wall, M. A., M. Socolich, and R. Ranganathan. 2000. The structural basis for red fluorescence in the tetrameric GFP homolog DsRed. Nature Structural Biology **7**:1133-1138.
Ward, W. W. 2002. Fluorescent proteins: who's got 'em and why? Pages 123–126 *in* P. E. S. a. L. J. Kricka, editor. Bioluminescence and chemiluminescence. World Scientific, Cambridge, UK.
Watt, W. B. 1990. The evolution of animal coloration- adaptive aspects from bioenergetics to demography. *in* M. Wicksten, editor. Adaptive coloration in invertebrates: proceedings of a symposium sponsored by the American Society of Zoologists. Texas A&M University Sea Grant College Program, College Station, Texas.
West, H. H. 1979. Pigmentation in the sea anemone *Corynactis californica*. Comparative Biochemistry and Physiology B **64**:195-200.
Wicksten, M. 1989. Why are there bright colors in sessile marine invertebrates? Bulletin of Marine Science **45**:519-530.
Wiedenmann, J., C. Elke, K.-D. Spindler, and W. Funke. 2000. Cracks in the beta-can: fluorescent proteins from *Anemonea sulcata*. Proc. Natl. Acad. Sci. **97**:14091-14096.
Wiedenmann, J., S. Ivanchenko, F. Oswald, F. Schmitt, C. Rocker, A. Salih, K.-D. Spindler, and a. N. G.U. 2004. EosFP, a fluorescent marker protein with UV-inducible green-to-red fluorescence conversion. Proc. Nat. Acad. Sci. **101**:15905-15910.
Wiedenmann, J., Ivanchenko, S, Oswald, F., and Nienhaus GU. 2004a. Identification of GFP-like proteins in non-bioluminescent, azooxanthellate Anthozoa opens new perspectives for bioprospecting. Marine Biotechnology **6**:270-277.

Wiedenmann, J., A. Schenk, C. Röcker, A. Girod, and K.-D. Spindler. 2002. A far-red fluorescent protein with fast maturation and reduced oligomerization tendency from *Entacmaea quadricolor* (Anthozoa, Actiniaria). Proc. Natl. Acad. Sci. **99**:11646-11651.

Wiedenmann, J., Vallone, B., Renzi, F., Nienhaus, K., Ivanchenko, S., Röcker, C., and Nienhaus, G.U. 2004b. Red fluorescent protein eqFP611 and its genetically engineered dimeric variants. Journal of Biomedical Optics **10**:1-7.

Williams, A. T. R., S. A. Winfield, and J. N. Miller. 1983. Relative fluorescence quantum yields using a computer-controlled luminsecence spectrometer. Analyst **108**:1067-1071.

Wilmann, P. G., J. Petersen, A. Pettikiriarachchi, A. M. Buckle, S. C. Smith, S. Olsen, M. A. Perugini, R. J. Devenish, M. Prescott, and J. Rossjohn. 2005. The 2.1 Å crystal structure of the far-red fluorescent protein HcRed: inherent conformational flexibility of the chromophore. J. Mol. Biol. **In Press**.

Woodbridge, R. G. I. 1961. A monograph on fluorescent mollusk shells of the seas, the shores, the fresh waters and the lands of the world. A review of the literature from earliest times to 1960 with original observations. Transspace Laboratory, Princeton Junction, N.J.

Yakhnin, A. V., L. M. Vinokurov, A. K. Surin, and Y. B. Alakhov. 1998. Green fluorescent protein purification by organic extraction. Protein Expression and Purification **14**:382–386.

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.

Yanushevich, Y. G., M. E. Bulina, N. G. Gurskaya, A. P. Savitskii, and K. A. Lukyanov. 2002. Key amino acid residues responsible for the color of green and yellow fluorescent proteins from the coral polyp *Zoanthus* sp. Russian Journal of Bioorganic Chemistry **28**:274-277.

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. Proceedings of the National Academy of Sciences USA **98**:462-467.

Zhang, J., R. E. Campbell, A. Y. Ting, and R. Y. Tsien. 2002. Creating new fluorescent probes for cell biology. Nature Reviews: Molecular Cell Biology **3**:906-918.

Zhu, Y. Y., E. M. Machleder, A. Chenchik, R. Li, and P. M. Siebert. 2001. Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction. BioTechniques **30**:892-897.

Table 1. Fluorescent protein genes cloned previously from tropical corallimorpharian species. U denotes that the information has not been published.

Species Name (Protein Name)	GenBank Accession Number	Excitation λ_{max} (nm)	Emission λ_{max} (nm)	Reference
Discosoma striata (dstrGFP)	AF168420	456	484	Matz et al. 1999
Discosoma sp. 1 (discRFP; DsRed)	AF168419	558	583	Matz et al. 1999
Discosoma sp. 2 (dis2RFP)	AF272711	573	593	Fradkov et al. 2000
Discosoma sp. 3 (dis3GFP)	AF420593	503	512	Labas et al. 2002
Ricordea florida	AY037774	506	512	Matz and Lukyanov 2001(U)
(rfloGFP)	AY037772	508	518	Labas et al. 2002
(rfloRFP)	AY037773	506; 566	517; 574	Labas et al. 2002

Table 2. Emission spectra taken from temperate non-bioluminescent azooxanthellate anthozoans from the Pacific Coast of the United States (MRY= Monterey; SB= Santa Barbara).

Species (Order)	Origin	Colors	Emission λ_{max} (nm)
Anthopleura artemisia (Actiniaria)	SB	Pink	595
Astrangia lajollaensis (Scleractinia)	MRY	Orange	563
Balanophyllia elegans (Scleractinia)	MRY	Orange	576
Corynactis californica (Corallimorpharia)	MRY	Red	605
Pachycerianthus fimbriatus (Ceriantharia)	SB	Orange	569
Paracyathus stearnsii (Scleractinia)	MRY	Green; Orange	510; 561

Table 3. Quantum yield (Q) of ccalRFP1 and ccalYFP1 in relation to known standards.

	Known Q	**Calc Q (Fluorescein)**	**Calc Q (EGFP)**	**Mean Q**
ccalYFP1	---	0.75	0.79	**0.77**
ccalRFP1	---	0.45	0.47	**0.46**
rEGFP	0.60	0.57	---	
Fluorescein	0.79	---	0.83	
DsRed	0.29	---	---	

Table 4. Nucleotide and amino acid similarity of ccalYFP1 and ccalRFP1 in relation to homologous FPs.

	Nucleotide % Similarity		Amino Acid % Similarity	
***Species* (Protein)**	**ccalYFP1**	**ccalRFP1**	**ccalYFP1**	**ccalRFP1**
Corynactis californica (ccalYFP1)	—	86	—	80
C. californica (ccalRFP1)	86	—	80	—
Anemonia majano (amajGFP)	70	70	61	64
Zoanthus sp. (zoanYFP)	63	62	55	55
Ricordea florida (rfloRFP)	59	59	47	46
Discosoma sp. (DsRed)	58	58	47	46
Aequorea victoria (GFP)	45	46	29	28

Table 5. Fluorescent protein genes included in the phylogenetic analyses. U denotes that the information has not been published. *Note: dendRFP was originally published as a green protein, dendGFP, by Labas et al. 2002.

Species Name **Protein Name (Other Protein Name)**	**GenBank Accession Number**	**Excitation λ_{max} (nm)**	**Emission λ_{max} (nm)**	**(Reference)**
Anemonia majano				
amajGFP	AF168421	458	486	(Matz et al. 1999)
amajCyFP	AX824719	U	U	(Patent WO02068459)
Anemonia sulcata				(Leutenegger and
asulGFP	AF545827	U	U	Wiedenmann 2002; U)
asuCP	AF322222	N/A	N/A	(Wiedenmann et al. 2000)
Astrangia lajollaensis				
alajGFP1	AY508123	509	517	(Bessette and Daugherty 2004)
alajGFP2	AY508124	509	517	(Bessette and Daugherty 2004)
alajGFP3	AY508125	494	504	(Bessette and Daugherty 2004)
Dendronephthya sp.				
dendRFP*	AF420591	U	U	(Shagin et al. 2004)
Discosoma sp. 1				
discRFP (DsRed)	AF168419	558	583	(Matz et al. 1999)
Discosoma sp. 2				
dis2RFP	AF272711	573	593	(Fradkov et al. 2000)
Discosoma sp. 3				
dis3GFP	AF420593	503	512	(Labas et al. 2002)
Discosoma striata				
dstrCyFP	AF168420	456	484	(Matz et al. 1999)
Entacmaea quadricolor				
equaRFP	AY130757	559	611	(Wiedenmann et al. 2002)
Goniopora tenuidens				
gtenCP	AF383156	N/A	N/A	(Gurskaya et al. 2001)
Lobophyllia hemprichii				
lhemRFP (EosFP)	AY765217	571	581	(Wiedenmann et al. 2004)
Montastraea cavernosa				
mcavGFP2	AY679111	U	U	(Carter et al. 2004; U)
mcavRFP	AY037770	572	580	(Labas et al. 2002)
mcavCyFP	AY056460	431	474; 497	(Sun et al. 2004)
Ricordea florida				
rfloGFP	AY037772	508	518	(Labas et al. 2002)
rfloRFP	AY037773	506; 566	517; 574	(Labas et al. 2002)
Scolymia cubensis				
scubGFP1	AY037767	497	506	(Labas et al. 2002)
scubRFP1	AY646064	U	576	(Meleshkevitch et al. 2004; U)
Trachyphyllia geoffroyi				
tgeoRFP (Kaede)	AB085641	572	582	(Ando et al. 2002)
Zoanthus sp.				
zoanGFP	AF168422	U	506	(Matz et al. 1999; U)
zoanYFP	AF168423	U	538	(Matz et al. 1999)
zoanRFP	AY059642	552	576	(Yanushevich et al. 2001; U)

Table 6. Collection location and depth information for seven different color morphs of *Corynactis californica.*

Color Morph	*Collection Location*	*Collection Depth (ft.)*
Pink2	Carmel, CA	0
Orange2	Carmel, CA	0
Lavender	Aumentos Reef, Pacific Grove, CA	50
Pink1	Aumentos Reef, Pacific Grove, CA	50
Orange1	Aumentos Reef, Pacific Grove, CA	50
Orange3	Aumentos Reef, Pacific Grove, CA	50
Red	Outer Pinnacles, Carmel, CA	90

Table 7. Primer sets, designed from *C. californica* fluorescent protein genes and gene fragments, used in this study. Primers are named based on ccalYFP1 and ccalRFP1 sequence numbering. T(°C) is the annealing temperature in degrees Celsius used in PCR with the primers. Sequences are given in a 5' to 3' orientation. IUPAC symbols are used for ambiguous nucleotides.

A) Degenerate primers used to amplify new FP gene fragments

Name	*T (°C)*	*Sequence*
ccal48F	62	TCATATGGATGGGTGTGTCAACG
ccal647R	62	GCAACAGCGTGTTCTTKGAKTTGA
ccal1F	50	ATGTCTCWWTCAAAGCARGTKMTC
ccal666R	50	TYRYRCTRTGAAGYRTTTYGCAAC

B) RACE primers provided with SMART RACE kit.

Name	*T (°C)*	*Sequence*
BD Smart II A	70	AAGCAGTGGTATCAACGCAGAGTACGCGGG
5′-CDS primer	70	AAGCAGTGGTATCAACGCAGAGTAC$(T)_{30}$VN
3′-CDS primer A	70	$(T)_{25}$VN
Universal Primer A Mix (UPM)	70	Long: CTAATACGACTCACTATAGGGCAAGCAGTGG TATCAACGCAGAGT Short: CTAATACGACTCACTATAGGGC

C) Gene-specific RACE primers based on E56 and G34 cDNA fragments

Name	*T (°C)*	*Sequence*
Acory510R-5′	70	CCAGCATGAGAAACATGGTCACATCGCC
Acory156F-3′	70	GGGAGGACCACTTCCGTTCGCCTTTGAT
Bcory472R-5′	70	CCATCACAGGGAGTTATTAGCTCGCTGGA
Bcory156F-3′	70	GGAGGGCCACTTCCATTCGCCTTTGAT

Table 7. (Continued)

D) Non-coding region degenerate primers based on regions flanking ccalYFP1 and ccalRFP1

Name	*T (°C)*	*Sequence*
OUT5'Cory	45, 43	CTTCSTTTTSYRACTYMAAACATG
OUT3'Cory	43	TTKRTAWYWRTTTACGYYTCCTTG

E) Gene-specific Ends primers based on ccalFP-A, -B, and -C cDNA fragments to generate cDNA ends

Name	*T (°C)*	*Sequence*
Acory375R-5'	45	GTGGATTGGTGCACAAAG
Bcory362R-5'	43	TAAAGCAGTCCTTATCGAGG
Ccory354R-5'	43	AAAGCCGTTCTTATCAAGG
Ccory336F-3'	43	CCTTGATAAGAACGGC

F) Gene-specific primers used to amplify full-length products for cloning into pCR2.1-TOPO.

Name	*T (°C)*	*Sequence*
ccalYFP-F	54	ATGTCTCATTCAAAGCAGGTGATC
ccalYFP-R	54	TGCTGTGAAGCATTTTGCAAC
ccalRFP-F	60	ATGTCTCTATCAAAGCAAGTTCTCC
ccalRFP-R	60	CACTATGAAGTGTTTTGCAACAGC
ccalFP-A-F	60	ATGTCTCTATCAAAGCAAGTTCTCCCACATG
ccalFP-A-R	60	CGTTTCCTTGACATCTACTGTGAAGTGTTTTG
ccalFP-B-F	60	ATGTCTCTATCTAAACAAGTTCTTCCAC
ccalFP-B-R	60	CGCCTCCTTGATTTGCAC
ccalFP-C-F	66	TCTCTATCAAAGCAAGTGGTCAAAGAAGACATGAAG
ccalFP-C-R	66	CATACTGTGAAGTATTTTGCAACAGCGTGTTCTTG

Table 7. (Continued)

G) Gene-specific primers for amplifying full-length products for cloning into pQE vectors. Gene-specific parts of the primer are underlined. NcoI and NotI enzyme sites are shown in bold.

Name	*T (°C)*	*Sequence*
		TACAGT**CCATGG**AA
ccalYFP-F-pQE	53	TCTCATTCAAAGCAGGTGATCACACAG
		ACATA**GCGGCCGC**ATTA
ccalYFP-R-pQE	53	TGCTGTGAAGCATTTTGCAACAGCG
		TACAGT**CCATGG**AA
ccalRFP-F-pQE	51	TCTCTATCAAAGCAAGTTCTCCCAC
		ACATA**GCGGCCGC**ATTA
ccalRFP-R-pQE	51	CACTATGAAGTGTTTTGCAACAGCG
ccalFP-A-F-pQE		TACAGT**CCATGG**AA
(ccalRFP-F-pQE)	51	TCTCTATCAAAGCAAGTTCTCCCAC
		ACATA**GCGGCCGC**ATTA
ccalFP-A-R-pQE	51	CGTTTCCTTGACATCTACTGTGAAG
		TACAGT**CCATGG**AA
ccalFP-B-F-pQE	57	TCTCTATCTAAACAAGTTCTTCCACGAGACGTGAAAATG
		ACATA**GCGGCCGC**ATTA
ccalFP-B-R-pQE	57	CGCCTCCTTGATTTGCACTGTGAAGTGTTTC
		TACAGT**CCATGG**AA
ccalFP-C-F-pQE	60	TCTCTATCAAAGCAAGTGGTCAAAGAAGACATGAAGATGAC
		ACATA**GCGGCCGC**ATTA
ccalFP-C-R-pQE	60	TACTGTGAAGTATTTTGCAACAGCGTGTTCTTGGATTTGAACAG

Table 7. (Continued)

H) Gene-specific primers for amplifying specific FPs from each color morph.

Name	*T (°C)*	*Sequence*
ccalYFP-F	53, 54	ATGTCTCATTCAAAGCAGGTGATC
ccalYFP-R	53, 54	TGCTGTGAAGCATTTTGCAAC
ccalRFP-F	60	ATGTCTCTATCAAAGCAAGTTCTCC
ccalRFP-R	60	CACTATGAAGTGTTTTGCAACAGC
ccalFP-A-25F	53	CCACATGACGTGCGGATG
ccalFP-A-620R	53	CCGTCGCTATCTTCGCCTTTC
ccalFP-B-23F	60, 64	TTCCACGAGACGTGAAAATGCGCTT
ccalFP-B-326R	60, 64	GCACTGGCCGTGCTACACGCC
ccalFP-C-21F	47	GGTCAAAGAAGACATGAAGATGAC
ccalFP-C-438R	47	GGAAGGTTCCCAATTCACTC

I) Sequencing primers used in this study

Name	*Sequence*
M13F	GTAAAACGACGGCCAG
M13R	CAGGAAACAGCTATGAC
pQEF	CGGATAACAATTTCACACAG
pQER	GTTCTGAGGTCATTACTGG

Table 8. Details of expression vectors used in this study. TEV protease is Tobacco Etch Virus protease and is tag from purified protein for downstream applications.

Name	*Features*	*Restriction Enzyme Cloning Sites*
pQE31XN-TEV-DsRed.T3	TEV protease site; N-terminal 6xHIS tag; IPTG inducible	NcoI/NotI
pQE31XN-DsRed.T3	N-terminal 6xHIS tag; IPTG inducible	NcoI/NotI

Table 9. Mean fluorescence emission wavelength from specific tissues of seven *C. californica* color morphs. A maximum emission wavelengths per peak per tissue location for each color morph with standard deviation (SI the sample size is given as (n). Major emission peaks are shown in bold; minor peaks are in plain typeface. En boxes were calculated from n = 175.

	Color Morph λ_{max} *(nm)* ± SD						
	Pattern A				*Pattern B*		
Tissue Location (n)	***Pink2***	***Orange1***	***Orange2***	***Red***	***Lavender***	***Pink1***	***Orang***
Mouth (5)	**599 ± 0.4**	**600 ± 1.5**		**601 ± 1.3**	**597 ± 1.2**		
			565 ± 1.1			**567 ± 1.5**	**568 ±**
	516 ± 1.0	518 ± 1.2	517 ± 1.8	515 ± 1.0	514 ± 0.9	515 ± 1.1	517 ±
Oral Disc (5)	**599 ± 0.8**	**599 ± 1.5**		**600 ± 1.7**	**596 ± 1.0**	596 ± 2.0	
		563	**562 ± 2.0**		563 ± 1.7	**565 ± 3.6**	560
	517 ± 1.1	520 ± 1.4	520 ± 1.2	515 ± 0.6	515 ± 1.7	517 ± 0.7	**520 ±**
Base of Tentacles (5)	**599 ± 1.8**	**599 ± 1.9**		**600 ± 1.5**		595	
		563	**561 ± 0.1**		**562 ± 1.4**	**563 ± 1.2**	**562 ±**
	519 ± 0.2	520 ± 1.0	521 ± 0.4	516 ± 1.7	518 ± 1.0	519 ± 1.5	517 ±
Tentacles (5)	**599 ± 1.0**	**599 ± 2.0**		600 ± 2.6	595		
		561	**562 ± 1.0**		**565 ± 8.2**	**562 ± 0.5**	**562 ±**
	519	522 ± 2.3	521 ± 1.0	518 ± 1.2	521 ± 3.4	519 ± 0.5	519 ±
Body Column (5)	**600 ± 1.8**	**600 ± 1.7**		**603 ± 1.6**	594 ± 4.1	**600 ± 2.8**	
			561 ± 0.8		**563 ± 1.8**	563	**564 ±**
	518 ± 1.2		521		521 ± 4.0	516 ± 1.4	520 ±
All Tissues (25)	**599 ± 1.3**	**599 ± 1.7**		**601 ± 2.2**	**596 ± 2.6**	**598 ± 3.0**	
		562 ± 0.5	**562 ± 2.0**		**563 ± 4.4**	**564 ± 2.5**	**564 ±**
	517 ± 1.6	520 ± 1.8	520 ± 1.8	516 ± 1.5	517 ± 3.4	517 ± 2.1	**518 ±**

Table 10. Details of the source color morph template used in PCR and the method that provided sequence information for novel FP genes. Products with an asterisk (*) were generated in Chapter 1.

PCR Product	*Color Morph*	*Method Used*
ccalYFP1 full-length gene*	Red	cDNA library screen
ccalRFP1 full-length gene*	Red	cDNA library screen
ccalFP-A partial gene fragment	Pink2	Degenerate PCR
ccalFP-A 5′ end	Pink1	Degenerate non-coding region PCR
ccalFP-A 3′ end	Pink1	SMART RACE
ccalFP-A full-length gene	Pink1	Gene-specific PCR
ccalFP-B partial gene fragment	Orange2	Degenerate PCR
ccalFP-B 5′ end	Orange2	Degenerate non-coding region PCR
ccalFP-B 3′ end	Orange2	SMART RACE
ccalFP-B full-length gene	Pink2	Gene-specific PCR
ccalFP-C partial gene fragment	Orange3	Degenerate PCR
ccalFP-C 5′ end	Lavender, Pink1	Degenerate non-coding region PCR
ccalFP-C 3′ end	Pink1	Degenerate non-coding region PCR
ccalFP-C full-length gene	Lavender	Gene-specific PCR

Table 11. Cloning results. Numbers indicate how many clones resulted from degenerate primer PCR using cDNA templates. The number of independent PCR reactions appears in parentheses after each number of clones. GS indicates that a certain FP gene was cloned from cDNA using gene-specific primers. An X indicates that all attempts to amplify a particular gene using degenerate and gene-specific primers on cDNA and specific primers on gDNA failed.

	FP gene							
Color Morph	***ccalRFP1***	***ccalYFP1***	***ccalRFP2***	***ccalOFP1***	***ccalGFP1***	***Total # of Clones Per Morph***	***Collection Depth (ft.)***	***Pattern***
Pink2	50 (4)	1	1	1	X	53	0	A
Orange2	7 (1)	6 (2)	3 (1)	2 (1)	X	18	0	A
Orange1	15 (3)	GS	X	GS	X	15	50	A
Red	45 (5)	GS	X	1	X	46	90	A
Lavender	GS	X	17 (2)	GS	17 (3)	34	50	B
Pink1	2 (1)	X	22 (3)	GS	2 (1)	26	50	B
Orange3	1	X	14 (3)	GS	3 (2)	18	50	B
Total # Clones per FP gene	120	7	57	4	22	210		

Table 12. Nucleotide and amino acid similarity of *Corynactis californica* fluorescent protein genes in relation to homologous FPs.

	Nucleotide % Similarity					*Amino Acid % Similarity*				
***Species* (Protein)**	**ccalYFP1**	**ccalRFP1**	**ccalRFP2**	**ccalOFP1**	**ccalGFP1**	**ccalYFP1**	**ccalRFP1**	**ccalRFP1**	**ccalOFP1**	**ccalGFP1**
Corynactis californica (ccalYFP1)	—	85	85	84	86	—	77	77	79	80
C. californica (ccalRFP1)	85	—	90	90	87	77	—	85	87	83
C. californica (ccalRFP2)	85	90	—	89	86	77	85	—	85	81
C. californica (ccalOFP1)	84	90	89	—	87	79	87	85	—	85
C. californica (ccalGFP1)	86	87	86	87	—	80	83	81	85	—
Anemonia majano (amajGFP)	70	70	68	67	70	61	63	60	60	62
Zoanthus sp. (zoanYFP)	62	62	61	61	63	56	54	53	52	55
Ricordea florida (rfloRFP)	58	58	56	57	60	48	49	47	47	49
Discosoma sp. (DsRed)	57	56	57	56	55	46	46	45	45	43
Entacmaea quadricolor (equaRFP)	57	56	56	56	56	47	47	46	46	47
Renilla reniformis (rrenGFP)	47	49	47	48	48	33	34	32	35	33

Figure 1a. Nucleotide sequence of ccalYFP1 (666 base pairs)

```
  1 ATGTCTCATT CAAAGCAGGT GATCACACAG GAGATGAAGA TGGTCTATCA TATGGATGGG
 61 TGTGTCAACG GACACTCCTT TACGATTGAA GGTGAAGGCA CTGGGAAACC ATACGAAGGA
121 AACCAGACTT TGAAACTGCG TGTCACCAAG GGAGGGCCAC TTCCATTCGC CTTTGATATT
181 TTGACGGCAA CGTTTTGTTA TGGAAACAGA TGCTTTTGTG AATATCCAGA AGACATGCCC
241 GACTACTACA AACAGTCATT CCCTGAAGGA TACTCATTTG AAAGGACTAT GATGTTCGAA
301 GACGGAGCGT GCTGCACTAC CAGTGTGCAT TTAAGCCTGA CTAAAAACTG CTTTGTGCAC
361 AACTCCACAT TTCACGGCGT CAACTTTCCT GCTAACGGAC CTGTGATGCA AAAGAAGACA
421 CTGAACTGGG AGCCTTCCAG CGAGAAAATA ACTCCCTTTG AGGGAAACTT GAAGGGCGAT
481 GTTACCATGT TTCTCAAGCT GGAAGGAGGT CAACAACACA GATGTCAATT CCAAACTACT
541 TACAAGGCAC ACAAGGCCGT CAAAATGCCA CCGAACCATA TCATAGAGCA CCGTCTTGTG
601 AGAAGCCAAG ATGGCGACGC AGTTCAACTC AAAGAACACG CTGTTGCAAA ATGCTTCACA
661 GCATGA
```

Figure 1b. Inferred amino acid sequence of ccalYFP1 (221 amino acids)

```
  1 MSHSKQVITQ EMKMVYHMDG CVNGHSFTIE GEGTGKPYEG NQTLKLRVTK GGPLPFAFDI
 61 LTATFCYGNR CFCEYPEDMP DYYKQSFPEG YSFERTMMFE DGACCTTSVH LSLTKNCFVH
121 NSTFHGVNFP ANGPVMQKKT LNWEPSSEKI TPFEGNLKGD VTMFLKLEGG QQHRCQFQTT
181 YKAHKAVKMP PNHIIEHRLV RSQDGDAVQL KEHAVAKCFT A*
```

Figure 2a. Nucleotide sequence of ccalRFP1 (666 base pairs)

```
  1 ATGTCTCTAT CAAAGCAAGT TCTCCCACGA GACGTCAAGA TGCGCTATCA TATGGATGGG
 61 TGTGTCAACG GACACCAGTT TATCATTGAA GGTGAAGGCA CTGGAAAACC ATACGAAGGA
121 AAAAAGATTT TGGAACTGCG AGTTACTAAA GGAGGGCCAC TTCCATTCGC CTTTGATATA
181 TTGTCCTCAG TGTTTACGTA TGGAAACAGA TGCTTTTGCG AGTATCCAGA AGACATGCCC
241 GACTATTTCA AACAGTCATT GCCTGAAGGA CACTCATGGG AACGAACTCT GATGTTCGAG
301 GACGGAGGGT GTGGCACAGC CAGTGCACAC ATAAGCCTTG ATAAAAACTG CTTTGTGCAC
361 AAATCCACAT TTCACGGCGT CAACTTTCCT GCTAACGGAC CTGTGATGCA AAAGAAGACC
421 CTGAACTGGG AGCCTTCCAG TGAGCTAATA ACTGCCGGTG ATGGAATACT GAAGGGCGAT
481 GTTACCATGT TTCTCATGCT GGAAGGAGGT CACCGCCTCA AATGTCAATT CACAACTTCT
541 TACAAGGCAA AGAAGGCTGT GAAAATGCCA CCGAACCATA TCATAGAACA CCGTCTTGTG
601 AGAAAGGAGG TTGCCGACGC TGTTCAAATC CAAGAACACG CTGTTGCAAA ACACTTCATA
661 GTGTGA
```

Figure 2b. Inferred amino acid sequence of ccalRFP1 (221 amino acids)

```
  1 MSLSKQVLPR DVKMRYHMDG CVNGHQFIIE GEGTGKPYEG KKILELRVTK GGPLPFAFDI
 61 LSSVFTYGNR CFCEYPEDMP DYFKQSLPEG HSWERTLMFE DGGCGTASAH ISLDKNCFVH
121 KSTFHGVNFP ANGPVMQKKT LNWEPSSELI TAGDGILKGD VTMFLMLEGG HRLKCQFTTS
181 YKAKKAVKMP PNHIIEHRLV RKEVADAVQI QEHAVAKHFI V*
```

Figure 3a and b. Excitation and emission spectra of ccalYFP1 and ccalRFP1

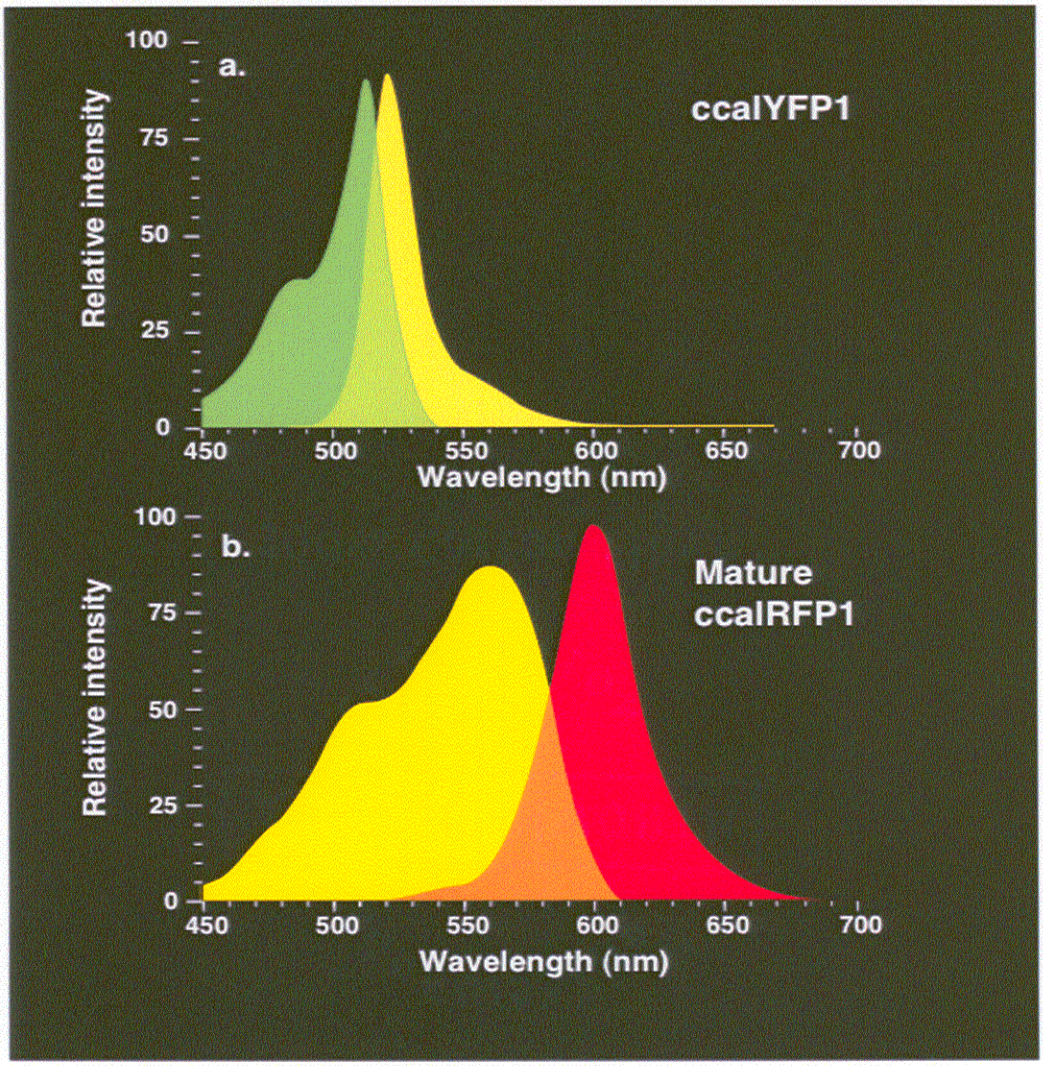


Figure 4. Maturation of ccalRFP1 emission spectra through time

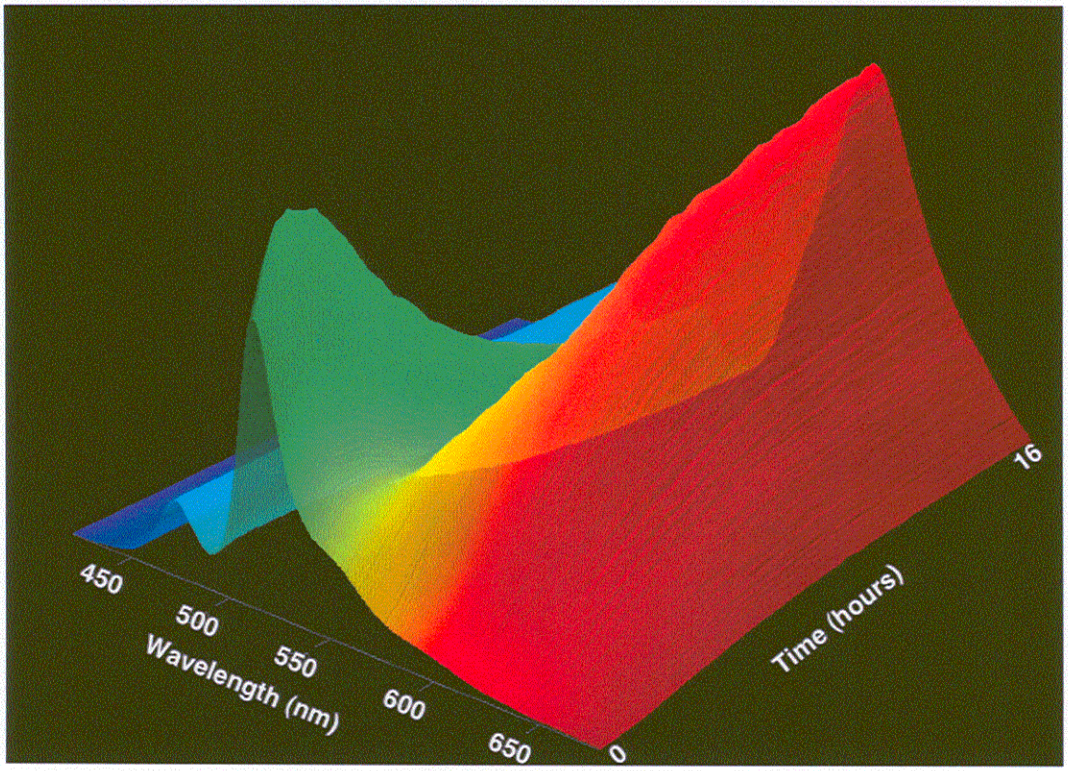
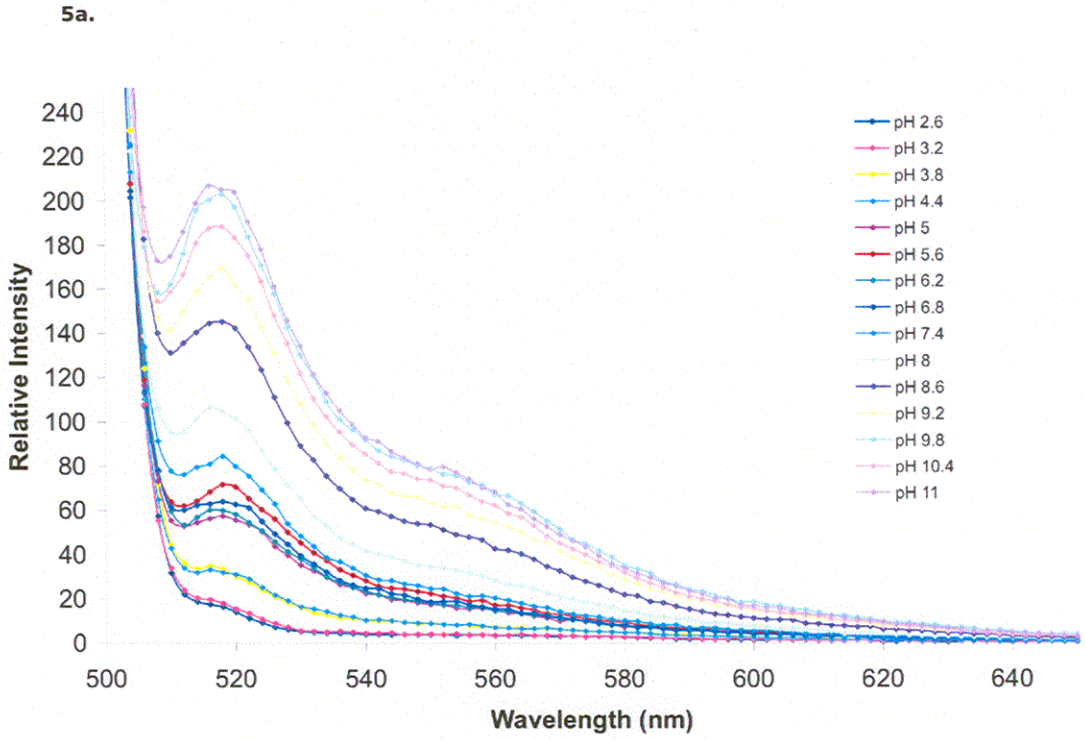


Figure 5a and b. Effect of pH on ccalRFP1 chromophore maturation – before (5a) and after (5b) overnight maturation.



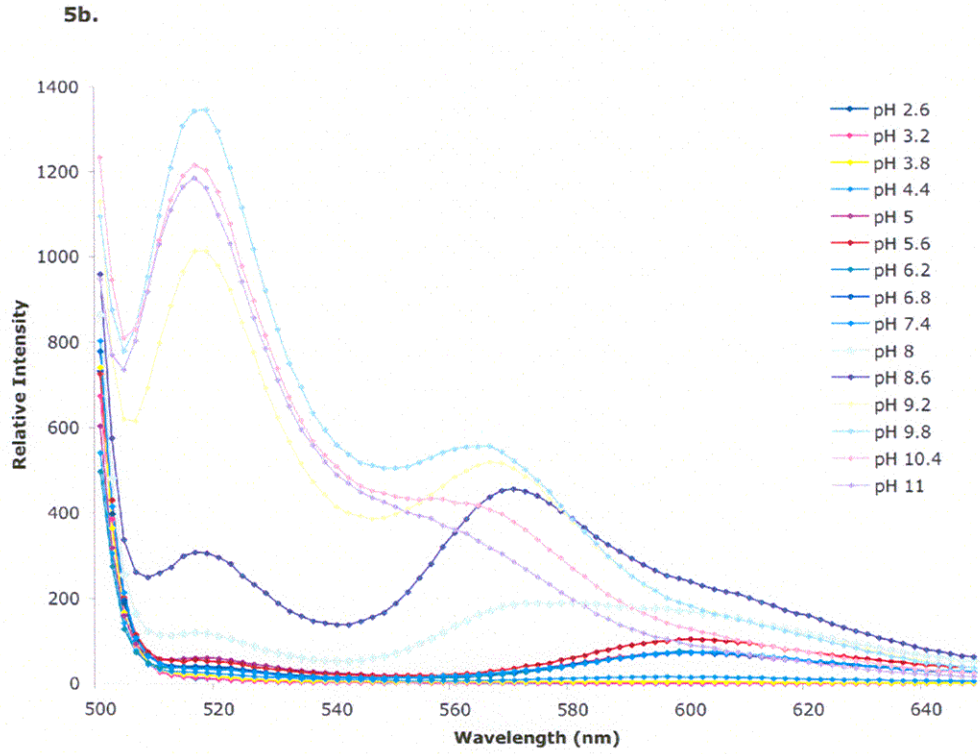


Figure 6a and b. Effect of salt concentration on ccalRFP1 chromophore maturation – before (6a) and after (6b) overnight maturation.

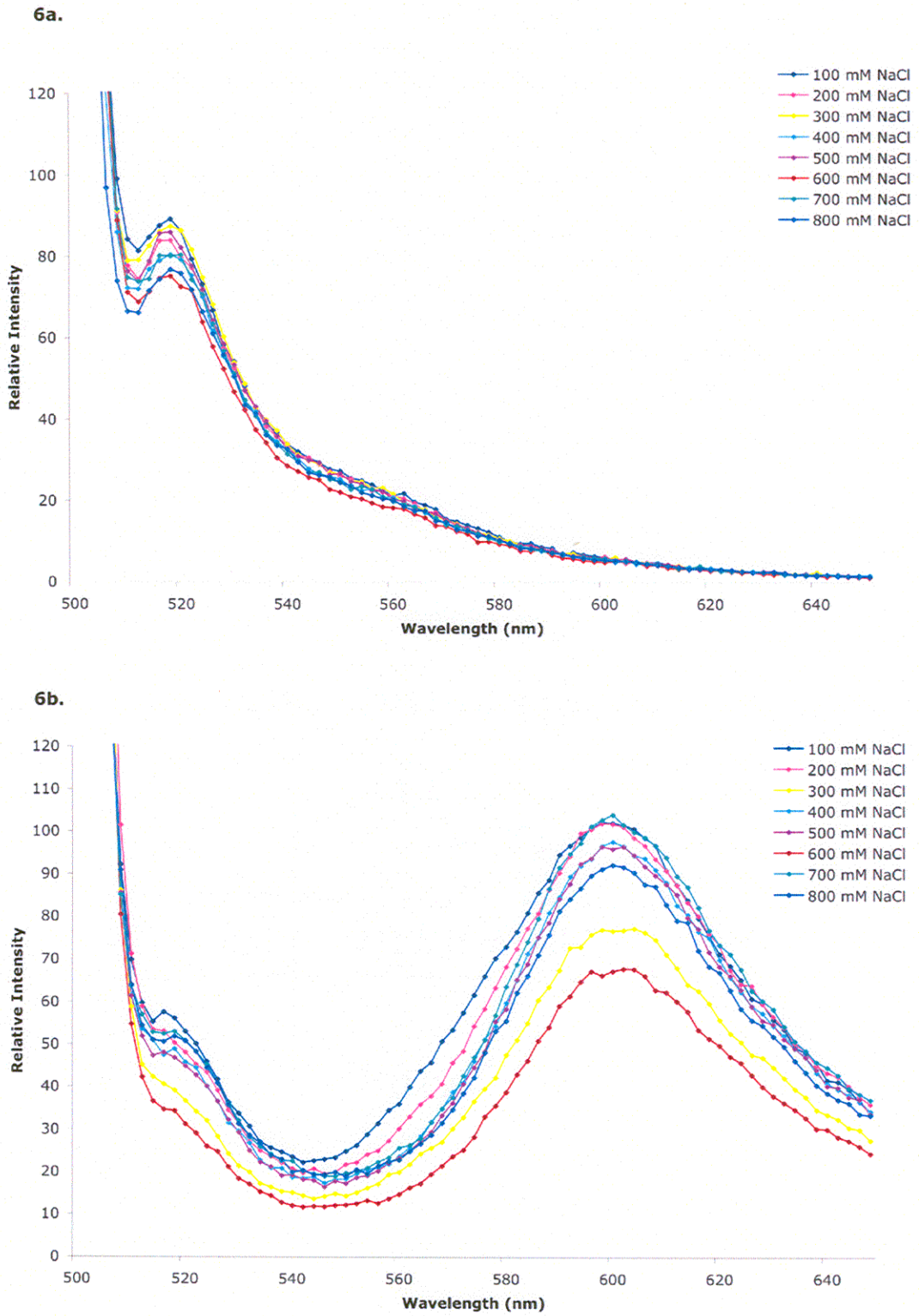


Figure 7. Amino acid alignment of ccalRFP1 and ccalYFP1 with homologous sequences shown relative to the original GFP sequence. A dot indicates identity with the GFP sequence and a dash indicates a gap in the alignment. Residues shaded in grey indicate the chromophore sequence.

```
                10        20        30        40        50        60        70        80        90
                |         |         |         |         |         |         |         |         |
GFP       MSKGEELFTG VVPILVELDG DVNGHKFSVS GEGEGDATY- GKLTLKFICT T---GKLPVP WPTLVTTFSY GVQCFSRYPD HMKQHDFFKS AMPEGYVQER
phiYFP    ..S.AL..H. KI.YV..ME. N.D..T..IR .K.Y...SV- ..VDAQ.... .---.DV... .S.....LT. .A...AK.GP EL.---.Y.. C.........
ccalRFP1  ..LSKQVLPR D.KMRYHM.. C....Q.IIE ...T.K-P.E ..KI.ELRV. K--G.P..FA FDI.SSV.T. .NR..CE..E D.P---.Y..Q SL...HSW..
ccalYFP1  ..HSKQVI.Q EMKMVYHM.. C....S.TIE ...T.K-P.E .NQ...LRV. K--G.P..FA FDI.TA..C. .NR..CE..E D.P---.YY.Q SF....SF..
alajGFP1  ..LSKQVIGK EMNMTYLM.. S....N.T.E ...D.K-P.E .HQC..LRI. K--GEP..FA FDI.TAA.C. .NR..VN..A EIA---.Y..Q .F...LTW..
amajGFP   .ALSNKFIGD DMKMTYHM.. C....Y.T.K ...N.K-P.E .TQ.ST.KV. MANG.P.AFS FDI.S.V.K. .NR..TA..T S.P---.Y..Q .F.D.MSY..
zoanYFP   .AHSKHGLKE EMTMKYHME. C......VIT ...I.Y-PFK ..Q.INLCVI E--G.P..FS EDI.SAG.K. .DRI.TE..Q DIV---.Y..N SC.A..TWG.
zoanRFP   .AHSKHGL.D DMTMHFRME. C.D....VIE .N.N.N-PFK ..QFINLCVI E--G.P..FS EDI.SAA.D. .NRL.TE..E GIV---.Y..N SC.A..TWH.
discRFP   .RSSKNVIKE FMRFK.RME. T....E.EIE .....R-P.E .HN.V.LKV. K--G.P..FA .DI.SPQ.Q. .SKVYVKH.A DIP---.YK.L SF...FKW..
tgeoRFP   ..-----.IKP EMK.KLLME. N....Q.VIE .D.K.H-PFE ..QSMDLVVK E--GAP..FA YDI.T.A.H. .NRV.AK... .IP---.Y..Q SF.K.FSW..

                100       110       120       130       140       150       160       170       180       190
                |         |         |         |         |         |         |         |         |         |
GFP       TIFFKDDGNY KTRAEVKFEG DTLVNRIELK GID--FKEDG NILGHKLEYN YNSHNVYIMA DKQKNGIKVN FKIRHNIEDG SVQLADHYQQ NTPIGDGPVL
phiYFP    ..T.EG..VF ......T..N GSVY..VK.N .QG---.K.. HV..KN..F. FTP.CL..WG .QANH.L.SA ...M.E.TGS KEDFIVADHT QMNTPI.GGP
ccalRFP1  .LM.E.G.CG TAS.HISLDK NCF.HK--ST FHGVN.PAN. PVMQK.TLNW EP.SELITAG .--GILKGDV TMFLMLEGGH RLKCQFTTSY KAKK---A.K
ccalYFP1  .MM.E.GACC T.SVHLSLTK NCF.HN--ST FHGVN.PAN. PVMQK.TLNW EP.SEKITPF E--G.LKGDV TMFLKLEGGQ QHRCQFQTTY KAHK---A.K
alajGFP1  SMT.E.G.FA AVS.DISLN. NCF.HQ--S. FVGVN.PA.. PVMQK.TVGW EP.TEKM.VR .--GILEGDV TMYLKLEGG. NYRCEYRTIY KAKK---D.K
amajGFP   .FTYE.G.VA TASW.ISLK. NCFEHK--ST FHGVN.PA.. PVMAK.TTGW DP.FEKMTVC .--GILKGDV TAFLMLQGG. NYRCQF.TSY K.KK---..T
zoanYFP   SFL.E.GAVC ICNVDITVSV KENCIYHKSI FNGMN.PA.. PVMKKMTTNW EA.CEKIMPV P..GILKGDV SMYLLLKDG. RYRCQFDTVY KAKS---VPS
zoanRFP   SFR.E.GAVC ICS.DITVNV RENCIYH.ST FYGVN.PA.. PVMKKMTTNW EP.CEKI.PI NS..ILKGDV SMYLLLKDG. RYRCQFDTIY KAKT---EPK
discRFP   VMN.E.G.VV TVTQDSSLQD GCF-IYKVKF I-GVN.PS.. PVMQK.TMGW EA.TERLYPR .--GVLKGEI H.ALKLKDG. HYLVEFKSIY MAKK---..Q
tgeoRFP   SLM.E.G.VC IATNDITLK. ..F-FNKVRF D-GVN.PPN. PVMQK.TLKW EA.TEKMYLR .--GVLTGDI TMALLLKG.V HYRCDFRTTY KSR----QEG

                200       210       220       230
                |         |         |         |
GFP       LPDNHYLSTQ SALSKDPNE- KRDHMVLLEF VTAAGITHGM DELY
phiYFP    VHVPE.HHIT YHVT-LSKDV TDHRDNMSLV E.VRAVDCRK TY--
ccalRFP1  M.P..IIEHR LVRKEVADAV QIQ------.H AV.KHFIV-- ----
ccalYFP1  M.P..IIEHR LVR.Q.GDAV QLK------.H AV.KCF.A-- ----
alajGFP1  M.KS.FVEHT LVRNNIKKDA .GNTI---.L IED.SAR--- ----
amajGFP   M.P..VVEHR I.RTDLDKGG NSVQLT--.H AV.HITSVVP ----
zoanYFP   KMPEWHFIQH KL.RE.RSDA .NQKWQ.T.H AI.FPSAL-- ----
zoanRFP   EMPDWHFIQH KLNRE.RSDA .NQKWQ.I.H AI.SRSAL-- ----
discRFP   ..GYY.VDSK LDITSHNEDY TIVEQY--.R TEGRHHLF-- ----
tgeoRFP   VKLPG.HFVD HCI.ILRHDK DYNEVK.Y.H AV.HSGLPDN V---
```

Figure 8. Consensus phylogram of all post burn-in trees of anthozoan fluorescent protein genes reconstructed by Bayesian analysis with posterior probabilities shown above or to the left of the branch where space permits.

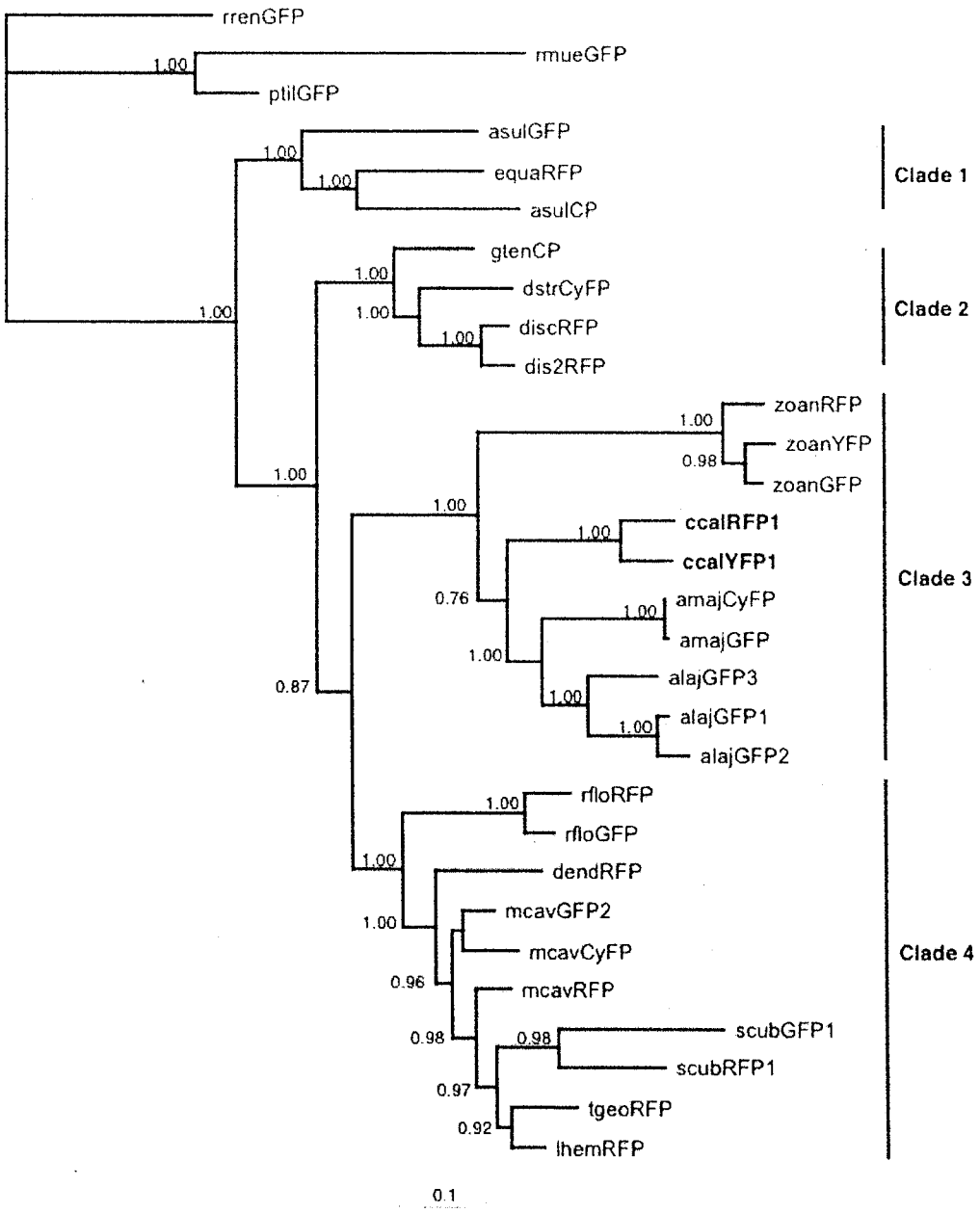


Figure 9. Consensus tree of all post burn-in trees of anthozoan fluorescent protein genes reconstructed by Bayesian analysis with posterior probabilities shown above or to the left of the branch and MP bootstrap support values shown in parentheses

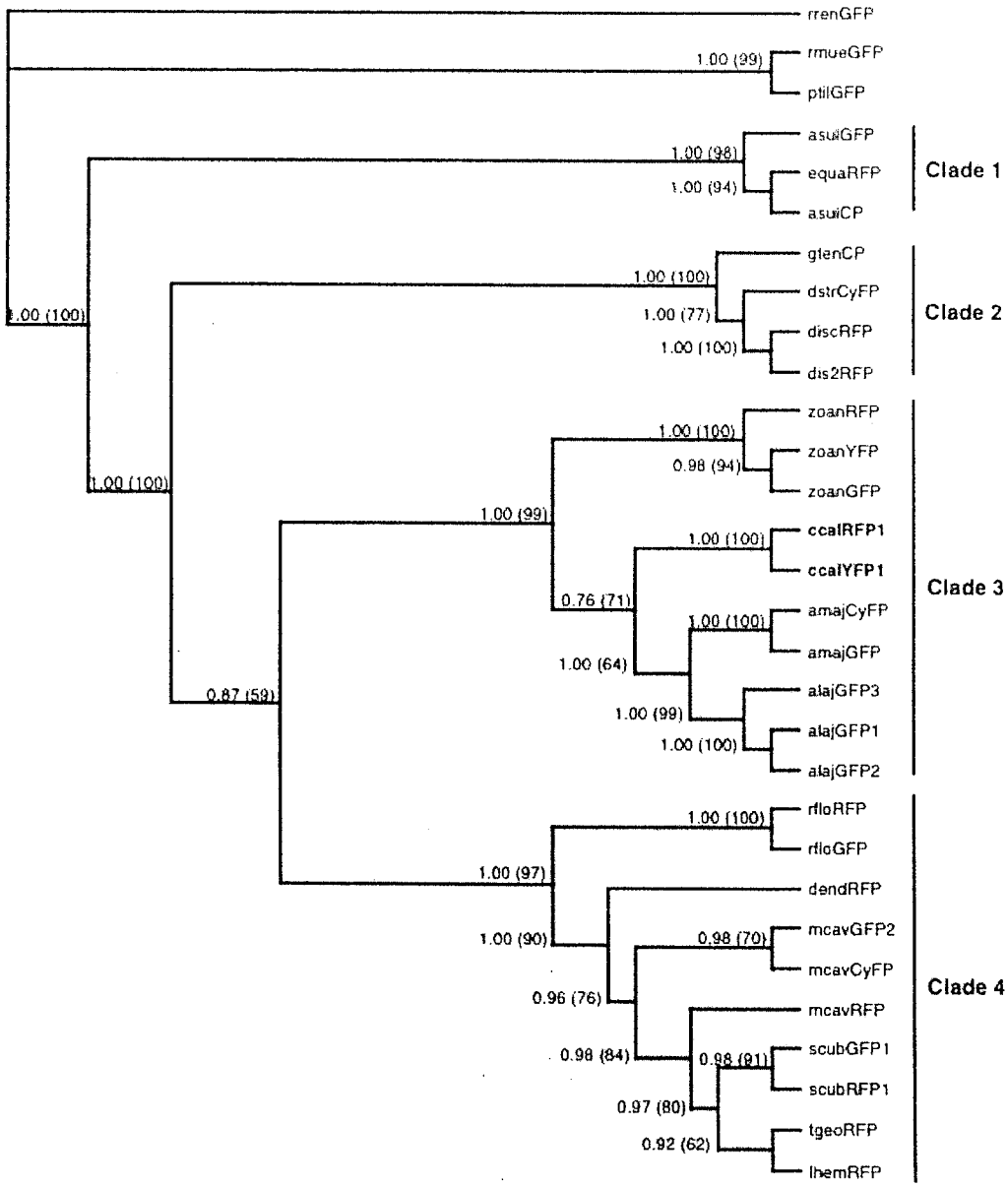
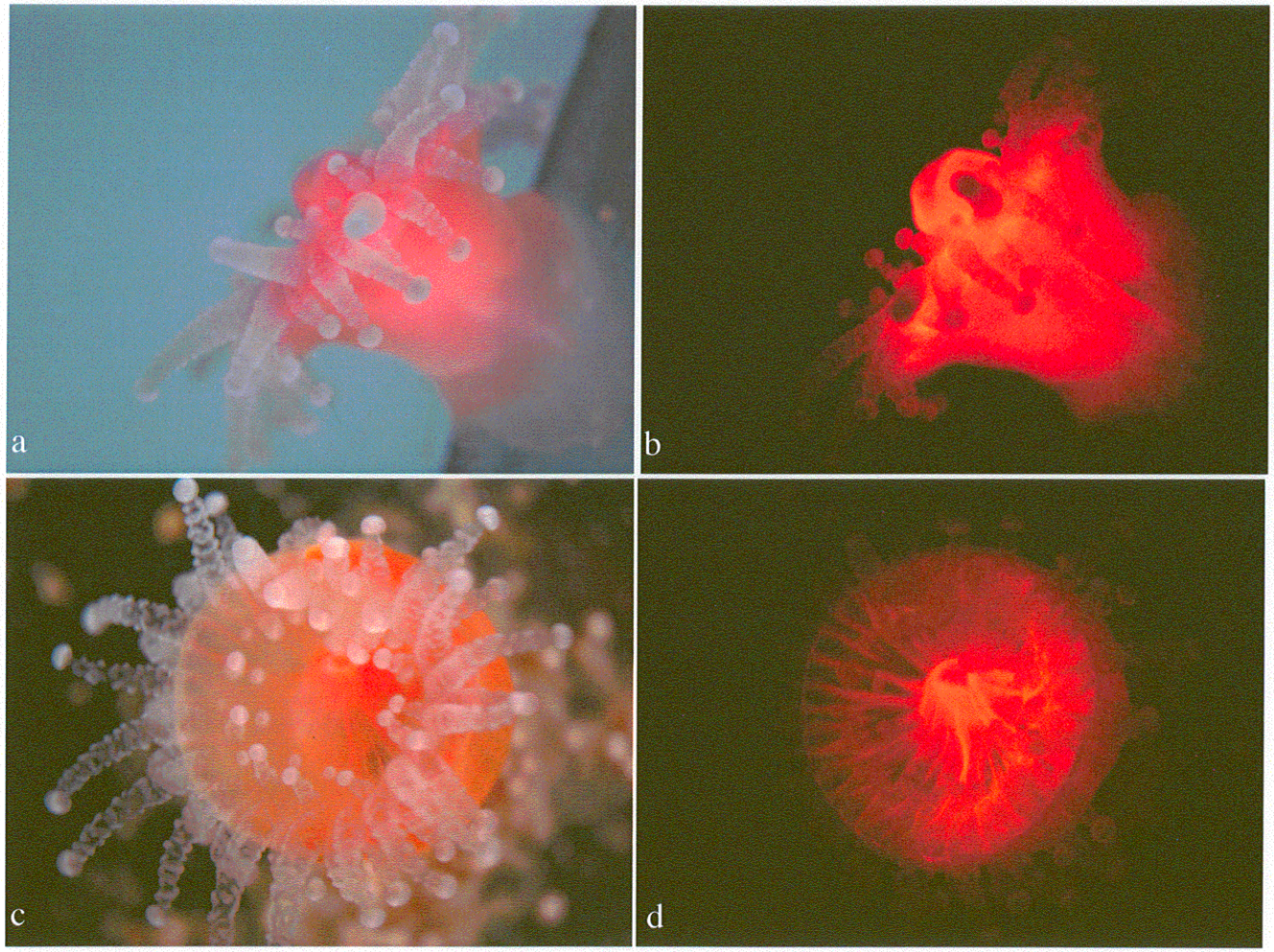
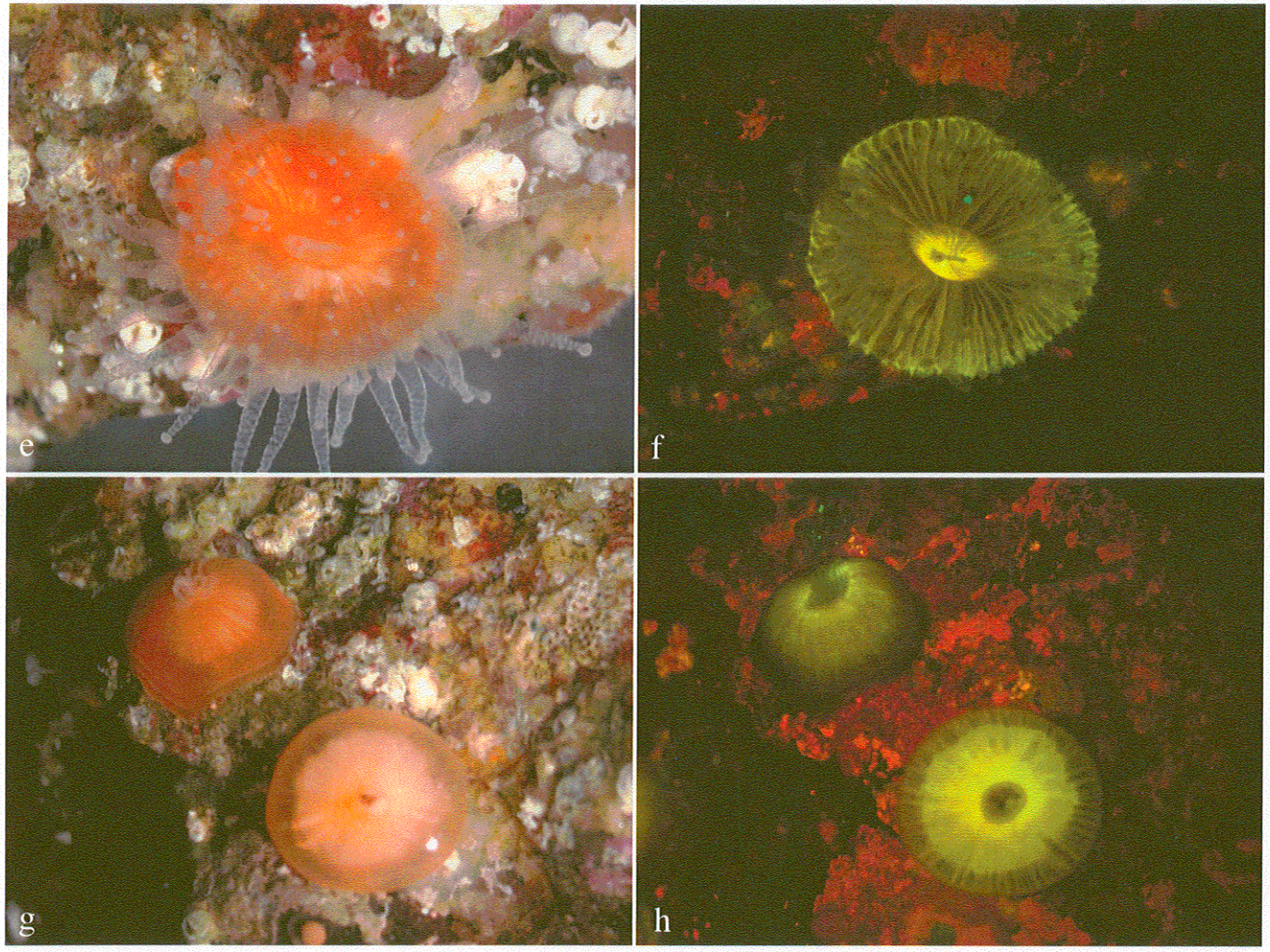


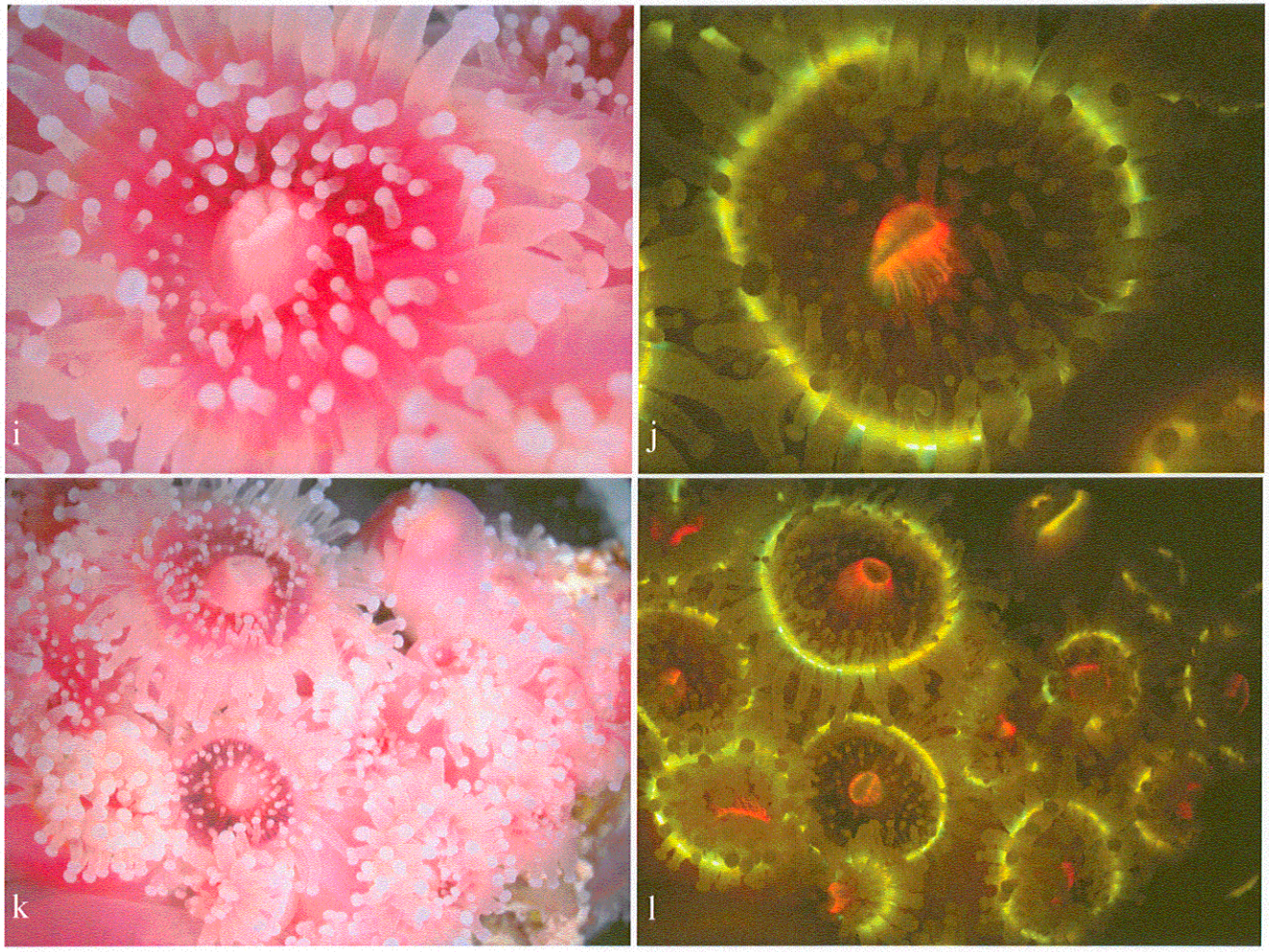
Figure 10. Paired images of *Corynactis californica* under white and blue light excitation. (a) Pink2 morph under white light (b) Pink2 morph fluorescing under blue light (c) Pink2 morph under white light (d) Pink2 morph fluorescing under blue light



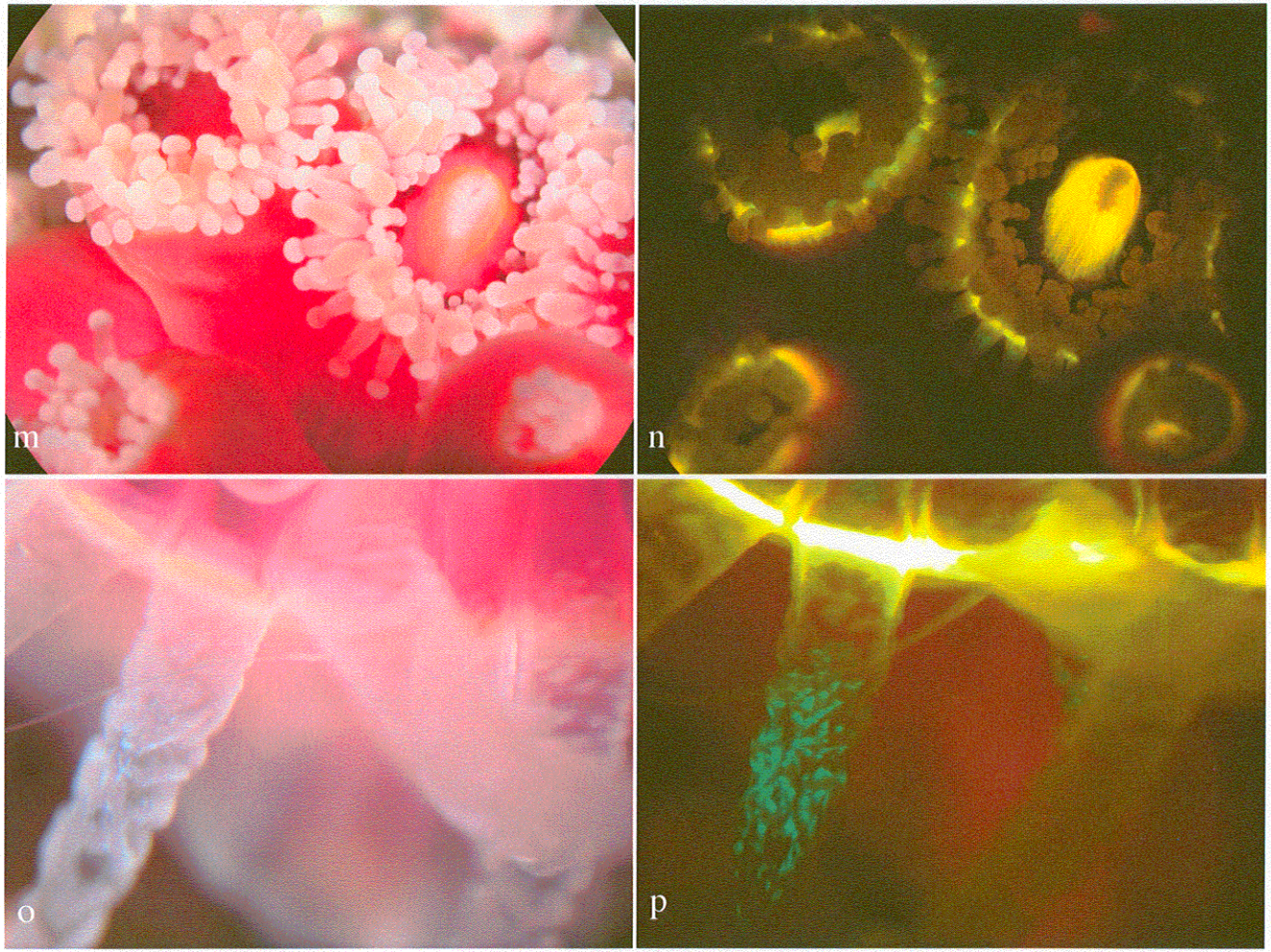
(e) Orange2 polyp under white light (f) Orange2 polyp fluorescing under blue light (g) Orange2 polyp under white light (h) Orange2 polyp fluorescing under blue light



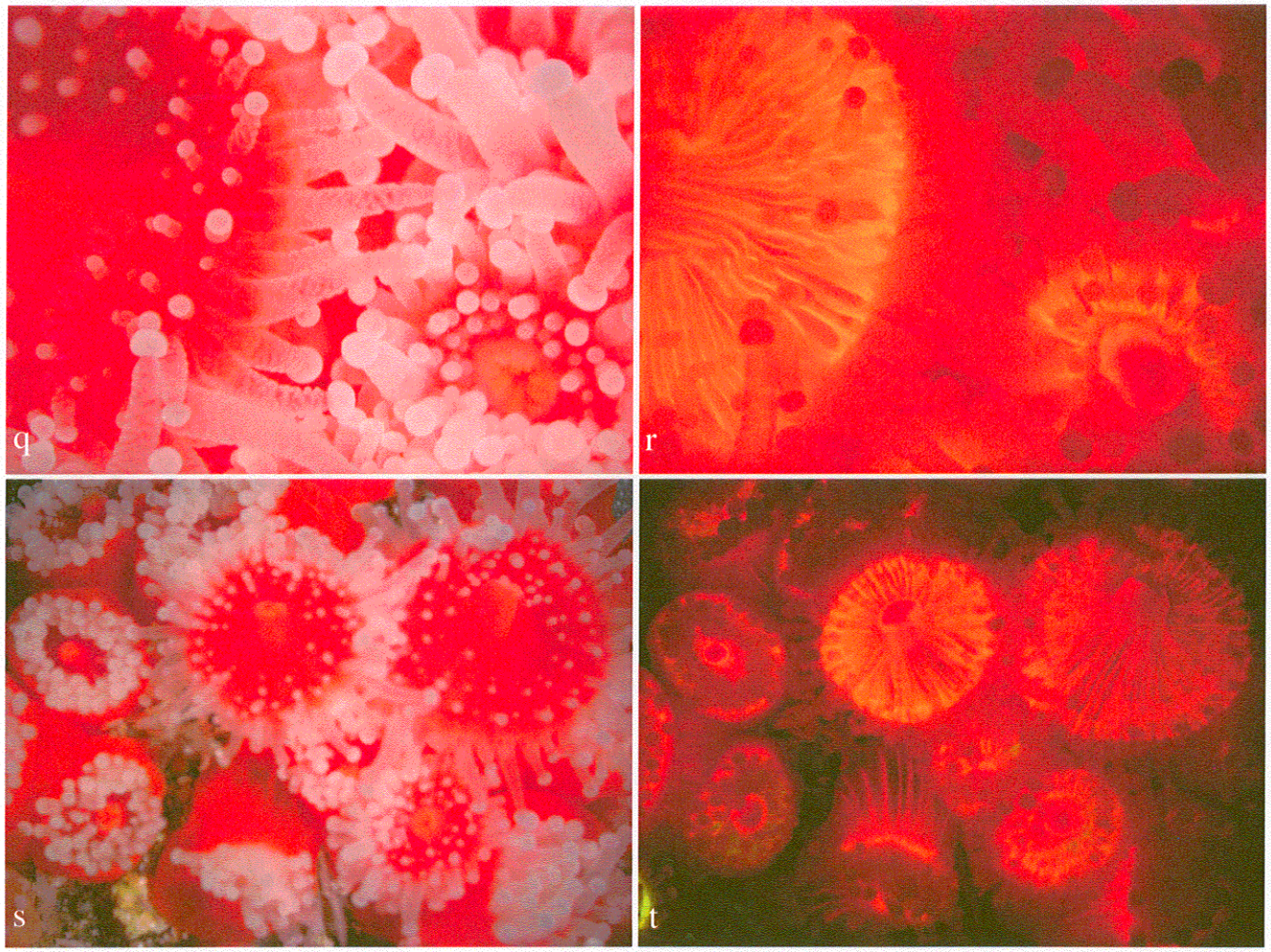
(i) Lavender polyp under white light (j) Lavender polyp fluorescing under blue light (k) Lavender polyps under white light (l) Lavender polyps fluorescing under blue light



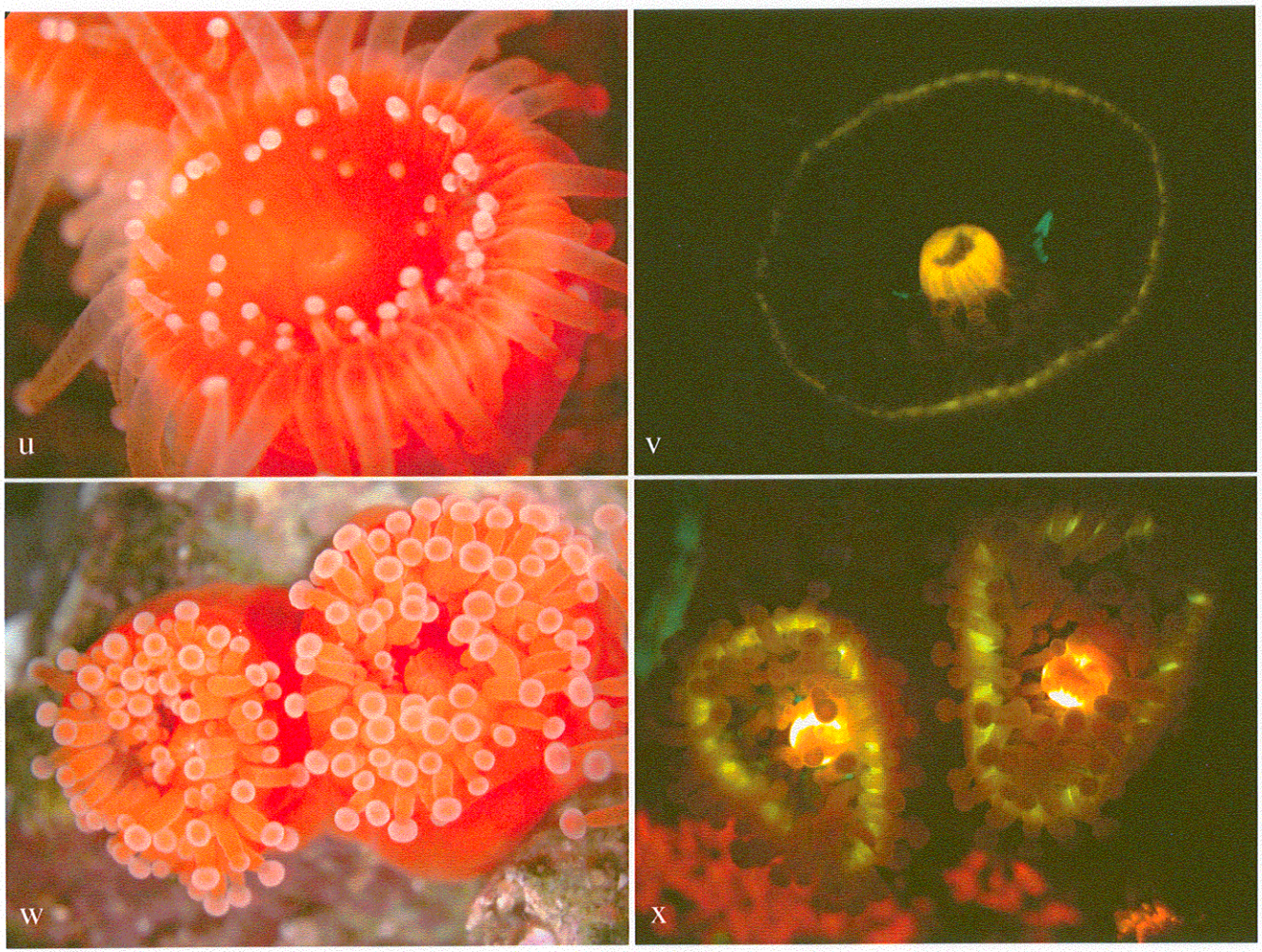
(m) Pink1 polyps under white light (n) Pink1 polyps fluorescing under blue light (o) Pink1 tentacle under white light (p) Pink1 tentacle fluorescing under blue light



(q) Orange1 polyps under white light (r) Orange1 polyps fluorescing under blue light (s) Orange1 polyps under white light (t) Orange1 polyps fluorescing under blue light



(u) Orange3 polyp under white light (v) Orange3 polyp fluorescing under blue light (w) Orange3 polyps under white light (x) Orange3 polyps fluorescing under blue light



(y) Red polyp under white light (z) Red polyp fluorescing under blue light (aa) Red polyp under white light (ab) Red polyp fluorescing under blue light

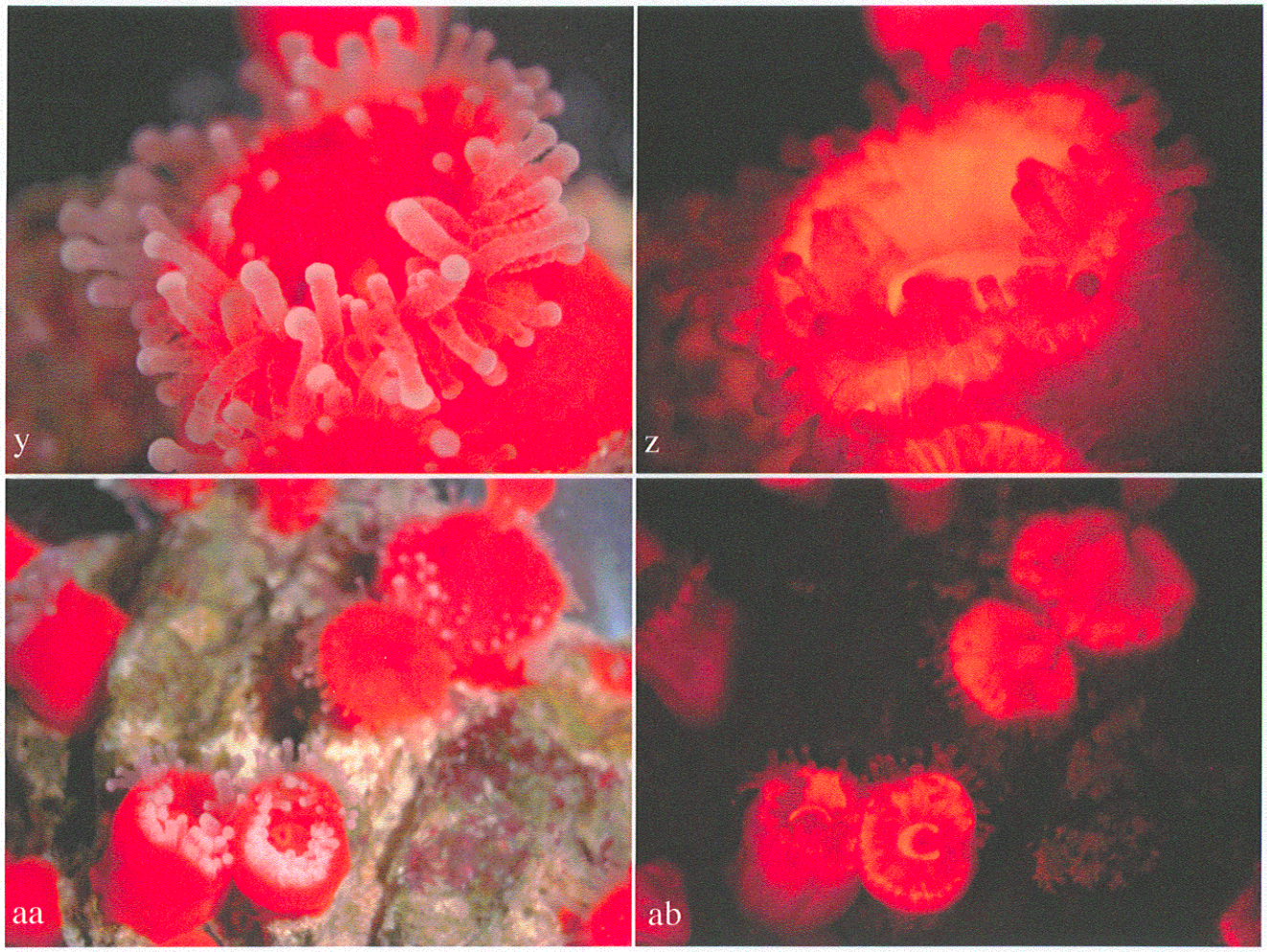


Figure 11. *In vivo* fluorescence emission spectra of all tissue types by individual color morph.

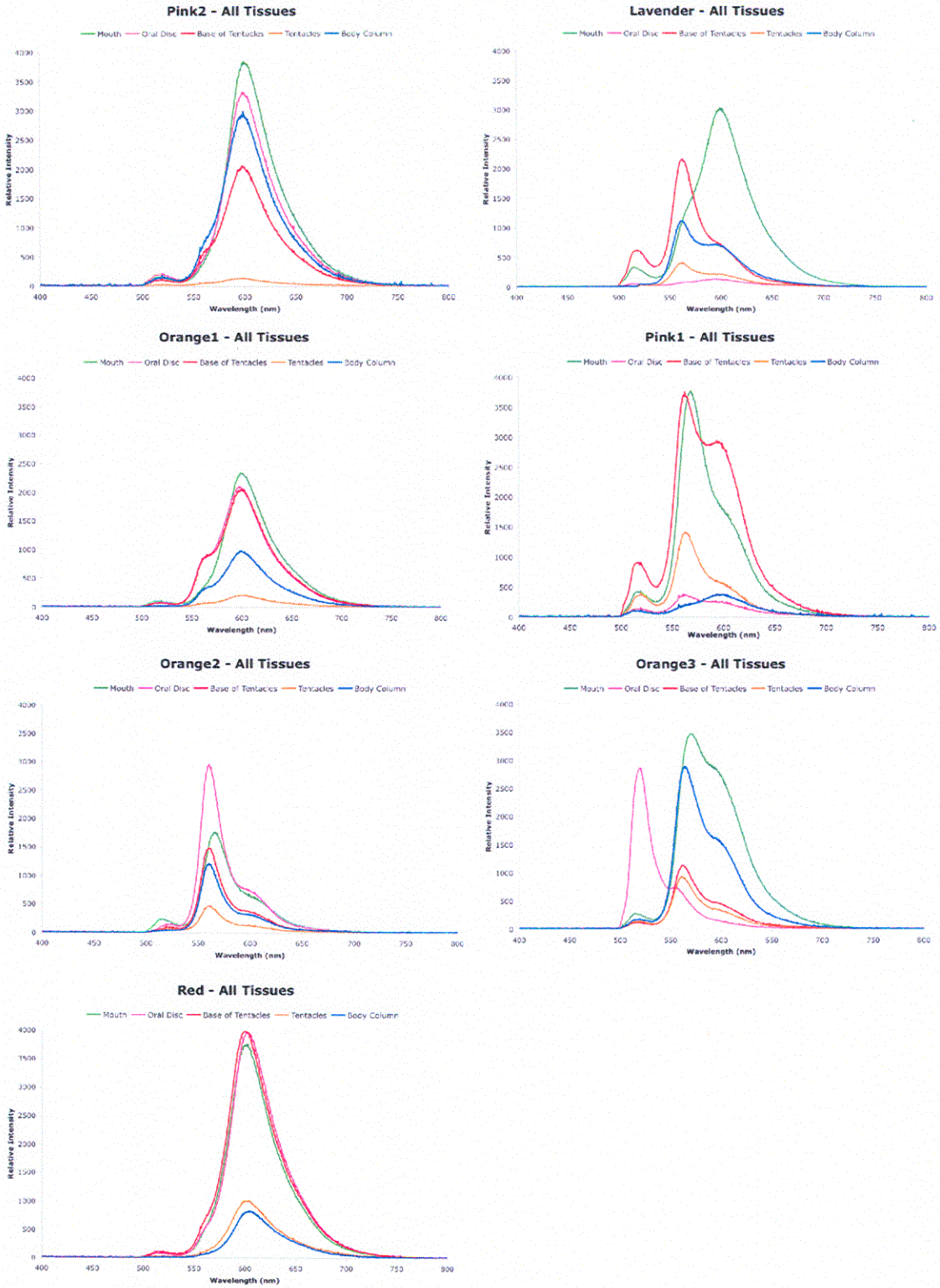


Figure 12. *In vivo* fluorescence emission spectra of all color morphs by individual tissue type.

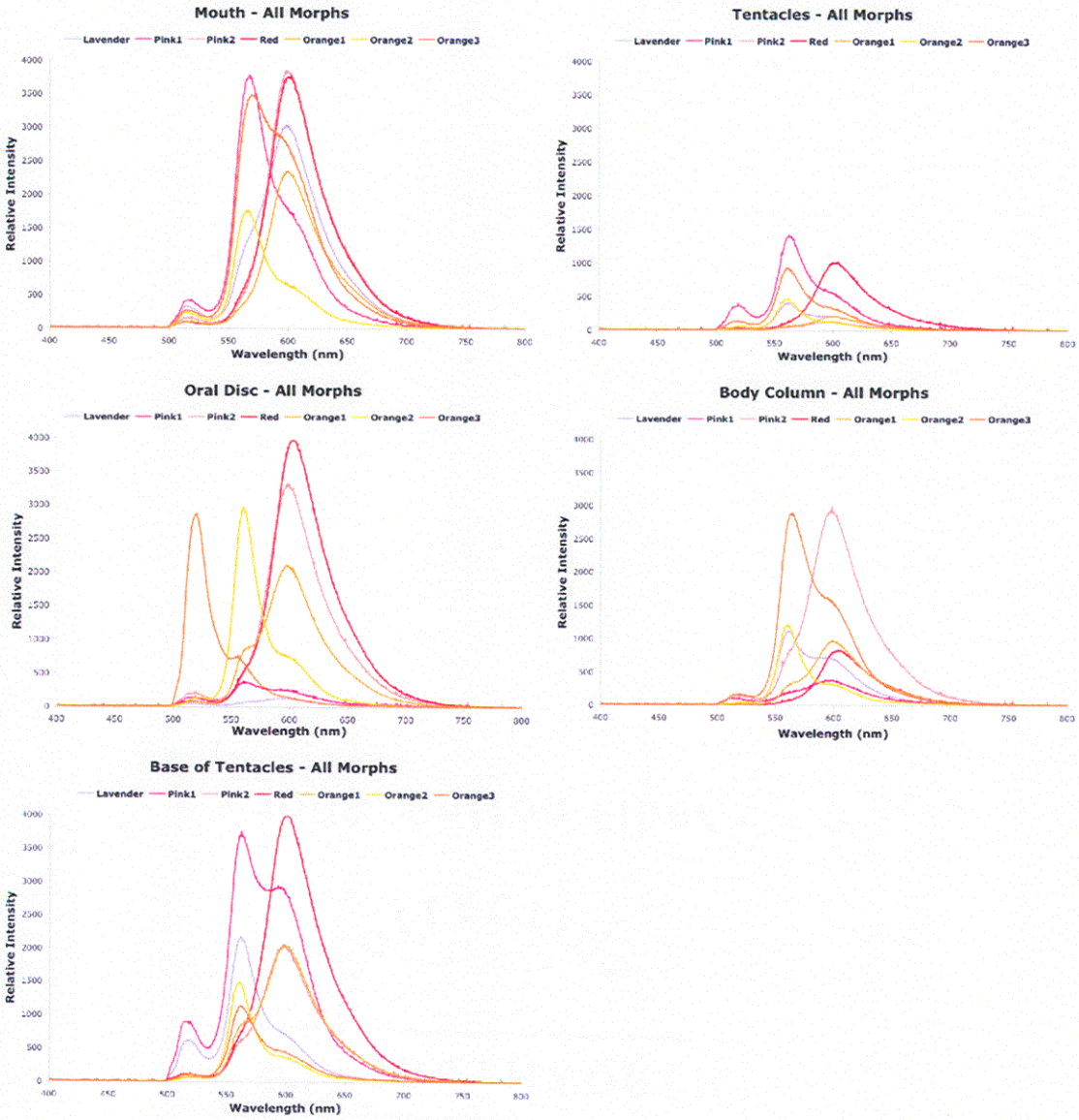


Figure 13. *In vivo* fluorescence emission spectra of all color morphs by two tissue types.

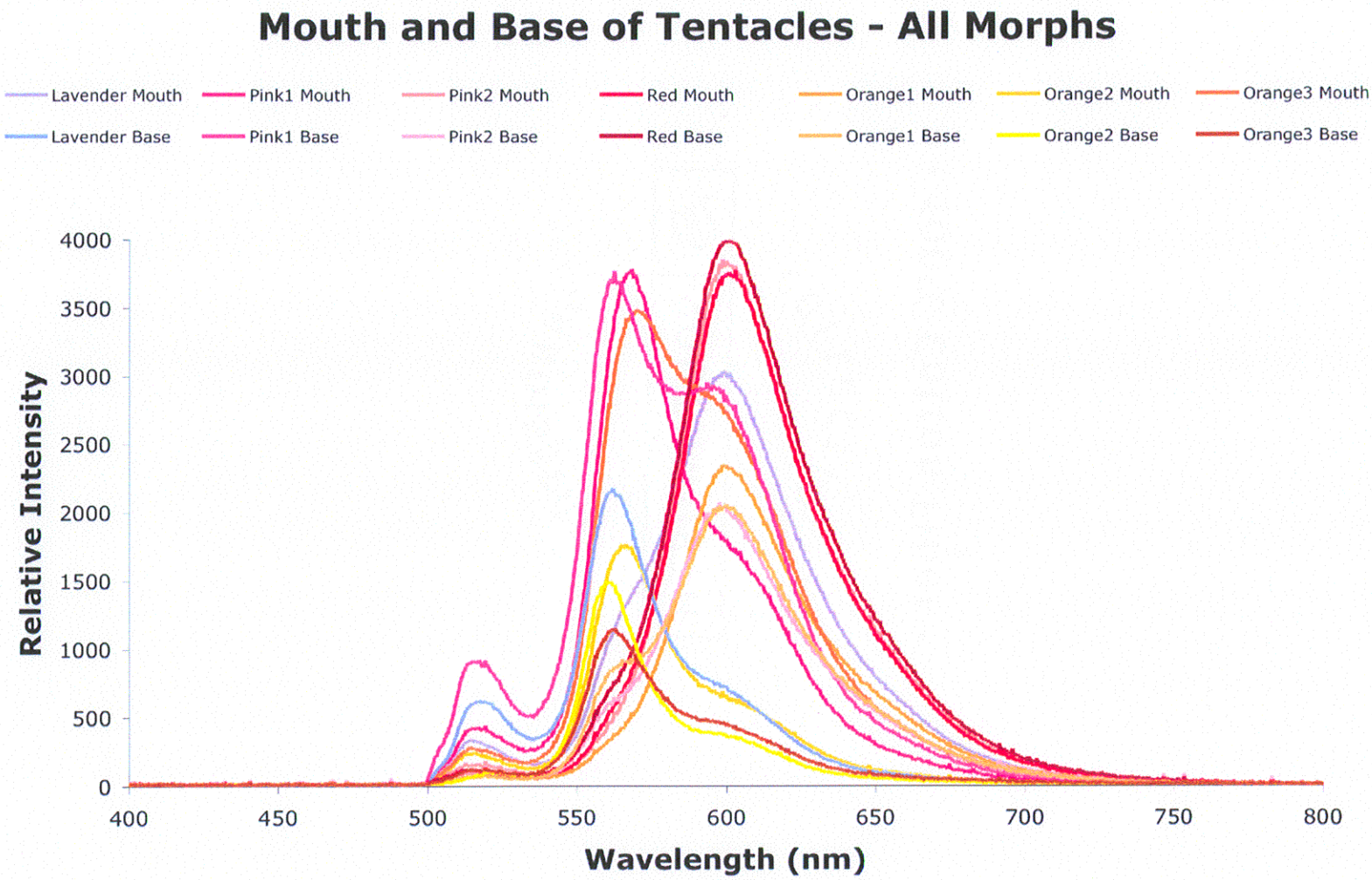


Figure 14. Fluorescence images and emission graphs of Pattern A morphs

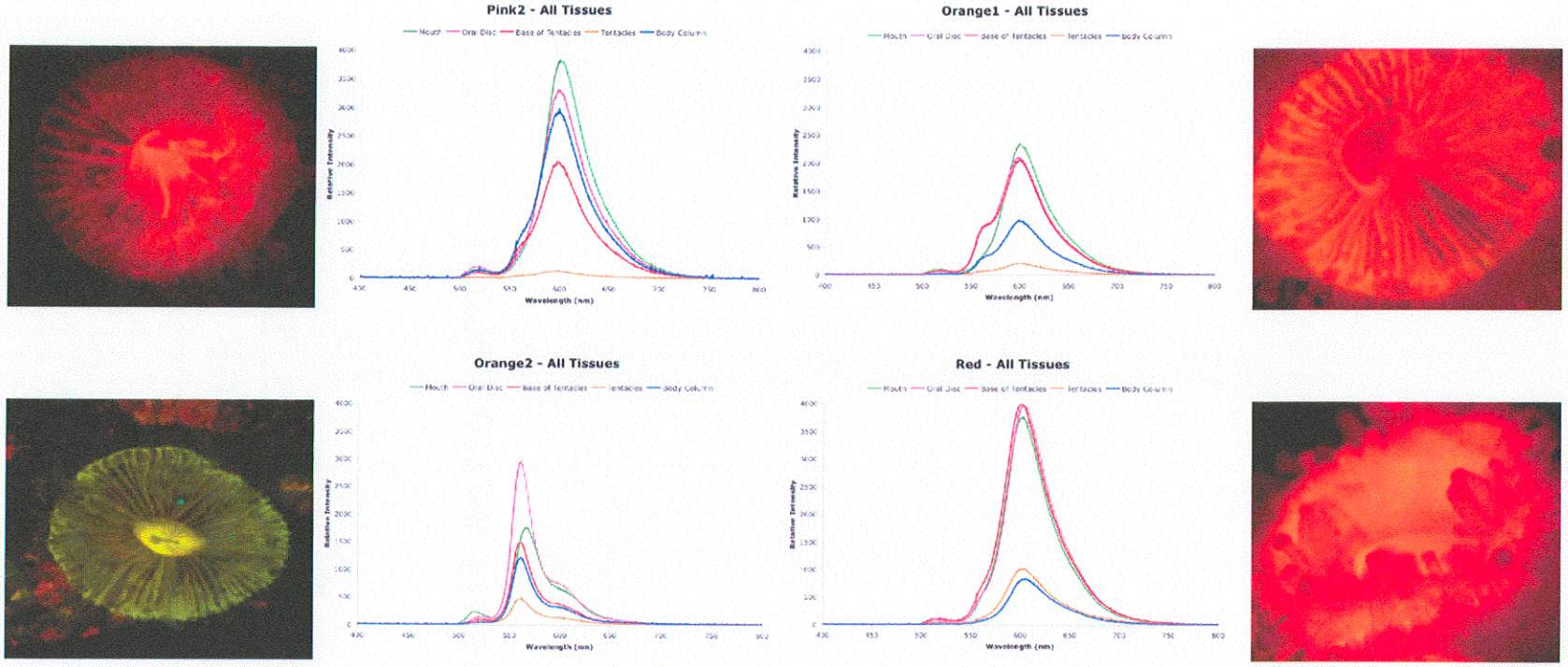


Figure 15. Fluorescence images and emission graphs of Pattern B morphs

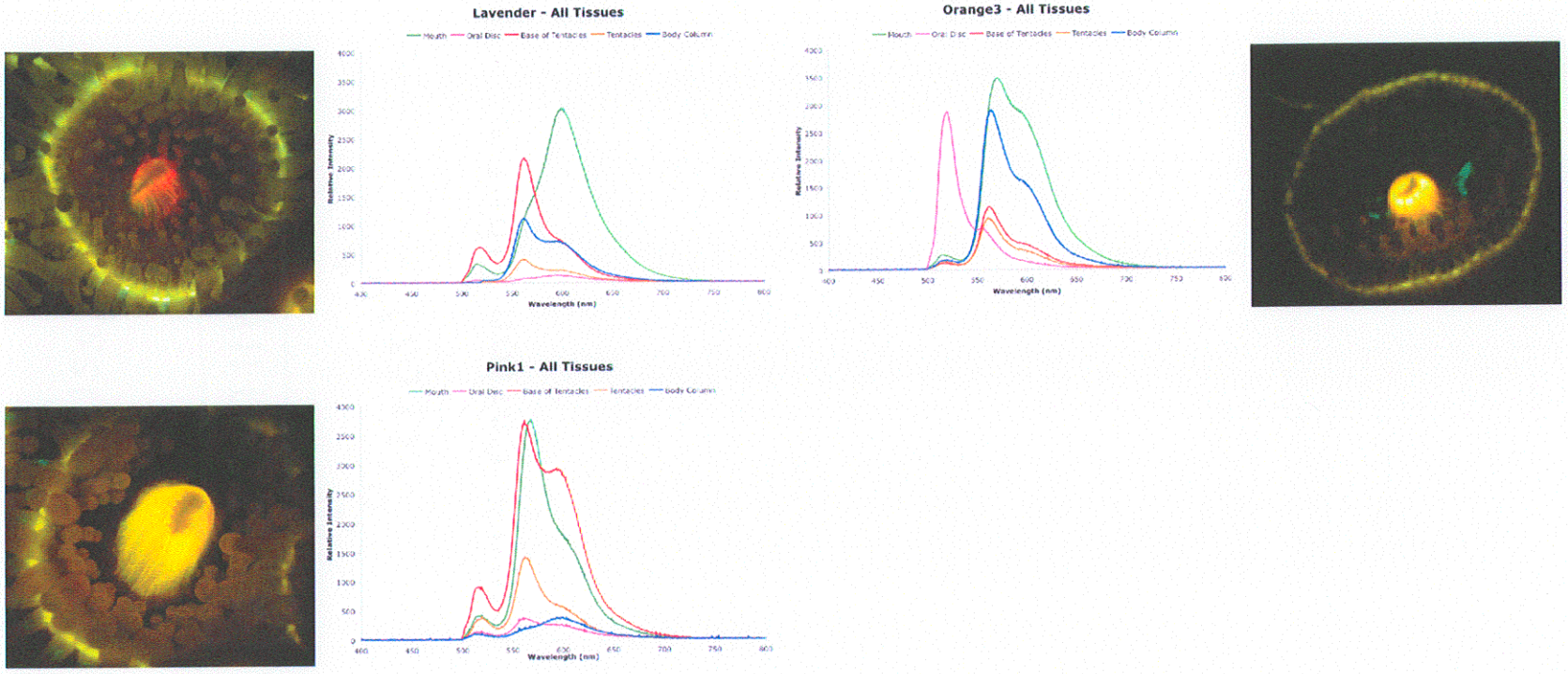


Figure 16a. Nucleotide sequence of ccalRFP2 (681 base pairs)

```
  1 ATGTCTCTAT CAAAGCAAGT TCTCCCACAT GACGTGCGGA TGCGCTATCA TATGGATGGG
 61 TGTGTCAACG GACACAGCTT TACCATTGAA GGAGAAGGCG CTGGAAAACC TTACGAAGGA
121 AAAAAGACTT TGAAATTGAG GGTCACCAAG GGAGGACCAC TTCCGTTCGC CTTTGATATA
181 TTGTCCGCAA CGTTTACGTA TGGAAACAGA TGCTTTTGCG AGTATCCAGA AGACATGCCC
241 GATTATTTCA AACAGTCATT GCCTGAAGGA TACTCATGGG AAAGGACTAT GATGTACGAG
301 GACGGAGGGT GTGGCACATC CAGTGCGCAC ATAAGACTCG AGAAAAACTG CTTTGTGCAC
361 CAATCCACAT TTCTCGGCGT CAACTTTCCT GCTAACGGAC CTGTGATGCA AAAAAAGGCC
421 CTGAACTGGG AGCCTTCCAG CGAGCTAATA ACTCCCTGTG ATGGAATACT GAAGGGCGAT
481 GTGACCATGT TTCTCATGCT GGAAGGAGGT CACCGACAGA AATGTCAATT CACAACTTCT
541 TACAAGGCAA GCAAGGCTGT CAAAATGCCA CCGAACCATA TCATAGAACA CGTTGTTTGG
601 AAAGGCGAAG ATAGCGACGG TTTTCAAATC AAAGAACACG CTGTTGCAAA ACACTTCACA
661 GTAGATGTCA AGGAAACGTA A
```

Figure 16b. Inferred amino acid sequence of ccalRFP2 (226 amino acids)

```
  1 MSLSKQVLPH DVRMRYHMDG CVNGHSFTIE GEGAGKPYEG KKTLKLRVTK GGPLPFAFDI
 61 LSATFTYGNR CFCEYPEDMP DYFKQSLPEG YSWERTMMYE DGGCGTSSAH IRLEKNCFVH
121 QSTFLGVNFP ANGPVMQKKA LNWEPSSELI TPCDGILKGD VTMFLMLEGG HRQKCQFTTS
181 YKASKAVKMP PNHIIEHVVW KGEDSDGFQI KEHAVAKHFT VDVKET*
```

Figure 17a. Nucleotide sequence of ccalOFP1 (681 base pairs)

```
  1 ATGTCTCTAT CTAAACAAGT TCTTCCACGA GACGTGAAAA TGCGCTTTCA TATGGATGGG
 61 TGTGTCAACG GACACAGCTT TACCATTGAA GGAGAAGGAA CTGGAAAACC ATACGAAGGA
121 AAAAAGACTT TGAAACTGCG TGTCACCAAG GGAGGGCCAC TTCCATTCGC CTTTGATATA
181 TTGTCCGCAA CGTTTACGTA TGGAAACAGA TGCTTTTGCG ACTATCCAGA AGAAATGCCC
241 GATTATTTCA AACAGTCATT GCCTGAAGGA TACTCCTGGG AAAGGACTAT GATGTACGAG
301 GACGGGGCGT GTAGCACGGC CAGTGCGCAC ATAAGCCTCG ATAAGGACTG CTTTATCCAC
361 AATTCCACAT TTCATGGCGT CAACTTTCCT GCTAACGGAC CTGTGATGCA AAAGAAGGCC
421 ATGAATTGGG AACCTTCCAG CGAGCTAATA ACTCCCTGTG ATGGAATACT GAAGGGCGAC
481 GTGACCATGT TTCTCTTGCA AGAAGGAGGT CACCGCCACA AATGCCAATT CACAACTTCT
541 TACAAGGCAC ACAAGGCTGT CAAAATTCCA CCGAACCACA TCATAGAACA CCGTCTTGTG
601 AGAAAGGAGG TTGGTGACGC TGTTCAAATC CAAGAACACG CTGTTGCGAA ACACTTCACA
661 GTGCAAATCA AGGAGGCGTA A
```

Figure 17b. Inferred amino acid sequence of ccalOFP1 (226 amino acids)

```
  1 MSLSKQVLPR DVKMRFHMDG CVNGHSFTIE GEGTGKPYEG KKTLKLRVTK GGPLPFAFDI
 61 LSATFTYGNR CFCDYPEEMP DYFKQSLPEG YSWERTMMYE DGACSTASAH ISLDKDCFIH
121 NSTFHGVNFP ANGPVMQKKA MNWEPSSELI TPCDGILKGD VTMFLLQEGG HRHKCQFTTS
181 YKAHKAVKIP PNHIIEHRLV RKEVGDAVQI QEHAVAKHFT VQIKEA*
```

Figure 18a. Nucleotide sequence of ccalGFP1 (666 base pairs)

```
  1 ATGTCTCTAT CAAAGCAAGT GGTCAAAGAA GACATGAAGA TGACCTATCA TATGGATGGG
 61 TGTGTCAACG GGCACTACTT TACGATTGAG GGTGAAGGCA CCGGAAAACC ATTCAAAGGA
121 CAAAAGACTT TGAAACTGCG TGTCACCGAG GGAGGGCCAC TTCCATTCGC TTTTGATATA
181 TTGTCCGCAA CGTTTACGTA TGGAAACAGA TGTTTTTGCG ATTATCCAGA GGACATGCCC
241 GACTATTTCA AACAGTCTCT CCCTGAAGGA TACTCATGGG AAAGGACCAT GATGTACGAG
301 GATGGAGCGT GTGGCACAGC CAGTGCGCAC ATAAGCCTTG ATAAGAACGG CTTTGTGCAT
361 AATTCTACAT TTCACGGTGT CAATTTTCCT GCTAACGGAC CTGTGATGAA AAAAAAAGGA
421 GTGAATTGGG AACCTTCCAG CGAGAAAATA ACTGCCTGTG ATGGAATACT GAAGGGTGAT
481 GTGACCATGT TTCTCGTGCT GGAAGGAGGT CACCGACTCA AATGTCTGTT CCAAACTACT
541 TACAAGGCAG ACAAGGTTGT CAAAATGCCA CCGAACCATA TCATAGAACA CCGTCTTGTG
601 AGAAGCGAAG ATGGCGACGC TGTTCAAATC CAAGAACACG CTGTTGCAAA ATACTTCACA
661 GTATGA
```

Figure 18b. Inferred amino acid sequence of ccalGFP1 (221 amino acids)

```
  1 MSLSKQVVKE DMKMTYHMDG CVNGHYFTIE GEGTGKPFKG QKTLKLRVTE GGPLPFAFDI
 61 LSATFTYGNR CFCDYPEDMP DYFKQSLPEG YSWERTMMYE DGACGTASAH ISLDKNGFVH
121 NSTFHGVNFP ANGPVMKKKG VNWEPSSEKI TACDGILKGD VTMFLVLEGG HRLKCLFQTT
181 YKADKVVKMP PNHIIEHRLV RSEDGDAVQI QEHAVAKYFT V*
```

Figure 19. Emission spectra of ccalRFP2 in the pQE31X-DsRed.T3 vector as it changes with overnight maturation

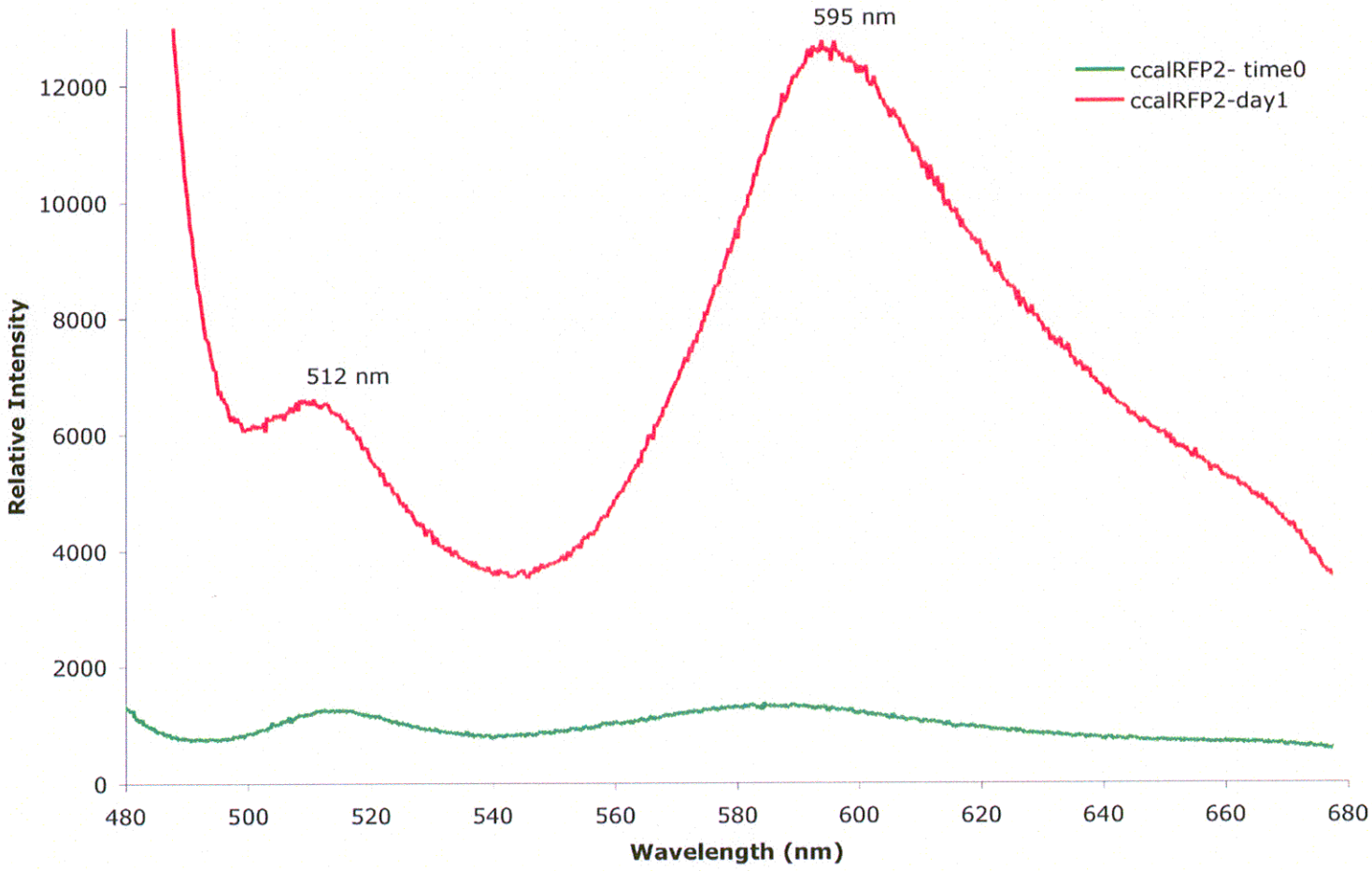


Figure 20. Emission spectrum of ccalOFP1 in the pQE31X-DsRed.T3 vector

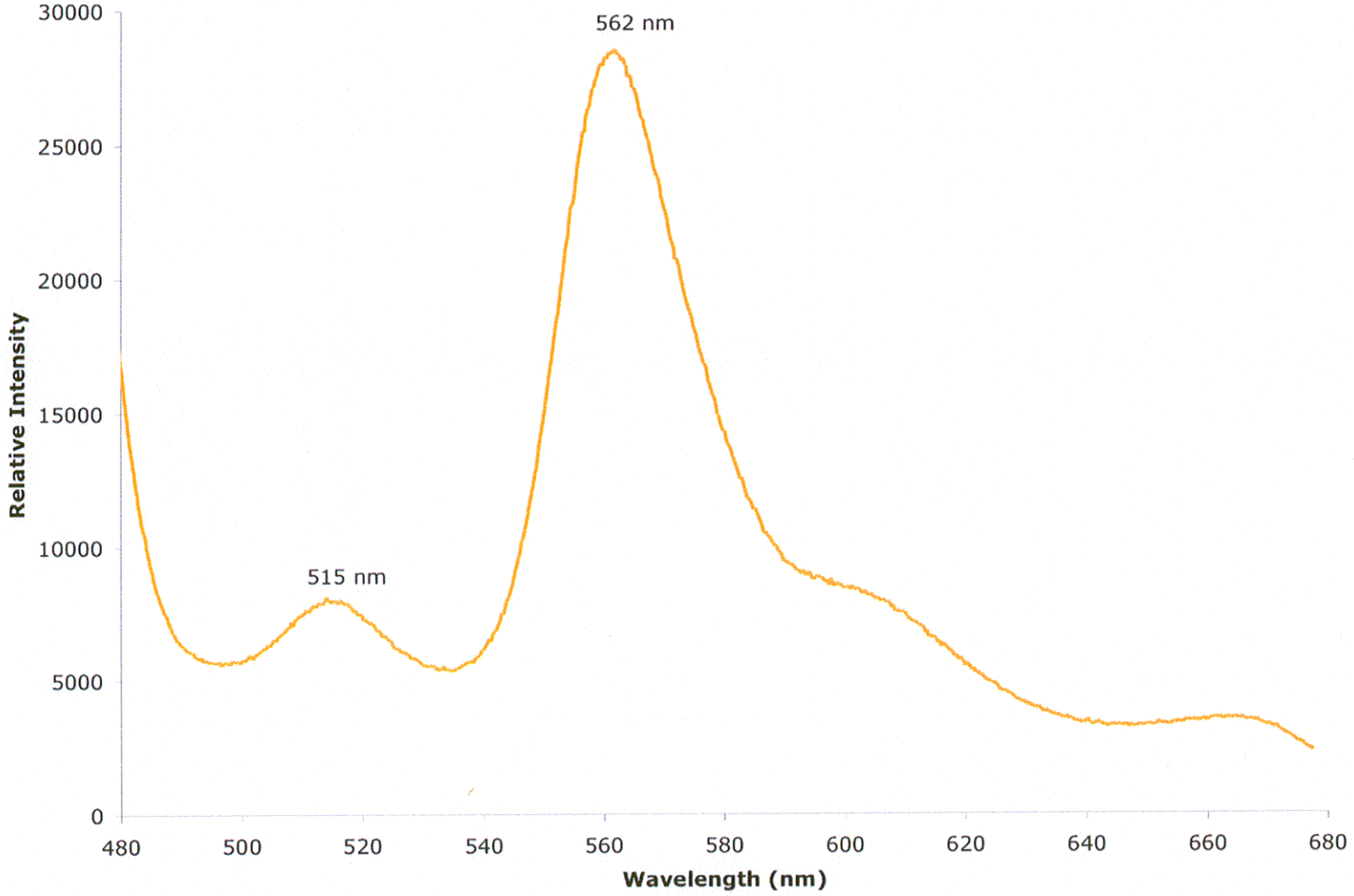


Figure 21. Emission spectrum of ccalGFP1 in the pQE31X-DsRed.T3 vector

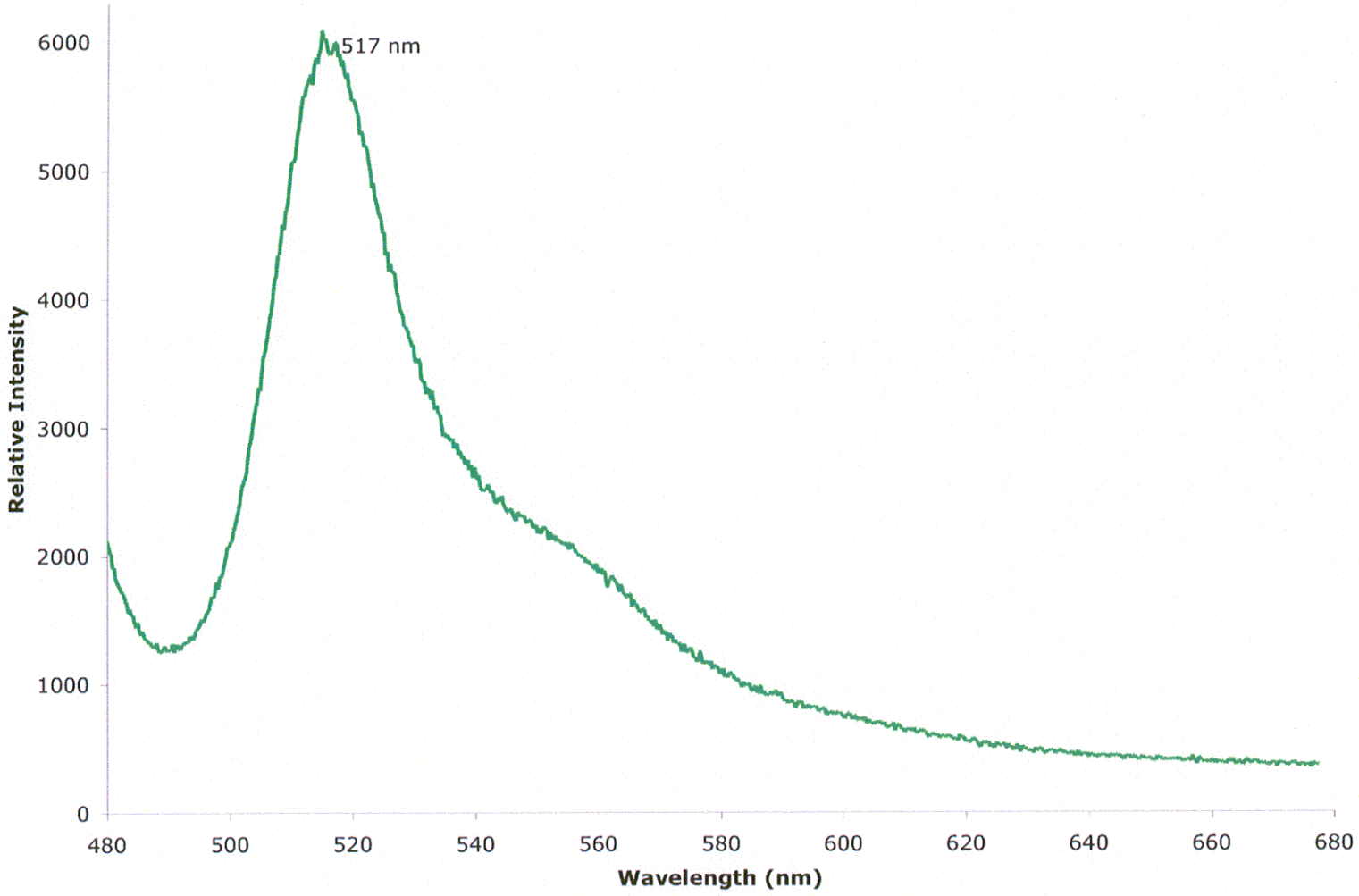


Figure 22. Emission spectra of ccalRFP1 in the pQE31X-DsRed.T3 vector as it matures through time.

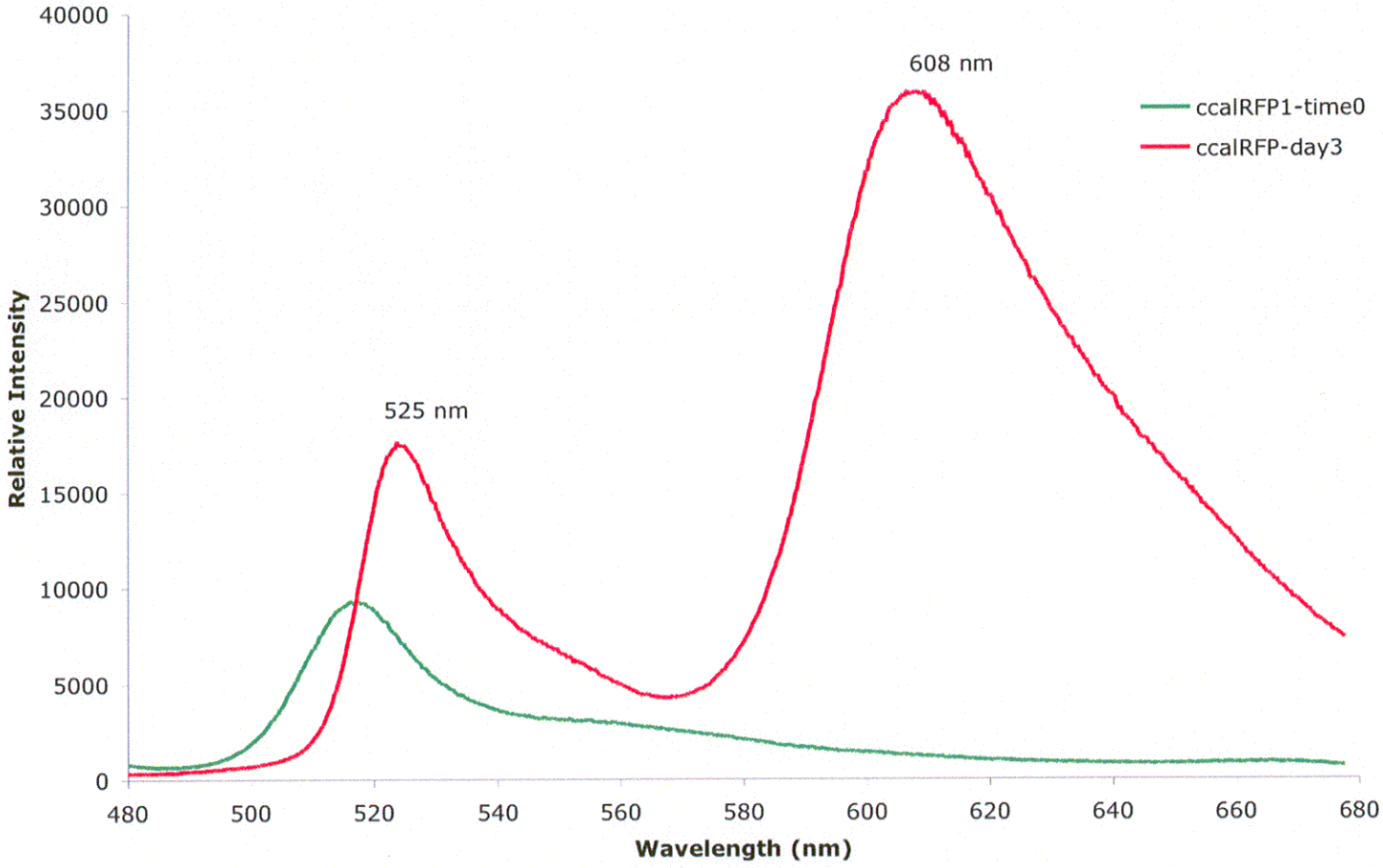


Figure 23. Amino acid alignment of *Corynactis californica* fluorescent proteins with homologous sequences shown relative to ccalRFP1. A dot indicates identity with ccalRFP1 and a dash indicates a gap in the alignment. Residues shaded in grey denote the chromophore sequence.

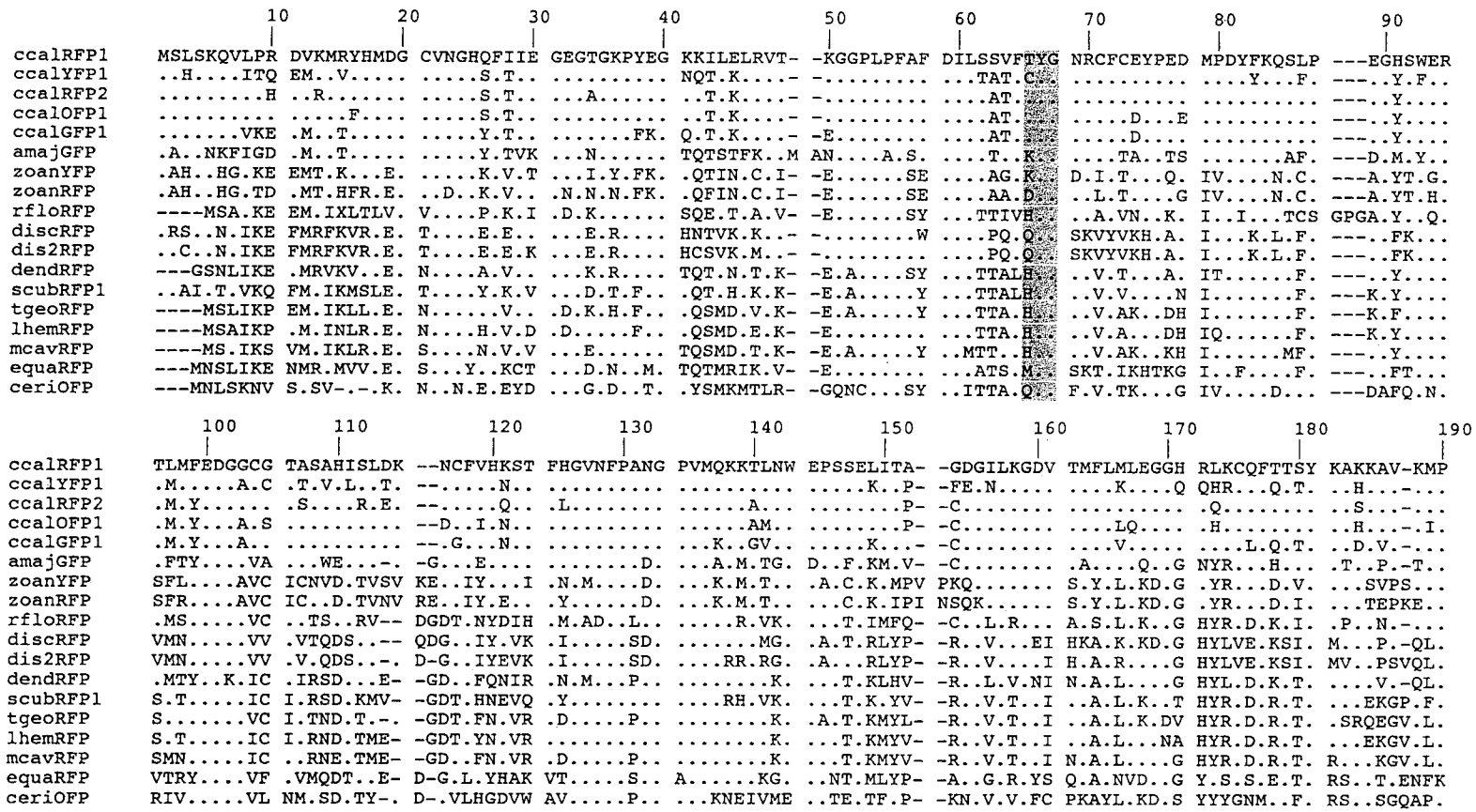


Figure 23. (continued)

```
                        200             210             220
                        |               |               |
ccalRFP1     PN--HIIEHR LVR-KEVADA --V-QIQEHA VAKHFIV--- ----
ccalYFP1     ..--..... ...-SQDG.. --.-.LK... ...C.TA--- ----
ccalRFP2     ..--.....V VWK-G.DS.G --F-..K... .....T.DVK ET--
ccalOFP1     ..--..... ...-...G.. --.-...... .....T.QIK EA--
ccalGFP1     ..--..... ...-S.DG.. --.-...... ...Y.T.--- ----
amajGFP      ..--.AV... IA.TDLDKGG NS.-.LT... ..HITS.VPF ----
zoanYFP      EW--.F.Q.K .L.EDRSDAK NQKW.LT.-- ---------- ----
zoanRFP      DW--.F.Q.K .N.EDRSDAK NQKW.LI..- ---------- ----
rfloRFP      GY--.FVD.C IEITSQQD.Y NV.ELY-.G. ..HYSPLQKP CQAK
discRFP      GY--YYVDSK .DITSHNE.Y TI.E.Y-.RT EGR.HLF--- ----
dis2RFP      GY--YYVDSK .DMTSHNE.Y TV.E.Y-.KT QGR.HPF--- ----
dendRFP      DY--.FVD.. IEILSNDS.Y NK.KLY-..G ..RYSPLPSQ AW--
scubRFP1     GY--.LVD.C IEITSHDK.Y NV.ELY-... ..HSGLPDSA NR--
tgeoRFP      GY--.FVD.C ISILRHDK.Y NE.KLY-... ..HSGLPDNV K---
lhemRFP      GY--.FVD.C IEILSHDK.Y NK.KLY-... ..HSGL---- ----
mcavRFP      DY--.FVD.S IEILRHDKEY TE.KLY-... E.HSGLPD-- ----
```

Figure 24. Consensus phylogram of all post burn-in trees of anthozoan fluorescent protein genes reconstructed by Bayesian analysis with posterior probabilities shown above or to the left of the branch where space permits.

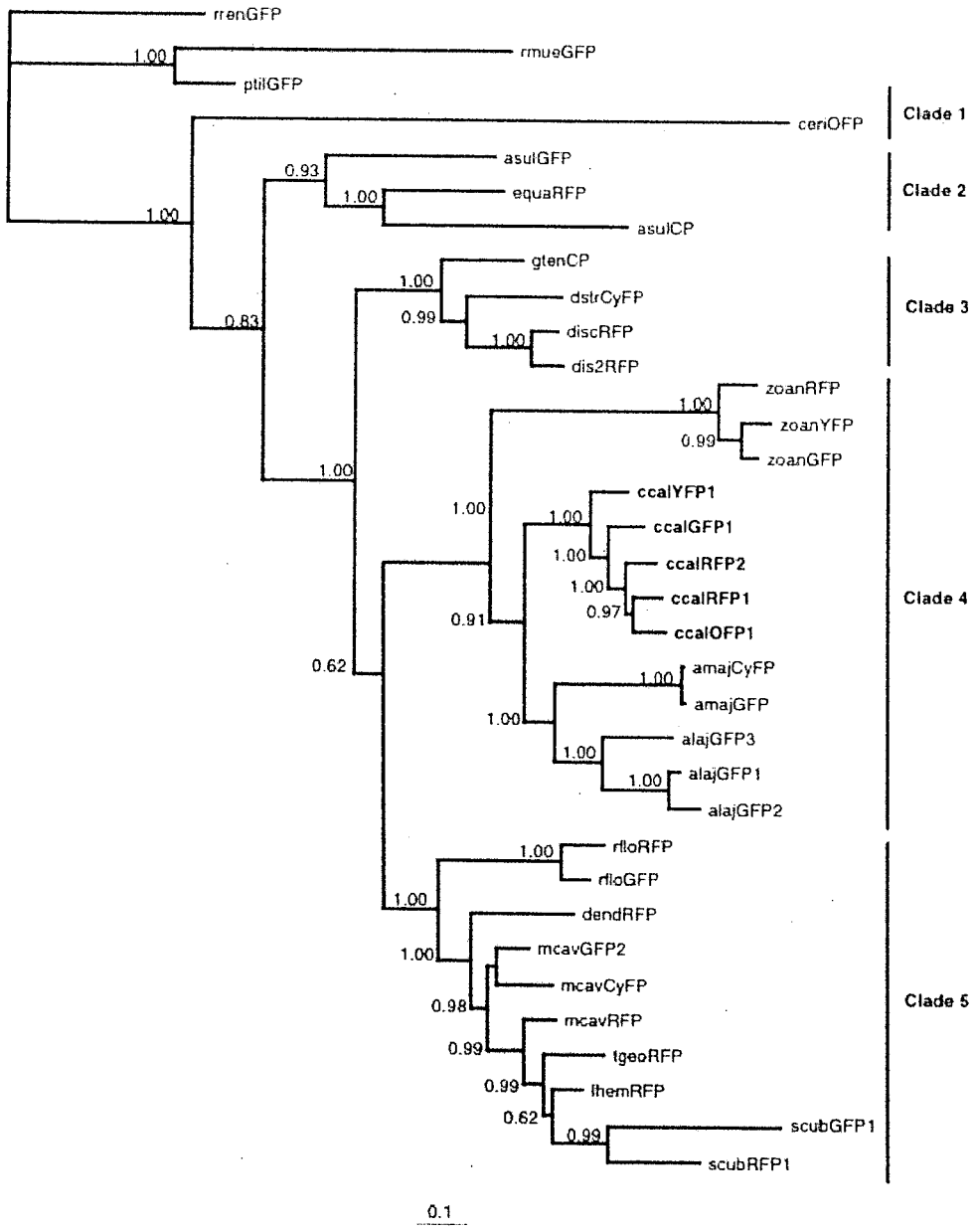


Figure 25. Consensus tree of all post burn-in trees of anthozoan fluorescent protein genes reconstructed by Bayesian analysis with posterior probabilities shown above or to the left of the branch and MP bootstrap support values shown in parentheses.

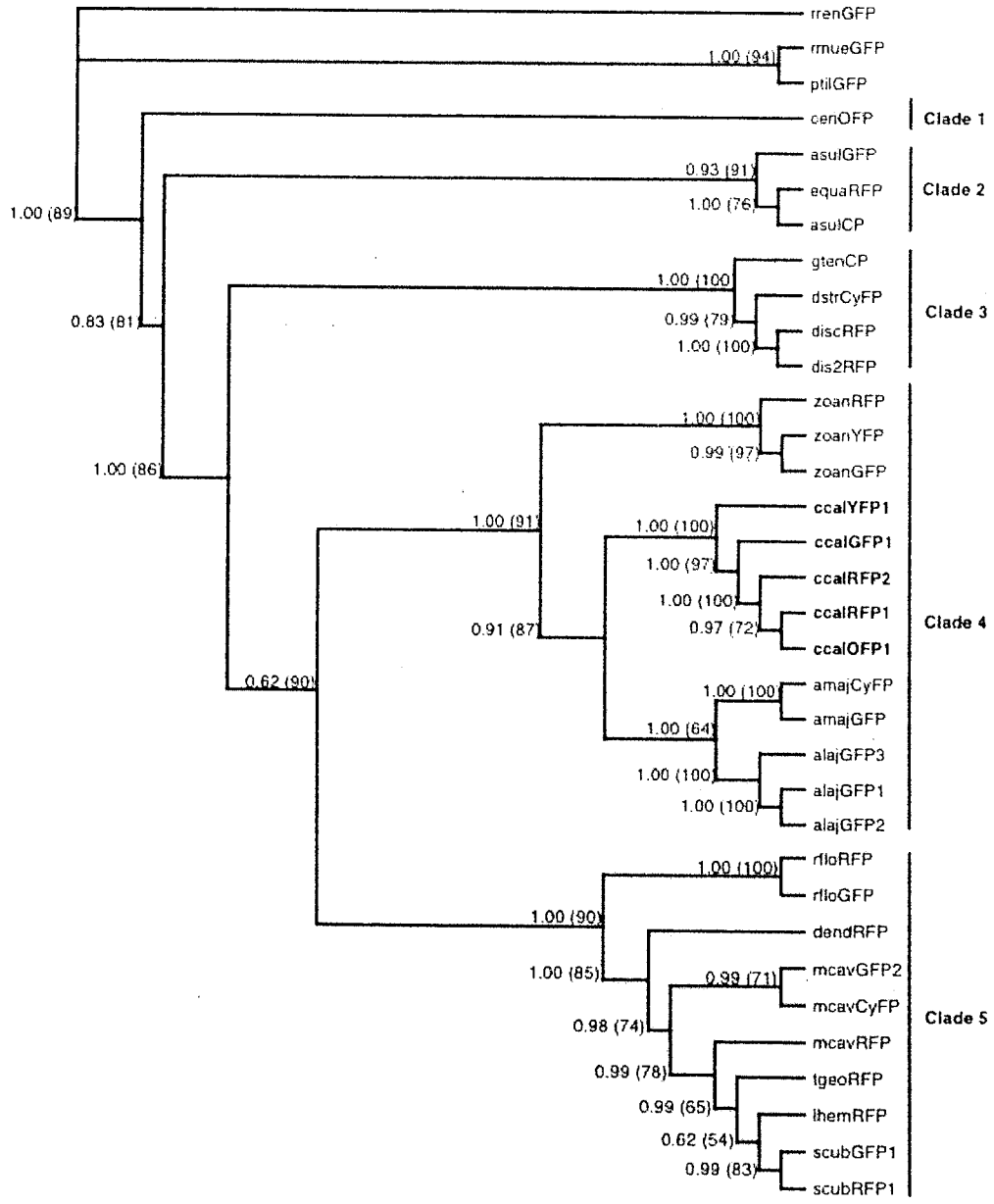


Figure 26. ML consensus tree of anthozoan fluorescent protein genes with bootstrap support shown above the branch.

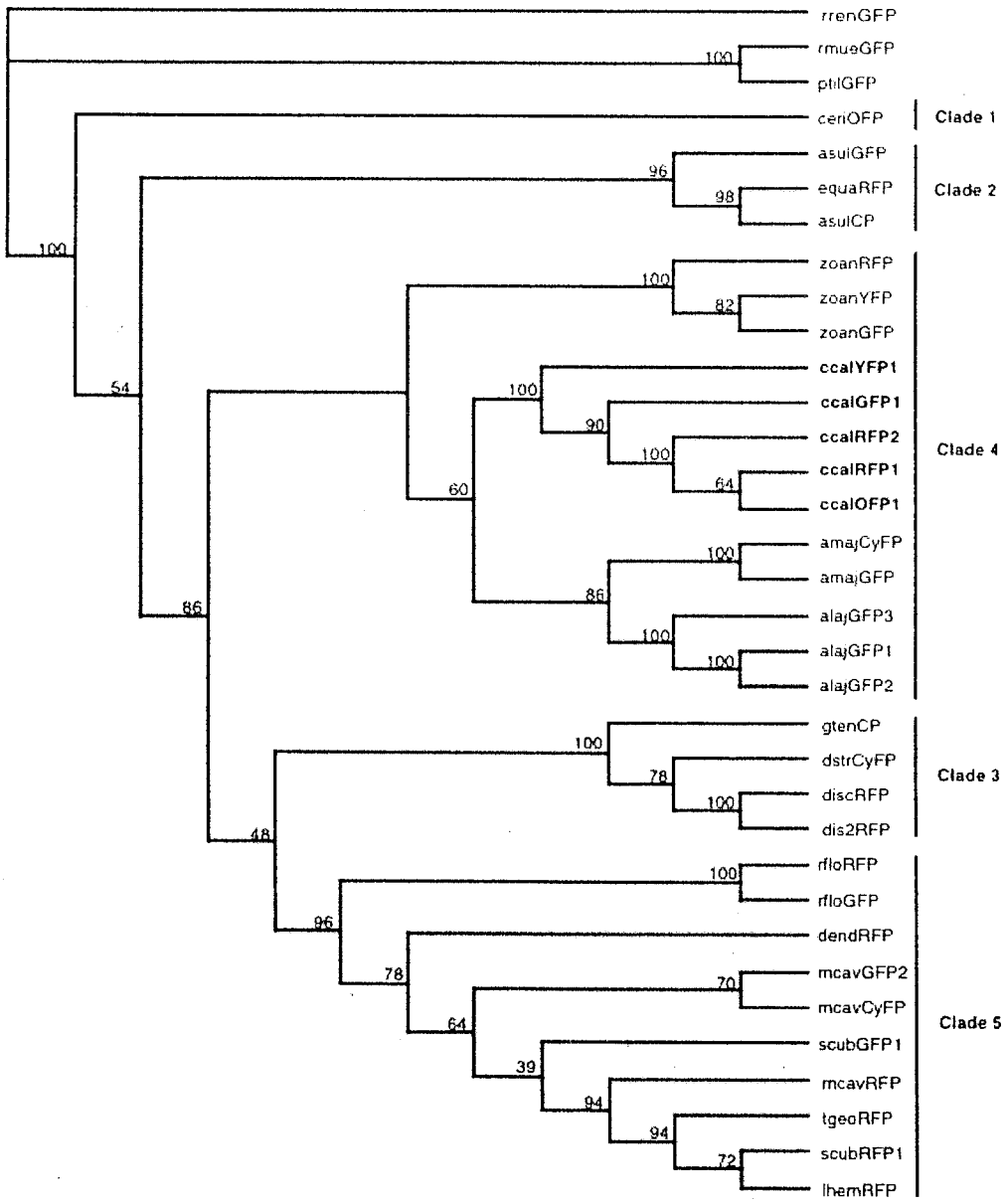


Figure 27. Histogram of emission maxima of 47 known native GFP-like proteins. ccalFPs are noted above their respective bins.