

RESEARCH ARTICLE

Open Access



Non-excitable fluorescent protein orthologs found in ctenophores

Warren R. Francis^{1,3}, Lynne M. Christianson¹, Meghan L. Powers¹, Christine E. Schnitzler^{2,4}
and Steven H. D. Haddock^{1*}

Abstract

Background: Fluorescent proteins are optically active proteins found across many clades in metazoans. A fluorescent protein was recently identified in a ctenophore, but this has been suggested to derive from a cnidarian, raising again the question of origins of this group of proteins.

Results: Through analysis of transcriptome data from 30 ctenophores, we identified a member of an orthologous group of proteins similar to fluorescent proteins in each of them, as well as in the genome of *Mnemiopsis leidyi*. These orthologs lack canonical residues involved in chromophore formation, suggesting another function.

Conclusions: The phylogenetic position of the ctenophore protein family among fluorescent proteins suggests that this gene was present in the common ancestor of all ctenophores and that the fluorescent protein previously found in a ctenophore actually derives from a siphonophore.

Keywords: Fluorescent protein, Ctenophore, Siphonophore, Transcriptome, Fluorescence, Haeckelia

Background

Fluorescent proteins (FPs) are abundant optical proteins in cnidarians [1–4] as well as other phyla, including arthropods [5–7] and chordates [8, 9]. They are structurally composed of a beta-barrel that surrounds the fluorophore [10]. The fluorophore is self-forming, requiring only molecular oxygen and the internal xYG residues of the protein. Because no other proteins or factors are required, FPs are used extensively in biotechnology for genetically encoded labels and reporters [11, 12].

In many cnidarians, such as the hydromedusa *Aequorea*, the fluorescent proteins are found as part of a binary system in conjunction with the bioluminescent photoproteins. In these natural resonant-transfer pairs, they modulate the color of the luminescence through energy transfer [1, 13–15]. On the other hand, many *non*-luminous cnidarians also exhibit a rainbow of fluorescence through diversification of the fluorescent protein set [4], potentially serving to attract prey [16].

It has long been known that most members of the phylum Ctenophora are bioluminescent [17–19] however it was not until more recently that a fluorescent protein was identified in the species *Haeckelia beehleri* [20]. The protein, which had the interesting property of photo-induced maturation, was clearly visible through the body, and the gene was reliably cloned from mRNA and expressed in bacteria. Because not all animal lineages had a sequenced member with an identified fluorescent protein, the position of this ctenophore protein in a phylogenetic tree had suggested it was very different from FPs of known cnidarians, particularly hydromedusa including *Aequorea victoria* [20]. However, in the published genome of *Mnemiopsis leidyi* [21], no fluorescent proteins were found, suggesting that ctenophores actually lack fluorescent protein genes [22].

Here we report evidence from the transcriptomes of 30 ctenophores that challenges both of these findings. Nearly every ctenophore examined expresses a fluorescent-protein-like (FPL) gene, including *Mnemiopsis leidyi* and *Pleurobrachia bachei*. These FPLs have substitutions in canonical residues involved in chromophore formation, suggesting that they either retained the ancestral function of the proto-FP, or may serve an entirely different function. Finally, by comparison to other FPs in

*Correspondence: haddock@mbari.org

¹ Monterey Bay Aquarium Research Institute, 7700 Sandholdt Rd., 95039 Moss Landing, USA

Full list of author information is available at the end of the article

the transcriptomes of more recently sequenced cnidarians, we determine that the FP from *Haeckelia beehleri* likely comes from a siphonophore through dietary uptake.

Results

Identification of a GFP-like protein in a ctenophore

We sequenced the transcriptomes of 28 ctenophores, and downloaded data for two other species. We had developed an automated protein search and identification strategy that searched for fluorescent proteins in hydromedusa transcriptomes, though this was automatically applied to the ctenophores as well. While most cnidarian fluorescent protein queries did not yield a BLAST hit in the ctenophore transcriptomes below the e-value threshold (10^{-6}), we were surprised to find that one ctenophore (undescribed species *spT*) appeared to have a BLAST hit when Azami-Green FP was used as the query [23].

Gene structure of the GFP-like protein

Based on this single ctenophore sequence, we re-examined the *Mnemiopsis leidyi* genome, which otherwise appeared to not have a fluorescent protein [22]. Using the undescribed species *spT* FPL as the query, we found an incomplete protein match (ML181711a) in the filtered gene models that had *Aequorea GFP* as the top BLAST hit. This protein appeared to have two exons missing at the N-terminus, although these exons were found in the unfiltered protein models (MLRB181734) (Fig. 1). Then, using the *M. leidyi* FPL as the query, orthologs in all other ctenophore species were easily found. In comparison to the transcriptome sequences, the *M. leidyi* protein model was missing a conserved N-terminal motif

(Fig. 2). Another in-frame methionine that was 10 residues upstream was manually identified as the start codon to better correspond to the conserved N-terminal motif (approximately MxxRMERxxxxFxG). This motif was consistently found in all other ctenophore FPLs. Because some GFPs in the hydromedusa *Clytia hemispherica* are targeted to the mitochondria [24], we examined this FPL for targeting peptides. Although the ctenophore FPLs have a conserved N-terminal motif that alternates between charged and non-polar residues, neither SignalP nor TPPred2 predicted secretory or mitochondrial targeting based on this motif [25, 26].

Ctenophore GFP-like proteins lack normal FP features

When the *M. leidyi* FPL was aligned to *A. victoria* FP, the most striking difference between the ctenophore FPL and cnidarian FP was the absence of the canonical xYG chromophore motif (Fig. 2, Additional files 1 and 2). In *M. leidyi*, this is replaced by QTN; it should be noted that the glutamine is completely conserved at this position in all other ctenophore species, and none of the other ctenophore sequences have the canonical chromophore residues. Furthermore, two critical residues involved in chromophore maturation, R96 and E222 (positions for *A. victoria*), are both substituted in all ctenophore FPLs. Based on known chromophore-formation mechanisms [27], we consider it highly unlikely that these proteins could form a fluorescent chromophore.

Origin of the *Haeckelia beehleri* FP

Further examination of the tree from the original *H. beehleri* FP publication [20] shows the ctenophore branch

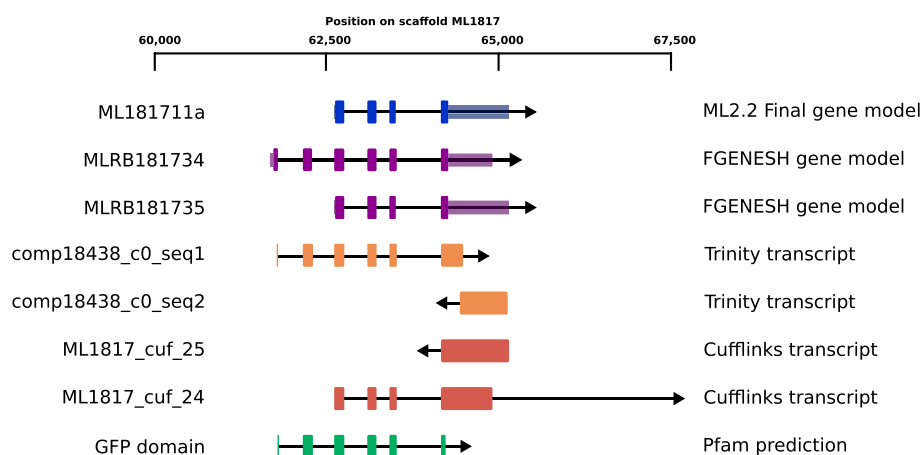
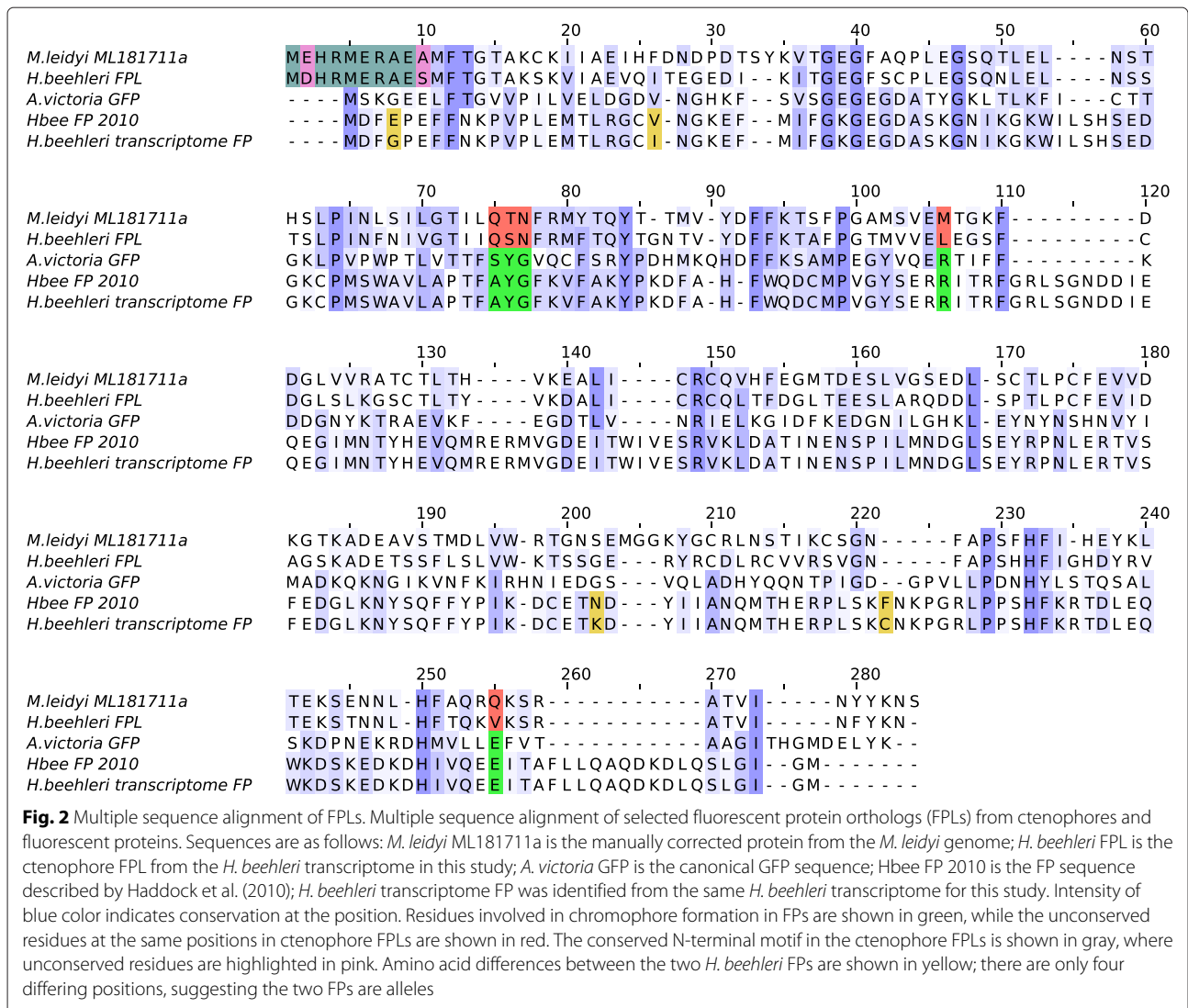


Fig. 1 Schematic of the gene structure of *Mnemiopsis leidyi* FPL. Thick bars represent coding sequences while thin bars represent untranslated exon regions in the gene models. The final gene model in the ML2.2 release was taken from the FGENESH gene model MLRB181735, while instead MLRB181734 most accurately depicts the putative structure of the *M. leidyi* FPL gene. Trinity transcripts comp18438_c0_seq1 and comp18438_c0_seq2 overlap for 24 bp in the middle, despite going different directions. The cufflinks transcript ML1817_cuf_24 extends downstream for almost 20 kbp before finishing at an incorrectly positioned 28 bp exon. The unfiltered gene model MLRB181733 (not shown for clarity) extends for thousands of bases in both directions, bridging several other genes, and is likely to be an artifact



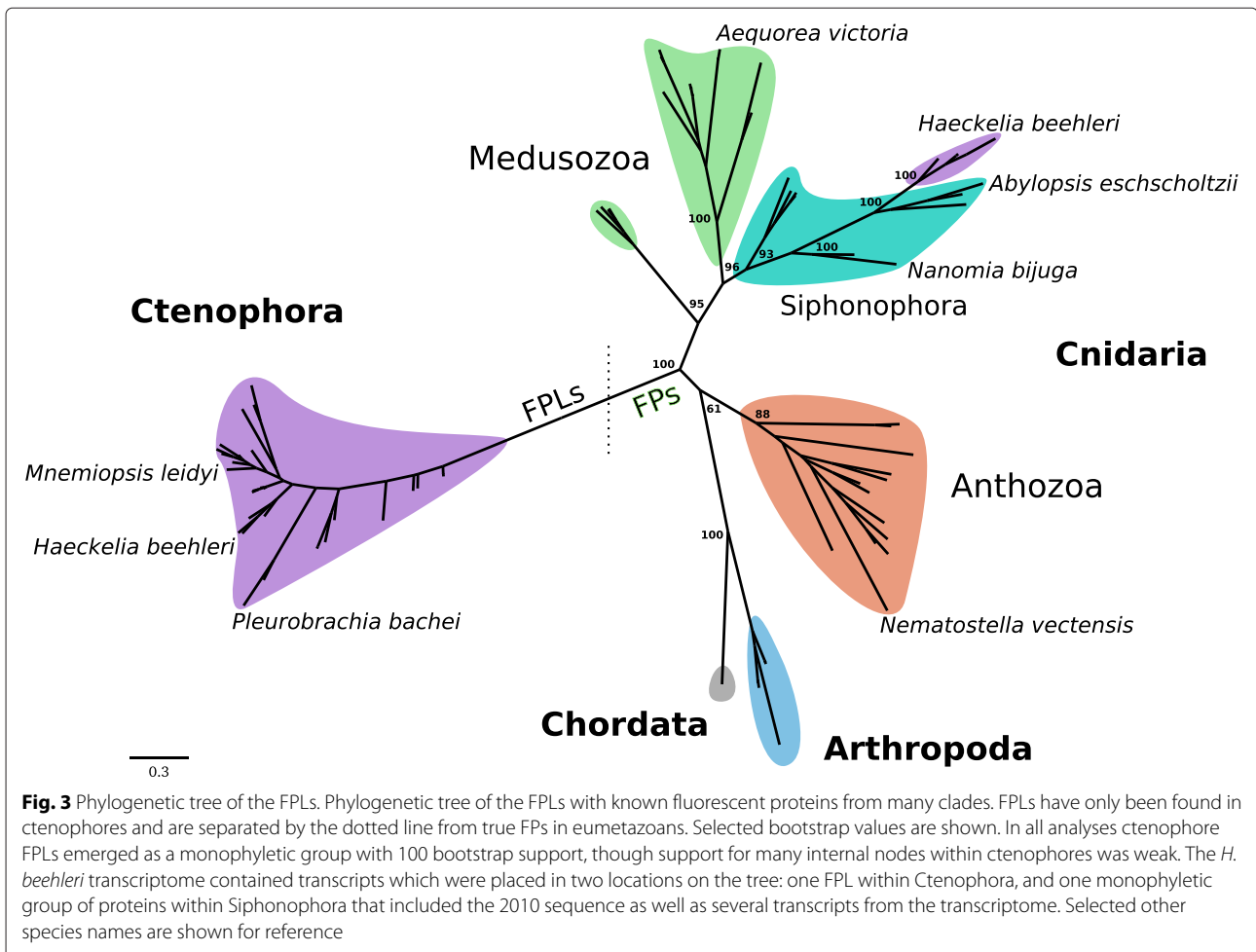
to be placed inside of the cnidarian clade, which is surprising given that ctenophores should be a separate clade. We carefully searched the transcriptome of *H. beehleri* and found two proteins, one more similar to the *M. leidy* FPL and one matching (98 % identity) the FP identified previously by Haddock et al. [19]. To examine the molecular evolution of these proteins, we generated a phylogenetic tree of all of the FPLs from this study with known FPs (Fig. 3, Additional file 3). We were surprised to find a remarkable difference between the positions of the previously cloned *H. beehleri* FP and the FPLs; the ctenophore FPLs formed a monophyletic group with 100 % bootstrap support at the base, while the previously cloned *H. beehleri* FP was included within the branch of siphonophores [28], a group of colonial cnidarians. This suggests that the FP in *H. beehleri* derives from a siphonophore, most likely as residual contamination from

ingested material. Nonetheless, the same FP was found in the transcriptome of samples collected years apart, suggesting that this process is ongoing and may demonstrate some other biological reason that the mRNA of the FPs, as well as the proteins themselves, are apparently not degraded.

Discussion

Function of the FPLs

We consider it unlikely that these proteins are fluorescent for several reasons. Although nearly all ctenophores are bioluminescent, most are not fluorescent. Thus, even though the gene is being expressed, that is, found from transcriptomes, the fact that most species are not fluorescent suggests that these proteins are not contributing to any fluorescent phenotype. Additionally, important amino acids known to be involved in chromophore formation are



absent in the ctenophore FPLs. While this does not strictly exclude the possibility that a chromophore could form in the native proteins, this is improbable.

The evolutionary origin of the GFP beta-barrel fold had been in question until a fragment of mouse nidogen was shown to have a remarkably similar structure [29]. Nidogen is a component of the extracellular matrix and is found in eumetazoans and the placozoan *T. adherens*, but no clear ortholog has been found in sponges or ctenophores. The G2 domain is involved in binding to perlecan and collagen, has the same secondary structure arrangement as GFP, and makes the same beta-barrel structure [29]. However, it was found that nidogens and FPSs do not align at the sequence level, showing that sequences can differ dramatically while generating similar tertiary structures [29].

Because of the sequence divergence between the G2 domain and FPSs, there is the remaining question of what would be required to generate a FPS from a non-fluorescent precursor. Some unpublished attempts to engineer a fluorescent version of nidogen G2F were unable to generate

a chromophore (Huiwang Ai and Robert Campbell, pers. comm.); their experiments included mutating the chromophore and maturation residues to those from GFP, swapping the alpha-helix of GFP into nidogen, and computational and random mutagenesis [30]. Their work suggests that many mutations are required to generate a chromophore and that the G2 domain of nidogen may very well have been optimized for orthogonal functions, thus making it unable to become a fluorescent protein. However, the divergence of the G2 domain from the pre-FPS may have occurred long before the most recent common ancestor (MRCA) of eumetazoans (bilateral animals and cnidarians) and ctenophores, which may also explain the dramatic sequence divergence between the two proteins. The ctenophore proteins, with higher percentage identity to FPSs than nidogen G2 domains, may be more amenable to such engineering experiments.

Evolution of FPSs and nidogen

It was argued that fluorescent proteins evolved once, rather than the G2 domain of a proto-nidogen becoming

fluorescent in two bilaterian lineages, arthropods (copepods) and chordates (lancelets, only *Branchiostoma*) [5, 8]. If FPs evolved one time, then either most bilaterian lineages have lost the fluorescent proteins or a few have acquired it horizontally from a cnidarian. The gene tree (Fig. 3) suggests the first scenario, as bilaterian FPs form a monophyletic cluster indicating a single origin. The presence of fluorescent proteins across all major cnidarian groups [31, 32] suggests that the last common ancestor of the cnidarian crown group already had a functional fluorescent protein, and indeed this would have been present in the MRCA of eumetazoans as well. The extreme paucity of FPs across bilaterians suggests that the benefits to these marine groups may not apply to other taxa. For instance, there may have been selection against fluorescent phenotypes across many bilaterian groups, or selection for highly pigmented surfaces for UV protection or camouflage that may have removed the need for fluorescence.

Because the phylogenetic tree (Fig. 3) was unrooted, the outgroup of the fluorescent proteins cannot be assumed purely from the sequences provided. Therefore, the tree is compatible with two hypotheses for the relationship of ctenophores to the rest of the tree: the “coelenterata” hypothesis, monophyly of cnidarians and ctenophores, or to two alternative hypotheses, the “ctenophore-sister” or “sponge-sister” hypotheses, where ctenophores or sponges are sister group to all other metazoans, respectively. For the “coelenterata” hypothesis, if it is still assumed that there was a single evolutionary event to create fluorescent proteins, then ctenophores must have kept the orthologous proteins but lost the fluorescent function. This scheme also would require both one addition of fluorescence for all eumetazoans and one loss at ctenophores, and still offers no explanation of how fluorescence evolved in the first place. Detailed morphological analysis [33] and some recent phylogenomic analyses of metazoan proteins [21, 34] or gene content [21] find no support for the “coelenterata” hypothesis, although another study recovers this grouping [36]. Alternatively, given that there are no known FPs or FPLs from sponges or placozoans, the two alternative hypotheses, “ctenophore-sister” and “sponge-sister”, are both topologically equivalent for our dataset. Although the order of the basal groups remains a matter of complex debate [21, 35, 36], in either of two “sister” hypotheses, the implication is that ctenophores retained the ancestral protein, which later became fluorescent in the branch leading to bilaterians and cnidarians.

It was suggested that the G2 domain (and ultimately the full nidogen protein with the modern domain structure) emerged by an ancient duplication before the MRCA of eumetazoans [5]. One copy of this ancient protein became

incorporated into nidogen, while the other changed to become fluorescent proteins in a pre-eumetazoan. The presence of nidogen with the G2 domain in a number of cnidarian genomes, such as *Nematostella vectensis* [37] and *Acropora digitifera* [38], and in bilateria indicates that both nidogen and fluorescent proteins were present at the MRCA of eumetazoans. Additionally, complete nidogen is also found in the genome of the placozoan *Trichoplax adherens* [39], showing that the G2 domain was already incorporated into nidogen before the emergence of eumetazoans. However, given that the G2 domain is a single exon in the placozoan *Trichoplax adherens* and the protein domain boundaries in the human gene nidogen-1 correspond to the exon boundaries as well, it is possible that the domain could have “jumped out” to become an isolated gene rather than individual domains joining together to form a larger protein, as is expected of nidogen. Thus, it is possible that FPs and FPLs were formed by copying out a single domain from a multidomain protein, even if that multidomain protein has since been lost multiple times. Further genomic analyses of non-bilaterian species and single-celled eukaryotes may uncover an older origin of this protein superfamily.

Conclusions

Here we have demonstrated that proteins from the FP family are found in transcriptomes of 30 ctenophores. These FPLs lack the normal features of canonical fluorescent proteins and are unlikely to form fluorescent chromophores. The ctenophore FPLs form a clade within a tree of FPs from cnidarians and bilaterians, and may represent the ancestral condition of this protein before it became fluorescent in eumetazoans.

Unexpectedly, we found siphonophore FP sequences in the transcriptome of *Haeckelia beehleri*. The siphonophore FP genes found in this study are nearly identical to those reported years before [20], likely from the same species. Better identification of the prey may help to reveal why the exogenous RNA and the proteins evade degradation in *H. beehleri*.

Methods

Specimens

Specimens were collected either by trawl net, during blue-water dives, or using remotely-operated-underwater vehicles (ROVs), as described previously [40]. Animals were collected in the region bounded by 36° 44' N 122° 02'W to the northeast and 35° 21'N 124° 00'W to the southwest. Operations were conducted under permit SC-4029 issued to SHD Haddock by the California Department of Fish and Wildlife. Species used are unprotected and unregulated, and no vertebrates or octopus were used, so the International and NIH ethics guidelines are not invoked. All samples were frozen in

liquid nitrogen immediately following collection. All specimens were sequenced at the University of Utah using the Illumina HiSeq2000 platform paired-end with 100 cycles.

Transcriptome assembly and analysis

All computations were done on a computer with two 2.5 GHz quad-core processors and 96 GB RAM. For each sample, raw RNAseq reads were processed as previously published [41]. Briefly, read order was randomized. Low-quality reads, adapters, and repeats were removed. For efficiency, subsets of reads were used to assemble transcriptomes. Assembly was done with both Velvet/Oases (v1.2.09/0.2.08) [42, 43] and Trinity (r2012-10-05) [44], though in nearly all cases better assemblies were obtained with Trinity. Transcripts from both assemblers were combined and redundant sequences were removed using the “sequiq” program in the GenomeTools package [45]. All BLAST searches were done using the NCBI BLAST 2.2.28+ package [46]. Ctenophore sequences used in analysis can be found at NCBI GenBank, with accessions: KT964712-KT964739, or found in Additional file 2.

Reference data

Gene models, scaffolds, and proteins for the *Mnemiopsis leidyi* genome [21] v2.2 were downloaded from the *Mnemiopsis* Genome Portal (<http://research.nhgri.nih.gov/mnemiopsis/>). Gene models and transcripts for *Pleurobrachia bachei* genome v1.1 [34] were downloaded from the the Moroz Lab (<http://moroz.hpc.ufl.edu/>). Because there was not a systematic correspondence of the transcripts to proteins in the draft genome of *Pleurobrachia bachei*, further nucleotide analyses were excluded. Transcriptomic raw reads for *Abylopsis tetragona*, *Nanomia bijuga*, and *Physalia physalis* were downloaded from the NCBI SRA bioproject PRJNA205486 with accessions: SRR871525, SRR871527, and SRR871528, respectively [47]. Reads were assembled as above for the ctenophore transcriptomes.

Sequence alignments and phylogenetic tree generation

Alignments for proteins sequences were created using MAFFT v7.029b, with L-INS-i parameters for accurate alignments [48]. The final alignment for tree generation contained 78 sequences with 428 positions, of which 44 % are gaps or undetermined positions. The phylogenetic tree for the fluorescent proteins was generated using RAXML-HPC-PTHREADS v7.2.8 [49], using the PROTCATWAG model for proteins and 1000 bootstrap replicates with the “rapid bootstrap” (-f a) algorithm and a random seed of 1234. Alternative models were tested (PROTGAMMAWAG and PROTCATLG,) though this resulted in nearly identical topology of the best tree with negligible changes to bootstrap support. Because some of the siphonophore sequences contained multiple tandem

domains, these domains were split and treated as separate proteins for alignment and tree building; in all cases FP domains formed monophyletic groups by species, suggesting that domain splitting had minimal impact on the overall tree topology. An alternate alignment was generated with only a single domain from each multidomain FP, though the resulting tree had no meaningful difference in topology and still maintained monophyly of ctenophore FPLs and bilaterian FPs with 100 % bootstrap support.

Additional files

Additional file 1: Fasta format of the alignment of all FPs and FPLs, for any alignment viewer such as SeaView. (TXT 34.4 kb)

Additional file 2: Unaligned proteins from ctenophores used to generate Figure 3. (TXT 19.5 kb)

Additional file 3: Newick format tree used for Figure 3, containing all bootstraps. (TXT 4.90 kb)

Acknowledgements

We would like to thank the ship operators and ROV pilots at MBARI for their careful assistance in capturing specimens, and the reviewers for helpful suggestions on the paper. WRF would like to thank M. Eitel and H. Belahbib for helpful discussions.

Funding

This work was supported by the David and Lucile Packard Foundation and funded by the National Institute of General Medical Sciences at the National Institutes of Health (NIGMS:5-R01-GM087198) to SHDH. This work was also supported in part by the Intramural Research Program of the National Human Genome Research Institute, National Institutes of Health to CES.

Availability of data and material

Ctenophore sequences used in analysis can be found at NCBI GenBank, with accessions: KT964712-KT964739, or found in Additional file 2.

Authors' contributions

WRF, LMC, MLP, SHDH designed experiments. WRF, LMC and MLP performed experiments. WRF, CES, and SHDH analyzed the data. WRF wrote the paper with contributions from the other authors. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

All authors have read and approved the final version of this manuscript.

Ethics approval and consent to participate

Not applicable.

Author details

¹Monterey Bay Aquarium Research Institute, 7700 Sandholdt Rd., 95039 Moss Landing, USA. ²National Human Genome Research Institute, National Institutes of Health, 50 South Drive, 20892 Bethesda, USA. ³Present address: Ludwig-Maximilians-Universität München, Munich, Germany. ⁴Present address: Whitney Laboratory for Marine Bioscience, University of Florida, St. Augustine, 32080 Florida, USA.

Received: 25 January 2016 Accepted: 9 August 2016

Published online: 24 August 2016

References

- Prasher DC, Eckenrode VK, Ward WW, Prendergast FG, Cormier MJ. Primary structure of the *Aequorea victoria* green-fluorescent protein. *Gene*. 1992;111(2):229–33. doi:10.1016/S0167-7799(98)01184-6.
- Matz MV, Fradkov aF, Labas Ya, Savitsky aP, Zaraisky aG, Markelov ML, Lukyanov Sa. Fluorescent proteins from nonbioluminescent Anthozoa species. *Nat Biotechnol*. 1999;17(10):969–73. doi:10.1038/13657.
- Labas YA, Gurskaya NG, Yanushevich YG, Fradkov AF, Lukyanov KA, Lukyanov SA, Matz MV. Diversity and evolution of the green fluorescent protein family. *Proc Natl Acad Sci U S A*. 2002;99(7):4256–61. doi:10.1073/pnas.062552299.
- Alieva NO, Konzen Ka, Field SF, Meleshkevitch Ea, Hunt ME, Beltran-Ramirez V, Miller DJ, Wiedenmann J, Salih A, Matz MV. Diversity and evolution of coral fluorescent proteins. *PLoS ONE*. 2008;3(7):2680. doi:10.1371/journal.pone.0002680.
- Shagin Da, Barsova EV, Yanushevich YG, Fradkov AF, Lukyanov Ka, Labas Ya, Semenova TN, Ugalde Ja, Meyers A, Nunez JM, Widder Ea, Lukyanov Sa, Matz MV. GFP-like proteins as ubiquitous metazoan superfamily: evolution of functional features and structural complexity. *Mol Biol Evol*. 2004;21(5):841–50. doi:10.1093/molbev/msh079.
- Masuda H, Takenaka Y, Yamaguchi A, Nishikawa S, Mizuno H. A novel yellowish-green fluorescent protein from the marine copepod, *Chiridius poppei*, and its use as a reporter protein in HeLa cells. *Gene*. 2006;372:18–25. doi:10.1016/j.gene.2005.11.031.
- Hunt ME, Scherrer MP, Ferreri FD, Matz MV. Very bright green fluorescent proteins from the Pontellid copepod *Pontella mimoceramii*. *PLoS ONE*. 2010;5(7):11517. doi:10.1371/journal.pone.0011517.
- Deheyndt DD, Kubokawa K, McCarthy JK, Murakami A, Porrachia M, Rouse GW. Endogenous Green Fluorescent Protein (GFP) in *Amphioxus*. *Biol Bull*. 2007;213(October):95–100.
- Baumann D, Cook M, Ma L, Mushegian A, Sanders E, Schwartz J, Yu CR. A family of GFP-like proteins with different spectral properties in lancelet *Branchiostoma floridae*. *Biol Direct*. 2008;3:28. doi:10.1186/1745-6150-3-28.
- Yang F, Moss LG, Phillips GN. The molecular structure of green fluorescent protein. *Nat Biotechnol*. 1996;14(10):1246–51. doi:10.1038/nbt1096-1246.
- Shaner NC, Patterson GH, Davidson MW. Advances in fluorescent protein technology. *J Cell Sci*. 2007;120(Pt 24):4247–60.
- Chudakov DM, Matz MV, Lukyanov S, Lukyanov K. a. Fluorescent proteins and their applications in imaging living cells and tissues. *Physiol Rev*. 2010;90(3):1103–63. doi:10.1152/physrev.00038.2009.
- Shimomura O, Johnson FH, Saiga Y. Extraction, Purification and Properties of Aequorin, a Bioluminescent Protein from the Luminous Hydromedusa, *Aequorea*. *J Cell Comp Physiol*. 1962;59(3):223–39. doi:10.1002/jcp.1030590302.
- Morise H, Shimomura O, Johnson FH, Winant J. Intermolecular energy transfer in the bioluminescent system of *Aequorea*. *Biochemistry*. 1974;13(12):2656–62.
- Prasher D, McCann RO, Cormier MJ. Cloning and expression of the cDNA coding for aequorin, a bioluminescent calcium-binding protein. *Biochem Biophys Res Commun*. 1985;126(3):1259–68.
- Haddock SHD, Dunn CW. Fluorescent proteins function as a prey attractant: experimental evidence from the hydromedusa *Olindias formosus* and other marine organisms. *Biol Open*. 2015:1–11. doi:10.1242/bio.012138.
- Haddock SHD, Case JF. Not All Ctenophores Are Bioluminescent: *Pleurobrachia*. *Biol Bull*. 1995;189(3):356. doi:10.2307/1542153.
- Haddock SHD, Case JF. Bioluminescence spectra of shallow and deep-sea gelatinous zooplankton: ctenophores, medusae and siphonophores. *Mar Biol*. 1999;133(3):571–82. doi:10.1007/s002270050497.
- Haddock SHD, Moline MA, Case JF. Bioluminescence in the Sea. *Annu Rev Mar Sci*. 2010;2(1):443–93. doi:10.1146/annurev-marine-120308-081028.
- Haddock SHD, Mastroianni N, Christianson LM. A photoactivatable green-fluorescent protein from the phylum Ctenophora. *Proc Biol Sci R Soc*. 2010;277(1685):1155–60. doi:10.1098/rspb.2009.1774.
- Ryan JF, Pang K, Schnitzler CE, Nguyen A-d, Moreland RT, Simmons DK, Koch BJ, Francis WR, Havlak P, Smith Sa, Putnam NH, Haddock SHD, Dunn CW, Wolfsberg TG, Mullikin JC, Martindale MQ, Baxeavanis AD, Comparative N, Program S. The Genome of the Ctenophore *Mnemiopsis leidyi* and Its Implications for Cell Type Evolution. *Science*. 2013;342(6164):1242592–1242592. doi:10.1126/science.1242592.
- Schnitzler CE, Pang K, Powers ML, Reitzel AM, Ryan JF, Simmons D, Tada T, Park M, Gupta J, Brooks SY, Blakesley RW, Yokoyama S, Haddock SH, Martindale MQ, Baxeavanis AD. Genomic organization, evolution, and expression of photoprotein and opsin genes in *Mnemiopsis leidyi*: a new view of ctenophore photocytes. *BMC Biol*. 2012;10(1):107. doi:10.1186/1741-7007-10-107.
- Karasawa S, Araki T, Yamamoto-Hino M, Miyawaki A. A green-emitting fluorescent protein from Galaxeidae coral and its monomeric version for use in fluorescent labeling. *J Biol Chem*. 2003;278(36):34167–71. doi:10.1074/jbc.M304063200.
- Foufrage C, Swann K, Gonzalez Garcia JR, Campbell AK, Houlston E. An endogenous green fluorescent protein-photoprotein pair in *Clytia hemisphaerica* eggs shows co-targeting to mitochondria and efficient bioluminescence energy transfer. *Open Biol*. 2014;4(4):130206. doi:10.1098/rsob.130206.
- Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods*. 2011;8(10):785–6. doi:10.1038/nmeth.1701.
- Savojarjo C, Martelli PL, Fariselli P, Casadio R. TPpred2: improving the prediction of mitochondrial targeting peptide cleavage sites by exploiting sequence motifs. *Bioinformatics*. 2014;30(20):2973–974. doi:10.1093/bioinformatics/btu411.
- Barondeau DP, Putnam CD, Kassmann CJ, Tainer Ja, Getzoff ED. Mechanism and energetics of green fluorescent protein chromophore synthesis revealed by trapped intermediate structures. *Proc Natl Acad Sci U S A*. 2003;100(21):12111–6. doi:10.1073/pnas.2133463100.
- Hunt ME, Modi CK, Aglyamova GV, Ravikant DVS, Meyer E, Matz MV. Multi-domain GFP-like proteins from two species of marine hydrozoans. *Photochem Photobiol Sci*. 2012;11(4):637. doi:10.1039/c1pp05238a.
- Hopf M, Göhring W, Ries A, Timpl R, Hohenester E. Crystal structure and mutational analysis of a perlecan-binding fragment of nidogen-1. *Nat Struct Biol*. 2001;8(7):634–40. doi:10.1038/89683.
- Nguyen X-TT, Zimmer M. Fast-forwarding Nidogen/GFP evolution by computational design. In: 229th ACS National Meeting. San Diego; 2005.
- Ugalde JA, Chang BSW, Matz MV. Evolution of coral pigments recreated. *Science*. 2004;305(5689):1433.
- Field SF, Bulina MY, Kelmanson IV, Bielawski JP, Matz MV. Adaptive evolution of multicolored fluorescent proteins in reef-building corals. *J Mol Evol*. 2006;62(3):332–9. doi:10.1007/s00239-005-0129-9.
- Harbison GR. On the classification and evolution of the Ctenophora. In: Conway Morris SC, George JD, Gibson R, Platt HM, editors. *The Origin and Relationships of Lower Invertebrates*. Oxford: Clarendon Press; 1985. p. 78–100.
- Moroz LL, Kocot KM, Citarella MR, Dosung S, Norekian TP, Povolotskaya IS, Grigorenko AP, Dailey C, Berezikov E, Buckley KM, Ptitsyn A, Reshetov D, Mukherjee K, Moroz TP, Bobkova Y, Yu F, Kapitonov VV, Jurka J, Bobkov YV, Swore JJ, Girardo DO, Fodor A, Gusev F, Sanford R, Bruders R, Kittler E, Mills CE, Rast JP, Derelle R, Solovyev VV, Kondrashov Fa, Swalla BJ, Sweedler JV, Rogaev EI, Halanych KM, Kohn AB. The ctenophore genome and the evolutionary origins of neural systems. *Nature*. 2014;17:1–123. doi:10.1038/nature13400.
- Dunn CW, Hejnol A, Matus DQ, Pang K, Browne WE, Smith Sa, Seaver E, Rouse GW, Obst M, Edgecombe GD, Sørensen MV, Haddock SHD, Schmidt-Rhaesa A, Okusu A, Kristensen RM, Wheeler WC, Martindale MQ, Giribet G. Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature*. 2008;452(7188):745–9. doi:10.1038/nature06614.
- Phillippe H, Derelle R, Lopez P, Pick K, Borchellini C, Boury-Esnault N, Vacelet J, Renard E, Houlston E, Quéinnec E, Da Silva C, Wincker P, Le Guyader H, Leys S, Jackson DJ, Schreiber F, Erpenbeck D, Morgenstern B, Wörheide G, Manuel M. Phylogenomics revives traditional views on deep animal relationships. *Curr Biol (CB)*. 2009;19(8):706–12. doi:10.1016/j.cub.2009.02.052.
- Putnam NH, Srivastava M, Hellsten U, Dirks B, Chapman J, Salamov A, Terry A, Shapiro H, Lindquist E, Kapitonov VV, Jurka J, Genikhovich G, Grigoriev IV, Lucas SM, Steele RE, Finnerty JR, Technau U, Martindale MQ, Rokhsar DS. Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science (New York)*. 2007;317(5834):86–94. doi:10.1126/science.1139158.
- Shinzato C, Shoguchi E, Kawashima T, Hamada M, Hisata K, Tanaka M, Fujie M, Fujiwara M, Koyanagi R, Ikuta T, Fujiyama A, Miller DJ, Satoh N.

Using the *Acropora digitifera* genome to understand coral responses to environmental change. *Nature*. 2011;476(7360):320–3. doi:10.1038/nature10249.

39. Srivastava M, Begovic E, Chapman J, Putnam NH, Hellsten U, Kawashima T, Kuo A, Mitros T, Salamov A, Carpenter ML, Signorovitch AY, Moreno Ma, Kamm K, Grimwood J, Schmutz J, Shapiro H, Grigoriev IV, Buss LW, Schierwater B, Dellaporta SL, Rokhsar DS. The Trichoplax genome and the nature of placozoans. *Nature*. 2008;454(7207):955–60. doi:10.1038/nature07191.
40. Francis WR, Shaner NC, Christianson LM, Powers ML, Haddock SHD. Occurrence of Isopenicillin-N-Synthase Homologs in Bioluminescent Ctenophores and Implications for Coelenterazine Biosynthesis. *Plos ONE*. 2015;10(6):0128742. doi:10.1371/journal.pone.0128742.
41. Francis WR, Christianson LM, Kiko R, Powers ML, Shaner NC, Haddock SHD. A comparison across non-model animals suggests an optimal sequencing depth for de novo transcriptome assembly. *BMC Genomics*. 2013;14(1):167. doi:10.1186/1471-2164-14-167.
42. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res*. 2008;18(5):821–9. doi:10.1101/gr.074492.107.
43. Schulz MH, Zerbino DR, Vingron M, Birney E. Oases: Robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics (Oxford)*. 2012;26(12):1509–15. doi:10.1093/bioinformatics/bts094.
44. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson Da, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol*. 2011;29(7):644–52. doi:10.1038/nbt.1883.
45. Gremme G, Steinbiss S, Kurtz S. GenomeTools: A Comprehensive Software Library for Efficient Processing of Structured Genome Annotations. *IEEE/ACM Trans Comput Biol Bioinforma*. 2013;10(3):645–56. doi:10.1109/TCBB.2013.68.
46. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. *BMC Bioinforma*. 2009;10:421. doi:10.1186/1471-2105-10-421.
47. Zapata F, Goetz FE, Smith SA, Howison M, Siebert S, Church SH, Sanders SM, Ames CL, McFadden CS, France SC, Daly M, Collins AG, Haddock SHD, Dunn CW, Cartwright P. Phylogenomic Analyses Support Traditional Relationships within Cnidaria. *Plos ONE*. 2015;10(10):0139068. doi:10.1371/journal.pone.0139068.
48. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013;30(4):772–80. doi:10.1093/molbev/mst010.
49. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics (Oxford)*. 2006;22(21):2688–90. doi:10.1093/bioinformatics/btl446.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

