The last common ancestor of bilaterian animals possessed at least 7 opsins

*Ramirez, MD[1], Pairett, AN[2], Pankey, MS[3], Serb, JM[2], Speiser, DI[4], Swafford, AJ[1], *Oakley, TH[1]*

May 12, 2016

1: Department of Ecology, Evolution and Marine Biology, University of California, Santa Barbara

2: Department of Ecology, Evolution and Organismal Biology, Iowa State University

3: Department of Molecular, Cellular and Biomedical Sciences, University of New Hampshire

4Department of Biological Sciences, University of South Carolina

*Author for Correspondence:
Todd Oakley, Department of Ecology, Evolution and Marine Biology, University of California, Santa Barbara, Santa Barbara, CA, oakley@lifesci.ucsb.edu

Opsins, the primary proteins animals use to sense light, have undergone a dramatic expansion since they originated early in animal evolution. Understanding the origins of opsin diversity can offer clues to how separate lineages of animals have repurposed different opsin paralogs for different light-detecting functions. However, the more we look for opsins outside of eyes and from additional animal phyla, the more opsins we uncover, suggesting we still do not know the true extent of opsin diversity, nor the ancestry of opsin diversity in animals. To estimate the number of opsin paralogs present in the last common ancestor of all bilaterians and Cnidaria + Bilateria, we reconstructed a reconciled opsin phylogeny using sequences from 15 animal phyla, including the traditionally poorly-sampled echinoderms and molluscs. Our analysis strongly supports a repertoire of nine opsin paralogs in the bilaterian ancestor and four opsin paralogs in the last common ancestor of cnidarians+bilaterians. Thus we have found a greater opsin diversity earlier in animal history than previously known. Further, opsins likely duplicated and were lost many times, with different lineages of animals maintaining different repertoires of opsin paralogs. This phylogenetic information can inform hypotheses about the functions of different opsin paralogs and be used to understand how and when opsins were incorporated into complex traits like eyes and extraocular sensors.

Keywords:

reconciled tree, eye evolution, extraocular photoreceptors, phototransduction, vision

Introduction

As the protein component of visual pigments, opsins are used in the majority of light-detecting cells found in animals (Nilsson 2013). Opsins are G-protein coupled receptors which bind a light-sensitive chromophore via a Schiff base linkage at a conserved lysine residue (Terakita 2005). When the chromophore absorbs a photon, conformational changes in the chromophore and opsin protein result in the activation of a G-protein based signal transduction cascade (Terakita 2005). Despite their widespread importance in animal photosensitivity, most work on the function and evolution of opsins focused initially on those expressed in the eyes of vertebrates and arthropods (O'Tousa et al. 1985); (Nathans & Hogness 1983). Only recently

has work on opsins included those expressed outside eyes or from other animal phyla (Velarde et al. 2005; Radu et al. 2008; Hering et al. 2012; D'Aniello et al. 2015; Hering & Mayer 2014). We now know the evolutionary history of opsins is one of many gains and losses of genes across time and among species (Colbourne et al. 2011; Henze & Oakley 2015; Davies et al. 2015; Liegertová et al. 2015; Feuda et al. 2016). This kind of high gene turnover requires broad taxonomic sampling of opsins to fully reconstruct their evolutionary origins, simply because we know that ancient losses may result in the complete absence of some opsin paralogs, even in major groups of animals. Previous large-scale opsin phylogenies have also found many sequences that fall outside of the well-known opsin groups, typically identified in phyla for which we have sparse data, e.g. arthropsins in *Daphnia* or Echinopsins B in echinoderms (e.g. Colbourne et al. 2011; D'Aniello et al. 2015). Most analyses do not address the nature of these orphaned sequences. While they may be recently-diverged, lineage-specific duplications, another possibility is that they represent entire opsin paralogs that are not found within the phyla that have been most heavily sampled, and have thus not been recognized. Without an accurate picture of how opsin paralogs are distributed among animals, it is challenging to address how diverse opsins really are, when that diversity arose, and how different opsins integrated into different kinds of light-detecting stuructures through evolution.

Opsins evolved very early in animals (Plachetzki et al. 2007; Oakley & Speiser 2015), likely first expressed in light-sensitive cells and later in more complex structures like eyes (Arendt & Wittbrodt 2001; Nilsson 2013). Historically, opsin diversity has been partitioned among three clades which we will refer to as 'canonical c-opsins', 'canonical r-opsins,' and 'tetraopsins', formerly 'ciliary', 'rhabdomeric,' and 'Group 4 opsins' *sensu* (Porter et al. 2012; Liegertová et al. 2015), respectively. A possible fourth clade of opsins, cnidops, are known only from cnidarians. To understand how many opsin paralogs were present in the last common eumetazoan and bilaterian ancestors, we need to understand when these major opsin clades arose and how they are related to each other.

Because cnidarians are one of the earliest branching animal lineages with opsins, the opsin repertoire of cnidarians is thought to represent the opsin paralogs present in the last common ancestor of eumetazoans. However, relating cnidarian opsins to the major animal opsin paralogs has proved difficult, and hypotheses on how cnidarian and bilaterian opsins relate vary widely between analyses. Overall, a prevailing view is that the most recent ancestor of eumetazoans had three opsin paralogs: c-opsins, r-opsins and tetraopsins (Roberto Feuda et al. 2014; Suga et al. 2008; Feuda et al. 2012). Cnidarian

genomes are hypothesized to encode either the cnidarian-specific cnidops alone (Porter et al. 2012; Liegertová et al. 2015), cnidops and c-opsins (Porter et al. 2012; Vopalensky & Kozmik 2009; Plachetzki et al. 2007) or c-opsins, r-opsins and Group 4 opsins in common with bilaterians (Roberto Feuda et al. 2014; Suga et al. 2008; Feuda et al. 2012). Based on an *in-vitro* assay, an opsin from the coral *Acropora palmata* couples with the G-protein q alpha subunit (Mason et al. 2012; Lee et al. 1994). Together with the hypothesized phylogenetic position of this opsin, the functional test suggests that some cnidarians may possess canonical r-opsins. Still, the exact placement of cnidarian opsins is in flux, highly sensitive to the specific substitution models and gene sampling regime used in each analysis, and yet a solid understanding of their placement is important for understanding the origins of bilaterian opsin diversity.

The reconstruction of opsin evolution in the bilaterians poses yet more challenges. Early estimates of opsin diversity in the last common bilaterian ancestor identified two (Nilsson 2005) or three (Plachetzki et al. 2007; Porter et al. 2012; Feuda et al. 2012; Roberto Feuda et al. 2014) paralogs, corresponding to the canonical c-opsins and canonical r-opsins, or canonical c-. r- and tetraopsins respectively. No bilaterians seemed to have direct orthologs of cnidops. Recent sampling efforts to survey new taxa and extraocular tissues have expanded our current view of opsin diversity, and we now recognize that multiple clades of opsins found in extant animals were present in the last common ancestor of bilaterians, based on their presence in both deuterostome (e.g. vertebrates and echinoderms) and protostome (e.g. arthropods and molluscs) genomes. This raises the minimum opsin paralog count in the last common ancestor of bilaterians to five (Terakita 2005; Vopalensky & Kozmik 2009; Suga et al. 2008) or six (Hering & Mayer 2014; Liegertová et al. 2015; Roberto Feuda et al. 2014), distributed between the bilaterian c-, r- and tetraopsins. With these additions, a pattern emerges -- as we catalog opsins in diverse phyla and from different types of light receptors, we uncover a greater diversity of opsin paralogs.

Thus, the goal of our analysis is to reconstruct a more taxonomically comprehensive evolutionary history of animal opsins to understand the origins of bilaterian opsin diversity. We achieve this in two ways. First, we include newly published opsin sequences from multiple studies that have yet to be synthesized in a large scale phylogenetic analysis. Second, we identify additional new opsins from both publicly available transcriptomes and nine unpublished mollusc transcriptomes, as molluscs are the second most speciose phylum but lag far behind other large taxa in terms of representation in opsin phylogenies to date (see Supplemental Table 1). With this more com-

prehensive data set, we produced the first large-scale reconciled opsin phylogeny to better estimate the number of opsins present in the last common bilaterian ancestor. This approach allows us to infer nine opsin paralogs were likely present in the last common bilaterian ancestor. Further, we find that all cnidarian opsins are sister to three opsin paralogs found in other animals, rather than forming cnidarian-specific paralogs. This distribution of cnidarian opsins allows us to infer the presence of four paralogs in the common ancestor of the eumetazoa. These results suggest a rapid radiation in opsin diversity prior to the origin of bilaterians, followed by unique patterns of duplications and losses specific to different animal lineages. Finally, these results urge a renewed focus on surveying opsins in understudied phyla (two prime remaining candidates are Annelida and Cnidaria), on including sufficiently diverse sequences when resolving opsin relationships, and on performing functional experiments to determine both the roles of non-visual opsins and the extent to which orthologous opsins in divergent phyla perform similar functions.

Methods

Data collection:

We searched both NCBI and UniProt using BLAST (Gish & States 1993) with a bait set of 5 opsin sequences (accession numbers: BAG80696.1; NP_001014890.1; CAA49906.1; O15974.1; P23820.1, see Suppl. Table 2 for more info) and an e-value cutoff of 1e-5. Our goal was to maximize the identification of potential opsins from understudied taxa, so we excluded vertebrates and arthropods from our BLAST search on NCBI and downloaded the top 250 hits per opsin bait. We searched Uniref90 with the bait sequences and same cutoff value, then downloaded only lophotrochozoan (NCBI taxonomic ID: 1206795) sequences/clusters. We combined all the sequences we recovered from NCBI and Uniref90 with sequences from other publications, which include tardigrades, arthropods, ambulcraria, cubozoan cnidarians and zebrafish (Hering et al. 2015; Davies et al. 2015). To this initial database of published sequences, we added mollusc opsins that we gathered by running Phylogenetically Informed Annotation, PIA, (Speiser et al. 2014) on transcriptomes from 7 cephalopods, 5 chitons, 1 gastropod, and 1 bivalve (see Suppl. Table 1 for species and sequence Genbank accession numbers).

Data grooming:

Because our initial data collection was permissive, our raw dataset (over 1,600 sequences) contained both duplicates as well as a num-

ber of non-opsin GPCRs. We used CD-HIT (Li & Godzik 2006; Fu et al. 2012) to cluster together sequences that were more than 90% similar to each other to remove duplicates and short sequences that were identical to longer sequences already in the dataset. This also allowed us to reduce the sample size in the alignment by cutting highly similar sequences, while maintaining overall diversity of sequences in the dataset. To remove non-opsin GPCRs, we first ran the dataset through SATé-II (Liu et al. 2012) using the automatic settings. SATe-II runs FastTree 2 (Price et al. 2010) on an initial MAFFT (Katoh & Standley 2013) alignment, then subdivides the alignment into subproblems (maximum size is 200 for the auto setting), which are each realigned with MAFFT. The realigned subproblems are then merged using MUSCLE (Edgar 2004), and a new tree produced by FastTree, and the maximum likelihood (ML) score is calculated. This process is iterated until a pre-defined stopping point. For multiple sequences alignments, SATe-II performs best overall compared to other alignment programs like MAFFT or MUSCLE (Pervez et al. 2014). We used FigTree (Rambaut 2007) to visualize the tree from our SATe run, rooted with melatonin receptors (Roberto Feuda et al. 2014). We then trimmed this tree to exclude non-opsins using a custom python script called Supercuts (Swafford 2016) and retained the ingroup clade for subsequent analyses. Next, we removed any sequences from the alignment that lacked the conserved lysine residue homologous to K296 of bovine rhodopsin. We also manually trimmed the beginning and end of the alignment to the first and last aligned blocks using Aliview (Larsson 2014). Finally, although they lack the conserved lysine, we added the *Trichoplax adherens* placopsins back to our dataset as a close outgroup to root our tree, as (Feuda et al. 2012) showed that placopsins are sister to all other animal opsins. In total, our groomed dataset had 789 opsins with the conserved K296 (plus three placozoan opsins without the lysine) from 368 species across 15 phyla.

Tree estimation and support values:

To create the final alignment for our dataset, we ran SATe on our dataset using the following configuration: a subproblem fraction of 0.025, stopping iterations after 5 unimproved ML scores and FastTree under the GTR gamma model. We used the MPI version of IQ-TREE 1.4.0 (Nguyen et al. 2014), to select a substitution model based on our SATe alignment, to infer a maximum likelihood tree and compute support values. IQ-TREE incorporates an approach for calculating ultrafast bootstraps (UFBoot), which have fewer biases compared to other bootstrapping methods (Minh et al. 2013). We were also

able to perform the SH-like approximate likelihood ratio test (SH-aLRT) and the approximate Bayes test (Anisimova et al. 2011) as implemented in IQ-TREE to assess support for single branches to complement our UFBoot analysis (Guindon et al. 2010; Anisimova et al. 2011). SH-aLRT branch supports are often more consistent and conservative than bootstrapping methods (Simmons & Randle 2014; Simmons & Norton 2014). The IQ-TREE substitution model test selected the LG+F+R8 model for our alignment based on BIC. Because we had a large number of relatively short sequences, we performed 50 ML tree searches varying the perturbation value (0.1-0.5). We also extended the number of trees IQ-TREE searched once it found a tree with a better ML score to 500. This helped ensure that the algorithm was exploring the tree parameter space and not getting stuck at a local maximum. Two trees had virtually identical high log-likelihood scores, and so we ran IQ-TREE again, setting each tree as the starting tree, to break the tie and to get UFBoot, SH-aLRT and aBayes values for the final, highest log-likelihood tree. The code used for this analysis, our dataset and the resultant tree are available on BitBucket (UCSB Phylogenetics).

Tree reconciliation and rearrangement:

We used NOTUNG 2.8 (Chen et al. 2000) to reconcile the gene tree with a metazoan species tree based on NCBI Taxonomy. This animal phylogeny places ctenophores sister to cnidarians, and sponges as sister to all other animals. While the order of branching in early metazoans is contentious, our results are unaffected by this uncertainty. To perform both a reconciliation and rearrangement of weakly supported branches, NOTUNG requires a fully resolved species tree. We used the ape package (Paradis et al. 2004) in R (R Core Team 2016) to randomly resolve polytomies present in the tree from NCBI. However, our analysis focuses on major splits in the animal phylogeny that are well supported, e.g. protostomes vs deuterostomes, and so the random resolution of more shallow nodes did not impact our results. We set the penalty for duplications to 1.5, losses to 1.0 and the edge weight threshold for rearrangement to 95.0. After reconciling the tree, we used NOTUNG to rearrange nodes with UFBoot supports that fell below the 95.0 threshold to create our finalized tree.

Tree visualization:

We used FigTree 1.4.2 (Rambaut 2007) and TreeGraph2 (Stöver & Müller 2010) to collapse opsin clades by hand according to major taxonomic group (chordates, echinoderms, lophotrochozoans or ecdysozoans), and Evolview (Zhang et al. 2012) to format the tree

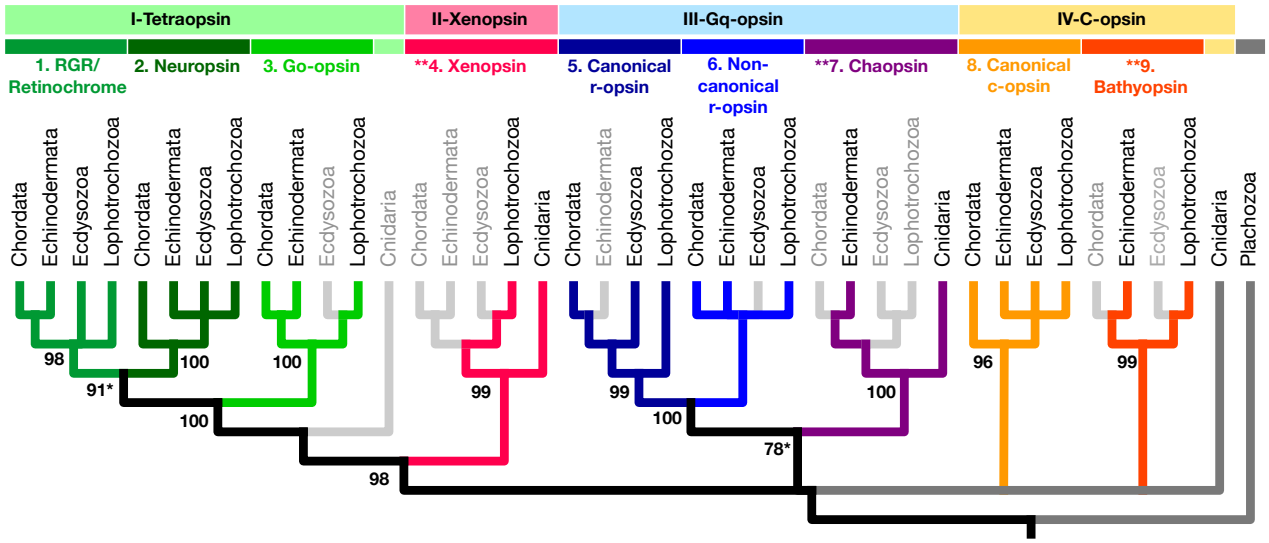

Figure 1: There are nine bilaterian opsin paralogs spread among four major eumetazoan opsin paralogs. The four major eumetazoan opsin paralogs are indicated at the top with roman numerals. The nine bilaterian opsin paralogs are indicated with arabic numerals and are color coded to match the corresponding branches. Colored branches indicate the presence of an opsin in at least one species within the major taxonomic group. Light gray branches indicate the absence of an opsin paralog from the taxa indicated at the tips. Ultrafast bootstrap (UFBoot) supports from IQ-TREE are given next to the branch they support. Bootstraps with asterisks were added from the gene tree after reconciliation analysis.

color, branch length, etc. for Figure 1. We used iTOL (Letunic & Bork 2011) to combine the tallies of opsin per phylum or molluscan class with animal and mollusc phylogenies (Figures 4 and 5). We made final adjustments to the outputs of these programs using OmniGraffle Pro (v. 6.5, Omni Group, Seattle, WA).

Results

From our reconciled tree containing 789 unique sequences, we infer nine bilaterian opsin paralogs spread across four major eumetazoan paralogs (Figures 1 & 2, complete tree in Suppl. Figure 1). We recover the six bilaterian opsin paralog groups described in previous publications: canonical c-opsin, canonical r-opsin, Go-opsin, RGR/retinochrome/peropsin, neuropsin and arthropsin. Our broader taxonomic sampling also allows us to infer three previously undescribed bilaterian paralogs, which we have named "xenopsins", "bathyopsins" and "chaopsins". Because adding so many new bilaterian opsins changes the relationships between paralogs, we establish new, named hypotheses for these relationships, as often done in species-level phylogenetic analyses (see Table 1). Also, to help clarify which opsin clades are inferred as eumetazoan versus bilaterian paralogs at a glance, we use roman numerals to refer to the eumetazoan paralogs and arabic numerals to refer to the bilaterian paralogs in the text and figures.

Table 1: Summary of opsins present before bilaterians and present in the last common bilaterian ancestor. We considered UFBoot values above 95 as strong support for the monophyly of that group, but allowed branch rearrangements below that threshold. Asterisks (**) indicate support for the node based on reconciliation with the species tree.

Last common eumetazoan ancestor opsin paralogs	Last common bilaterian ancestor opsin paralogs	Previously named clades within each group	UFBoot support
Tetraopsins	RGR/Peropsins/Retinochromes	chordate Rrh/RGR/peropsin, echinoderm RGR-like, mollusc retinochrome/peropsin-like/arthropod peropsin-like	98
	Go-opsins	echinoderm, cephalochordate and lophotrochozoan Go-opsins	100
	Neuropsins	chordate, non-mammalian vertebrate, ambulacrarian, lophotrochozoan and arthropod neuropsins/opn5	100
C-opsins	Canonical c-opsins	chordate TMT, chordate encephalopsins, echinoderm encephalopsin-like, arthropod pteropsins, Platynereis c-opsin, vertebrate visual c-opsins	96
	Bathyopsins	echinoderm and inarticulate brachiopod bathyopsins	99
Xenopsins	Xenopsins	Cnidarian cnidops, lophotrochozoan xenopsins	99
Gq-opsins	Canonical r-opsins	lophotrochozoan visual r-opsins, platyhelminthes r-opsin, chordate melanopsins, arthropod visual r-opsins, arthropod arthropsins	99
	Non-canonical r-opsins	lophotrochozoa, ambulacrarian and cephalochordate 'arthropsins'	**
	Chaopsins	echinoderm Echinopsins B and Anthozoa I	100

The eumetazoan ancestor likely had 4 opsin paralogs

We find four eumetazoan opsin paralog: the tetraopsins, the xenopsins, the c-opsins, and the Gq-opsins. The tetraopsins consist of bilaterian RGR/retinochromes/peropsins, neuropsins, and Go-opsins. Xenopsins consist of the cnidarian cnidops and a new clade of lophotrochozoan xenopsins. Tetraopsins are well supported as the sister clade to the xenopsins (see Figure 1). The eumetazoan Gq-opsins include the canonical visual opsins of protostomes (e.g. arthropods and molluscs) plus the chordate melanopsins, bilaterian non-canonical r-opsins, and new clade chaopsins. Chaopsins consist of sequences from echinoderms and cnidarians. Finally, the eumetazoan c-opsins are the canonical c-opsins, (which include vertebrate visual and brain opsins, arthropod pteropsins and the annelid brain c-opsin), 'Anthozoa II' (Hering & Mayer 2014), as well as the new clade bathyopsins.

All cnidarian opsins fell within three of the inferred four eumetazoan opsin paralog groups, as cnidarian xenopsins, c-opsins and Gq-opsins (see Figure 1 and Suppl. Fig 1). We did not recover a cnidarian clade of the tetraopsins. Unlike previous analyses which did not include lophotrochozoan xenopsins, we find that cnidarian cnidops (Plachetzki et al. 2007) form the xenopsins with sequences from lophotrochozoans with high bootstrap support. Cnidarian

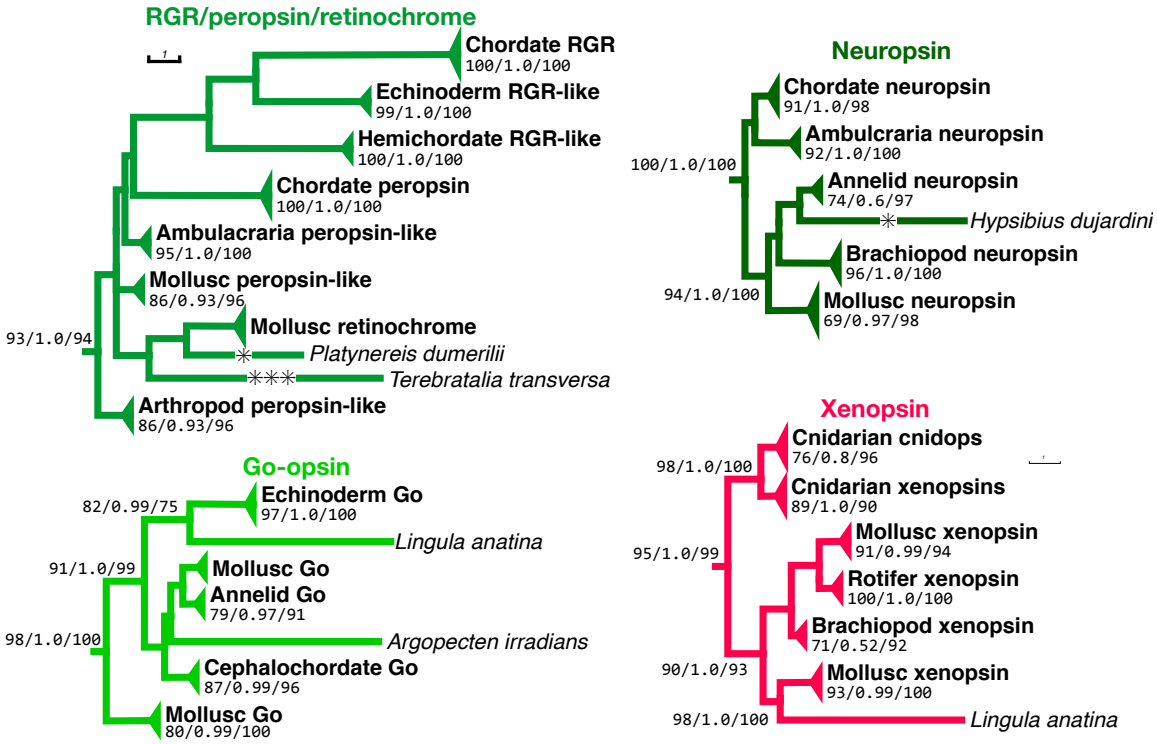

Figure 2: Opsin paralog trees for the tetraopsins and xenopsins, representing the relationships between opsins by phylum. Each tree shows opsin sequences collapsed by clade. Values below the clade name representSH-aLRT/aBayes/UFBoots. Only clades with bootstrap supports above 75% are shown. The full gene tree can be found in Supplemental Figure 1. Each asterisk '*' on a branch represents a shortening by 5 branch length units.

c-opsins are a group of anthozoan opsins, Anthozoa II (Hering & Mayer 2014), which were placed within the eumetazoan c-opsins based on our reconciliation analysis (Plachetzki et al. 2007; Suga et al. 2008; Feuda et al. 2012; Roberto Feuda et al. 2014). Finally, cnidarian Gq-opsins are another set of anthozoan sequences, Anthozoa I (Hering & Mayer 2014), which fall sister to the echinoderm chaopsins with high bootstrap support.

The bilaterian ancestor likely had 9 different opsin paralogs

I- Tetraopsins

Similar to previous analyses (Porter et al. 2012; Hering & Mayer 2014; Roberto Feuda et al. 2014), we recover the tetraopsins, formerly "RGR/Go" or "Group 4" opsins, as a monophyletic group with strong support (UFBoot=100). They consist of RGR/retinochromes /peropsins, Go-opsins, and neuropsins. Because our tree shows strong support for these opsins as most closely related to each other, we have renamed this clade of opsins tetraopsins. Further, we find that each of the previously recognized, major splits within tetraopsins has representatives from both protostomes and deuterostomes. So while RGR/retinochromes/peropsins, Go-opsins, and

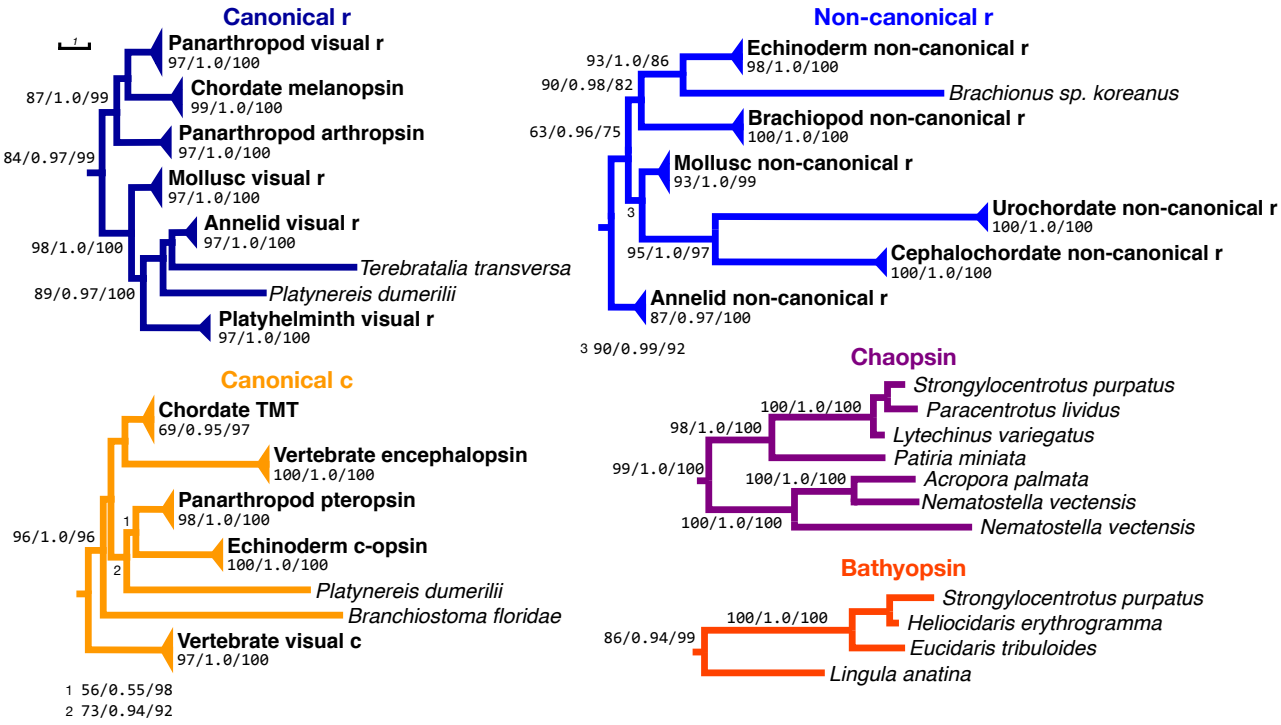

Figure 3: Opsin paralog trees for the Gq-opsins and C-opsins, representing the relationships between opsin orthologs by phylum. Each tree shows opsin sequences collapsed by clade. Values below the clade name represent SH-aLRT/aBayes/UFBoots. Only clades with bootstrap supports above 75% are shown. The full gene tree can be found in Supplemental Figure 1.

neuropsins form a well supported clade, they also each represent distinct bilaterian opsin clades with high bootstrap support (see Figure 1).

1. RGR/retinochromes/peropsins

The RGR/retinochrome/peropsin clade is well-supported by our tree (UFBoot=98, see Figure 2). Deuterostome RGRs include the original RGRs identified in vertebrates, as well as RGR-like sequences in cephalochordates, hemichordates, and echinoderms (Jiang et al. 1993; Holland et al. 2008; D'Aniello et al. 2015). Deuterostome peropsins include RRH from vertebrates as well as peropsin-like sequences from cephalochordates, hemichordates and echinoderms (Sun et al. 1997; Holland et al. 2008; D'Aniello et al. 2015). Protostome retinochromes include the originally described retinochromes from cephalopods, plus retinochrome-like sequences in bivalve and gastropod molluscs (Hara & Hara 1967; Katagiri et al. 2001). We recovered an additional 3 retinochrome-like sequences from mollusc transcriptomes, including 1 from the gastropod *Bithynia siamensis goniomphalos* and 2 from the chitons *Stenoplax conspicua* and *Chiton virgulatus*. In addition to the molluscs, we found retinochrome-like sequences in

the brachiopod *Terebratalia transversa* , previously described as a Go-opsin (Passamaneck & Martindale 2013) and a sequence previously described as a peropsin in the annelid *Platynereis dumerilli* (Marlow et al. 2014). We also found a small clade of protostome sequences that fell outside of the protostome retinochromes, including 4 sequences from the genomes of the mollusks *Crassostrea gigas, Lottia gigantea* and *Octopus bimaculoides* (Albertin et al. 2015). Finally, non-insect arthropod peropsin-like sequences (Henze & Oakley 2015) also belonged in the clade of protostome retinochromes. It is unclear from our analysis whether RGR/retinochromes and peropsins are separate bilaterian paralogs. We did recover distinct groups, suggestive of of two bilaterian clades, but had low support values at these nodes, and so we collapsed the groups together (see Suppl. Figure 1).

2. *Neuropsins*

The split between the protostome and deuterostome neuropsins is well supported (UFBoot=100, see Figure 2). Deuterostome neuropsins/opn5 sequences include a large clade of vertebrate and cephalochordate neuropsins, a large clade of non-mammalian vertebrate neuropsins, plus neuropsin-like sequences from the Ambulacraria (including those from both hemichordates and echinoderms). Neuropsins from protostomes include sequences from annelids, both *Platynereis dumerilli* (Gühmann et al. 2015) and *Capitella teleta*(Simakov et al. 2012), bivalve and gastropod molluscs, and from the brachiopod *Lingula anatina* (previously annotated as a peropsin). We recovered an additional bivalve neuropsin from the transcriptome of the scallop *Argopecten irradians* . We also found two pan-arthropod neuropsin-like sequences from water flea *Daphnia pulex* (Hering & Mayer 2014; Brandon 2015) and the tardigrade *Hypsibius dujardini* (Hering & Mayer 2014).

3. *Go-opsins*

The deuterostome and protostome Go-opsins form a well supported clade of bilaterian opsins (UFBoot=100, see Figure 2). We recovered the same deuterostome Go-opsins from echinoderms and cephalochordates as identified from previous analyses (D'Aniello et al. 2015). From protostomes, we found previously described sequences of Go-opsins from both bivalve and gastropod mollusc, and also sequences from brachiopods and annelids. We also recovered a new Go-opsin from the transcriptome of the scallop *A. irradians*.

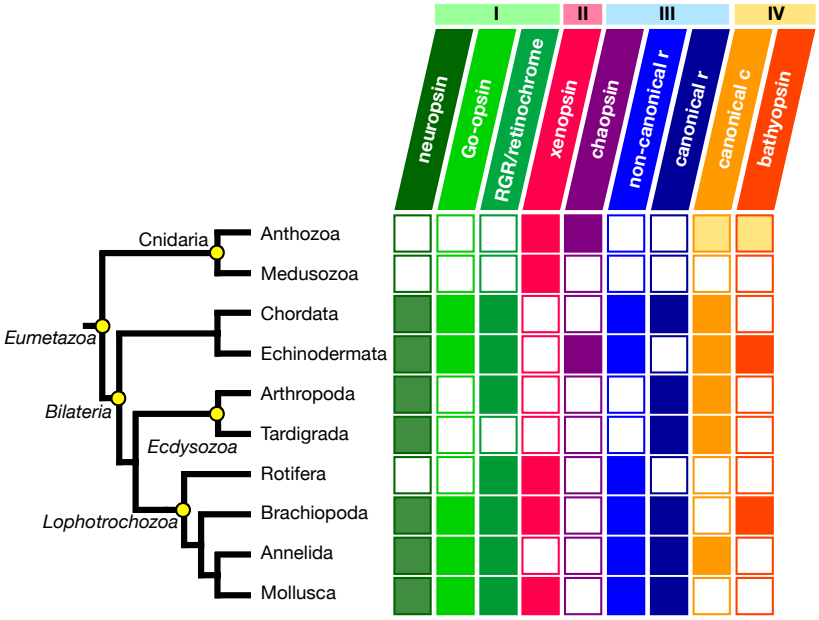

Figure 4: The history of opsins is marked by ancient diversity and subsequent losses of paralogs along different animal lineages. Summary of known opsin complements in major animal phyla. Major subdivisions of metazoans are indicated on the phylogeny as yellow dots with italic labels. Phyla are represented at the tips, except for cnidarians, which are broken down into the two major cnidarian splits. Colored bars with roman numerals indicate opsin paralogs present in the most recent ancestor of eumetazoans. The nine bilaterian opsin paralogs are indicated by slanted colored bars and full opsin names. Filled squares represent presence, empty squares absence of at least one sequence from the opsin paralog group for each phylum listed. Note that no extant phylum included in our analysis seems to have the full complement of bilaterian opsins. The maximum is 7 opsin paralogs in both echinoderms and brachiopods. The anthozoan c-opsins are indicated by yellow boxes under the eumetazoan c-opsin, but fall outside of either bilaterian paralog group.

II-4. New opsin group: Xenopsins

The opsin group we call xenopsins consists of sequences from a variety of lophotrochozoan protostomes (molluscs, rotifers and brachiopods) and cnidarian cnidops. This clade is well supported in our tree (see Figure 2, UFBoot=99; aLRT=95.4; aBayes=1.0). We did not find support for any other protostome (e.g. ecdysozoan) or deuterostome xenopsins. We recovered both the lophotrochozoan xenopsins and cnidarian cnidops with strong support (see Suppl Figure 1). The xenopsins are well supported as sister to the tetraopsins (UFBoot=98; aLRT=90.8; aBayes=1.0, see Figure 1 and Suppl. Figure 1)

Many xenopsins were initially described as c-opsin-like in previous analyses, including sequences from the genomes of the mollusks *Crassostrea gigas* and *Lottia gigantea* and the rotifer *Branchionus sp.*, and those from gene expression data generated from the larval eyes of the articulate brachiopod *Terebratalia transversa* , the optic lobes of *Octopus bimaculoides* and the adult eyes of *Idiosepius paradoxus* (Passamaneck et al. 2011; Albertin et al. 2015; Yoshida et al. 2015). However, we believe that the limited taxonomic scope of previous analyses lead to the incorrect classification of these sequences as c-opsin-like. Our tree is the first to include all of these sequences into a single analysis, and our results clearly support them as a monophyletic clade. Finally, in addition to xenopsins that were previously described, we found 7 new mollusc xenopsins from combing through

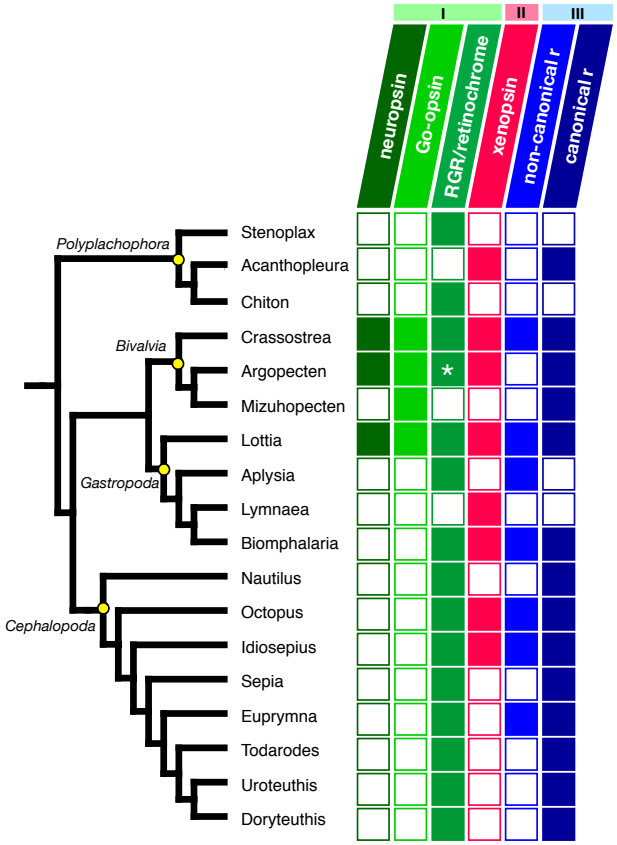

Figure 5: The ancestral mollusc likely had 7 opsins from six of the bilaterian paralog groups. Summary of known opsin complements within the molluscs. Colored bars with roman numerals indicate opsin paralogs present in the most recent ancestor of eumetazoans. The nine bilaterian opsin paralogs are indicated by slanted colored bars and full opsin names. Filled squares represent presence, empty squares absence of at least one sequence from the opsin paralog for each genus listed. The major classes of molluscs are noted with yellow dots and italic labels. Argopecten irradians retinochrome was not included our original analysis, but is present, noted by an asterisk (See Suppl. Figure 2 for RGR/retinochrome gene tree that includes this sequence).

transcriptomes (see Suppl. Table 1).

III-Gq-opsins

We recover a monophyletic clade of r-opsins, consisting of the invertebrate visual r-opsins, chordate melanopsins, arthropod arthropsins, and non-canonical r-opsins from mollusks, annelids and echinoderms. The sister clade to the canonical and non-canonical r-opsins with moderate support (UFBoot=78, see Suppl Fig 2; aLRT=17; aBayes= 0.697) are the chaopsins, described below.

5. Canonical visual r-opsins

This opsin group is well supported by our tree (UFBoot=99, see Figure 3) and includes the following four clades: the canonical visual opsins of arthropods; the chordate melanopsins and arthropod arthropsins; the canonical visual opsins of mollusks; and the (presumably) visual r-opsins from annelids, brachiopods and platyhelminths.

6. Non-canonical r-opsins

The non-canonical r-opsins are sister to the canonical r-opsins with high support (UFBoot=100, see Figure 1), though we do not have strong support for the monophyly of the non-canonical r-opsins, even after reconciliation (see Figures 1 & 3; Suppl. Figure 1). The non-canonical r-opsins contain sequences from deuterostome lineages like echinoderms, hemichordates and cephalochordates, and previously unannotated sequences from protostomes groups that include annelids, brachiopods and molluscs.

7. New opsin group: chaopsins

The opsin group we have named chaopsins consists of two previously described clades of opsins, a group of cnidarian opsins called Anthozoa I (Hering & Mayer 2014) and the echinoderm Echinopsins B (D'Aniello et al. 2015). The grouping of these anthozoan and echinoderm sequences as monophyletic chaopsins is well supported (see Figure 3, UFBoot=100; aLRT=98.6; aBayes=1.0).

IV- C-opsins

This opsin paralog group consists of the canonical c-opsins, a new opsin paralog group named "bathyopsins" and a clade of cnidarian opsins from anthozoans previously named as Anthozoa II (Hering & Mayer 2014). This paralog group had low bootstrap support, but was supported by our reconciliation analysis.

8. Canonical c-opsins

We have renamed as "canonical c-opsins" the monophyletic clade of opsins containing sequences from deuterostomes such as vertebrate visual and brain c-opsins and sequences from protostomes like the arthropod pteropsins (Velarde et al. 2005) and *Platynereis* c-opsin (Arendt et al. 2004). We recovered the canonical c-opsins with high support (UFBoot=96, see Figure 3). Despite mining numerous mollusk transcriptomes for opsin sequences (n=14), we did not recover any additional lophotrochozoan or protostome c-opsins that clustered with the canonical c-opsins besides the single c-opsin reported from the annelid *Platynereis dumerilii* .

9. New hypothesized opsin group: Bathyopsins

The opsin paralog we have named bathyopsins is a small clade with bilaterian support. Sequences from the echinoderms, Echinopsins A (D'Aniello et al. 2015), represent deuterostomes, and sequences from

the genome of the brachiopod *Lingula* represent the prostostomes. Bathyopsins are well supported (see Figure 3, UFBoot=99) as a monophyletic group.

Discussion

Reconstructing the evolutionary history of opsins is vital for understanding how evolution produced light-detecting structures like eyes. Unfortunately, the problem of how and when opsin diversity arose is made difficult by the large number of duplications and losses that have occurred within their evolutionary history. While most analyses of opsin diversity to date have focused on understanding opsin complements within a set of focal taxa, we included multiple poorly-sampled phyla to ensure the broadest phylogenetic scope to date, for a total of 324 species from 15 phyla. Our analysis reveals three previously unrecognized opsin paralogs in extant animals, and the surprising result that these three additional opsin paralogs likely arose early in the evolution of bilaterians, followed by losses and duplications within those opsins that remained.

Our first major finding is that the diversity of opsins in extant animals suggests the presence of at least seven separate opsin paralogs in the bilaterian ancestor, and we infer a total of 9 bilaterian opsin paralogs. In addition to the 6 previously identified bilaterian opsins (c-opsin, r-opsin, melanopsin, Go-opsin, peropsin/RGR/retinochrome and neuropsin), we propose three additional bilaterian opsins--xenopsins, bathyopsins and chaopsins. While we acknowledge the need for additional sequence and expression data to confirm the monophyly of these clades of opsin paralogs, our results are consistent with the hypothesis that these opsin paralogs were all present in the last common ancestor of bilaterians. Hints of the three new opsin groups we have identified can be seen in previous opsin phylogenies (Hering & Mayer 2014; D'Aniello et al. 2015), but hypotheses for how these orphaned sequences relate to other better-studied opsins remained obscure with less broad taxonomic coverage.

For example, cnidarian cnidops have historically been difficult to place consistently within opsin phylogenies. They are sister to the c-opsins in some analyses (Porter et al. 2012; Hering & Mayer 2014; Liegertová et al. 2015) and the tetraopsins in others (Feuda et al. 2012; Roberto Feuda et al. 2014). The fact that cnidops changed positions between analyses suggests that the sequences in the clade are divergent compared to others in the dataset. We found that cnidops fall sister to the lophotrochozoan xenopsins with high bootstrap and branch support, suggesting they are a monophyletic clade. Further, the hypothesis of xenopsins as the sister clade to the tetraopsins is

also well supported both by UFBoot and single branch tests. If our reconciled gene tree is correct, the grouping of lophotrochozoan and cnidarian xenopsins suggests that xenopsins were present in both the bilaterian and eumetazoan ancestors. This differs significantly from previous opsin phylogenies, since those did not include bilaterian xenopsin-like sequences, and so found that cnidops was its own eumetazoan opsin paralog that was lost from bilaterians entirely.

Lophotrochozoan xenopsins are a well-supported monophyletic clade, suggesting that xenopsins were present in the lophotrochozoan ancestor. Interestingly, xenopsins are absent from publicly available *Platynereis* opsins and the *Capitella* and *Helobdella* genomes. However, because our sampling from annelids is so sparse given the large number of species in the phylum, it seems likely that annelid xenopsins could be uncovered after broader sampling. Xenopsins are also absent from both the ecdysozoan and deuterostome taxa included in our analysis. Given that arthropods, chordates, and echinoderms are now well-sampled for opsin diversity, it seems unlikely, though possible, that xenopsins could be uncovered from unsampled species belonging to these phyla. Thus we hypothesize that the absence of xenopsins from these groups in our dataset reflects true losses of xenopsins from ecdysozoan and deuterostomes lineages. Given this hypothesis, we infer that xenopsins were lost at least three times in bilaterians: from ancestors of the annelids, Panarthropoda, and the deuterostomes.

Increased taxon sampling also allows us to hypothesize the bathyopsins and chaopsins as paralogs present in the last common ancestor of bilaterians. These opsins are are unusual because of their extreme phylogenetic sparseness, suggesting that if our gene tree inference is correct, these opsin paralogs were lost in the majority of bilaterians. However, we interpret this sparseness more as an indication that even our inclusive dataset may still be under-sampling true opsin diversity in animal phyla, rather than representing an accurate distribution of these opsins among animals. Bathyopsins are found in only two phyla so far, Echinodermata and Brachiopoda, and are well supported as a monophyletic clade in our tree, though their position as paralog of the eumetazoan c-opsin group is only supported by our reconciliation analysis. Given that bathyopsins are represented by one deuterostome and one protostome, we must infer that bathyopsins were present in the last common bilaterian ancestor. We have not yet found chordate or hemichordate representatives. In protostomes, we infer that the lophotrochozoan ancestor had bathyopsins, but since bathyopsins are unknown in ecdysozoa entirely, it is possible that they were lost or became rare in ecdysozoa after the lophotrochozoan/ecdysozoan split. Because opsins from chordates

and arthropods are well sampled, it is unlikely, but possible, that these phyla possess bathyopsins. We have not uncovered annelid, mollusc or rotifer bathyopsins. However, because lophotrochozoans, and especially annelids, are underrepresented even in our analysis, it is possible that opsin surveys from lophotrochozoans will reveal additional members of the bathyopsins in these phyla.

Chaopsins are found in only two phyla, echinoderms and cnidarians. We find support for monophyletic chaopsins from both high bootstrap and single branch tests, and moderate support of their relationship as sister to the bilaterian r-opsins. Given our data set and analyses, we hypothesize that chaopsins were lost up to three times in bilaterians: twice from deuterostomes (chordates and hemichordates) and once in the ancestor of all protostomes (including both ecdysozoans and lophotrochozoans). We also find that anthozoans are the only cnidarians that have chaopsins, which suggests another potential loss of chaopsins from the ancestor of hydrozoans and cubozoans. As with the other new opsin types we have described, we are more confident that chaopsins are truly lost from chordates and arthropods because of the extent to which those phyla have been sampled. We are much less confident regarding the loss of chaopsins from the lophotrochozoans, which are extremely poorly sampled, especially considering the amount of animal diversity found in the group.

The second major finding is that the eumetazoan ancestor likely had four opsin paralogs, based on the distribution of cnidarian opsins in our analyses. This differs from previous reports which divide cnidarian opsins into one ((Suga et al. 2008; Porter et al. 2012; Feuda et al. 2012; R. Feuda et al. 2014; Hering & Mayer 2014; Liegertová et al. 2015), two (Plachetzki et al. 2007) or three groups (Suga et al. 2008; Porter et al. 2012; Feuda et al. 2012; R. Feuda et al. 2014; Hering & Mayer 2014; Liegertová et al. 2015). We find that all of the cnidarian opsins fall sister to bilaterian sequences. Similar to previous studies, we found that Anthozoa II (Hering & Mayer 2014) is sister to the c-opsins, although with low support. We also found that the anthozoan chaopsins, Anthozoa I (Hering & Mayer 2014), together with the echinoderm chaopsins , Echinopsins B, are sister to the r-opsins with moderate bootstrap and aBayes support, and low aLRT support. The final extant group of cnidarian opsins are the cnidarian xenopsins, composed of the cnidarian cnidops (Plachetzki et al. 2007), which fall sister to the lophotrochozoan xenopsins. In the past, cnidops have been described as their own eumetazoan opsin paralog, but the addition of numerous lophotrochozoan sequences has revealed the bilaterian orthologs of this group. Unlike the other cnidarian opsin paralog groups which contain only anthozoan se-

quences, the cnidarian xenopsins have representatives from all the major classes of cnidarians. Finally, from our phylogeny, we infer the loss of cnidarian opsins belonging to the tetraopsins. However, because cnidarians are not well sampled, it is possible that a cnidarian ortholog of the bilaterian tetraopsins may be uncovered.

We used a traditional animal phylogeny to reconcile our gene tree, with ctenophores placed sister to cnidarians, and these two phyla together as sister to bilaterians. However, the traditional view of ctenophores as sister to cnidarian (Coelenterata hypothesis) is challenged by multiple studies that instead place ctenophores sister to all other animals (ctenophore-out hypothesis) (Dunn et al. 2008; Ryan et al. 2013; Moroz et al. 2014; Borowiec et al. 2015; Pisani et al. 2015; Halanych et al. 2016). There seem to be two opsin paralogs in ctenophores, but the relationship between those opsin paralogs and opsins from other animals is contentious, particularly the placement of Mnemiopsis 3 (Roberto Feuda et al. 2014; Schnitzler et al. 2012). Although Mnemiopsis 3 does have the conserved lysine that aligns at bovine rhodopsin position 296, it was excluded from (Hering & Mayer 2014) because there is an additional insertion that is absent from the other *Mnemiopsis* opsins. Its placement in the metazoan opsin phylogeny is also highly sensitive to outgroups as seen in (Schnitzler et al. 2012; Roberto Feuda et al. 2014). For these reasons, we did not include Mnemiopsis 3 in our analysis. Overall, the results of our analysis are not affected the current controversy about the relationship between ctenophores and other animals (Borowiec et al. 2015; Pisani et al. 2015; Halanych et al. 2016; Pisani et al. 2016). Our results also cannot support either hypothesis, as the ctenophore opsins we included were not placed in the animal opsin phylogeny with high support.

Our opsin dataset includes more sequences from more phyla than any previously published opsin phylogeny. These conditions meant that many nodes were difficult to resolve with high statistical bootstrap support (see Supp. Figure 1) and so we used a unique combination of methods to produce our final tree, incorporating both maximum-likelihood (SATe and IQ-TREE) and gene tree-species tree reconciliation (NOTUNG). Because we were interested in knowing how many opsin paralogs were present in the common ancestor of bilaterians, we needed a better understanding of how different opsins were related to one another. However, because the history of duplications and losses of opsins can make a gene tree by itself difficult to interpret, we reconciled the opsin gene tree to a species tree using NOTUNG, which looked for the most parsimonious pattern of duplications and losses to account for known species relationships. This tells us how opsin paralogs might be related to each other, specifi-

cally whether any pair of opsins found between species arose from duplication events (paralogs) or speciation events (orthologs), and allowed us to determine which groups may have bilaterian or eumetazoan origins. It is worth noting that while reconciliation can bias counts of duplicates and losses (Hahn 2007), we estimate the same minimum number of opsin paralogs in the most recent ancestor of bilaterians with both our reconciled tree and the IQ-TREE gene tree (Suppl. Figures 1). Overall, where we have overlapping data, our final reconciled tree is generally consistent with other large-scale opsin phylogenies (Porter et al. 2012; Hering & Mayer 2014; R. Feuda et al. 2014).

Opsin evolution is surprisingly complex, and this complexity hints at just how much we have yet to learn about how animals use opsins, how these functions shaped the evolution of the gene family and the physiology and behaviors that require opsins. It is not yet clear to what extent the loss of an opsin paralog within an animal lineage suggests the concomitant loss of the organismal function, or whether other opsin paralogs can take over that function. At present, we have no functional data for the majority of the 700+ opsins included in this analysis. While we can make some inferences about the function of a particular opsin based on what we know about orthologous opsins, the data we do have suggests that different animal phyla use related opsins for different purposes, e.g. r-opsins likely mediate vision in many protostome eyes, but the related orthologous melanopsins in vertebrate retinal ganglion cells only have roles in non-visual tasks.

The discovery of opsins restricted to traditionally understudied groups like echinoderms, molluscs, annelids, and brachiopods strongly underscores the need to look outside well-sampled phyla like vertebrates and arthropods for opsins. Additionally, many of these traditionally understudied animals are not well known for possessing eyes, and yet our analysis (among others) shows rich repertoires of opsins in these clades, urging us to look beyond animal eyes for opsin expression. By now it is exceedingly clear that opsins are expressed across the bodies of animals (Ramirez et al. 2011), and our results show that there may still be opsin paralogs to discover outside of animal eyes. Further, many analyses, including ours, use the presence of a conserved lysine residue (bovine rhodopsin K296) as diagnostic for opsins (Terakita 2005). However, we found that keeping this requirement eliminated ~500 sequences recovered by BLAST. While many of these are likely closely related, but non-opsin GPCRs, some may be opsin duplicates that have lost the conserved lysine (Henze & Oakley 2015). Future analyses of these opsin-like sequences may reveal even more ancient diversity than what we have recovered so far.

Finally, our analysis raises an urgent and intriguing question--what was the last common ancestor of Bilateria are animals doing with 9 opsin paralogs? Cataloging opsin sequence diversity alone is insufficient to understand why animals have so many different opsins. We must also take what we learn from an analysis like our own to understand how changes to an opsin's sequence alter its function and how animals use opsins for different tasks. Besides spatial vision, opsins are used for myriad purposes, as depth-gauges, for circadian rhythms or enabling private communication between conspecifics (Bennett 1979; Lythgoe 1979; Cummings et al. 2003; Bybee et al. 2012). Further, while opsins are canonical light detectors, two recent studies have shown roles for opsins in both heat sensing and detecting mechanical stimuli in *Drosophila* (Shen et al. 2011; Senthilan et al. 2012). These studies provide a tantalizing glimpse into opsin functions in sensory modalities besides light detection. Without understanding the true extent of opsin diversity, we cannot understand opsin evolution, the evolution of eyes and other light sensors, or even how a complex trait like eyes can evolve.

Acknowledgments

This work was supported by the National Science Foundation [grant numbers IOS-1457148 to D.I.S & T.H.O.; DMR-1121053, CNS-0960316 to the California NanoSystems Institute and Materials Research Laboratory through the Center for Scientific Computing]. We would like to thank Davide Pisani and Roberto Feuda for comments on preprint version of this manuscript.

References

Albertin CB et al. 2015. The octopus genome and the evolution of cephalopod neural and morphological novelties. Nature. 524:220–224. doi: 10.1038/nature14668.

Anisimova M, Gil M, Dufayard J-F, Dessimoz C, Gascuel O. 2011. Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. Syst. Biol. 60:685–699. doi: 10.1093/sysbio/syr041.

Arendt D, Tessmar-Raible K, Snyman H, Dorresteijn AW, Wittbrodt J. 2004. Ciliary photoreceptors with a vertebrate-type opsin in an invertebrate brain. Science. 306:869–871. `http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=15514158&retmode=ref&cmd=prlinks`.

Arendt D, Wittbrodt J. 2001. Reconstructing the eyes of Urbilateria. Philos. Trans. R. Soc. Lond. B Biol. Sci. 356:1545–1563. `http://rstb.`

royalsocietypublishing.org/cgi/doi/10.1098/rstb.2001.0971.

Bennett MF. 1979. Extraocular Light Receptors and Circadian Rhythms. In: Comparative Physiology and Evolution of Vision in Invertebrates.Handbook of Sensory Physiology Springer Berlin Heidelberg pp. 641–663. doi: 10.1007/978-3-642-66999-6_11.

Borowiec ML, Lee EK, Chiu JC, Plachetzki DC. 2015. Extracting phylogenetic signal and accounting for bias in whole-genome data sets supports the Ctenophora as sister to remaining Metazoa. BMC Genomics. 16:987. doi: 10.1186/s12864-015-2146-4.

Brandon CS. 2015. The Evolutionary Biology of Vision in Daphnia. University of South Carolina http://scholarcommons.sc.edu/etd/3100/?utm_source=scholarcommons.sc.edu%2Fetd%2F3100&utm_medium=PDF&utm_campaign=PDFCoverPages (Accessed May 10, 2016).

Bybee SM et al. 2012. UV photoreceptors and UV-yellow wing pigments in Heliconius butterflies allow a color signal to serve both mimicry and intraspecific communication. Am. Nat. 179:38–51. doi: 10.1086/663192.

Chen K, Durand D, Farach-Colton M. 2000. NOTUNG: a program for dating gene duplications and optimizing gene family trees. J. Comput. Biol. 7:429–447. doi: 10.1089/106652700750050871.

Colbourne JK et al. 2011. The ecoresponsive genome of Daphnia pulex. Science. 331:555–561. doi: 10.1126/science.1197761.

Cummings ME, Rosenthal GG, Ryan MJ. 2003. A private ultraviolet channel in visual communication. Proc. Biol. Sci. 270:897–904. doi: 10.1098/rspb.2003.2334.

D'Aniello S et al. 2015. Opsin evolution in the Ambulacraria. Mar. Genomics. doi: 10.1016/j.margen.2015.10.001.

Davies WIL et al. 2015. An extended family of novel vertebrate photopigments is widely expressed and displays a diversity of function. Genome Res. doi: 10.1101/gr.189886.115.

Dunn CW et al. 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. Nature. 452:745–749. doi: 10.1038/nature06614.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1792–1797. doi: 10.1093/nar/gkh340.

Feuda R, Hamilton SC, McInerney JO, Pisani D. 2012. Metazoan opsin evolution reveals a simple route to animal vision. Proc. Natl. Acad. Sci. U. S. A. 109:18868–18872. http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=23112152&retmode=ref&cmd=prlinks.

Feuda R, Marletaz F, Bentley MA, Holland PWH. 2016. Conservation, duplication and divergence of five opsin genes in insect evolution. Genome Biol. Evol. doi: 10.1093/gbe/evw015.

Feuda R, Rota-Stabelli O, Oakley TH, Pisani D. 2014. The comb jelly opsins and the origins of animal phototransduction. Genome Biol. Evol. 6:1964–1971. doi: 10.1093/gbe/evu154.

Feuda R, Rota-Stabelli O, Oakley TH, Pisani D. 2014. The Comb Jelly Opsins and the Origins of Animal Phototransduction. Genome Biol. Evol. 6:1964–1971. doi: 10.1093/gbe/evu154.

Fu L, Niu B, Zhu Z, Wu S, Li W. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics. 28:3150–3152. doi: 10.1093/bioinformatics/bts565.

Gish W, States DJ. 1993. Identification of protein coding regions by database similarity search. Nat. Genet. 3:266–272. doi: 10.1038/ng0393-266.

Gühmann M et al. 2015. Spectral Tuning of Phototaxis by a Go-Opsin in the Rhabdomeric Eyes of Platynereis. Curr. Biol. 25:2265–2271. doi: 10.1016/j.cub.2015.07.017.

Guindon S et al. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59:307–321. doi: 10.1093/sysbio/syq010.

Hahn MW. 2007. Bias in phylogenetic tree reconciliation methods: implications for vertebrate genome evolution. Genome Biol. 8:R141. doi: 10.1186/gb-2007-8-7-r141.

Halanych KM, Whelan NV, Kocot KM, Kohn AB, Moroz LL. 2016. Miscues misplace sponges. Proc. Natl. Acad. Sci. U. S. A. 113:E946–7. doi: 10.1073/pnas.1525332113.

Hara T, Hara R. 1967. Rhodopsin and retinochrome in the squid retina. Nature. 214:573–575. http://www.ncbi.nlm.nih.gov/pubmed/6036171.

Henze MJ, Oakley TH. 2015. The Dynamic Evolutionary History of Pancrustacean Eyes and Opsins. Integr. Comp. Biol. doi: 10.1093/icb/icv100.

Hering L et al. 2012. Opsins in onychophora (velvet worms) suggest a single origin and subsequent diversification of visual pigments in arthropods. Mol. Biol. Evol. 29:3451–3458. http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=22683812&retmode=ref&cmd=prlinks.

Hering L, Mayer G. 2014. Analysis of the opsin repertoire in the tardigrade Hypsibius dujardini provides insights into the evolution of opsin genes in panarthropoda. Genome Biol. Evol. 6:2380–2391. doi: 10.1093/gbe/evu193.

Holland LZ et al. 2008. The amphioxus genome illuminates vertebrate origins and cephalochordate biology. Genome Res. 18:1100–1111. doi: 10.1101/gr.073676.107.

Jiang M, Pandey S, Fong HK. 1993. An opsin homologue in the retina and pigment epithelium. Invest. Ophthalmol. Vis. Sci.

34:3669–3678. http://www.ncbi.nlm.nih.gov/pubmed/8258527.

Katagiri N, Terakita A, Shichida Y, Katagiri Y. 2001. Demonstration of a rhodopsin-retinochrome system in the stalk eye of a marine gastropod, Onchidium, by immunohistochemistry. J. Comp. Neurol. 433:380–389. doi: 10.1002/cne.1146.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30:772–780. doi: 10.1093/molbev/mst010.

Larsson A. 2014. AliView: a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics. 30:3276–3278. doi: 10.1093/bioinformatics/btu531.

Lee YJ et al. 1994. The Drosophila dgq gene encodes a G alpha protein that mediates phototransduction. Neuron. 13:1143–1157. http://www.ncbi.nlm.nih.gov/pubmed/7946351.

Letunic I, Bork P. 2011. Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. Nucleic Acids Res. 39:W475–8. doi: 10.1093/nar/gkr201.

Liegertová M et al. 2015. Cubozoan genome illuminates functional diversification of opsins and photoreceptor evolution. Sci. Rep. 5:11885. doi: 10.1038/srep11885.

Liu K et al. 2012. SATé-II: Very Fast and Accurate Simultaneous Estimation of Multiple Sequence Alignments and Phylogenetic Trees. Syst. Biol. 61:90–106. doi: 10.1093/sysbio/syr095.

Li W, Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics. 22:1658–1659. doi: 10.1093/bioinformatics/btl158.

Lythgoe JN. 1979. *Ecology of vision*. Clarendon Press; Oxford University Press http://agris.fao.org/agris-search/search.do?recordID=US201300573075.

Marlow H et al. 2014. Larval body patterning and apical organs are conserved in animal evolution. BMC Biol. 12:7. doi: 10.1186/1741-7007-12-7.

Mason B et al. 2012. Evidence for Multiple Phototransduction Pathways in a Reef-Building Coral. PLoS One. 7:e50371. http://dx.plos.org/10.1371/journal.pone.0050371.

Minh BQ, Nguyen MAT, von Haeseler A. 2013. Ultrafast approximation for phylogenetic bootstrap. Mol. Biol. Evol. 30:1188–1195. doi: 10.1093/molbev/mst024.

Moroz LL et al. 2014. The ctenophore genome and the evolutionary origins of neural systems. Nature. 510:109–114. doi: 10.1038/nature13400.

Nathans J, Hogness DS. 1983. Isolation, sequence analysis, and intron-exon arrangement of the gene encoding bovine rhodopsin. Cell. http://www.sciencedirect.com/science/article/pii/0092867483905378.

Nguyen L-T, L.-T. N, Schmidt HA, von Haeseler A, Minh BQ. 2014. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Mol. Biol. Evol. 32:268–274. doi: 10.1093/molbev/msu300.

Nilsson D-E. 2013. Eye evolution and its functional basis. Vis. Neurosci. 30:5–20. http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=23578808&retmode=ref&cmd=prlinks.

Nilsson D-E. 2005. Photoreceptor evolution: ancient siblings serve different tasks. Curr. Biol. 15:R94–6. doi: 10.1016/j.cub.2005.01.027.

Oakley TH, Speiser DI. 2015. How Complexity Originates: The Evolution of Animal Eyes. Annu. Rev. Ecol. Evol. Syst. 46:null. doi: 10.1146/annurev-ecolsys-110512-135907.

O'Tousa JE, Baehr W, Martin RL, Hirsh J, Pak WL. 1985. The Drosophila ninaE gene encodes an opsin. Cell. http://www.sciencedirect.com/science/article/pii/0092867485903435.

Paradis E, Claude J, Strimmer K. 2004. APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics. 20:289–290. doi: 10.1093/bioinformatics/btg412.

Passamaneck YJ, Furchheim N, Hejnol A, Martindale MQ, Lüter C. 2011. Ciliary photoreceptors in the cerebral eyes of a protostome larva. Evodevo. 2:6. doi: 10.1186/2041-9139-2-6.

Passamaneck YJ, Martindale MQ. 2013. Evidence for a phototransduction cascade in an early brachiopod embryo. Integr. Comp. Biol. 53:17–26. doi: 10.1093/icb/ict037.

Pervez MT et al. 2014. Evaluating the accuracy and efficiency of multiple sequence alignment methods. Evol. Bioinform. Online. 10:205–217. doi: 10.4137/EBO.S19199.

Pisani D et al. 2015. Genomic data do not support comb jellies as the sister group to all other animals. Proc. Natl. Acad. Sci. U. S. A. doi: 10.1073/pnas.1518127112.

Pisani D et al. 2016. Reply to Halanych et al.: Ctenophore misplacement is corroborated by independent datasets. Proc. Natl. Acad. Sci. U. S. A. 113:E948–9. doi: 10.1073/pnas.1525718113.

Plachetzki DC, Degnan BM, Oakley TH. 2007. The Origins of Novel Protein Interactions during Animal Opsin Evolution Snel, B, editor. PLoS One. 2:e1054. http://dx.plos.org/10.1371/journal.pone.0001054.t001.

Porter ML et al. 2012. Shedding new light on opsin evolution. Proc. R. Soc. Lond. B Biol. Sci. 279:3–14. http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=22012981&retmode=ref&cmd=prlinks.

Price MN, Dehal PS, Arkin AP, Others. 2010. FastTree 2--approximately maximum-likelihood trees for large alignments. PLoS One. 5:e9490.

http://dx.plos.org/10.1371/journal.pone.0009490.

Radu RA et al. 2008. Retinal pigment epithelium-retinal G protein receptor-opsin mediates light-dependent translocation of all-trans-retinyl esters for synthesis of visual chromophore in retinal pigment epithelial cells. J. Biol. Chem. 283:19730–19738. doi: 10.1074/jbc.M801288200.

Rambaut A. 2007. FigTree, a graphical viewer of phylogenetic trees. See http://tree. bio. ed. ac. uk/software/figtree.

Ramirez MD, Speiser DI, Pankey MS, Oakley TH. 2011. Understanding the dermal light sense in the context of integrative photoreceptor cell biology. Vis. Neurosci. 28:265–279. doi: 10.1017/S0952523811000150.

R Core Team. 2016. R: A Language and Environment for Statistical Computing. https://www.R-project.org.

Ryan JF et al. 2013. The genome of the ctenophore Mnemiopsis leidyi and its implications for cell type evolution. Science. 342:1242592. doi: 10.1126/science.1242592.

Schnitzler CE et al. 2012. Genomic organization, evolution, and expression of photoprotein and opsin genes in Mnemiopsis leidyi: a new view of ctenophore photocytes. BMC Biol. 10:107. doi: 10.1186/1741-7007-10-107.

Senthilan PR et al. 2012. Drosophila auditory organ genes and genetic hearing defects. Cell. 150:1042–1054. http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=22939627&retmode=ref&cmd=prlinks.

Shen WL et al. 2011. Function of Rhodopsin in Temperature Discrimination in Drosophila. Science. 331:1333–1336. http://www.sciencemag.org/cgi/doi/10.1126/science.1198904.

Simakov O et al. 2012. Insights into bilaterian evolution from three spiralian genomes. Nature. 1–6. doi: 10.1038/nature11696.

Simmons MP, Norton AP. 2014. Divergent maximum-likelihood-branch-support values for polytomies. Molecular phylogenetics and evolution. doi: 10.1016/j.ympev.2014.01.018.

Simmons MP, Randle CP. 2014. Disparate parametric branch-support values from ambiguous characters. Mol. Phylogenet. Evol. http://www.sciencedirect.com/science/article/pii/S1055790314001572.

Speiser DI et al. 2014. Using phylogenetically-informed annotation (PIA) to search for light-interacting genes in transcriptomes from non-model organisms. BMC Bioinformatics. 15:350. doi: 10.1186/s12859-014-0350-x.

Stöver BC, Müller KF. 2010. TreeGraph 2: combining and visualizing evidence from different phylogenetic analyses. BMC Bioinformatics. 11:7. doi: 10.1186/1471-2105-11-7.

Suga H, Schmid V, Gehring WJ. 2008. Evolution and functional diversity of jellyfish opsins. Curr. Biol. 18:51–55. doi: 10.1016/j.cub.2007.11.059.

Sun H, Gilbert DJ, Copeland NG, Jenkins NA, Nathans J. 1997. Peropsin, a novel visual pigment-like protein located in the apical microvilli of the retinal pigment epithelium. Proc. Natl. Acad. Sci. U. S. A. 94:9893–9898. http://www.ncbi.nlm.nih.gov/pubmed/9275222.

Swafford A. 2016. Supercuts. https://bitbucket.org/swafford/supercuts (Accessed April 28, 2016).

Terakita A. 2005. The opsins. Genome Biol. 6:213. doi: 10.1186/gb-2005-6-3-213.

Velarde RA, Sauer CD, O Walden KK, Fahrbach SE, Robertson HM. 2005. Pteropsin: A vertebrate-like non-visual opsin expressed in the honey bee brain. Insect Biochem. Mol. Biol. 35:1367–1377. http://linkinghub.elsevier.com/retrieve/pii/S0965174805001748.

Vopalensky P, Kozmik Z. 2009. Eye evolution: common use and independent recruitment of genetic components. Philos. Trans. R. Soc. Lond. B Biol. Sci. 364:2819–2832. doi: 10.1098/rstb.2009.0079.

Yoshida MA et al. 2015. Molecular Evidence for Convergence and Parallelism in Evolution of Complex Brains of Cephalopod Molluscs: Insights from Visual Systems. Integr. Comp. Biol. doi: 10.1093/icb/icv049.

Zhang H, Gao S, Lercher MJ, Hu S, Chen W-H. 2012. EvolView, an online tool for visualizing, annotating and managing phylogenetic trees. Nucleic Acids Res. 40:W569–72. doi: 10.1093/nar/gks576.