McCulloch *et al. EvoDevo* (2023) 14:14
https://doi.org/10.1186/s13227-023-00218-8

EvoDevo

RESEARCH

Open Access

Nematostella vectensis exemplifies the exceptional expansion and diversity of opsins in the eyeless *Hexacorallia*

Kyle J. McCulloch[1], Leslie S. Babonis[2,5], Alicia Liu[3,4], Christina M. Daly[3,4], Mark Q. Martindale[5] and Kristen M. Koenig[3,4,6*]

Abstract

Background Opsins are the primary proteins responsible for light detection in animals. Cnidarians (jellyfish, sea anemones, corals) have diverse visual systems that have evolved in parallel with bilaterians (squid, flies, fish) for hundreds of millions of years. Medusozoans (e.g., jellyfish, hydroids) have evolved eyes multiple times, each time independently incorporating distinct opsin orthologs. Anthozoans (e.g., corals, sea anemones,) have diverse light-mediated behaviors and, despite being eyeless, exhibit more extensive opsin duplications than medusozoans. To better understand the evolution of photosensitivity in animals without eyes, we increased anthozoan representation in the phylogeny of animal opsins and investigated the large but poorly characterized opsin family in the sea anemone *Nematostella vectensis*.

Results We analyzed genomic and transcriptomic data from 16 species of cnidarians to generate a large opsin phylogeny (708 sequences) with the largest sampling of anthozoan sequences to date. We identified 29 opsins from *N. vectensis* (*NvOpsins*) with high confidence, using transcriptomic and genomic datasets. We found that lineage-specific opsin duplications are common across Cnidaria, with anthozoan lineages exhibiting among the highest numbers of opsins in animals. To establish putative photosensory function of *NvOpsins*, we identified canonically conserved protein domains and amino acid sequences essential for opsin function in other animal species. We show high sequence diversity among *NvOpsins* at sites important for photoreception and transduction, suggesting potentially diverse functions. We further examined the spatiotemporal expression of *NvOpsins* and found both dynamic expression of opsins during embryonic development and sexually dimorphic opsin expression in adults.

Conclusions These data show that lineage-specific duplication and divergence has led to expansive diversity of opsins in eyeless cnidarians, suggesting opsins from these animals may exhibit novel biochemical functions. The variable expression patterns of opsins in *N. vectensis* suggest opsin gene duplications allowed for a radiation of unique sensory cell types with tissue- and stage-specific functions. This diffuse network of distinct sensory cell types could be an adaptive solution for varied sensory tasks experienced in distinct life history stages in Anthozoans.

Keywords *Nematostella*, Sea anemone, Opsin, Rhodopsin, Photoreceptor, Cnidaria, Anthozoa, Hexacorallia

*Correspondence:
Kristen M. Koenig
kmkoenig@utexas.edu
Full list of author information is available at the end of the article

Background

Opsins are a family of G protein-coupled receptors (GPCRs) associated with animal visual systems [1, 2]. Previous phylogenetic analysis has shown that bilaterian and cnidarian opsin groups are interspersed and sister to each other, suggesting the major branches of this protein family already diversified in the common ancestor of these lineages, with 3–4 opsins thought to be already present [1, 3–6]. In all characterized opsins, the protein binds to a vitamin-A derived chromophore, forming a rhodopsin complex, responsible for absorbing light [2, 7]. The transduction of light by rhodopsin into a cellular signal occurs via a G protein signaling pathway known as the phototransduction cascade, leading to downstream ion exchange across the membrane and dictating the cell's physiological response [2]. Opsins from distinct subclades bind to specific G protein alpha subunits (Gt, Gq, Gs, etc.), which in turn signal via distinct phototransduction cascades [4, 8, 9]. Eye-associated opsins and their phototransduction cascades are well-studied in vertebrates and insects [2, 10–12]. The emergence of expanded sequencing resources across the animal tree of life has led to the discovery of new types of opsins. This has changed our understanding of opsin evolution and has modified historic interpretations of opsin function [13–16]. Although eye-related opsins have received most attention, the importance of non-ocular opsin gene expansions, from which eye-associated opsins have repeatedly evolved, is relatively unexplored [5, 17].

The best studied opsins are those expressed in vertebrate or fly eyes. Ciliary-associated (c-) opsins were once thought to be exclusive to the vertebrate eye, and microvillar associated rhabdomeric (r-) opsins were thought to be exclusive to protostome eyes [8, 18–20]. C- and r-opsins are now known to be expressed in both deuterostome and protostome lineages [14, 20–23] and new opsin clades have been found throughout Bilateria (xenopsins and tetraopsins, [24–29]). Cnidarian opsins are grouped in three clades: cnidopsins, which are found throughout Cnidaria, and two **a**nthozoan-**s**pecific **o**psin clades (ASO-I and ASO-II) [6, 29]. Cnidopsins are sister to the bilaterian-specific xenopsins. The positions of ASO-I and ASO-II are not consistently supported in published available opsin phylogenies [1, 6, 29–31]. ASO-I has been found either to be sister to r-opsins or sister to all animal opsins [1, 6, 29–31]. ASO-II is considered sister to ciliary opsins, although this is not well supported [1, 6, 30, 31]. However, ASO-II shares intronic structure with ciliary opsins, which supports this placement [29]. Together these data suggest at least 3 major opsin clades existed in the ancestor of both Bilateria and Cnidaria [1, 4, 6, 15, 32]. Evidence from both eyed and eyeless cnidarians shows that cnidarian opsins function as photoreceptors using canonical phototransduction cascades [3, 31, 33–46]. Both sequence homology and physiological evidence suggest cnidarian opsin biochemistry and function is similar to bilaterian opsins [47].

Among anthozoan opsins, gene family expansions have been observed, yet the phylogenetic history and potential adaptive consequences remain unknown [6, 29]. Although anthozoans lack eyes, large numbers of opsins have been reported in corals and sea anemones [6]. Recent next-generation sequencing in Cnidaria has improved phylogenetic sampling and clarified relationships among all animal opsins. However, cnidarian opsins have undergone large gene family expansions and transcriptomic studies have failed to consistently identify the numbers of opsin genes in each lineage, especially within anthozoans [5, 6, 47]. This variability is likely due to similar sequences of recent duplicates, spatiotemporal restriction or low expression. Without complete genomes and manual annotation, the extent of cnidarian opsin expansions and potential functional diversification remains obscured.

The starlet sea anemone, *Nematostella vectensis,* is an eyeless anthozoan and an ideal model for interrogating the evolutionary history and potential functional diversity of opsins. *N. vectensis* exhibits multiple light-mediated behaviors including spawning, larval swimming, and circadian locomotor activity [48, 49]. Previous work has shown *N. vectensis,* like other cnidarians, has many opsins; however, estimates of the precise number of opsins encoded in the genome of this animal have ranged from 30 to 52 [5, 6, 47]. Furthermore, key requisites for inferring distinct functions of each *NvOpsin*, such as genomic architecture, complete sequence, and whether opsin genes are silent or expressed, have not been analyzed. With the release of two chromosome-quality genomes, we are now able to corroborate transcriptomic evidence with genomic loci in this highly duplicated gene family [50, 51]. In this study, we reduce the most recent estimate of NvOpsins from 52 [6] to 29, closer to previous estimates [47]. Using evidence from genomic, transcriptomic, phylogenetic, and in situ mRNA expression analyses, we characterize a diverse repertoire of opsins from *N. vectensis* and suggest unique populations of opsin-expressing sensory cells have diverse roles across life stages in this animal.

Results

Phylogenetic identification and genomic architecture of the 29 opsins in *N. vectensis*

Using multiple genomic and transcriptomic sources of evidence, we were able to identify 159 additional opsin sequences from 15 recently reported anthozoan transcriptomic and genomic datasets, with each species

having between 4 and 22 paralogs (Fig. 1A). We report the chromosomal location of all previously reported *NvOpsin* gene models while removing erroneous predictions and adding new loci, identifying a total of 29 distinct genes (Fig. 1B). We made a maximum-likelihood gene tree for the opsin protein family, adding *N. vectensis* sequences and additional new anthozoan data to published cnidarian opsin alignments (Additional file 1 and Additional file 2). To facilitate comparisons across previously published investigations of opsin diversification, we have provided a table of *N. vectensis* opsin IDs from previous studies with new genomic identification and an updated naming convention (Additional file 3: Table S1). We used this tree to investigate the orthology of NvOpsins (Figs. 1, 2, 3). ASO-I opsins are highly supported (99.9/100 SH-aLRT/UFboot support) within the opsin clade, sister to all other opsins (Fig. 1C, for full tree support values see Additional File 2) [5, 6, 29]. Cnidopsins and xenopsins form a monophyletic group (88.6/90 SH-aLRT/UFboot support) and the ASO-II group is found sister to c-opsins, however this relationship is not highly supported (Fig. 1C). Our tree shows all NvOpsins fall into ASO-I, ASO-II, and cnidopsin subclades (Fig. 1C).

To further investigate possible mechanisms of opsin diversification, we assessed genomic structure and conservation of intron–exon boundaries of opsin genes in the *N. vectensis* genome (Fig. 1D, Additional File 2). Investigating chromosomal locations of all *NvOpsin* loci, we found that chromosomes containing multiple opsins largely contain genes from the same subclade and not from other subclades (Fig. 1B). We further examined intron/exon structure in *NvOpsins* and found that both *NvASO-I* genes and 5 of the 12 *NvCnidopsins* lack introns, a hallmark of duplication by retrotransposition (Fig. 1B, Additional File 3: Table S1) [29, 38, 40]. By contrast, all *NvASO-II* genes and the remaining 7 *NvCnidopsins* have introns [50]. Together, these results suggest tandem duplication and retrotransposition have been major drivers of opsin expansion in *N. vectensis*.

To further examine the evolutionary relationships within the xenopsins/cnidopsins clade, we examined intronic structure in *N. vectensis* and representative bilaterians. Cnidopsins together with xenopsins form a single clade suggesting one ortholog was present in the cnidarian/bilaterian ancestor and that each lineage (anthozoa, medusozoa, and bilateria) experienced extensive lineage-specific duplication of this ancestral gene. We compared NvCnidopsin intron/exon boundaries to bovine c-opsin and a *Crassostrea gigas* xenopsin. Anthozoan cnidopsins are split into two subclades (see below), and all *NvCnidopsins* in Subclade 2 contain introns (Additional file 3: Table S1). None of these cnidopsins share intron/exon boundaries with c-opsin, but all Group 2 *NvCnidopsins* share one intron/exon boundary position and phase with the oyster xenopsin (Fig. 1D). This conserved intron–exon structure further supports a common ancestor of all cnidopsins and xenopsins and suggests Subgroup 2 *NvCnidopsins* are structurally similar to the ancestral cnidopsin.

Extensive duplication and divergence of opsins is a common feature of Hexacorallia

The addition of more anthozoan sequences in our phylogeny reveals opsin gene family expansions are common in Hexacorallia (Fig. 2). In general, all anthozoans have two ASO-I opsin paralogs that group into two monophyletic subclades, suggesting that two ASO-I opsins were present in the common ancestor of this clade. Very few lineages exhibit more than two ASO-I paralogs although one or both duplicates are absent in several lineages, and both *NvASOI* duplicates are intronless, suggesting gene loss, rather than duplication, shaped the evolution of ASO-I opsins (Fig. 2A). In contrast, ASO-II opsin paralogs are expanded, and divergence within and between lineages is high (see branch lengths, Fig. 1C). ASO-II duplications already occurred in the lineage leading to Hexacorallia (Fig. 2B), while additional recent duplications greatly expanded ASO-II numbers within Hexacorallia (Fig. 2A, B). No ASO-II members have been

(See figure on next page.)

Fig. 1 Phylogenetic placement and genomic architecture of the 29 *N. vectensis* opsins. **A** Simplified cnidarian phylogeny adapted from [91]. The range of known opsins found in a single species in each group is indicated on the phylogeny (from literature review). *N. vectensis* is in Actiniaria, Hexacorallia, in the Anthozoa (in blue) and has 29 opsins, the most so far identified of any anthozoan. **B**, **C** Light blue corresponds to ASO-I, purple to ASO-II, and green to cnidopsins. **B** *N. vectensis* chromosomes with opsin loci are shown. Nearly all *N. vectensis* opsins segregate on chromosomes by clade. Numbers below each chromosome are length in megabases, (+) indicates recent tandem duplicates with highly similar sequences. Arrowheads indicate direction the gene is found in the genome. **C** Maximum likelihood tree of 708 opsins, with major animal opsin clades labeled (Tetraopsins include RGR/Go opsins/Group 4 opsins). Cnidarian-specific clades are colored and in bold. IQtree branch support is defined by ultrafast bootstraps (Ufboot) and likelihood ratio test (SH-aLRT). The number of *N. vectensis* opsins in each clade is listed (blue numbers). **D** Conservation of intron structure in the cnidops/xenops clade. A representative xenopsin from the oyster *Crassostrea gigas* shares an intron/exon boundary with all other xenopsins investigated [29] and several *N. vectensis* cnidopsins. Red box shows intron/exon boundary mapped on to amino acid alignment. Gray bars represent aligned sequence, black lines are gaps in the alignment

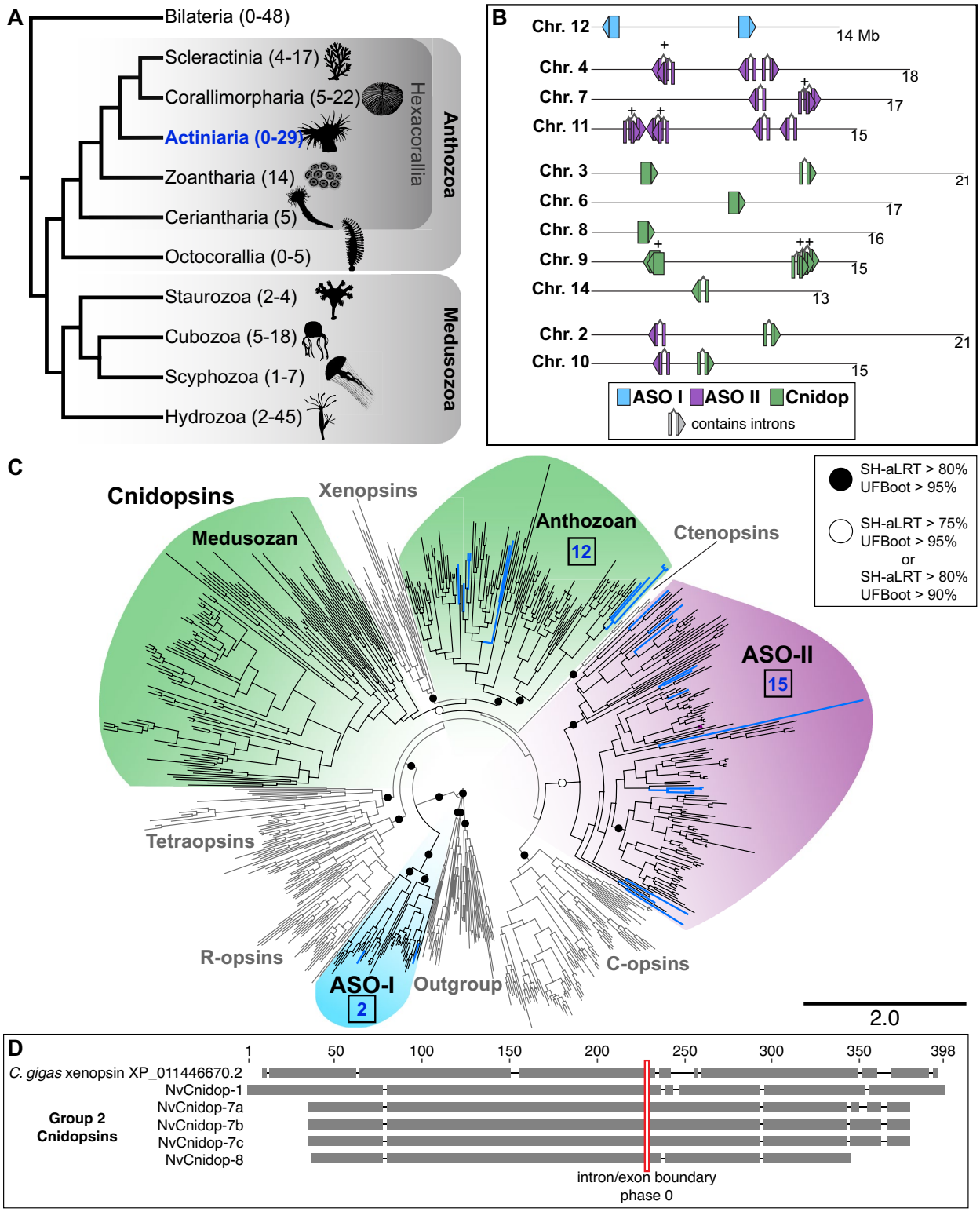

Fig. 1 (See legend on previous page.)

identified in any Octocorallia species (soft corals, sea pens), suggesting a loss in this lineage. The presence of two ASO-II sequences from ceriantharians (tube anemones) suggests a minimum of two ASO-II opsins were present in the last common ancestor of Hexacorallia. Within the ASO-II clade, previously identified ASO-II 2.1 and 2.2 subgroups [6] are well supported. Previously identified Subgroup 1 is less well supported, and this large group shows four duplication events occurred in the lineage leading to Scleractinia and Corallimorpharia. Among sea anemone opsins, parallel duplications have occurred between the major anemone subfamilies (Anenthemonae—to which *N. vectensis, Scolanthus, and Edwardsiella* belong, and Enthemonae to which all other included species belong) (Figs. 2A, 3B). Subgroup 2.1 has fewer duplications in hexacorallian lineages than Subgroup 1. Sister to Subgroup 2.1 is a newly identified group including the highly divergent NvASOII-4 sequence (Figs. 3B, 4B), orthologs of which were identified in Corallimorpharia and Scleractinia but not found in any other sea anemone. Previously Subgroup 2.2 was thought only to contain sea anemone sequences [6]. The addition of new data identifies cerianthid, zoanthid, and corallimorpharian sequences, suggesting this subgroup of opsins arose in the last common ancestor of Hexacorallia (Fig. 3B).

Within the cnidopsin clade, we identify one well-supported subclade containing only Hexacorallia and another clade containing all anthozoan lineages (Figs. 2, 3C). This suggests one cnidopsin ortholog was present in the common ancestor of anthozoans and a duplication event occurred in the common ancestor of Hexacorallia. Like the ASO-II group, multiple lineage and species-specific duplications have expanded cnidopsin numbers in anthozoans. Cnidopsin Subgroup 1 contains all the octocorallian sequences in a single well supported clade, with at least one octocorallian-specific duplication event or multiple species-specific events. No cerianthid sequences were identified as cnidopsins, though one is likely placed improperly in the bilaterian-specific xenopsins, and likely to be the cerianthid cnidopsin ortholog (Additional file 2). The ancestor of Scleractinia and Corallimorpharia duplicated their Group 1 and Group 2 opsins at least once, followed later by multiple species-specific duplications (*A. digitifera,* six opsins, *Discosoma,* eight opsins). In *N. vectensis*, lineage-specific expansions led to 12 total cnidopsins, in parallel to multiple independent duplications in the other sea anemone group, Enthemonae. Together our phylogeny reveals extensive duplication of opsins throughout the diversification of anthozoans contributed to the expansive repertoire of opsins in *N. vectensis* and other cnidarians.

To better understand this evolutionary history Anthozoan gene loss and gain, we inferred ancestral opsin numbers through multiple methods. We first assessed the pattern by parsimony considering the gene tree structure and the minimum number of gains and losses for both each opsin clade and each anthozoan lineage (Fig. 2B). To support this work with more formal analysis, we used Mesquite to reconstruct ancestral states using opsin counts as both continuous and meristic data (Additional file 4: Fig. S1). Mesquite predicted similar patterns of opsin gains and losses in the anthozoan species tree, but missed losses that we infer from phylogenetic relationships in the gene tree. To include phylogenetic signal we used Notung to predict gains and losses (Additional file 4: Fig. S2). Notung predicted major losses that we infer through gene tree topology. Surprisingly, Notung predicted far more opsin duplicates early in animal evolution (14 at the ancestor of Eumetazoa) and far more losses in the leaves of the species tree (Additional file 3: Table S2, Additional file 4: Fig. S2, 3).

NvOpsins are intact GPCRs with a potential diversity of functions

In addition to phylogenetic placement, we assessed whether the 29 NvOpsins were capable of functioning as GPCRs and the protein components of photopigments by analysis of conserved functional residues. The most basic criteria for any opsin to function as a photopigment are the presence of seven transmembrane domains, which form a chromophore binding pocket, and the presence of a lysine at position 296 (bovine rhodopsin numbering), which binds to the chromophore (Fig. 4A) [2]. The presence of a pair of cysteines to stabilize the opsin structure via a disulfide bridge is additionally required (Fig. 4A) [52]. A negatively charged residue known as

(See figure on next page.)

Fig. 2 Current and ancestral anthozoan-specific opsin duplications. **A** Anthozoan species with opsin data are listed. Each column is the subclade number within the three major anthozoan opsin clades. "Un" is ASO-II opsins that are unspecified, or not found in previously identified clades. The number of opsins for each species and opsin subclade is listed in the boxes. Gray boxes with 0 signify genomic evidence of no opsins, while white boxes signify no opsin identified from transcriptomic evidence. Species names in bold have genomic evidence available, asterisks are for species with newly reported opsins in this study. **B** Anthozoan lineage tree, with numbers of opsins represented by line thickness and estimated by parsimony based on opsin numbers and phylogenetic positions (Fig. 3). Numbers in boxes represent the estimated number of opsins present in the last common ancestor of each anthozoan lineage. An increase in thickness of lines signifies an opsin duplication while an x signifies loss of the gene. The dashed line signifies uncertainty due to limited sampling in Octocorallia

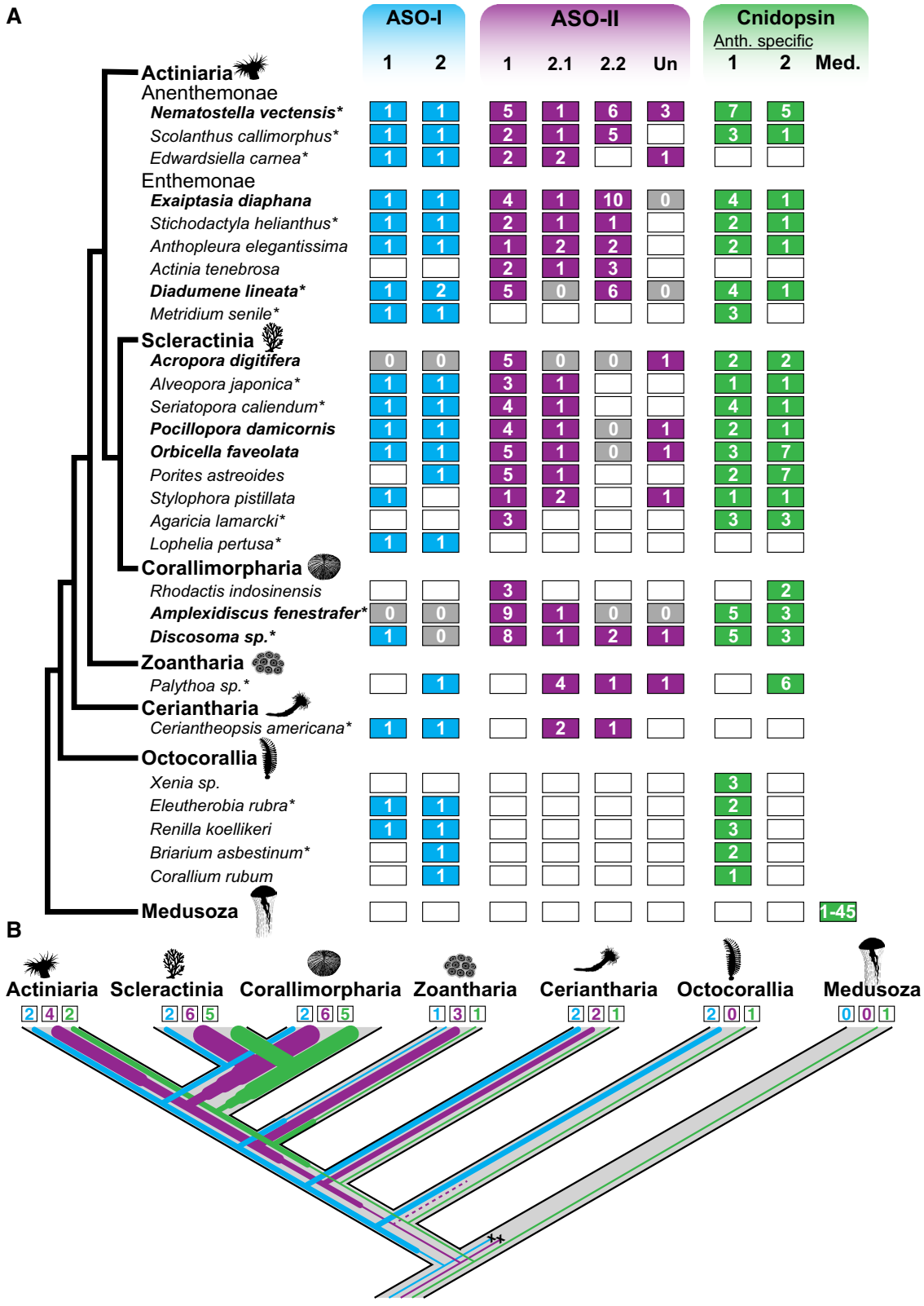

Fig. 2 (See legend on previous page.)

the counterion is thought to be required to stabilize the interaction with the chromophore, but the specific site and residue of this counterion can vary across opsins clades [2]. Beyond these basic functional features, conserved sites on the cytoplasmic tail/loops bind to specific G protein alpha subunits, which are required for activating the phototransduction cascade [53, 54]. Opsins within a clade are usually specialized in binding one or limited number of G protein alpha subunit subtypes (Gt, Gq, Gs, etc.) [8]. The few bilaterian and jellyfish opsins that have been characterized in detail have identified specific functionally relevant amino acid sites and identities required for G protein signaling [41, 42] (Fig. 4A).

All NvOpsins have the canonical seven transmembrane domains and all but one have the conserved lysine, Lys-296 (Fig. 4B). Interestingly, NvASOII-4 has a Lys296His substitution at this site, suggesting a non-photoreceptive function (Fig. 4B). Previously identified counterion sites included Glu-113 in vertebrate ciliary opsins and Glu-181 in invertebrate Go-coupled opsins and Gq-coupled r-opsins [8, 55]. There is some evidence that Asp-83 is the counterion site for Gq-coupled melanopsin, the r-opsin found in vertebrates, and Glu-94 is the counterion in cubozoan Gs-coupled cnidopsin [8, 41]. No NvOpsin has Glu-94 (cubozoan) or similar negatively charged amino acid at this position, however some have either Glu-181 or Glu-113 and NvASOII-1 have both (Fig. 4B). All NvCnidopsins, most NvASO-II, and one NvASO-I sequence has Asp-83, while other NvASO-II and NvASO-I sequences has an Asn substitution at the Asp-83 site. Only NvASOII-4 has neither aspartic acid nor asparagine in this position, instead having Cys-83 (Fig. 4B). Together sequence analysis suggests NvOpsin paralogs tend to have bilaterian-like counterion residues while the Glu-94 counterion appears to be an evolutionary novelty in box jellies.

Spectral tuning—the shift in the spectrum of wavelengths an opsin-based visual pigment absorbs—can be mediated by amino acid substitutions that affect the interaction with the chromophore [56]. Two substitutions, Asp83Asn and Ala292Ser, result in opsin absorbance shifts from green toward blue, and have evolved independently in multiple vertebrates [57–59] and insects [60, 61]. We find that in many *N. vectensis* sequences, one or both substitutions are also present, and evolved independently within *N. vectensis* opsins multiple times (Fig. 4B). This suggests potential blue shifts in wavelength absorption compared to related opsins, although the absorbance spectra of these opsins remain to be measured. The functional relevance of other amino acids at either of these sites is unclear. The conservation of these sites in opsins as distantly related as c- and r- opsins suggests their presence is also involved in spectral tuning of *N. vectensis* opsins, but further functional evidence is needed for confirmation.

We also investigated conservation of sites known to be important in G-protein signaling. In the rhodopsin-like GPCR family, the conserved tripeptide Asp/Glu-Arg-Tyr/Trp at sites 134–136 is known to be important for receptor activation, located after the third transmembrane domain in the second intracellular loop [62–64]. Only NvASOI-1 and NvCnidop-4 have all three of any of these residues. Most NvCnidopsins have a similarly positively charged lysine in place of arginine at site 135, which may be a conservative substitution. We also searched for three known motifs that directly interact with the G protein alpha subunit and are important for G protein signaling activation [65]. Most *N. vectensis* opsins have NPXXY at these sites, but some of the NvASO-II sequences are divergent at this highly conserved motif. Another known functional motif is the tripeptide at sites 310–312 followed by a conserved FR sequence (313–314). The canonical c-opsin and r-opsin tripeptide is NKQ and HMK, respectively, while the functional box jelly photopigment, Tcop13, has HKQ at this site [40]. In the box jelly opsin, in vitro substitution with other tripeptides did not result in inactivation of Gs signaling, but instead altered the dynamics of light response in cell lines [40]. The FR motif is conserved in NvOpsins, but none of the known tripeptide motifs are found in any NvOpsin. Given the *T. cystophora* evidence, these data may suggest functional diversity in binding dynamics of G proteins across NvOpsins.

(See figure on next page.)

Fig. 3 Anthozoan-specific opsin evolutionary patterns of duplication and loss. Trees are zoomed subsets of the maximum-likelihood tree (Fig. 1C) for each anthozoan opsin group. Species names and branches are color coded according to lineage. Support for branches is denoted with a black circle or a white circle. **A** ASO-I group opsins are split into two main subclades, with most anthozoans having an opsin duplicate in each subclade. **B** The ASO-II group is sister to c-opsins and comprised exclusively of opsins from hexacorals. Previously identified ASO-II Groups 2.1 and 2.2 [6] are well-supported but ASO-II Group 1 is not. **C** Anthozoan cnidopsins form a single well-supported clade within the larger cnidopsin/xenopsin clade. Within Anthozoan cnidopsins there are two sister clades, one of which is well-supported while the other is not. For clarity, branch lengths are transformed, and branch support is not shown for branches leading to the two shallow-most nodes on these trees (For full tree topology and support values see Fig. 1C, Additional File 2)

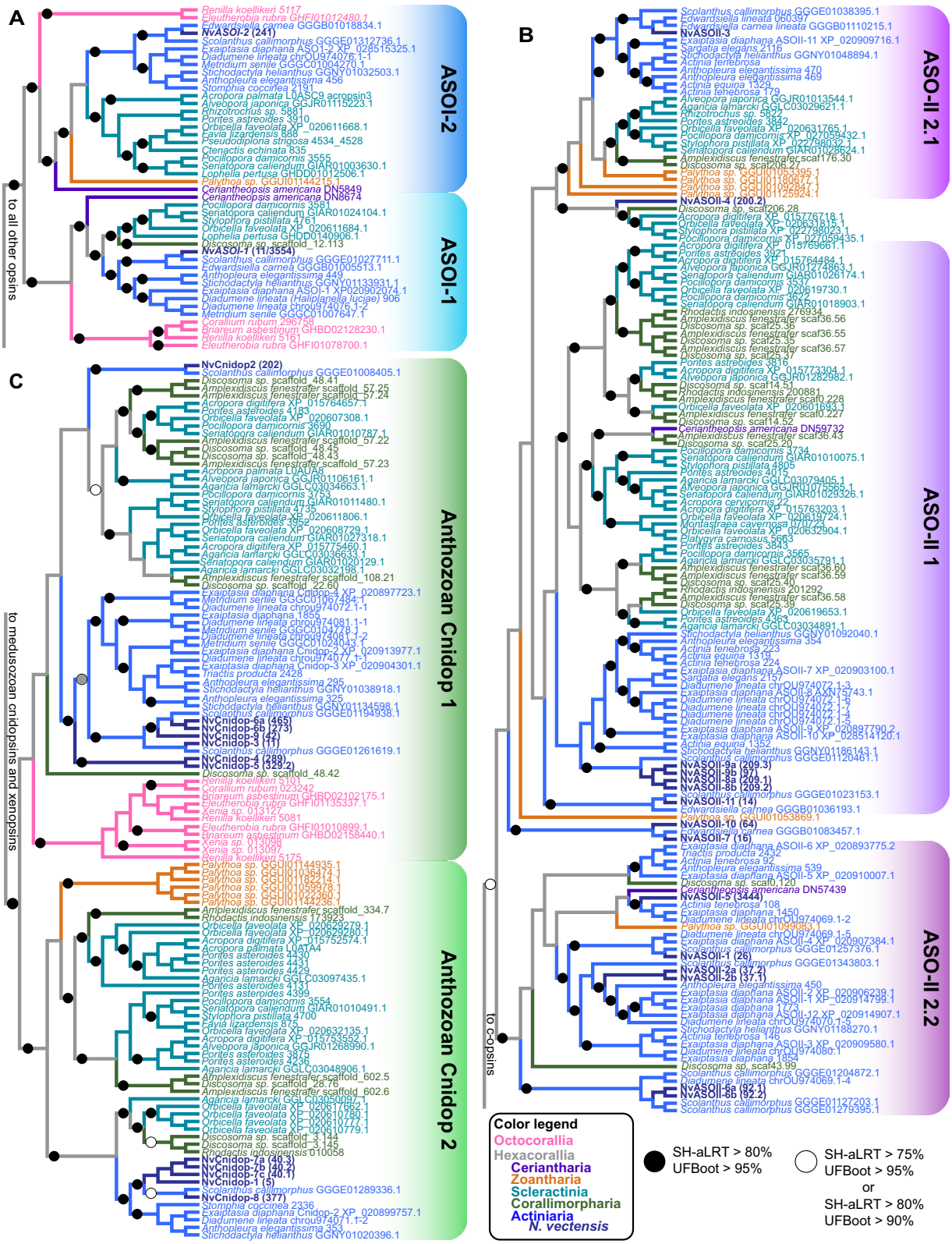

Fig. 3 (See legend on previous page.)

Opsin expression is dynamic throughout *N. vectensis* tissues during development

To investigate which tissues may play a role in light-sensitive behaviors, we analyzed spatiotemporal patterns of opsin expression in *N. vectensis* using a combination of RNA-seq and in situ hybridization. The lifecycle of *N. vectensis* proceeds from fertilization to blastula stage by 12 h and gastrula stage by 24 h. By 48 h a swimming planula larva stage develops [66]. This stage is characterized by an aboral ciliated sensory organ known as the apical organ, which faces forward when swimming. By 10 days post-fertilization mesenteries and four tentacles form and elongate and the planula metamorphoses into a primary polyp.

We generated a high-quality de novo assembled transcriptome from four developmental stages (blastula, gastrula, mid-planula, and primary polyp) as well as male and female adults. Out of 221,245,806 total reads, 96.29% were properly paired and used for downstream applications with a BUSCO completeness of 97.9%. Replicate libraries of the same stage were highly concordant with one another (Additional file 4: Fig. S4). The RNA-seq time course analysis revealed that *NvASOI-2* is highly expressed at blastula stage relative to all other opsins (Fig. 5A, top scale). We show the rest of the expression data scaled without this highly abundant transcript (Fig. 5A, bottom scale). The time course analysis allows for qualitative comparison across developmental stages for each opsin transcript. Many NvOpsins peak in expression at different stages, including blastula (*NvASOII-8a, -9a*), gastrula (*NvASOII-9b*), planula (*NvASOII-3; NvCnidop-1, -3, -9*), and primary polyp (*NvASOII-2a, -2b -3, -6a; NvCnidop-2b, -4, -6b*) (Fig. 5A). We also compared our data with a published developmental time course (NvERTx) for all opsins found in that dataset [67], which generally agreed with our expression levels and provided additional time resolution at early stages (Additional file 3: Table S3; Additional file 4: Fig S5). Light is known to induce spawning in *N. vectensis* and other cnidarians, so we wanted to know whether some opsins were specific to adults or specific to sex. Our comparison between adult males and females showed several opsins that were expressed in both adult sexes as well as six opsins that suggests variable expression by sex. The data suggest that *NvASOI-2, NvCnidop-2*, and *NvCnidop-8* are upregulated in females, while *NvASOII-6b* and *NvCnidop-4* are more highly expressed in males (Fig. 5B). Together the RNA-seq data suggest dynamic expression of distinct NvOpsins throughout development and across sexes.

To corroborate our RNA-seq analysis and better understand tissue specificity, we performed in situ hybridization for select opsin genes at peak expression levels (Fig. 6). Our expression studies show that tissue- and stage-specific expression patterns do not correlate with phylogenetic signal. Opsins of distinct clades are found in both germ layers and sometimes have overlapping expression patterns. Opsins from each major clade were expressed in both embryo and polyp stages (Figs. 5, 6).

In embryonic and larval stages, we observed opsin expression in both ectoderm and endomesoderm. *NvASOII-8b* is expressed by gastrula stage in scattered cells throughout the ectoderm (Fig. 6A, A′) and by late planula, there is also expression in the sensory apical organ (Fig. 6B, B′). Both *NvCnidop-9* and *NvASOI-1* are expressed in the ectoderm at planula stages (Fig. 6C–E′) and are expressed in scattered cells of the aboral ectoderm but are not expressed in the apical organ. *NvASOI-2,* the opsin with the highest expression in the RNA-seq dataset, is detected in scattered cells at late blastula stage and its expression expands into the ectoderm in larval stages (Fig. 6F–H′). Expression is concentrated orally as development continues, eventually forming a ring around the oral region in the late planula stages (Fig. 6H, H′).

(See figure on next page.)

Fig. 4 *N. vectensis* opsins have high levels of sequence diversity at canonically conserved functional sites. **A** Left, cartoon bovine rhodopsin structure showing 7 transmembrane domains surrounding the vitamin A derived chromophore and the highly conserved lysine at position 296 (black) required for chromophore binding. The glutamic acid counterion (orange) is conserved among vertebrate c-opsins and important for chromophore binding. Right, cartoon diagram of bovine rhodopsin sequence with select functional sites color coded by functional category, matching the numbered sites in **B**. Cartoon G protein subunits are shown bound to rhodopsin. The specific G protein alpha subunit can vary (letters in parentheses) depending on opsin sequence, with functional implications for type of signaling cascade activated. **B** Left, maximum-likelihood phylogeny of *N. vectensis* opsins, with IQtree2 support. Right, select *N. vectensis* opsin amino acids are aligned with bovine rhodopsin (c-opsin), *D. melanogaster* r-opsin, and *T. cystophora* cnidopsin. Canonically conserved functional residues and positions follow bovine rhodopsin numbering and correspond to **A**. From left to right, the first box contains structural features minimally required for function in all opsins. 7TM indicates protein sequence has seven transmembrane domains. Black is the conserved Lys(K)296; yellow shows conserved cysteine residues that form a stabilizing disulfide bridge. Second box lists known counterions from bilaterian and box jelly opsins (orange). Third box shows a conserved spectral tuning site across opsin clades (blue). A second spectral tuning site is also found at counterion site Asp(D)83. Substitutions known to cause a blue shift in both sites are shaded blue. Fourth box contains conserved sites known to be important for G-protein signaling (light gray). NvOpsin residues that are conserved at the known functional sites are shaded in dark gray. Rightmost, NvOpsins are labeled by subclade within each major anthozoan opsin clade

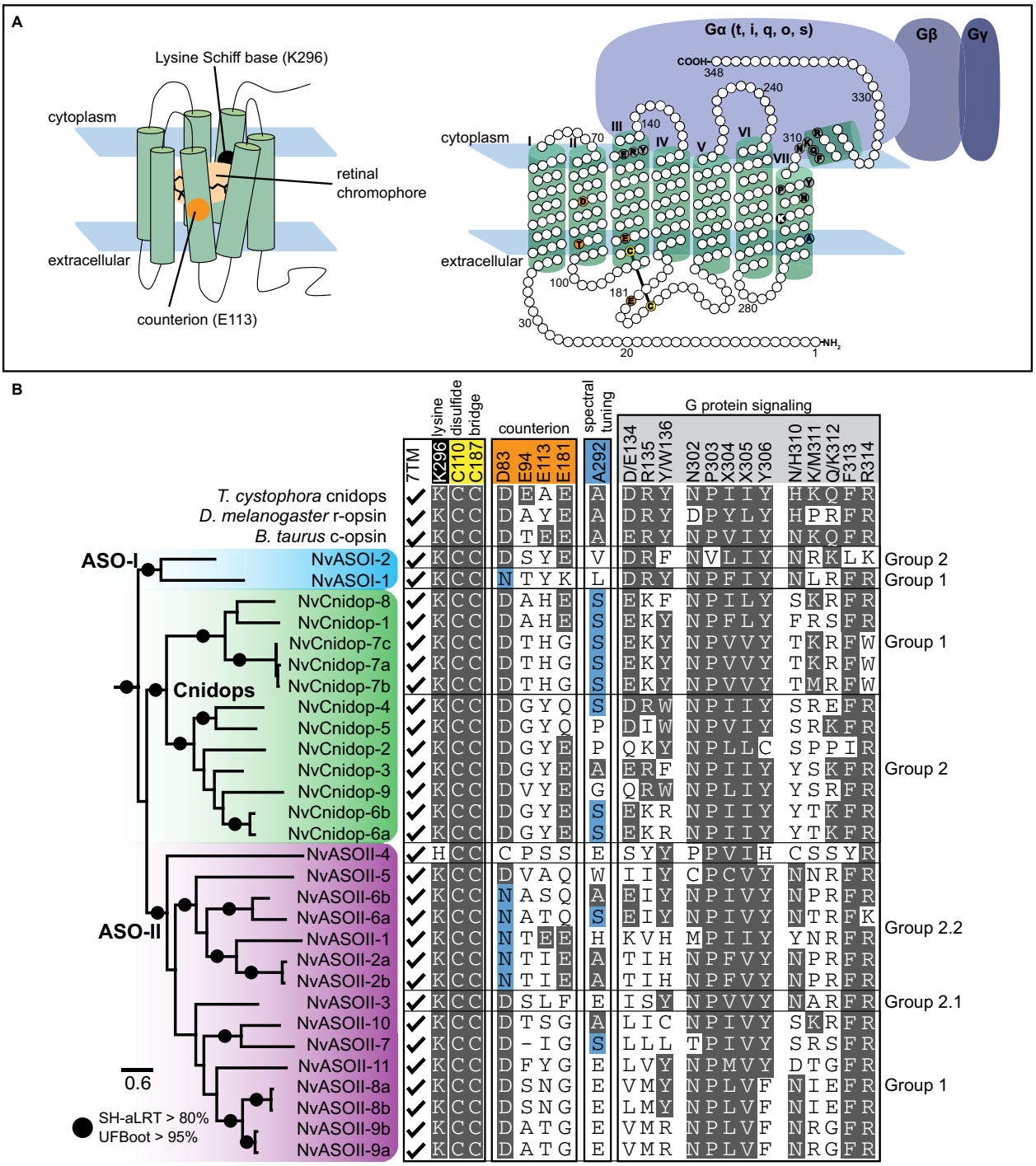

Fig. 4 (See legend on previous page.)

NvASOI-2, which is expressed in ectoderm in larval stages, is found in endomesodermal tissue surrounding the pharynx at the time of tentacle bud formation and into the primary polyp stage (Fig. 6I–L′). *NvCnidop-6b* is also expressed in clusters of cells within the pharyngeal endomesoderm and extends aborally into the mesenteries (Fig. 6M–N′). Distinct tissue- and stage-specific expression of opsins from distinct subclades suggests different functions throughout development.

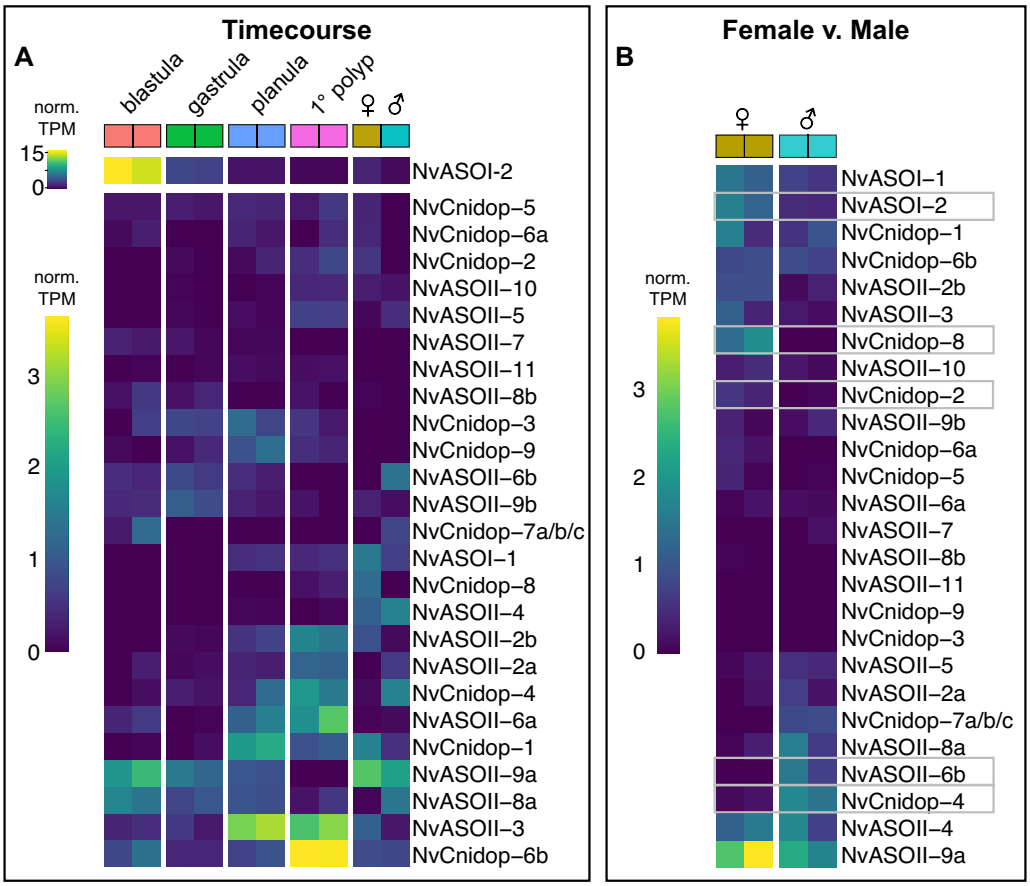

Fig. 5 NvOpsins are expressed dynamically throughout development and between sexes. **A** Expression levels from time course analysis are plotted for opsins across all stages and both sexes. Each box is from a single replicate library. For visual clarity, highly expressing *NvASOI-2* is shown using a separate scale. Heatmap illustrates variable opsin expression across development in all opsin transcripts. *NvCnidop7a/b/c* loci are highly similar in coding sequence, such that transcripts are not distinguished between the three sequences. *NvASOII-1* had no high-identity transcript in our transcriptome or in NvERTx. **B** A separate analysis comparing male and female adult opsin expression is visualized in a heatmap. The gray boxes indicate opsins that are qualitatively distinct between the sexes

Discussion

We have identified 29 opsins in the genome of *N. vectensis,* the most of any anthozoan so far investigated [6, 38]. Maintenance of complete coding sequences, evidence of distinct genomic loci for similar paralogs, and expression patterns lead us to conclude with high confidence these encode bona fide opsin proteins. Our addition of recent anthozoan phylogenetic data has revealed new patterns of duplication and evolution, both in understudied cnidarian lineages and poorly characterized opsin clades. *N. vectensis* shows large lineage-specific expansions in both cnidopsins and ASO-II opsins. While *N. vectensis* opsins are conserved with related bilaterian opsins in structural domains, significant divergence at canonically conserved putative sites suggests high functional diversity. The investigation of anthozoan opsin clades not found in medusozoans potentially leads to novel opsin functional diversity. This is also reflected in mRNA expression, where multiple opsins are expressed dynamically throughout development, differentially expressed between adult sexes, and spatially restricted to specific cell types and tissues.

A prime example of this observed diversity is opsin NvASOII-4, which has a His296Lys amino acid substitution. This lysine is almost universally conserved across all animal opsins and it is essential for the covalent bond with a vitamin-A derived chromophore, necessary for absorbing photons. The presence of His at this site could suggest a novel chromophore for NvASOII-4. Alternatively, this opsin may not have a light-related function. It is becoming increasingly clear animal opsins are not exclusive for light sensing [68]. It has been shown in *D. melanogaster* that opsins are involved in thermal sensing [69], proprioception [70], taste [71], and more. Though the biochemical mechanism and signaling dynamics have yet to be understood, ion channels such as TRP channels

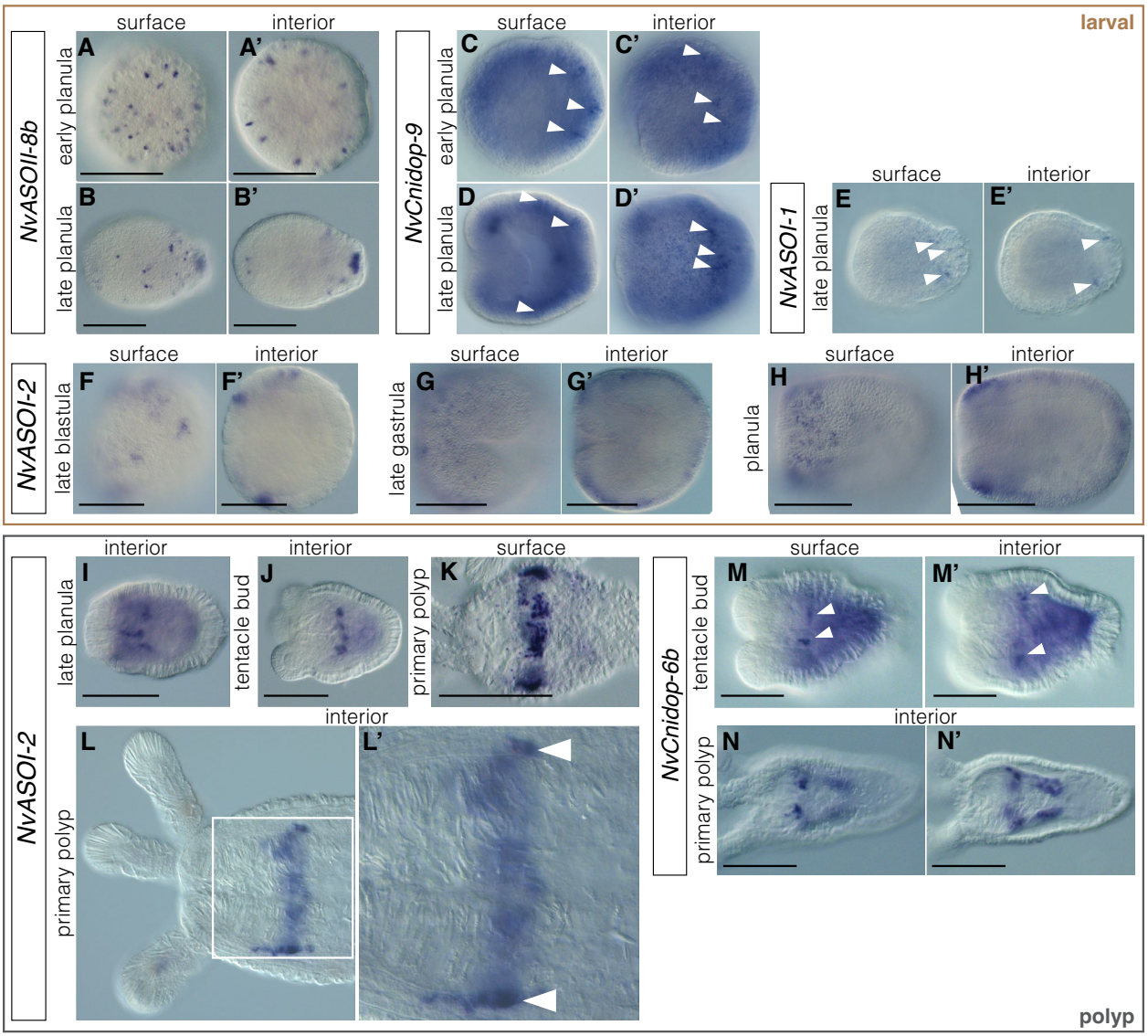

Fig. 6 NvOpsin spatial expression patterns suggest a diversity of functions throughout development. **A**–**B**′ *NvASOII-8b* is only expressed at the swimming planula stage. Expression is ectodermal, in individual cells scattered throughout the ectoderm in early planula. In later stage planula (**B**′) the expression is also concentrated at the aboral end in the sensory apical organ. **C**–**D**′ *NvCnidop-8* is expressed in a subset of aboral ectodermal cells at planula stage (arrowheads). **E**–**E**′ *NvASOI-1* is also expressed in a subset of cells in the aboral ectoderm. **F**–**H**′ *NvASOI-2* is expressed in the ectoderm at early stages. Starting at blastula stage (**F**) through gastrula (**G**), expression is patchy in clusters of cells which tend to be concentrated more orally later in development (**H**). By late planula/tentacle bud stage expression has concentrated in an endomesodermal ring at the aboral end of the pharynx, where it can be seen into the primary polyp stage (**I**–**L**′). **M**–**N**′ *NvCnidop-6b* begins to express at late planula/tentacle bud stage in the pharyngeal endomesoderm, and by polyp stage is expressed in specific cells throughout the mesenteries. Scale bar 100 um; in all images oral is left, aboral is right

involved in visual transduction play a role in these sensory modalities as well. Many of the canonical opsin amino acid motifs are not conserved in NvOpsins. These proteins may function as photoreceptors and interact with their binding partners in different ways. It is also possible these opsins have non-canonical functions. In an eyeless animal with an abundance of opsin proteins, *N. vectensis* may be a prime candidate to study the evolution of opsin novelty and structure–function relationships in this protein family.

Our opsin tree was able to capture more anthozoan-specific opsin evolutionary history than previous work due to increased sampling with recently available sequencing. Our data show expansions in both ASO-II

and cnidopsins are concentrated in Hexacorallia and duplications have occurred at both ancient and recent timescales (Fig. 3). In contrast, genomic data and analysis of the gene tree suggest loss of opsin genes over time. ASO-I and Medusozoan cnidopsins are intronless suggesting that they are a result of duplication by retrotransposition. The ancestral gene, has either been subsequently lost or is highly diverged in another clade. Medusozoan ASO-II loss is inferred if ASO-II is indeed sister to C-opsins, as this suggests this clade was present before the split of Cnidaria and Bilateria. Xenopsins are only found in Spiralia but are in a clade with Cnidopsins, suggesting a loss of related opsins in other bilaterian lineages. Our data suggest that while recent opsin duplication appears to be the norm, loss of multiple orthologs from an early repertoire of opsins could have occurred while early animal lineages diverged.

A limitation to this and other trees is that mRNA sequencing tends to miss opsins expressed at a restricted stage, within a restricted tissue, or those that are too similar to be distinguished by assembly algorithms. Our analysis of two chromosome-level genomes was able to add and remove erroneously annotated or previously missed *N. vectensis* opsins. More high-quality long-read genomic sequencing would greatly increase confidence in opsin numbers for many species and help define patterns of opsin duplication and loss in Cnidaria. Genomic sequence would also help with cross species contamination from wild caught transcriptomic references. The addition of our anthozoan sequences placed five anthozoan opsin sequences in the R-opsin clade, and two in the medusozoan-specific cnidopsins, which are likely contaminations. In addition, variable phylogenetic support for deep opsin relationships between Bilateria and Cnidaria requires methods such as intron analysis to show evidence of shared ancestry. All medusozoan opsins are cnidopsins and nearly all are intronless [40]. The hydrozoan *Clytia hemisphaerica* is an exception, with two intron-containing opsins. However, neither of these intron-containing opsins share intron/exon boundaries with any other cnidarian opsin or xenopsin [31, 72]. We show for the first time using genomic data from *N. vectensis* that the cnidopsin locus shares a common intron/exon structure with xenopsins in bilaterians. The lack of introns in medusozoan and some *N. vectensis* cnidopsins suggests duplication by retrotransposition may have occurred early in the expansion of cnidopsins or in parallel after Anthozoa and Medusoza split. Furthermore, this suggests a medusozoan-specific loss of the original intron-containing genes, while some Anthozoa have maintained this ancestral gene structure shared by xenopsins.

Expression of NvOpsins is dynamic, found across life stages, tissue types, and sexes. *NvASOII-8b,* found only at the swimming stage and in the sensory apical organ, may function in larval light detection and control of swimming. Others, such as *NvCnidop-6b,* are found in the developing mesenteries and upregulated in adult males, relative to females, suggesting a role in reproduction. This would be expected, as spawning is light-induced in *N. vectensis* [48]. *NvASOI-2* is highly expressed early and persists through adult stages, its expression domain changes over time, and it is sexually dimorphic in adults. Primordial germ cells develop in the mesenteries near the pharynx, at a similar location to *NvASOI-2* expression [73]. Together this suggests important functional roles throughout development and a possibly distinct adult role in reproduction (Fig. 7). Functional studies in other anthozoans have suggested many light-dependent behaviors may be mediated by specific opsins though none have directly shown this in any anthozoan. In corals, some opsins are expressed at the aboral pole of the swimming larva—similar to *NvASOII-8b*—and were proposed to be involved in light detection while swimming [34]. In coral, some opsins have been proposed to be important for broadcast spawning or algal symbiosis though neither has been shown directly [43, 74]. Unlike corals, where husbandry and functional genetics can be challenging, *N. vectensis* is amenable to laboratory study and genetic manipulation. By molecularly characterizing the opsins in this animal, it will now be possible to perform experiments to better characterize light-mediated behaviors and link them with cell physiology and specific opsin functions.

Although the divergence from canonically conserved sites suggests new and unidentified functions, we hypothesize that *N. vectensis* has multiple functional opsin photopigments because the animal exhibits several light-mediated behaviors [48, 49]. It is possible that shallow-water species like *N. vectensis* and other anthozoans like reef corals are exposed to more variable light environments, using light sensory information for spawning, substrate settling, defense, and predation. It remains unknown whether opsin gene family expansion in the Hexacorallia is adaptive for these animals, however, our findings suggest a diversity of opsin-expressing cell types involved in multiple light-mediated functions in *N. vectensis*.

Methods

Animal care

N. vectensis adults were kept in 1/3 concentration artificial sea water in glass dishes on a 12 h light/dark cycle at 18 °C. Animals were fed freshly hatched *Artemia* five times per week. Spawning was induced by placing

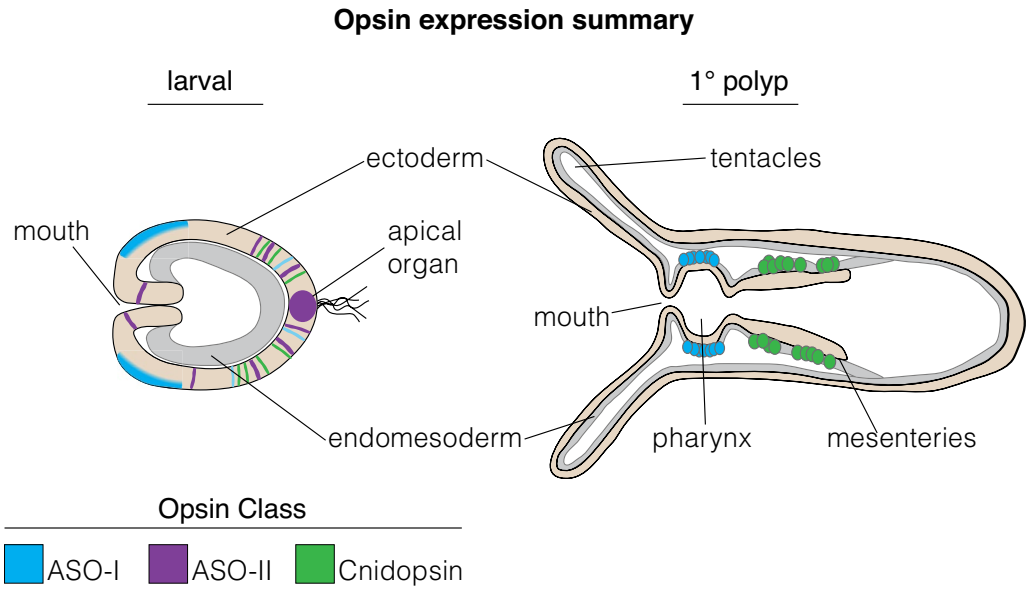

Fig. 7 Summary of opsin expression patterns in development. Two summary stages are shown for the primarily early larval expression, and the primarily later primary polyp expression patterns of *N. vectensis* opsins. Opsin expression is color-coded by clade, ectoderm is beige, endomesoderm is gray. Opsins are expressed throughout developmental stages and tissue types

animals at 24 °C overnight with light. Eggs and sperm were collected separately within 2 h after spawning and fertilized in a new Petri dish and were left to develop at room temperature.

Opsin identification and phylogenetics

Initial *N. vectensis* opsin sequences were collected from three published datasets [5, 6, 47]. All potential *N. vectensis* sequences were aligned in Geneious Prime v.11.0.12, and identical matches and fragments were discarded. BLAST was used with the final set of opsins as bait to search against the *Nematostella vectensis* Embryogenesis and Regeneration Transcriptomics database (NvERTx) [67], our own reference transcriptome (generated for this study), and two publicly available chromosome-level genomes: v2 genome hosted by the Stowers Institute [50, 75], and the genome from the Darwin Tree of Life Programme at the Wellcome Sanger Institute (ENA submission accession: ERA9667479) [51]. Hits were added inclusively and combined with a modified alignment including non-redundant sequences from Vöcking et al. 2017, Picciani et al. 2018, and Gornik et al. 2021 opsin phylogenies [5, 6, 29]. These included outgroups from closely related GPCRs, including melatonin receptors and *Trichoplax* opsin-like proteins. We then added opsin sequences from new anthozoan transcriptome and genome data. We added opsins from 12 species with new transcriptomes, and 3 species from new genomic data (Additional file 3: Table S4). This was done using *N. vectensis* and *Acropora digitifera* sequences from ASOI, ASOII, and Cnidops groups as bait for BLAST searches. The top 100 BLAST hits were initially conservatively kept. All sequences were aligned in Geneious using MAFFT v7.450 with default settings [76]. FastTree was used in Geneious to check the tree, removing unannotated outgroups from the multiple cnidarian and *N. vectensis* BLAST searches. Trimmed (using TrimAl) and untrimmed alignments were generated, but untrimmed sequences generated the best supported phylogeny and are reported here.

Final maximum-likelihood gene trees were constructed using IQtree2 with the following command: iqtree2 –s <alignment.phy> -st AA -nt AUTO -v -m TEST -bb 1000 -alrt 1000 [77] on the Minnesota Supercomputing Institute's (MSI) High Performance Computing system at the University of Minnesota. The LG + G4 model of protein evolution was auto-calculated from this command, and 1000 ultrafast bootstrap (UFboot) replicates and SH-aLRT tests were used for evidence of support. Clades were considered supported only with > 80% SH-aLRT / > 95% UFboot support. The tree was rooted based on known outgroups from previous phylogenies.

Predictions of opsin gene counts in the early branching lineages of Anthozoa were assessed three ways. We assessed manually by parsimony considering the topology of the gene tree and the species relationships. Subclades with species within shared lineages were counted for the minimum number of duplications required for each anthozoan lineage and mapped to each node. Mesquite v3.81 was used with the species tree shown

in Additional File 4: Figs. S1, S2, S3, and opsin counts from Fig. 2B. Default analyses were performed assessing the data as both continuous and meristic count data. Notung analysis was performed according to default settings, with the same species tree and the entire gene tree (Fig. 1), which was automatically pruned to include only species in the gene tree [78].

Genomic and opsin sequence analysis

To identify distinct opsin paralogs, we used BLAST with transcripts identified as NvOpsins above as bait to identify genomic loci, resulting in the identification of 29 opsin genes. We manually identified intron/exon boundaries at all cnidopsins with introns by aligning coding sequences to the genome and translated these to identify their location and phase on the protein. We then aligned the cnidopsin amino acid sequences with a xenopsin and bovine c-opsin in Geneious and mapped exon boundaries manually onto the alignment. Opsin genomic coordinates for both new Nvec genomes [50, 51], and previous Nvecv1 scaffold coordinates [79] are found in Additional file 3: Table S1. We used InterPro scan in Geneious to predict the seven transmembrane domains. Using bovine rhodopsin as a reference, we aligned our 29 opsins, *Drosophila* r-opsin and *T. cystophora* cnidopsin sequences to identify residues at canonically conserved positions.

A few tandem loci next to *NvASOII-8a, NvASOII-8b, NvASOII-9a, NvASOII9b, NvASOII-7,* and *NvASOII-4* genes were found in the v2 genome [50] whose nucleotide sequences were nearly identical (>97% similarity). We noticed adjacent non-coding intra- and intergenic regions were also near-identical, suggesting these may be haplotype differences and not real paralogs. One of these adjacent pairs also was split by a section of Ns in the genome (Additional file 4: Fig. S6), an indicator of difficulty with assembly in this region. We compared these trouble spots with the Wellcome-Sanger genome [51] and for each, found only one of the two near-identical pairs, further evidence that these copies are a result of assembly error, while other high-similarity but distinct loci were left in our final opsin count.

RNA-sequencing and analysis

Libraries were generated from two independent spawns each at blastula, gastrula, planula, primary polyp stages, two adult males and two adult females. Adult animals had been kept in normal laboratory conditions at 18°C with a 12 h light/12 h dark photoperiod and at least one week since they were last spawned. Live animals were placed directly into Trizol reagent and tissue was homogenized in a microcentrifuge tube with a pestle before performing RNA extraction. Total RNA was extracted from Trizol (Invitrogen), gDNA removed by gDNA cleanup kit (Qiagen), RNA was precipitated using isopropanol and then cleaned up using ethanol precipitation. Libraries were prepared using Kapa Stranded RNA Hyperprep (Illumina), quality control and size selection were performed by the Bauer Core at Harvard University and paired end, 150-bp reads were sequenced on an Ilumina NovaSeq.

Raw sequencing libraries were processed for erroneous k-mers and unfixable reads using rCorrector [80]. Adapter sequences were trimmed using Trim Galore (Babraham Bioinformatics, UK), and rRNA reads were mapped to the *N. vectensis* mitochondrial genome and removed using Bowtie2 [81]. All libraries were combined to generate a single de novo transcriptome assembly using Trinity v2.12.0 [82]. Transcriptome quality was assessed by checking alignment statistics and quantifying BUSCO completeness [83]. We used EvidentialGene to combine gene and transcript evidence from multiple sources and reduce the number of genes in our transcriptome [84]. We used our own transcripts, and three other assemblies [50, 67, 85], and the coding sequences from v2 genome [50], as input. We used the EvidentialGene output for our final transcriptome. mRNA quantification was done using Kallisto for each of our 12 libraries using our de novo assembled transcriptome [86]. Stages were normalized and compared using the time course analysis in Sleuth run in R, and adult libraries only were additionally pairwise compared using Sleuth [87]. The data for heatmaps were extracted from the Sleuth object using normalized transcripts per million (TPM) in either the timecourse analysis or male v. female comparison. Heatmaps were made with the heatmap command in gplots package [88] using R v4.2.3 run in RStudio v2022.12.0+353 [89].

RNA probe synthesis

Tissue for probe synthesis was pooled from the same developmental stages as RNA-seq and homogenized in TRIzol (Invitrogen). Total RNA was then extracted with chloroform and cleaned up using a Qiagen RNeasy mini kit. cDNA was made from total RNA using the iScript advanced kit (Bio-Rad). Primers were designed in Geneious (modified Primer3 v2.3.7) (Additional file 3: Table S5). PCR was used to amplify opsin sequences from cDNA and size checked by gel electrophoresis. Properly sized products were ligated to the pGEM T-Easy plasmid (Promega), transformed into chemically competent DH5-alpha *E. coli* cells (New England Biolabs). Plasmid DNA was isolated using the Qiagen miniprep kit and sequenced for confirmation and orientation of insertion into plasmid. Restriction digests linearized the plasmid, were cleaned up using the Qiagen PCR clean up kit, and antisense probes were generated using either Ambion T7

or SP6 megascript RNA polymerases with digoxigenin-labeled (DIG) nucleotides. Probes were cleaned up with Qiagen RNeasy micro kit and eluted in 14 µL. Proper transcription of the probe was checked on a gel. Clean probe was stored at a 50/50 concentration with formamide at − 20° C. When ready to use, probe was diluted 1:1000 in hybridization buffer.

In situ hybridization

In situ hybridization was performed according to Wolenski et al. 2013, with slight modification [90]. Before fixation, embryos at blastula and gastrula stages were dejellied by rocking in 40 mg/mL L-cysteine in 1/3×artificial sea water (Instant Ocean) for 10 min. Planula and polyp stage animals were immobilized by gently adding 6.5% magnesium chloride in 1/3 seawater. Embryos were fixed in glass vials for 90 s in ice cold 0.25% glutaraldehyde/4% paraformaldehyde (PFA) in phosphate buffered saline (PBS). This was removed, 4% PFA was added, and embryos were fixed on a rocker for 1 h at 4 °C. PFA was washed out with PBS with 0.1% Tween-20 (PTw) and embryos were stored in methanol at − 20 °C. To begin in situ hybridization, embryos were stepped out of methanol and into PTw, then treated with 0.01 mg/mL proteinase K, followed by two glycine washes (2 mg/mL), and two 1% triethanolamine washes in PTw. Embryos were then washed first at 3 µL/mL then 6 µL/mL of acetic anhydride/1% triethanolamine in PTw. These were washed out in PTw then fixed in 4% PFA for one hour on a rocker at room temperature. PFA was washed out with PTw and embryos were pre-washed in hybridization buffer (hyb) (50% formamide, 5×SSC, 50 µg/mL heparin, 0.1% Tween-20, 1% SDS, 100 µg/mL herring sperm DNA, 370 µL 1 M citric acid) for 10 min at room temperature. This was replaced with a second hyb wash, and sealed vials were kept at 63 °C for at least 12 h.

Following pre-hybridization, DIG-labeled RNA probes were diluted 1:1000 in hybridization buffer and preheated to 90 °C. Animals were transferred to mesh baskets in 24-well plates at 63 °C. All subsequent washes were performed by quickly transferring baskets to new wells with fresh solution. Probe was added to wells and baskets were transferred quickly into probe and left at 63 °C for at least 48 h. Probe was removed and stored for reuse, embryos were washed twice with hybridization buffer, and stepped 25%, 50%, and 75% SSC concentrations in hybridization buffer up to 100% 2×SSC solution, at 63°C. Next embryos were washed in 0.02×SSC, removed from heat and stepped into 25%, 50%, and 75% PTw in 0.02×SSC, up to 100% PTw. Embryos were then washed in 1×Roche blocking reagent for 1–2 h at room temperature. Block was replaced with alkaline phosphatase-labeled (AP) anti-DIG Fab fragment antibody (Roche) in Roche blocking reagent at a concentration of 1:5000, rocking overnight at 4 °C. The next day, antibody was removed, and embryos were washed 10 times for 15 min each in phosphate buffered saline with 0.1% Triton-X, then AP reaction buffer (100 mM NaCl, 50 mM $MgCl_2$, 100 mM Tris, pH 9.5, 0.5% Tween-20) minus $MgCl_2$, followed by two washes of AP buffer with $MgCl_2$. Finally, embryos were placed in BCIP-NBT (Promega) with AP buffer to react. Solution was checked for color change and replaced every half hour for the first several hours then once a day up to several weeks at 4 °C, depending on speed of the reaction. When the chromogenic reaction was finished, embryos were washed in order with: PTw, water, ethanol, water, and PTw, then postfixed with PFA and cleared with 90% glycerol. Embryos were mounted on slides and imaged with DIC optics on a Zeiss Axioimager microscope.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13227-023-00218-8.

Additional file 1. Opsin alignment in Phylip format.

Additional file 2. Opsin gene tree file.

Additional file 3: Table S1. Opsin counts and recently published Anthozoan accessions. **Table S2.** Notung analysis summary of duplications and losses. **Table S3.** Genomic and cross-reference evidence for *N. vectensis* opsins. **Table S4.** Opsin counts from each dataset from NvERTx. **Table S5.** Primers used for *in situ* hybridization.

Additional file 4: Fig. S1. Mesquite ancestral state reconstruction. **Fig. S2.** Notung species tree-gene tree reconciliation summary. **Fig. S3.** Notung gene tree raw duplicates and losses. **Fig. S4.** RNA-seq library analysis. Fig. S5. Opsin expression heatmaps from NvERTx datasets. Fig. S6. Sequence similarity of similar opsin genomic loci.

Acknowledgements

The authors acknowledge the Minnesota Supercomputing Institute (MSI) at the University of Minnesota for providing resources that contributed to the research results reported within this paper. URL: http://www.msi.umn.edu. We'd like to thank Cassandra Extavour and Amaneet Lochab for help with microscope use.

Author contributions

KJM, LSB, and KMK conceived of and designed the study. Funding acquisition by KMK and MQM. KJM, AL, LSB, and CD generated in situ probes and performed experiments. KJM and AL performed phylogenetics and sequence analysis. KJM wrote the manuscript. All authors read and provided feedback on the manuscript.

Funding

This work was supported by funding from the NIH Director 1DP5OD023111-01 to KMK and the John Harvard Distinguished Science Fellowship to KMK and the National Aeronautics and Space Administration NNX14AG70G to MQM.

Availability of data and materials

All RNA-seq raw reads generated during the current study are available in NCBI under BioProject ID PRJNA962884. Full length opsin sequences, transcriptome assembly, alignment and tree files are publicly available on DRYAD: https://datadryad.org/stash/share/qyUWZYiB-QKUMyLYuw0uMvLeOIkA-mAuNPU5_8NeeDQ. All other data are found in the supplement included with this publication.

Declarations

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Department of Ecology, Evolution and Behavior, University of Minnesota, St. Paul, MN 55108, USA. [2]Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY 14853, USA. [3]John Harvard Distinguished Science Fellowship Program, Harvard University, Cambridge, MA 02138, , USA. [4]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA. [5]Whitney Lab for Marine Bioscience, University of Florida, St. Augustine, FL 32080, USA. [6]Department of Molecular Biosciences, University of Texas at Austin, Austin, TX 78712, USA.

Received: 17 May 2023 Accepted: 11 September 2023
Published online: 21 September 2023

References
1. Feuda R, Hamilton SC, McInerney JO, Pisani D. Metazoan opsin evolution reveals a simple route to animal vision. Proc Natl Acad Sci. 2012;109:18868–72. https://doi.org/10.1073/pnas.1204609109.
2. Terakita A. The opsins. Genome Biol. 2005;6:1–9.
3. Kozmik Z, Ruzickova J, Jonasova K, Matsumoto Y, Vopalensky P, Kozmikova I, et al. Assembly of the cnidarian camera-type eye from vertebrate-like components. Proc Natl Acad Sci. 2008;105:8989–93. https://doi.org/10.1073/pnas.0800388105.
4. Ramirez MD, Pairett AN, Pankey MS, Serb JM, Speiser DI, Swafford AJ, et al. The last common ancestor of most bilaterian animals possessed at least nine opsins. Genome Biol Evol. 2016;8:3640–52.
5. Picciani N, Kerlin JR, Sierra N, Swafford AJM, Ramirez MD, Roberts NG, et al. Prolific origination of eyes in cnidaria with co-option of non-visual opsins. Curr Biol. 2018;28:2413-2419.e4. https://doi.org/10.1016/j.cub.2018.05.055.
6. Gornik SG, Bergheim BG, Morel B, Stamatakis A, Foulkes NS, Guse A. Photoreceptor diversification accompanies the evolution of *Anthozoa*. Mol Biol Evol. 2021;38:1744–60.
7. Shichida Y, Matsuyama T. Evolution of opsins and phototransduction. Philos Trans R Soc B Biol Sci. 2009;364:2881–95.
8. Porter ML, Blasic JR, Bok MJ, Cameron EG, Pringle T, Cronin TW, et al. Shedding new light on opsin evolution. Proc R Soc B Biol Sci. 2011;279:3–14.
9. Oakley TH, Speiser DI. How complexity originates: the evolution of animal eyes. Annu Rev Ecol Evol Syst. 2015;46:237–60.
10. Arshavsky VY, Lamb TD, Pugh EN. G proteins and phototransduction. Annu Rev Physiol. 2002;64:153–87.
11. Hardie RC, Raghu P. Visual transduction in *Drosophila*. Nature. 2001;413(6852):186–93.
12. Katz B, Minke B. Drosophila photoreceptors and signaling mechanisms. Front Cell Neurosci. 2009;3:2.
13. Koyanagi M, Kubokawa K, Tsukamoto H, Shichida Y, Terakita A. Cephalochordate melanopsin: evolutionary linkage between invertebrate visual cells and vertebrate photosensitive retinal ganglion cells. Curr Biol. 2005;15:1065–9.
14. Velarde RA, Sauer CD, Walden KKO, Fahrbach SE, Robertson HM. Pteropsin: a vertebrate-like non-visual opsin expressed in the honey bee brain. Insect Biochem Mol Biol. 2005;35:1367–77.
15. Döring CC, Kumar S, Tumu SC, Kourtesis I, Hausen H. The visual pigment xenopsin is widespread in protostome eyes and impacts the view on eye evolution. Elife. 2020;9:1–23.
16. Roberts NS, Hagen JFD, Johnston RJ. The diversity of invertebrate visual opsins spanning *Protostomia*, *Deuterostomia*, and *Cnidaria*. Dev Biol Prep. 2022;492:187–99. https://doi.org/10.1016/j.ydbio.2022.10.011.
17. Fernald RD. Casting a genetic light on the evolution of eyes. Science. 2006;313:1914–8.
18. Eakin RM. Evolution of photoreceptors. Cold Spring Harb Symp Quant Biol. 1965;30:363–70. https://doi.org/10.1101/SQB.1965.030.01.036.
19. Eakin RM. Evolution of photoreceptors. Evol Biol. 1968. https://doi.org/10.1101/SQB.1965.030.01.036.
20. Fain GL, Hardie R, Laughlin SB. Phototransduction and the evolution of photoreceptors. Curr Biol. 2010;20:R114–24. https://doi.org/10.1016/j.cub.2009.12.006.
21. Hattar S, Liao H-W, Takao M, Berson DM, Yau K-W. Melanopsin-containing retinal ganglion cells: architecture, projections, and intrinsic photosensitivity. Science. 2002;295:1065–70. https://doi.org/10.1126/science.1069609.
22. Arendt D, Tessmar-Raible K, Snyman H, Dorresteijn AW, Wittbrodf J. Ciliary photoreceptors with a vertebrate-type opsin in an invertebrate brain. Science. 2004;306:869–71.
23. Lesser MP, Carleton KL, Bottger SA, Barry TM, Walker CW. Sea urchin tube feet are photosensory organs that express a rhabdomeric-like opsin and PAX6. Proc R Soc B Biol Sci. 2011;278:3371–9. https://doi.org/10.1098/rspb.2011.0336.
24. Shen D, Jiang M, Hao W, Li T, Salazar M, Fong HKW. A human opsin-related gene that encodes a retinaldehyde-binding protein. Biochemistry. 1994;33:13117–25.
25. Koyanagi M, Terakita A, Kubokawa K, Shichida Y. Amphioxus homologs of Go-coupled rhodopsin and peropsin having 11-cis-and all-trans-retinals as their chromophores. FEBS Lett. 2002;531:525–8.
26. Tarttelin EE, Bellingham J, Hankins MW, Foster RG, Lucas RJ. Neuropsin (Opn5): a novel opsin identified in mammalian neural tissue. FEBS Lett. 2003;554:410–6.
27. Yamashita T, Ohuchi H, Tomonari S, Ikeda K, Sakai K, Shichida Y. Opn5 is a UV-sensitive bistable pigment that couples with Gi subtype of G protein. Proc Natl Acad Sci USA. 2010;107:22084–9.
28. Gühmann M, Jia H, Randel N, Verasztó C, Bezares-Calderón LA, Michiels NK, et al. Spectral tuning of phototaxis by a Go-opsin in the rhabdomeric eyes of platynereis. Curr Biol. 2015;25:2265–71.
29. Vöcking O, Kourtesis I, Tumu SC, Hausen H. Co-expression of xenopsin and rhabdomeric opsin in photoreceptors bearing microvilli and cilia. Elife. 2017;6:1–26.
30. Feuda R, Rota-Stabelli O, Oakley TH, Pisani D. The comb jelly opsins and the origins of animal phototransduction. Genome Biol Evol. 2014;6:1964–71.
31. Quiroga Artigas G, Lapébie P, Leclère L, Takeda N, Deguchi R, Jékely G, et al. A gonad-expressed opsin mediates light-induced spawning in the jellyfish Clytia. Elife. 2018;7:1–22.
32. Kozmik Z. The role of Pax genes in eye evolution. Brain Res Bull. 2008;75:335–9.
33. Leach WB, Reitzel AM. Decoupling behavioral and transcriptional responses to color in an eyeless cnidarian. BMC Genom. 2020;21:1–15.
34. Mason B, Schmale M, Gibbs P, Miller MW, Wang Q, Levay K, et al. Evidence for multiple phototransduction pathways in a reef-building coral. PLoS ONE. 2012;7:1–9.
35. Hendricks WD, Byrum CA, Meyer-Bernstein EL. Characterization of Circadian Behavior in the Starlet Sea Anemone *Nematostella vectensis*. PLoS ONE. 2012;7:1–10.
36. Foo SA, Liddell L, Grossman A, Caldeira K. Photo-movement in the sea anemone Aiptasia influenced by light quality and symbiotic association. Coral Reefs. 2020;39:47–54. https://doi.org/10.1007/s00338-019-01866-w.
37. Plachetzki DC, Fong CR, Oakley TH. Cnidocyte discharge is regulated by light and opsin-mediated phototransduction. BMC Biol. 2012;10:17.
38. Macias-Munõz A, Murad R, Mortazavi A. Molecular evolution and expression of opsin genes in *Hydra vulgaris*. BMC Genomics BMC Genomics. 2019;20:1–19.
39. Plachetzki DC, Fong CR, Oakley TH. The evolution of phototransduction from an ancestral cyclic nucleotide gated pathway. Proc R Soc B Biol Sci. 2010;277:1963–9.
40. Liegertová M, Pergner J, Kozmiková I, Fabian P, Pombinho AR, Strnad H, et al. Cubozoan genome illuminates functional diversification of opsins and photoreceptor evolution. Sci Rep. 2015;5:1–19. https://doi.org/10.1038/srep11885.
41. Gerrard E, Mutt E, Nagata T, Koyanagi M, Flock T, Lesca E, et al. Convergent evolution of tertiary structure in rhodopsin visual proteins from vertebrates and box jellyfish. Proc Natl Acad Sci. 2018;115:6201–6. https://doi.org/10.1073/pnas.1721333115.
42. Koyanagi M, Takano K, Tsukamoto H, Ohtsu K, Tokunaga F, Terakita A. Jellyfish vision starts with cAMP signaling mediated by opsin-Gs cascade. Proc Natl Acad Sci USA. 2008;105:15576–80.

43. Kaniewska P, Alon S, Karako-Lampert S, Hoegh-Guldberg O, Levy O. Signaling cascades and the importance of moonlight in coral broadcast mass spawning. Elife. 2015;4:1–14.
44. Levy O, Dubinsky Z, Achituv Y. Photobehavior of stony corals: Responses to light spectra and intensity. J Exp Biol. 2003;206:4041–9.
45. Picciani N, Kerlin JR, Jindrich K, Hensley NM, Gold DA, Oakley TH. Light modulated cnidocyte discharge predates the origins of eyes in Cnidaria. Ecol Evol. 2021;11:3933–40.
46. Anctil M, Hayward DC, Miller DJ, Ball EE. Sequence and expression of four coral G protein-coupled receptors distinct from all classifiable members of the rhodopsin family. Gene. 2007;392:14–21.
47. Suga H, Schmid V, Gehring WJ. Evolution and functional diversity of jellyfish opsins. Curr Biol. 2008;18:51–5.
48. Fritzenwanker JH, Technau U. Induction of gametogenesis in the basal cnidarian *Nematostella vectensis* (*Anthozoa*). Dev Genes Evol. 2002;212:99–103.
49. Tarrant AM, Helm RR, Levy O, Rivera HE. Environmental entrainment demonstrates natural circadian rhythmicity in the cnidarian *Nematostella vectensis*. J Exp Biol. 2019;222(Pt 21):205393.
50. Zimmermann B, Robb SMC, Genikhovich G, Fropf WJ, Weilguny L, He S, et al. Sea anemone genomes reveal ancestral metazoan chromosomal macrosynteny. BioRxiv. 2020. https://doi.org/10.1101/2020.10.30.359448.
51. Fletcher C, Pereira da Conceicoa L. The genome sequence of the starlet sea anemone *Nematostella vectensis*. Wellcome Open Res. 2023;8(79):79.
52. Karnik SS, Sakmar TP, Chen HB, Khorana HG. Cysteine residues 110 and 187 are essential for the formation of correct structure in bovine rhodopsin. Proc Natl Acad Sci. 1988;85:8459–63. https://doi.org/10.1073/pnas.85.22.8459.
53. Franke RR, Sakmar TP, Graham RM, Khorana HG. Structure and function in rhodopsin. Studies of the interaction between the rhodopsin cytoplasmic domain and transducin. J Biol Chem. 1992;267:14767–74.
54. Fritze O, Filipek S, Kuksa V, Palczewski K, Hofmann KP, Ernst OP. Role of the conserved NPxxY(x) 5,6 F motif in the rhodopsin ground state and during activation. Proc Natl Acad Sci. 2003;100:2290–5. https://doi.org/10.1073/pnas.0435715100.
55. Sakmar TP, Franke RR, Khorana HG. Glutamic acid-113 serves as the retinylidene Schiff base counterion in bovine rhodopsin. Proc Natl Acad Sci USA. 1989;86:8309–13.
56. Hunt DM, Carvalho LS, Cowing JA, Davies WL. Evolution and spectral tuning of visual pigments in birds and mammals. Philos Trans R Soc B Biol Sci. 2009;364:2941–55.
57. Sugawara T, Terai Y, Imai H, Turner GF, Koblmüller S, Sturmbauer C, et al. Parallelism of amino acid changes at the RH1 affecting spectral sensitivity among deep-water cichlids from Lakes Tanganyika and Malawi. Proc Natl Acad Sci USA. 2005;102:5448–53.
58. van Hazel I, Dungan SZ, Hauser FE, Morrow JM, Endler JA, Chang BSW. A comparative study of rhodopsin function in the great bowerbird (*Ptilonorhynchus nuchalis*): spectral tuning and light-activated kinetics. Protein Sci. 2016;25:1308–18.
59. Dungan SZ, Kosyakov A, Chang BSW. Spectral tuning of killer whale (*Orcinus orca*) rhodopsin: evidence for positive selection and functional adaptation in a cetacean visual pigment. Mol Biol Evol. 2016;33:323–36.
60. Salcedo E, Huber A, Henrich S, Chadwell LV, Chou WH, Paulsen R, et al. Blue- and green-absorbing visual pigments of *Drosophila*: ectopic expression and physiological characterization of the R8 photoreceptor cell-specific Rh5 and Rh6 rhodopsins. J Neurosci. 1999;19:10716–26.
61. Salcedo E, Zheng L, Phistry M, Bagg EE, Britt SG. Molecular basis for ultraviolet vision in invertebrates. J Neurosci. 2003;23:10873–8.
62. Oliveira L, Paiva ACM, Sander C, Vriend G. A common step for signal transduction in G protein-coupled receptors. Trends Pharmacol Sci. 1994;15:170–2.
63. Scheer A, Fanelli F, Costa T, De Benedetti PG, Cotecchia S. Constitutively active mutants of the alpha 1B-adrenergic receptor: role of highly conserved polar amino acids in receptor activation. EMBO J. 1996;15:3566–78. https://doi.org/10.1002/j.1460-2075.1996.tb00726.x.
64. Bockaert J, Pin JP. Molecular tinkering of G protein-coupled receptors: an evolutionary success. EMBO J. 1999;18:1723–9.
65. Fritze O, Filipek S, Kuksa V, Palczewski K, Hofmann KP, Ernst OP. Role of the conserved NPxxY(x)5,6F motif in the rhodopsin ground state and during activation. Proc Natl Acad Sci U S A. 2003;100:2290–5.
66. Layden MJ, Rentzsch F, Röttinger E. The rise of the starlet sea anemone Nematostella vectensis as a model system to investigate development and regeneration. Wiley Interdiscip Rev Dev Biol. 2016;5:408–28.
67. Warner JF, Guerlais V, Amiel AR, Johnston H, Nedoncelle K, Röttinger E. NvERTx: a gene expression database to compare embryogenesis and regeneration in the sea anemone *Nematostella vectensis*. Development. 2018;145:dev162867. https://doi.org/10.1242/dev.162867.
68. Feuda R, Menon AK, Göpfert MC. Rethinking Opsins. Mol Biol Evol. 2022;39(3):msac033.
69. Sokabe T, Chen HC, Luo J, Montell C. A switch in thermal preference in *Drosophila* larvae depends on multiple rhodopsins. Cell Rep. 2016;17:336–44. https://doi.org/10.1016/j.celrep.2016.09.028.
70. Zanini D, Giraldo D, Warren B, Katana R, Andrés M, Reddy S, et al. Proprioceptive opsin functions in drosophila larval locomotion. Neuron. 2018;98:67-74.e4.
71. Leung NY, Thakur DP, Gurav AS, Kim SH, Di Pizio A, Niv MY, et al. Functions of opsins in *Drosophila* taste. Curr Biol. 2020;30:1367-1379.e6. https://doi.org/10.1016/j.cub.2020.01.068.
72. Leclère L, Horin C, Chevalier S, Lapébie P, Dru P, Peron S, et al. The genome of the jellyfish *Clytia hemisphaerica* and the evolution of the cnidarian life-cycle. Nat Ecol Evol. 2019;3:801–10.
73. Chen CY, McKinney SA, Ellington LR, Gibson MC. Hedgehog signaling is required for endomesodermal patterning and germ cell development in the sea anemone nematostella vectensis. Elife. 2020;9:1–27.
74. Levy S, Elek A, Grau-Bové X, Menéndez-Bravo S, Iglesias M, Tanay A, et al. A stony coral cell atlas illuminates the molecular and cellular basis of coral symbiosis, calcification, and immunity. Cell. 2021;184:2973-2987.e18.
75. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215:403–10.
76. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30:772–80.
77. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 2020;37:1530–4.
78. Chen K, Durand D, Farach-Colton M. NOTUNG: A program for dating gene duplications and optimizing gene family trees. J Comput Biol. 2000;7:429–47.
79. Putnam NH, Srivastava M, Hellsten U, Dirks B, Chapman J, Salamov A, et al. Sea anemone genome reveals the gene repertoire and genomic organization of the eumetazoan ancestor. Siencemag. 2007;317(5834):86–94.
80. Song L, Florea L. Rcorrector: Efficient and accurate error correction for Illumina RNA-seq reads. Gigascience. 2015;4:1–8. https://doi.org/10.1186/s13742-015-0089-y.
81. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357–9.
82. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011;29:644–52.
83. Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. Mol Biol Evol. 2021;38:4647–54.
84. Gilbert DG. Longest protein, longest transcript or most expression, for accurate gene reconstruction of transcriptomes? BioRxiv. 2019. https://doi.org/10.1101/829184.
85. Fredman D, Schwaiger M, Rentzsch F, Technau U. Nematostella vectensis transcriptome and gene models v3.0. Berlin: Springer; 2013.
86. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol. 2016;34:525–7.
87. Pimentel H, Bray NL, Puente S, Melsted P, Pachter L. Differential analysis of RNA-seq incorporating quantification uncertainty. Nat Methods. 2017;14:687–90.
88. Warnes G, Bolker B, Bonebakker L, Gentleman R, Huber W, Liaw A, et al. Gplots: various r programming tools for plotting data R package version 3.1.3. Berlin: Science open; 2022.

89. Team RC. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2023.
90. Wolenski FS, Layden MJ, Martindale MQ, Gilmore TD, Finnerty JR. Characterizing the spatiotemporal expression of RNAs and proteins in the starlet sea anemone *Nematostella vectensis*. Nat Protoc. 2013;8:900–15.
91. Kayal E, Bentlage B, Sabrina Pankey M, Ohdera AH, Medina M, Plachetzki DC, et al. Phylogenomics provides a robust topology of the major cnidarian lineages and insights on the origins of key organismal traits. BMC Evol Biol. 2018;18:1–18.

Publisher's Note