Molluscan Genomes Reveal Extensive Differences in Photopigment Evolution Across the Phylum

Kyle E. McElroy [1],*[,1] Jorge A. Audino [1,2] and Jeanne M. Serb [1]

[1]Ecology, Evolutionary, and Organismal Biology, Iowa State University, Ames, IA, USA
[2]Department of Zoology, University of São Paulo, São Paulo, Brazil

***Corresponding author:** E-mail: address: kmcelroy@iastate.edu.

Associate editor: Fabia Ursula Battistuzzi

Abstract

In animals, opsins and cryptochromes are major protein families that transduce light signals when bound to light-absorbing chromophores. Opsins are involved in various light-dependent processes, like vision, and have been co-opted for light-independent sensory modalities. Cryptochromes are important photoreceptors in animals, generally regulating circadian rhythm, they belong to a larger protein family with photolyases, which repair UV-induced DNA damage. Mollusks are great animals to explore questions about light sensing as eyes have evolved multiple times across, and within, taxonomic classes. We used molluscan genome assemblies from 80 species to predict protein sequences and examine gene family evolution using phylogenetic approaches. We found extensive opsin family expansion and contraction, particularly in bivalve xenopsins and gastropod G_o-opsins, while other opsins, like retinochrome, rarely duplicate. Bivalve and gastropod lineages exhibit fluctuations in opsin repertoire, with cephalopods having the fewest number of opsins and loss of at least 2 major opsin types. Interestingly, opsin expansions are not limited to eyed species, and the highest opsin content was seen in eyeless bivalves. The dynamic nature of opsin evolution is quite contrary to the general lack of diversification in mollusk cryptochromes, though some taxa, including cephalopods and terrestrial gastropods, have reduced repertoires of both protein families. We also found complete loss of opsins and cryptochromes in multiple, but not all, deep-sea species. These results help set the stage for connecting genomic changes, including opsin family expansion and contraction, with differences in environmental, and biological features across Mollusca.

Key words: **opsins, evolution, mollusks, cryptochromes.**

Introduction

Light is an important cue for many biological processes. Generalized photoreception is used for food production, entraining circadian rhythm, setting of photoperiodism, and regulating physiological processes, such as vasorelaxation and gonadal growth (Bradshaw and Holzapfel 2007; Nakane et al. 2010; Sikka et al. 2014; Yu and Fischer 2019; Hu et al. 2022). In animals, proteins from 2 major gene families bind to light-absorbing chromophores to form photopigments that transduce light signals: opsin and cryptochrome (CRYs). Opsin, a type of G-protein coupled receptor (GPCR) that forms a photopigment with retinal, can absorb light at specific wavelengths across the UV-visible spectrum into near infrared. Though opsins are a diverse family of GPCRs they have a highly conserved 7-transmembrane (TM) structure and lysine residue necessary for retinal binding, homologous to K296 in Bovine Rhodopsin (reviewed in Smith 2010). Opsins are classified based on the type of photoreceptors they were discovered in (e.g. rhabdomeric "r-opsins" and ciliary "c-opsins"), the G-protein they couple with (e.g. G_q vs. G_t), and phylogenetic relationship (e.g. "tetraopsins") (Shichida and Matsuyama 2009; Porter et al. 2011). The most well-studied opsins are those used in vision (referred to as "visual opsins"), such as c- and r-opsins, which are densely packed in membrane specializations of photoreceptor cells (Nilsson 2021). But opsin proteins are not just found in the eyes. Opsins are expressed in dermal melanophores and chromatophores (Ramirez and Oakley 2015), light organs (Tong et al. 2009), brains, and adrenal glands (Ohuchi et al. 2012) (Table 1). While most opsins are used for light-dependent functions, members of the opsin family have been co-opted for light-independent sensory modalities including taste (Leung et al. 2020), auditory (Senthilan et al. 2012), mechanoreception (Katana et al. 2019), and temperature reception (Shen et al. 2011) (reviewed in Leung and Montell 2017). Therefore, opsins are a remarkably diverse GPCR group with a myriad of sensory functions.

Even though visual opsins have been largely characterized in some animal lineages, such as vertebrates, we know little about opsin expression, function, and evolution in invertebrates, particularly for nonvisual opsins. Mollusks are a great group of animals to explore questions about the evolution of light sensing. Eyes have evolved multiple times across, and within, taxonomic classes (Serb and Eernisse 2008;

Received: June 29, 2023. **Revised:** October 14, 2023. **Accepted:** November 14, 2023
© The Author(s) 2023. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Open Access

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msad263/7512 by guest on 20 December 2023

Table 1 Summary of key information associated with opsin types and current unknowns

Opsin type	G-protein	Signaling cascade	Cell response	Stability	Photoreceptor cell type	Functions
r-opsin (canonical)	G_q	Phosphoinositol	Depolarize	Bistable	Rhabdomeric	Visual opsin mainly expressed in the eyes of protostomes, but also found in extraocular tissues and in some deuterostomes. Involved in phototaxis, circadian clock, and photoentrainment.
r-opsin (noncanonical)	Likely G_q	Unknown	Unknown	Unknown	Nonvisual cells but largely unknown	Functions largely unknown, but evidence of expression in nonvisual cells of planarians suggests a role in phototaxis.
c-opsin	G_t/G_i	Cyclic nucleotide	Hyperpolarize	Monostable	Ciliary	Visual opsin mainly expressed in the eyes of vertebrates, but also found in extraocular tissues and in some invertebrates.
Xenopsin	G_i	Cyclic nucleotide	Unknown[a]	Bistable	Ciliary/ Rhabdomeric	Poorly characterized with unknown functions in protostomes.
Cnidopsin	$G_s/G_q/G_c$	Cyclic nucleotide	Depolarize	Monostable	Ciliary and nonvisual cells	Visual opsin in cnidarian ocular organs, also mediates light-dependent reproductive control.
Retinochrome	Not coupled to G-proteins	Unknown	Unknown	Bistable	Rhabdomeric	Photoisomerization of all-*trans* retinal to the 11-*cis* form in mollusks. Function poorly characterized in other invertebrates.
RPE-retinal G-protein receptor (RGR)	Likely not coupled to G-proteins	Unknown	Unknown	Bistable	Pigment cells	Photoisomerization of all-*trans* retinal to the 11-*cis* form in vertebrates. Function poorly characterized in invertebrates.
Peropsin	G_i/G_s	Cyclic nucleotide	Unknown	Bistable	Visual and nonvisual cells	Likely function includes light-dependent regulation of retinal cycling in vertebrates and invertebrates, however not characterized yet.
Go-opsin	G_o	Cyclic nucleotide	Hyperpolarize	Bistable	Ciliary/ Rhabdomeric	Function poorly characterized with expression in ocular tissues of invertebrates.
Neuropsin (opsin-5)	G_i	Cyclic nucleotide	Hyperpolarize	Bistable	Ciliary and nonvisual cells	Light-dependent functions, including circadian photoentrainment, in ocular and extraocular tissues of vertebrates. Function poorly characterized in invertebrates.

References in supplementary table S1, Supplementary Material online
See also Fig. 1 in Vöcking et al. (2022) for additional details.
[a]Cell response is likely depolarization.

Audino et al. 2020). These organs represent a range of eye complexity, from a simple pit eye to elaborate structures such as the compound eyes analogous to those in arthropods, to the camera-type eyes, and eyes with mirror-based optics. Light sensing structures also vary across ontogeny, where most molluscan larvae have a pair of simple eyespots that are lost during metamorphosis (Raven 1966). Even without eyes, nearly all adult mollusks have a photosensitive mantle (Kennedy 1960), the membranous organ that lines both shell valves (Kennedy 1960; Carter et al. 2012). In gastropods and bivalves, these epithelial tissues of the mantle can develop into sensory extensions as papillate, lobes, or tentacles that perform a variety of modalities, including chemo-, photo-, and mechano-sense (e.g. Audino and Marian 2016; Audino and Marian 2020). Likely, the ancestral mollusk had 7 types of opsins: canonical r-opsins, noncanonical r-opsins, xenopsins, retinochromes, G_o-opsins, and neuropsins (Ramirez et al. 2016) (Table 1). However, outside of the canonical r-opsins used for vision, little is known about the function or spatial expression of these other opsin clades (but see Kingston et al. 2015).

In contrast to opsins, the second photopigment-forming protein family, the cryptochromes, are restricted to blue-light sensitivity. Cryptochromes are a class of flavoproteins that bind to a flavin cofactor, flavin adenine dinucleotide (FAD). Cryptochromes primarily entrain the circadian clock and are highly conserved at the level of sequence and function (reviewed in Öztürk et al. 2007; Öztürk 2017). Cryptochromes belong to a larger protein family with the photolyases (PLs) that share sensitivity to blue light and are found ubiquitously across cellular life (Lin and Todo 2005; Mei and Dvornyk 2015; Öztürk 2017). While PLs repair UV-damaged DNA, cryptochromes have evolved other light directed-functions in plants and animals (see Table 2 for summary of CRY-photolyase (PL) functions). Regardless of

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msad263/7457512 by guest on 20 December 2023

Table 2 Summary of taxonomic distribution and known key functions for cryptochrome (CRY)/ PL protein families

CRY-PL family	Taxa	Functions
CPD-I	M, F, P, E, B, A	UV-light induced CPD DNA-damage repair
CPD-II (phr)	M, F, P, E, B, A	UV-light induced CPD DNA-damage repair
CPD-III	M, P, E, B, A	UV-light induced CPD DNA-damage repair
PCRY	P	Light-sensitive plant cryptochrome, involved with circadian clock regulation and various photomorphogenic roles
PCRY-like	M, P, E	Largely unknown
CRY-DASH	M, F, P, E, B, A	UV-light induced ssDNA-damage repair. Also has circadian clock functions; regarded as "intermediate" or "link" between DNA-damage repair PLs and cryptochromes
PPL	P, E	UV-light induced CPD DNA-damage repair
DCRY (CRY-I)	M	Light-sensitive animal cryptochrome. Photoreceptor activated by blue light, resets circadian clock.
CRY4	M	Magnetoreception in migratory vertebrates
6-4 PL	M, F, P, E, B, A	UV-light induced 6-4 DNA-damage repair
MCRY (CRY-II)	M	Light-insensitive animal cryptochrome, involved with circadian clock regulation.

References in supplementary table S2, Supplementary Material online.
Metazoa (M), Fungi (F), Viridiplantae (P), other unicellular eukaryotes (E), Bacteria (B), Archaea (A), pyrimidine-pyrimodone (6-4), cyclobutane-pyrimidine dimers (CPD), single-stranded DNA (ssDNA).

taxonomic lineage or habitat, all cryptochrome-based photopigments absorb light in the near UV to blue range (λ_{max} between 370–440 nm; Sancar 2000).

Here, we conducted a phylum-wide characterization of opsin and CRY-PL genomic content based on de novo gene prediction for these 2 protein families. We use these data to explore how different photopigment gene families evolve in this diverse phylum. Recent work characterized opsin genomic content to describe ancestral states for opsin number in different mollusk classes and contrast the rich molluscan opsin repertoire to other, opsin-poor, lophotrochozoan phyla, like Platyhelminthes (De Vivo et al. 2023). Our work adds to these findings by sampling more molluscan taxa (80 vs. 22 species), enabling a deep account of cross-taxa differences and similarities in opsin evolution. Furthermore, we evaluated differences in evolutionary patterns across opsin clades, which may provide insight into functional properties of the poorly understood opsin groups common to mollusks. Using a phylogenetic framework, we were able to identify dramatic differences in evolutionary dynamics within opsins and between 2 light-sensing gene families (opsin vs. cryptochrome). Several of the opsin gene lineages repeatedly were duplicated in a taxon-specific manner and were in stark contrast to other genes that were evolutionarily static. Because our results only reflect genomic presence, with many sequences identified as fragments, we stress that our findings do not have direct functional interpretations and instead should be viewed as a broad description of opsin and cryptochrome genomic content in mollusks that we hope will prompt future investigation into the causes for the huge variation in opsin abundance we found across Mollusca.

Materials and Methods

We downloaded genome assemblies for 83 molluscan species from NCBI GenBank (where available, see supplementary table S3, Supplementary Material online for exceptions), including 39 bivalves, 8 cephalopods, 35 gastropods, and 1 polyplacophoran—the chiton, *Acanthopleura granulata*. We collected, as summary statistics for these genome assemblies, the number of scaffolds, scaffold N50, and total assembly sequence length using bbmap version 37.36 statswrapper.sh (https://github.com/BioInfoTools/BBMap) (supplementary table S3, Supplementary Material online).

Molluscan Species Tree

To create a phylogenetic framework to examine opsin and cryptochrome evolution, we generated a molluscan species tree based on the genome assemblies. We identified orthologs for tree inference using the Benchmarking Universal Single-Copy Orthologs (BUSCO) software (Manni et al. 2021). The BUSCO single-copy genes exhibit high sequence conservation and, with nearly 1,000 sequences, represent a large dataset. We ran BUSCO (version v5.2.2) using the obd10 metazoan and molluscan databases on each of the 83 assemblies. We found that the "complete" scores were generally higher with the metazoan database compared with the molluscan database (supplementary fig. S1, Supplementary Material online) and therefore used the metazoan results for our phylogenetic analysis. To remove assemblies of low quality for gene discovery, we plotted (R version 3.6.2; ggplot2 version 3.3.5) the BUSCO % complete scores as a histogram to determine natural breaks in results (supplementary fig. S1, Supplementary Material online). Three assemblies less than 73% complete were removed from subsequent analyses (*Limnoperna fortunei*: GCA_003130415.1; *Panopea generosa*: GCA_902825435.1; *Pinna nobilis*: GCA_016161895.1).

From the remaining 80 genomes (1 polyplacophoran, 8 cephalopods, 35 cephalopods, and 36 bivalves), we found 899 complete single copy BUSCO genes present in at least 75% (60 out of 80) genomes and these were used to estimate the species tree. Genes were aligned with MAFFT (version 7.453; Kuraku et al. 2013) using the "auto" parameter. Alignments were quality trimmed using trimal (version 1.4.rev22; Capella-Gutiérrez et al. 2009) with the "-automated1" parameter. We ran IQ-TREE2 (version 2.1.3; Minh et al. 2020) on all 899 protein sequence alignments to search for the best-fitting model of protein sequence evolution for each gene with ModelFinder

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msad263/7457512 by guest on 20 December 2023

(Kalyaanamoorthy et al. 2017). We then used catsequences (https://zenodo.org/record/4409153#.Y8gnmXbMJPY) to concatenate the 899 trimmed amino acid alignments and combined the accompanying partition configuration file with the ModelFinder results as input for the partitioned analysis in IQ-TREE2 (Chernomor et al. 2016). Branch support of the maximum-likelihood results was determined with 1,000 replicates for each: ultrafast bootstrap, SH approximate likelihood ratio test, and an approximate Bayes test (parameter: –alrt 1000 -B 1000 –abayes) (Anisimova et al. 2011; Hoang et al. 2018).

Reference Opsin set for Molluscan Gene Annotation

Genome annotation strategy and quality may differ widely across genomes, which can lead to the false appearance of "lineage-specific" genes (Weisman et al. 2022), thus, weakening comparative genomic analyses like gene family evolution. Furthermore, many genome assemblies lack publicly available annotations entirely, including 18 of the 80 genomes analyzed here. For these reasons, we opted for de novo prediction of opsin and CRY-PL genes to have a consistent process in generating datasets (i.e. protein sequences) for phylogenetic analysis. We applied the targeted gene annotator pipeline BITACORA (version 1.3; Vizueta et al. 2020), which incorporates the homology-based gene predictor, Gene Model Mapper (GeMoMa) (Keilwagen et al. 2016, 2018) and leverages its parameter for annotating specific genes (vs. genome-wide annotation) to predict opsin sequences. Because BITACORA uses a reference set of protein sequences and Hidden Markov Models (HMMs) as queries in tblastn and hmmer searches, high-quality input data is needed for accurate gene prediction. We developed a high-quality reference set of molluscan opsin sequences from 9 molluscan species based on the following criteria: (i) identification of 7 TM domains, (ii) presence of lysine residue for chromophore binding, and (iii) complete coding sequence. Protein sequences were downloaded from genome annotation of 9 mollusk species representing a diverse sampling of molluscan lineages (1 cephalopod, 4 gastropods, and 4 bivalves; bolded in supplementary table S3, Supplementary Material online). We used blastp (BLAST+ version 2.6.0; Camacho et al. 2009) with an e-value of 1e-20 to query opsins from the metazoan-wide opsin gene collection in Ramirez et al. (2016) against each of 9 molluscan protein sets. Next, we used HMMER version 3 .1b1 hmmscan (hmmer.org) to search the Ramirez et al. (2016) hits against the protein families (Pfam) database version 35.0 (Mistry et al. 2021) and retained sequences with hits to the 7tm_1 (PF00001.24) domain "7-TM receptor (rhodopsin family)." We then visually screened sequences for the conserved lysine residue (K296 in Bovine rhodopsin) necessary for retinal binding in MEGA X (Kumar et al. 2018). Finally, we manually curated gene models using GeneWise (Birney et al. 2004; Madeira et al. 2019) based on tblastn hit coordinates in each species' genome assembly. For all 9 species, we ensured that start and stop codons were present in an open reading frame of concatenated exons.

We found that the protein names from public databases NCBI and PROSITE were often not specific enough (e.g. "rhodopsin-like") for our purposes and several of the species' datasets lacked functional annotations. Therefore, we phylogenetically analyzed this panel of molluscan opsins to assign specific clade-level names so that we could use these sequences downstream in our classification of de novo predicted opsin sequences. To classify the opsin sequences in our molluscan reference panel, we used MAFFT to align these putative opsin sequences to known molluscan opsins (Ramirez et al. 2016) and the outgroup sequences used in Vöcking et al. (2017), which includes opsin-like sequences from the placozoan, *Trichoplax adhaerens*, known as "placopsins," melatonin receptors, along with other GPCRs such as adrenergic, dopamine, and octopamine receptors (also used in Döring et al. 2020). Next, the sequence alignment was manually trimmed, and a maximum likelihood tree was generated with IQ-TREE2 (—alrt 1000 -B 1000 –abayes). We then assigned the following classifications to each sequence in our reference set based on their phylogenetic placement: r-opsin, nonconical r-opsin, xenopsin, retinochrome/RGR/peropsin, G_o-opsin, neuropsin. We added to the reference set 2 neuropsin sequences from the slug *Ambigolimax valentianus* and a xenopsin from the squid *Idiosepius paradoxus* to complement missing/truncated sequences from heterobranch gastropods and cephalopods (see supplementary table S4, Supplementary Material online for list of sequences in opsin reference set).

We were able to classify 107 of the 109 molluscan opsin reference sequences as canonical and noncanonical r-opsins (n = 11, n = 17, respectively), neuropsin (n = 12), G_o-opsin (n = 11), xenopsin (n = 40), peropsin (n = 5), and retinochrome (n = 10) types (supplementary fig. S2 and supplementary table S4, Supplementary Material online). The phylogenetic distinction of "noncanonical" vs. "canonical" r-opsin follows Ramirez et al. (2016). All opsin type clades had high support values and a topology similar to previous studies (e.g. (Rawlinson et al. 2019). Unlike Vöcking et al. (2017) and Döring et al. (2020), in our results, the outgroup sequences form a strongly supported monophyletic clade and included placopsins, versus the placopsins as the sister lineage to opsins, then melatonin receptors and the other GPCRs further outside. The xenopsins were divided into 2 major clades that we assigned "a" and "b" based on the literature (supplementary fig. S2, Supplementary Material online) (Rawlinson et al. 2019; Döring et al. 2020). We also found that a group of 7 xenopsins from the bivalve *Sinovacula constricta* formed a distinct clade outside the "a" and "b" groups (90.1/81 branch support), which were named "Scon-opnGxS#" (supplementary fig. S2, Supplementary Material online). The 2 sequences in our reference set that did not clearly belong to a specific opsin clade were from *Lottia gigantea* (XP_009051446.1, labeled as Lgig-opnUNK) and *Sinovacula constricta* (evm.model.Chr17.756, labeled Scon-opnUNK). This reference gene set covers the 3 main classes of mollusks analyzed here and, for bivalves and gastropods, includes 4 distinct orders, and appears to account for all

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msad263/7457512 by guest on 20 December 2023

opsin groups found in Mollusca. Therefore, these sequences represent a powerful dataset for both homology-based opsin discovery in molluscan genomes and the subsequent phylogenetic classification of opsin sequences.

Opsin Search in Molluscan Genome Assemblies and Phylogenetic Analysis

The reference set of molluscan opsins was used as input for BITACORA to generate opsin gene models from the 80 molluscan genome assemblies represented in our species tree. First, we generated an HMM profile of the reference opsins from an alignment of the 109 reference opsins using HMMR. Then, we used the "runBITACORA_genome_mode.sh" script to run BITACORA in genome mode (i.e. no input genome annotation) for de novo gene prediction using GeMoMa (parameter GEMOMA = T) based on tblastn results. We screened all the resulting protein predictions for the presence of a K296 retinal binding site (supplementary table S5, Supplementary Material online for BITACORA opsin gene counts).

We produced opsin phylogenies that included: (i) exclusively mollusk sequences identified in this study, along with outgroups (Vöcking et al. 2017); (ii) a more diverse panel of opsins, including opsin sequences from the light-interacting toolkit (Speiser et al. 2014) along with additional xenopsin and cnidopsin sequences from (Gühmann et al. 2022); and (iii) an extensive opsin tree that also includes chaopsin, ctenopsin, bathyopsin sequences (defined in Ramirez et al. 2016) and anthozoan specific opsins (ASO-I, ASO-II) from Gornik et al. (2020) (see supplementary table S6, Supplementary Material online for details). These additional opsin datasets were included to ensure that we were not inadvertently forcing the predicted molluscan sequences into certain groups, as our reference opsin panel included 2 "unknown" opsins after our initial molluscan-only phylogenetic analysis. Furthermore, additional nonmolluscan sequences could help distinguish divisions within opsin types such as the "a" vs. "b" clades of xenopsin, and clarify how the molluscan peropsins and retinochrome relate to each other. In each phylogenetic analysis, we aligned the amino acid sequences with MAFFT using the E-INS-i iterative refinement method and trimmed the sequences with trimal using "gappyout" mode. Finally, we generated a maximum likelihood tree with IQ-TREE2 using ModelFinder to determine the best-fit protein substitution model according to Bayesian information criterion scores (LG + F + R10 selected each time). Ultrafast bootstrap (1,000 replicates), SH approximate likelihood ratio test (1,000 replicates), and approximate Bayes test were used to evaluate branch support.

To characterize patterns of gain and loss of opsins across the molluscan phylogeny, we used GeneRax (version 2.0.4; Morel et al. 2020) to generate reconciled gene trees for each major opsin type with the BUSCO-based species tree from this study (supplementary fig. S3, Supplementary Material online). We used the UndatedDL probabilistic model for computing the reconciliation likelihood and the SPR tree search mode. We used ThirdKind to visualize the reconciliation results (Penel et al. 2022).

Cryptochrome Identification and Phylogenetic Analysis

We employed a similar strategy for identifying cryptochrome sequences as for opsins. We queried the same 9 genome annotations for potential cryptochrome sequences by blastp hits to cryptochrome amino acid sequences from Rivera et al. (2012). From the blast hits, we ran hmmscan against the pfam database and retained sequences with FAD binding domain of DNA PL (PF03441.17) and DNA PL (PF00875.21). We then combined our hits with cryptochrome amino acid sequences used in Deppisch et al. (2022) as landmarks to classify the sequences for our molluscan reference cryptochrome set. We aligned sequences with MAFFT (LINSI), trimmed the alignment with trimal (automated1), and produced a maximum-likelihood tree with IQ-TREE2 (LG + R5). We then classified the mollusk sequences as phr (*photorepair* gene in *Drosophila*, belonging to CDP-II), CRY-DASH (*Drosophila*, *Arabidopsis*, *Synechocystis*, human cryptochromes), (6-4)PLs, CRY-I (animal cryptochrome-1, the "mammal"-type, i.e. MCRY), and CRY-II (animal cryptochrome-2, the "Drosophila"-type, i.e. DCRY) based on their phylogenetic groupings (supplementary fig. S4, Supplementary Material online). For the CRY-PL molluscan reference sequences, we classified all 43 sequences into CDP-II (n = 9), CRY-DASH (n = 8), 6-4PL (n = 6), CRY-I (n = 9), CRY-II (n = 9) from the same 9 mollusk species (summarized in supplementary table S7, Supplementary Material online). We used the 43 classified sequences as a database for BITACORA to search for cryptochromes in the rest of the 80 molluscan genome assemblies. The predicted genes from BITACORA were then screened against the pfam database with hmmscan for FAD binding and DNA PL domains (see supplementary table S5, Supplementary Material online for BITACORA CRY-PL counts). As with the opsins, we produced an initial CRY-PL tree and searched for possible "missing" sequences first with blastp against the BITACORA output in case our filtering step removed a positive CRY-PL sequence and then using tblastn and GeneWise to generate protein models (16 CRY-PL sequences total, 51 opsins). We also removed apparent contaminant sequences that grouped outside of the mollusk clades. These 4 sequences were located on short scaffolds in their respective genome assemblies and aligned to bacterial sequences via blastp. Our final CRY-PL phylogeny was produced with IQ-TREE2 (LG + R8) and included amino acid sequenced from Gornik et al. (2020), the LIT (cry2_default_clock.fas from LIT_1.1 included with PIA2 at https://github.com/MartinGuehmann/PIA2), and landmark sequences used in (Deppisch et al. 2022) to aid in broader CRY-PL classification.

Photic Environment and Eye Type

To complement our investigation with general information on molluscan ecology and visual systems, we gathered data from the literature for all 80 species, focusing on the presence

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msad263/7457512 by guest on 20 December 2023

of eyes in the adult stage. Even though most mollusks have a photosensitive mantle, eyeless species are defined by the absence of a structure capable of creating images or detecting light direction. We also gathered information for eye type, optical components, environment, and aquatic habitat depth. All traits, states, and respective references are listed in supplementary table S8, Supplementary Material online.

Results and Discussion

Identification and Classification of Mollusk Opsins From Genome Assemblies

Across the phylum, mollusk genomes contain opsins from as many as 7 distinct clades but lack c-opsins and cnidopsins (Fig. 1). We were able to phylogenetically place 1,174 out of 1,196 predicted opsin sequences into 1 of these 7 opsin groups (canonical or noncanonical r-opsins, neuropsin, G_o-opsin, xenopsin, peropsin, and retinochrome) with high support values (supplementary fig. S5, Supplementary Material online). These classifications held whether we produced a mollusk-only tree (supplementary fig. S5, Supplementary Material online) or included sequences from other taxa (e.g. deuterostome, arthropod, annelid RGR/peropsins; details in supplementary table S9, Supplementary Material online) (supplementary fig. S6, Supplementary Material online). In fact, the phylogeny that also included c-opsins, cnidopsins, and non-molluscan xenopsins, had clearer separation of "a" and

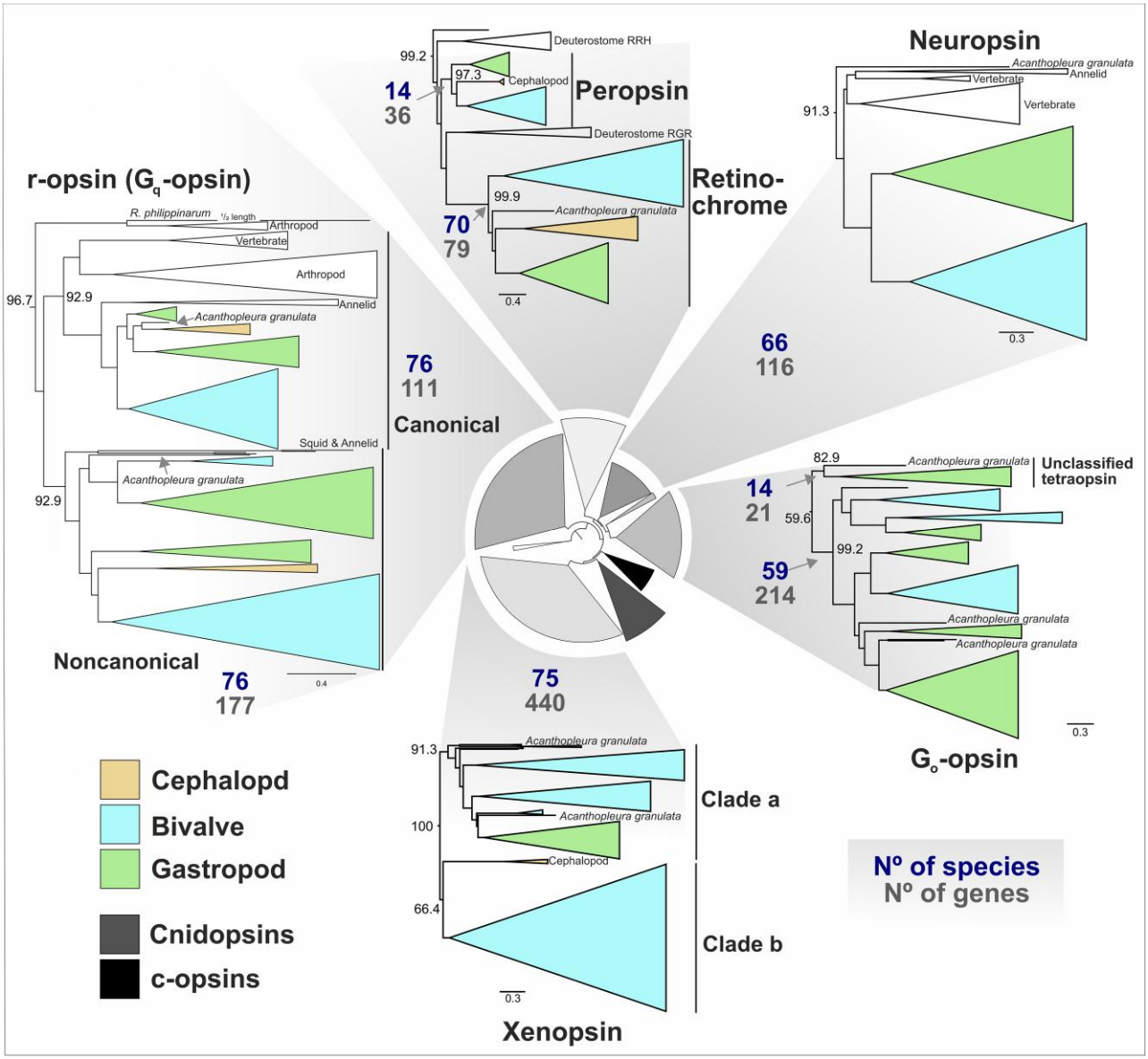

Fig. 1. Opsin clades differ in total size and number of species represented. Subtrees of mollusk containing opsin clades displayed from the IQ-TREE2 ML opsin phylogeny (LG + F + R10 amino acid substitution), shown at center. For each of 7 opsin types found in mollusks (plus the "Unclassified tetraopsin group), the total number of sequences in the tree (N° of genes) and species (N° species) with that opsin are displayed. Clades are collapsed according to taxonomic clades (e.g. bivalves) and color coded by mollusk class. UF-bootstrap support values shown at base of named of opsin clades. Support values for inner circular tree are shown in supplementary fig. S12a, Supplementary Material online and mollusk-specific clades in supplementary fig. S6, Supplementary Material online; full tree with support values is in supplementary fig. S7, Supplementary Material online.

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msad263/7457512 by guest on 20 December 2023

"b" xenopsin subclades (e.g. as seen in Rawlinson et al. 2019). In the mollusk-only tree (supplementary fig. S5, Supplementary Material online), a group of sequences from heteroconch bivalves forms an additional xenopsin clade like the sequences from *Sinovacula constricta* in the reference opsin panel (supplementary fig. S2, Supplementary Material online). Those heteroconch-specific sequences fall under clade "b" in the more diverse opsin tree (see "Scon-opnGxS" sequences in supplementary fig. S7, Supplementary Material online). Our findings do not reveal any other major groups of opsins in mollusks that were not already identified from gene expression data (Ramirez et al. 2016).

Of the remaining 22 gene sequences that were not phylogenetically placed into 1 of the 7 main opsin classes, 20 sequences formed a distinct clade of "unclassified tetraopsins" (Fig. 1, supplementary fig. S4, Supplementary Material online). The unclassified tetraopsins are typically single copy, restricted to gastropods and the lone chiton representative (Fig. 1), and are the sister group to either neuropsin (supplementary fig. S5, Supplementary Material online), G_o-opsin (supplementary figs. S6, and S8 Supplementary Material online), or G_o-opsin + neuropsin (supplementary fig. S9, Supplementary Material online). So, it is not clear whether this clade of opsins belongs to neuropsin, G_o-opsin, or forms its own group outside of them. Given that no nonmolluscan sequences in our extended phylogenies grouped with these 22 sequences it seems more likely to neuropsin or G_o-opsin, as opposed to a clade that split before the neuropsin-G_o split. The 2 remaining unclassified opsins are from the chiton *Acanthopleura granulata* (supplementary fig. S5, Supplementary Material online).

To classify these chiton opsins, conducted a phylogenetic analysis that included the reference molluscan opsins, opsins from the previously extended phylogeny (e.g. c-opsins and cnidopsins) and added ctenophore-specific "ctenopsins", anthozoan-specific opsins, bathyopsins and chaopsins (defined in Ramirez et al. 2016) and produced a new maximum likelihood tree (see supplementary table S9, Supplementary Material online for dataset details). We found that the 2 chiton sequences grouped with the bathyopsins (supplementary fig. S9, Supplementary Material online), a small clade reported as the sister clade to c-opsins (Ramirez et al. 2016; Vöcking et al. 2017; Rawlinson et al. 2019). This topology was consistent when the total molluscan BITACORA opsin dataset was included, as well (supplementary fig. S8, Supplementary Material online). Bathyopsins are distinct from c-opsins because they include genes from echinoderms and brachiopods, thus representing a clade that would have predated the protostome-deuterostome split (Ramirez et al. 2016). While others (De Vivo et al. 2023) have recently reported the same chiton sequences as a c-opsin, when we compared key functional motifs between these chiton sequences to c-opsins we found the tripeptide sequence in the fourth cytoplasmic loop important for specific G-protein binding varies. Vertebrate c-opsins use the tripeptide sequence NKQ, located in the fourth extracellular loop (EC4), for G_t binding (Marin et al. 2000). Most of the c-opsins we analyzed, including many outside of the vertebrate c-opsins, have NKQ (Fig. 2a). The 2 chiton sequences have NSR and NST, indicating that if they function as photoreceptors, it is unlikely that they drive similar phototransduction pathways, if at all, to the c-opsins. These sequences do demonstrate, though, that c-opsin-like sequences were present in early molluscan evolution and subsequently lost for most of the phylum.

While most bivalves, gastropods, and the chiton *Acanthopleura granulata* have at least 1 representative sequence for each of 7 opsin types (Fig. 1 and 3), cephalopods have the fewest number of opsins and have lost multiple opsin types. Genomes of octopus and squid cephalopods typically have 5 opsins (canonical and noncanonical r-opsins, xenopsin, and 2 sequences from the RGR/retinochrome clade), while *Nautilus* only have 3 opsins (canonical and noncanonical r-opsins, and retinochrome), indicating that neuropsin and G_o-opsin were likely lost in the cephalopod ancestor (Fig. 1). The most opsin-rich genomes, based on our analysis, are the small freshwater *Dreissena* zebra and quagga mussels (Myida: Dreissenidiae) with 54 and 63 opsins, respectively. Such high numbers of opsins are not unprecedented, as dozens of opsins have been identified from genomes of the crustacean *Daphnia* (n = 46) (Colbourne et al. 2011) and dragonflies (n = 15 to 33) (Futahashi et al. 2015; Brandon et al. 2017).

Dramatic Differences in Evolutionary Dynamics Across Opsin Type and Mollusk Species

To understand the broad differences in opsin type number among lineages (Fig. 3), we next examined the evolutionary dynamics of opsin gene family expansion by comparing species tree topology with and gene trees. We used the species tree derived from 899 BUSCO protein sequences. The resulting BUSCO-based species tree was largely consistent with recent molecular phylogenies (Kocot et al. 2011; Smith et al. 2011; Combosch et al. 2017; Cunha and Giribet 2019) and branch support was high (ultrafast bootstrap/approximate Bayes test/SH approximate likelihood ratio test) in all but 1 location: a bivalve clade with Adapedonta as the sister group to Myida-Venerida (49/1/71) (supplementary fig. S3, Supplementary Material online). We then generated gene trees for each major opsin type defined in Fig. 1. Species and gene trees were compared under reconciled tree analyses to characterize gain and loss patterns for each major opsin type across the molluscan phylogeny. We found that opsin types differed greatly in their size relative to one another, with some families having more of a "static" evolutionary history versus other families with "dynamic" changes, and some opsin types vary in size within particular taxonomic lineages (Table 3). In mollusks, the "static" opsin families include the canonical and noncanonical r-opsins, retinochrome, peropsin, and, to a lesser extent, neuropsin. The G_o-opsins and xenopsins

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msad263/7457512 by guest on 20 December 2023

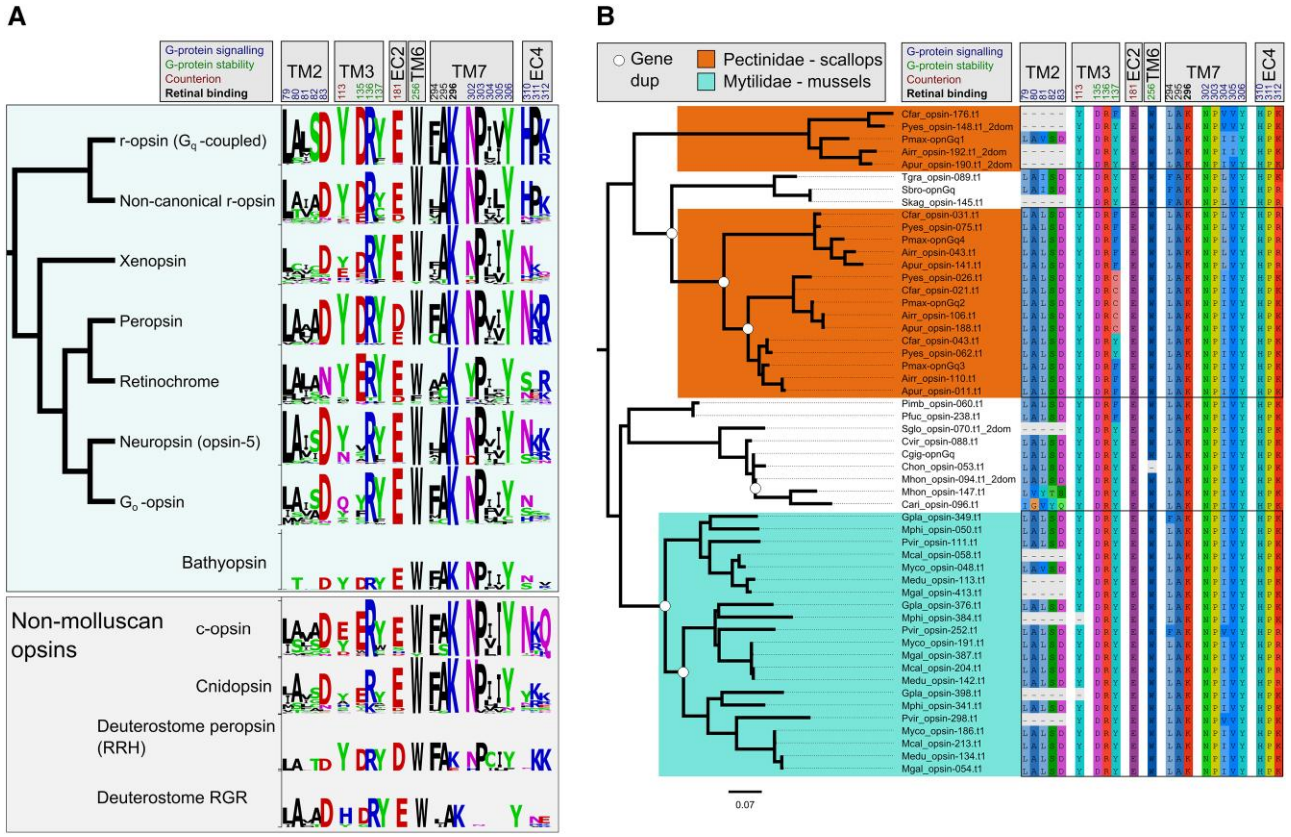

Fig. 2. Important amino acid sites for opsin structure and function. a) Molluscan opsins sequence logos for important motifs. Logos generated with WebLogo3 (Schneider and Stephens 1990; Crooks et al. 2004). Tree of molluscan opsin relationships based on Fig. 1, supplementary figs. S5 and S6, Supplementary Material online. Additional nonmolluscan opsin logos generated from sequences used in supplementary fig. S6, Supplementary Material online and bathyopsins in supplementary figs. S7 and S8, Supplementary Material online. Heights reflect ranked abundance of amino acids at each site. b) Duplications in Pectinidae and Mytilidae canonical r-opsins (G_q-coupled) differ in E/DRY motif. Pteriomorphia portion of GeneRax reconciled canonical r-opsin tree. Orange and teal shading and boxes highlight duplication in Pectinidae and Mytilidae, respectively. The E/DRY motif in TM3 varies across G_q-opsin duplicates in Pectinidae [DRF (9/20), DRY (6/20), DRC (5/20)] vs. DRY in all (21/21) Mytilidae G_q-opsins, and nearly every other molluscan G_q-opsin. The NPxxY motif in TM7 also varies more in Pectinidae [NPVVY (2/20), NPJIV (2/20), NPLVY (5/20), NPJVY (10/20)] than Mytilidae [NPJVY (2/21), NPVVY (19/21)]. In a) and b), positions included above are based on Bovine Rhodopsin (NP_001014890.1) numbering. At top are locations relative to the protein structure (TM = transmembrane, EC = extracellular loop). Color coding of residue positions reflects function to the opsin protein, broadly based on literature from GPCRs. The retinal binding site K296 and counterion positions are critical for photopigment formation.

are the "dynamic" opsins families, with large changes in gene abundance across the phylum. A potential limitation to these results is that incomplete lineage sorting (ILS) can limit the performance of gene tree reconciliation analyses, including GeneRax, which can handle moderate but not high degrees of ILS (Morel et al. 2020). At present, little is known about ILS in molluscan genome evolution and evidence for broad effects of ILS in mollusks is lacking; therefore, we do not necessarily expect ILS to influence our results in a meaningful way. However, with more molluscan genomes becoming available, characterizing the extent of ILS at varying depths of divergence in mollusks will be a valuable avenue of research.

The GeneRax reconciliation results indicate that most of the opsins identified from molluscan genomes are the result of speciation events rather than duplications (Table 3). In part, this result reflects the "stasis" of some opsin clades and that most duplications occur at nodes on our species tree (speciation events), as opposed to tips. Over 90% of sequences in the canonical r-opsin, peropsin, and retinochrome clades are from speciation events (Table 3). This result highlights the extreme rarity of duplications of these opsins to be retained. The noncanonical r-opsins, neuropsins, and G_o-opsins had speciation events accounting for 78% to 80% of the sequences in our reconciled trees (Table 3). Xenopsins were by far the most dynamic group of opsins, with only 560 of the sequences in our reconciled tree coming from speciation events (Table 3). We illustrate 2 extreme cases of xenopsin vs. retinochrome from the marine bivalve species in Pteriomorphia in Fig. 4. Aside from a single loss in the blood clam, *Scapharca kagoshimensis*, retinochrome is maintained at speciation events as a single-copy gene. In contrast, xenopsin duplications and losses occur regularly throughout the Pteriomorphia (discussed below).

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msad263/7457512 by guest on 20 December 2023

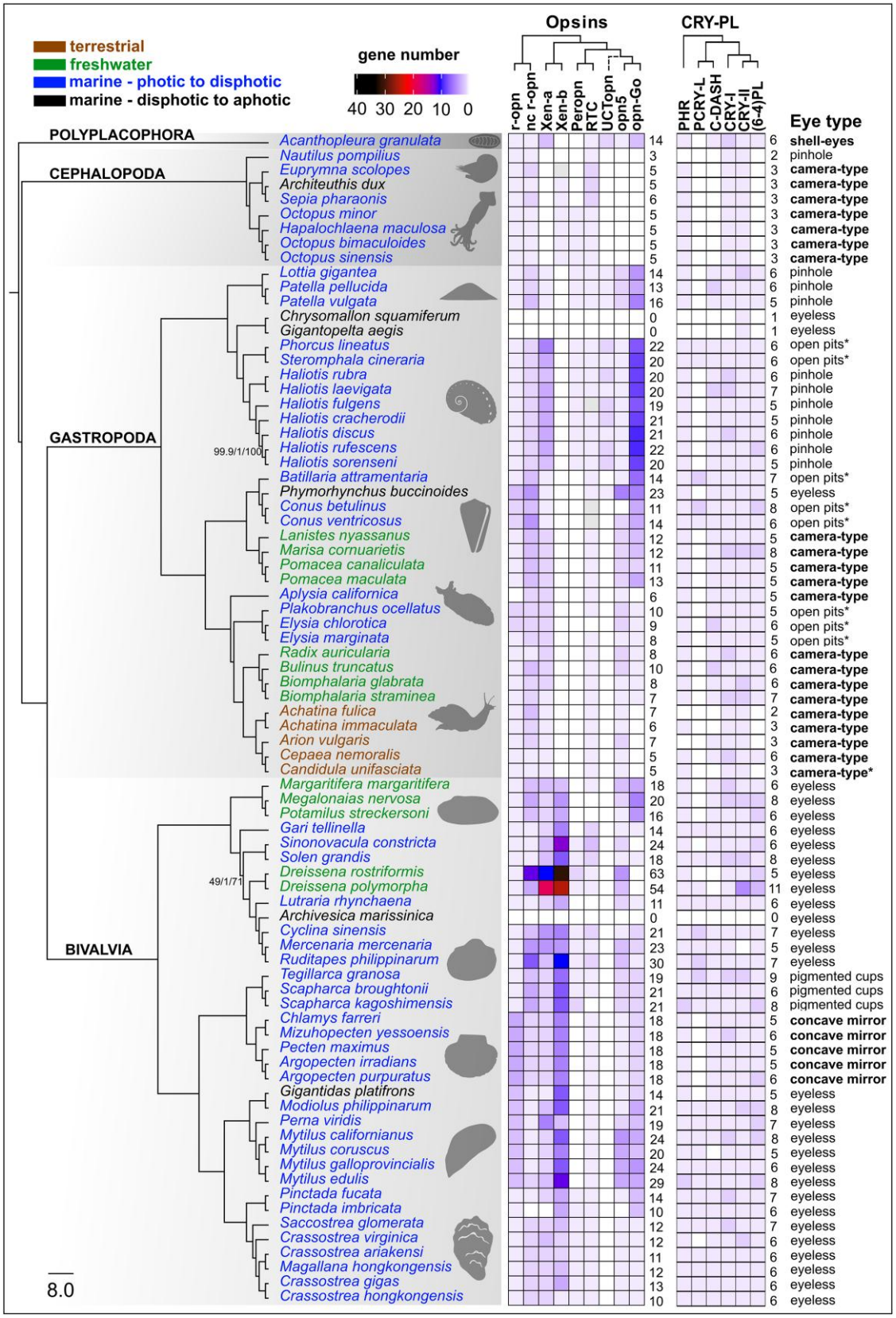

Fig. 3. Summary of molluscan opsin and cryptochrome genomic content from 80 species. Maximum likelihood species tree generated with IQ-TREE2 based on a partitioned amino acid supermatrix from 899 complete single copy BUSCO sequences (metazoan_obd10) recovered from at least 60/80 species. Branch values are SH-aLRT % support (with 1,000 replicates)/aBayes probability/UFBoot support % (with 1,000 replicates). Branch supports shown only where any of the 3 values is less than 100%. Species tree rooted with the polyplacophoran chiton.

(continued)

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msad263/7457512 by guest on 20 December 2023

Table 3 Summary of "events" from GeneRax reconciliation of opsin families

Event	r-opsin	Noncanonical r-opsin	Xenopsin	Peropsin	Retinochrome	Neuropsin	Go-opsin
S	101	142	263	33	74	91	172
L	10	35	145	10	17	28	73
D	11	35	175	2	4	24	48
Total events	102	142	293	25	61	87	147
S%	0.90	0.80	0.60	0.94	0.95	0.79	0.78
D%	0.10	0.20	0.40	0.06	0.05	0.21	0.22

S, speciation: L, loss; D, duplication; T, total. Rows below the double line are the proportion of total events from S and D.

Lineage-specific opsin expansions have been seen in a variety of taxa, but these duplications are generally limited to one type of opsin, e.g. c-opsins in vertebrates (e.g. Borges et al. 2015, but see Beaudry et al. 2017 as relating to whole genome duplication), and r-opsins in insects (e.g. Feuda et al. 2021). Major differences in opsin content across cnidarians are accounted for by ASO-II opsins and cnidopsins, the latter being the only opsin type in Medusozoa (Gornik et al. 2020). Mollusks, due to lineage-specific expansions, exhibit huge variation for both xenopsin and G_o-opsin abundance. While neuropsin and r-opsin tend to be more evolutionarily "static" in mollusks, we found some cases of expansion including noncanonical r-opsins in heteroconch bivalves and neuropsins in mytilids (Fig. 3). These results paint mollusks as having among the most, if not *the* most, variable opsin repertoires across the metazoan phyla.

Retinochrome and Peropsin are Rarely Duplicated in Mollusks

Retinochrome was among the most consistently recovered opsin, and most consistently single copy (Fig. 1, Fig. 4b). Unlike most opsins, retinochrome does not initiate phototransduction but instead functions in the visual cycle as a photoisomerase for converting all-*trans* retinal back into 11-*cis* retinal (Terakita 2005; Terakita and Nagata 2014). Originally characterized in squid (Hara and Hara 1965, 1967), retinochrome was eventually identified across Mollusca from transcriptomes (Ramirez et al. 2016) and more recently in other Lophotrocozoa phyla, including annelids and platyhelminthes (Vöcking et al. 2021; Gühmann et al. 2022). Given its critical role in the visual cycle, it is not surprising that retinochrome is rarely lost in mollusks. Also, if it primarily functions to recycle and store retinal, we might not expect much functional diversification for this opsin. In fact, we only see 2 examples of retinochrome duplication, in bivalves and squid (supplementary fig. S9, Supplementary Material online).

The bivalve duplication of retinochrome, based on GeneRax reconciliation, appears to be ancestral for the class but only retained in the heteroconch species (supplementary fig. S9, Supplementary Material online). Why a second retinochrome is retained in these bivalves is unclear and no other information exists about this duplication. A necessary first step toward understanding its biological relevance will be demonstrating if, when, and where it is expressed. The squid retinochrome duplication was first recognized in *Idiosepius* (Yoshida et al. 2015). RNA-seq from several tissues point toward 1 squid retinochrome specializing in eye expression and the other copy having a broader range, including the gut (Yoshida et al. 2015). Octopus species have a single retinochrome, but they have a peropsin, which squids lack. Given the view of retinochrome and peropsins as primarily photoisomerases (e.g. (Vöcking et al. 2021), it is tempting to consider the octopus peropsin and duplicated squid retinochrome having similar functions. Exactly how functionally similar the molluscan peropsin is to retinochrome or other peropsins (e.g. retinal pigment epithelium-derived rhodopsin homolog, "RRH" in humans) is unclear. Retinochrome and molluscan peropsin differ in key amino acids for opsin function, such as the NPxxY motif and a tripeptide important for G-protein signaling (supplementary fig. S8, Supplementary Material online). Interestingly, based on the conservation of these motifs, the molluscan peropsin may be able to drive phototransduction, as has been proposed for other peropsins but not retinochrome or RGR (Vöcking et al. 2022).

A major complication to inferring function of mollusk peropsin is that "peropsin" does not appear to be a monophyletic group, but rather a name given to several taxon-specific clades within the RGR/retinochrome group of tetraopsins: deuterostome peropsins, arthropod peropsin-like, and mollusk peropsin-like (Ramirez et al. 2016; Vöcking et al. 2021). Furthermore, phylogenies with mollusk-peropsin sequences differ in their placement

Fig. 3. (Continued)
Branches transformed with Figtree as "proportional." Taxon names are color coded by habitat type. Opsin and cryptochrome abundances ("gene number") reflected by heatmap. Empty squares indicate gene absence, gray likely presence. Total counts for opsins and cryptochromes included to right of each matrix. Unclassified tetraopsin indicated as "UCTopn" dashed lines indicate ambiguous phylogenetic placement. Habitat and photic environment of each species indicated by text color and eye type listed to right of figure. R script for heatmap adapted from Gornik et al. (2020). Species silhouettes obtained from PhyloPic (https://www.phylopic.org/).

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msad263/7457512 by guest on 20 December 2023

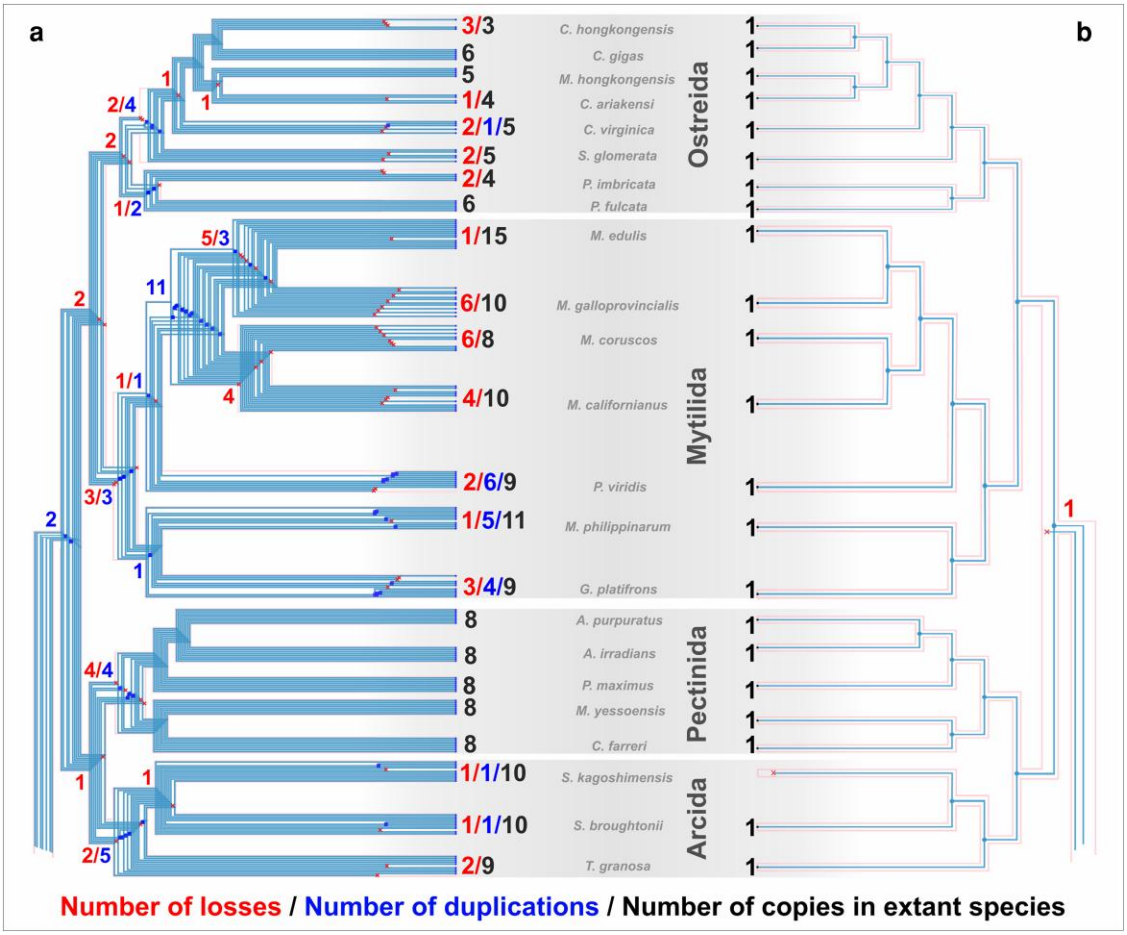

Fig. 4. Comparison of GeneRax reconciliation for a) xenopsin and b) retinochrome evolution in Pteriomorphia. Subsets of Thirdkind visualization from opsin reconciliation with locations of gene duplication and loss given at nodes and tips of species tree along with the total observed number of genes in each species for these 2 opsins. Blue line traces evolutionary history for each gene. Xenopsin is recurrently expanded via duplication in different orders of Pteriomorphia while retinochrome remains single copy, only evolving via speciation.

(Ramirez et al. 2016; Vöcking et al. 2021; Gühmann et al. 2022) and our lack of strong statistical support for the relationships between different groups within the RGR/retinochrome/peropsin group (supplementary fig. S6 and supplementary fig. S8, Supplementary Material online) do not resolve this issue. Ultimately, our results do support retinochrome and mollusk peropsin as distinct clades and likely not sister clades within the broader RGR/retinochrome/peropsin lineage. Still, these opsins are very similar in their extreme rarity for duplication. We found that peropsin in mollusks is almost always single copy—the only exception here is the 2 copies in the blood clam, *S. kagoshimensis*—but prone to loss (Fig. 3). Peropsin has been lost at least 7 times in mollusks, including in squid and nautilus, a gastropod clade containing Caenogastropoda and Heterobranchia, and numerously across bivalves: Pectinida, Mytiloida, Unionida, Cardiida (as evidenced by *Gari tellinella*), and some venerids (Fig. 3). The chiton genome also lacks peropsin, which suggests gene loss but a more robust phylogeny of the RGR/retinochrome/peropsin opsin clade would be needed for clarity and help reconstructing the history of functional divergence between these different photoisomerase opsins.

Repeated Xenopsin Expansions Contribute to Large Opsin Repertoires in Mollusks

We estimated 175 xenopsin duplication events, 3 to nearly 90 times higher than any other opsin type (Table 3). Xenopsins appear to be particularly prone to expansion in bivalves. For example, in Pteriomorphia, the ark clams (Arcida), scallops (Pectinida), mussels (Mytilida), and oysters (Ostreida) all experienced distinct, multiple rounds of xenopsin duplication (Fig. 4a). In another bivalve lineage, xenopsins account for most of the opsin sequences we recovered from the 2 *Dreissena* genomes (44 out of 54; 42 out of 63). The vestigastropods also have multiple rounds of xenopsin duplication, not shared by other gastropods.

This little-known opsin clade has only recently been recognized as a distinct opsin type, in part, because they are absent in chordates and arthropods. Xenopsin expression has been observed in larval eyes of the chiton *Leptochiton asellus* (Vöcking et al. 2017), flatworm

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msad263/7457512 by guest on 20 December 2023

(Rawlinson et al. 2019), and the bryozoan (Döring et al. 2020). Xenopsin is also expressed in adult eye rhabdomeric photoreceptors of the land slug, *Limax valentianus* (Nishiyama et al. 2019). In all four of these studies, xenopsin was co-expressed with r-opsins, and notably c-opsin in the case of the flatworm (Rawlinson et al. 2019), demonstrating that xenopsin and c-opsins are not mutually exclusive groups of opsins, as had initially been hypothesized (Vöcking et al. 2017). Rawlinson et al. (2019) also showed, in vitro, that xenopsins couple with $G_{\alpha i}$ to drive phototransduction, which was supported by (Sakai et al. 2022) who also demonstrated blue-light sensitivity and bistability in xenopsin from a chiton, brachiopod, platyhelminth, and chaetognath. These results point toward a scenario where xenopsin photopigments play an important role in visual processing in protostomes, including mollusks, but the exact nature of their function remains unclear, particularly for eyeless species.

Go-Opsin Numbers Differ Greatly Across Gastropods

The G_o-opsins were also prone to expansions in gastropods and mytiloid bivalves, though to a lesser extent than xenopsins (Fig. 3, Table 3). This type of opsin was first discovered in scallop eyes and was notably expressed in the ciliary cells of the distal retina (Kojima et al. 1997). These opsins differ in their coupled G-protein, but also in that G_o-opsin-based photoreception leads to cell membrane hyperpolarization, rather than depolarization found in r-opsin systems (Table 1). Other than their effect on cell membrane potential and G-protein, very little is known about G_o-opsins, but they have been tied to digestion regulation via control of pyloric opening in sea urchin larvae (Yaguchi and Yaguchi 2021), shadow response in annelid (Ayers et al. 2018), and detection of moving objects in scallops (Speiser et al. 2016).

The G_o-opsins were typically present in high numbers in various gastropod lineages. The species with the highest number of G_o-opsins (8–10 sequences) were the Vetigastropoda, with true limpets and abalone sharing numerous rounds of duplication (Fig. 3). In addition to xenopsins in some bivalves, the G_o-opsins in abalone represent some of the most abundant opsins in mollusks. Our results also point to strong clade-level discrepancies between G_o-opsins abundance within the gastropods (Fig. 3). We found 3 to 7 G_o-opsins in caenogastropods, while Heterobranchia had a reduced set with both Neogastropoda and Planorbidae having a single G_o-opsin per species and a loss of G_o-opsin entirely in the Stylommatophora land snails. Species in this clade of land snails have the smallest opsin repertoire that we observed, outside of the cephalopods. The reduced opsin diversity in land snails could be related to transitions to terrestrial habitats that have different light availability relative to aquatic relatives. Terrestrialization in gastropods has occurred upwards of 30 times (Vermeij and Watson-Zink 2022), yet genomic data has thus far only been collected from this single transition. A much greater genomic sampling of gastropods will be needed to formally test whether transitions to land are related to opsin repertoire in gastropods.

Rhabdomeric Opsin Duplication is Rare in Mollusks But Not Restricted to Scallops

The opsins most consistently recovered in the genomes from this phylum-wide search were the 2 clades of rhabdomeric or r-opsins: the "canonical" (G_q-coupled, invertebrate visual opsin) and "noncanonical" (Fig. 1). The "noncanonical" clade was only recently described in Ramirez et al. (2016) as a distinct group from the r-opsins that include arthropod and molluscan visual opsins and chordate melanopsin (collectively referred to as "canonical" in Ramirez et al. 2016). The noncanonical r-opsins have been found in mollusks and other lophotrochozoans, hemichordates, and nonvertebrate chordates, indicating that this r-opsin split predates metazoan—or at least bilaterian—diversification, lending support for nomenclatural distinction. Whether the noncanonical r-opsins have similar functional properties as the better described canonical r-opsins is unknown. Other than the deep-sea lineages that completely lacked opsins, only 1 molluscan species appears to not have a noncanonical r-opsin, the pearl oyster *Pinctada imbricata*, though we found a noncanonical r-opsin in the congener *Pinctada fucata*. Most species in our dataset had 1 to 3 noncanonical r-opsins, but we observed 12 copies in *Dreissena* and a similarly large expansion in some venerid bivalves (Fig. 3). Most of the noncanonical r-opsins have the HPK tripeptide motif located in the fourth extracellular loop (Fig. 2a), which is a hallmark $G_{\alpha q}$-coupled opsins (Plachetzki and Oakley 2007), so it is likely that these opsins function similarly to the canonical r-opsins (see Table 1 for details).

Canonical r-opsin was recovered from every genome—excluding the deep-sea lineages with total opsin loss—except for the sea slug *Aplyisa californica*, in which we found evidence of a pseudogenized remnant sequence. Given the established role of canonical r-opsins in visual processing for invertebrate bilaterians, the evolution of these opsins may be particularly tied to eye complexity and photic environment. Duplication and specialization of canonical r-opsins in arthropods are very common (Cronin and Porter 2014). In jumping spiders, for example, different r-opsin genes are expressed in the principal eye and 3 secondary eyes (Nagata et al. 2012). In the most extreme example yet, mantis shrimp have been shown to express 33 r-opsins in their compound eyes, with several distinct combinations of opsins sensitive to ultraviolet, long, short, and medium wavelengths of light observed across numerous distinct photoreceptor types (Porter et al. 2020). This degree of r-opsin diversification is not seen in mollusks. In fact, the majority (58/80) of mollusk species examined here have a single canonical r-opsin sequence. However, it is worth noting that arthropods lack xenopsins and G_o-opsins (Ramirez et al. 2016), the 2 most expansion-prone groups of opsins in mollusks according to our findings.

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msad263/7457512 by guest on 20 December 2023

Despite the overall rarity of canonical r-opsin duplication in mollusks, an expansion in scallops represents a compelling case for opsin recruitment in the evolution of novel eyes. A single r-opsin duplication was initially identified in the bay scallop, *Argopecten irradians*, a notable exception for mollusks at the time (Serb et al. 2013). With transcriptomic sequencing, four r-opsins were eventually identified in *A. irradians*, which differed in expression levels and had some notable sequence differences at critical amino acid sites (Porath-Krause et al. 2016). Analyses of additional scallop genome assemblies have also identified four r-opsins, which tend to have eye-biased gene expression levels (Li et al. 2017; Wang et al. 2017).

We recovered four r-opsins in all 5 species of scallops surveyed here (Fig. 2b, Fig. 3). The 3 intronless r-opsins in scallops form their own clade, demonstrating that the first r-opsin duplication in scallops (apparently a retrotransposition event) likely occurred in the ancestor for all scallops (Porath-Krause et al. 2016). We also observed a second series of r-opsin duplication events exclusive in sea mussels (Fig. 2b, Fig. 3). The four species of *Mytilus* along with *Gigantidas platifrons*, *Modiolus philippinarum*, and *Perna viridis* all have 3 r-opsins and none of these species have eyes as adults, bringing the hypothesis of r-opsin recruitment for visual processes into question. A notable difference between the scallop and mussel r-opsin expansions is the possible functional divergence of the scallop duplicates, as indicated by their E/DRY motif, a sequence in the third TM important for stabilizing GPCR inactive-state confirmation (Rovati et al. 2007). In the scallop r-opsin clades, we see the following motifs: DRY, DRC, DRF, while all the mussel r-opsins have DRY (Fig. 2b). In fact, DRY was found in all other r-opsins recovered from mollusk genomes in this study (see logo in Fig. 2a), except for the 2 *Pinctada* pearl oyster species (Fig. 2b).

Mollusks Have Up To 6 Types of CRY-PL Proteins

Using curated classified reference protein sets and the BITACORA protein prediction pipeline, we identified 440 CRY-PL gene models from 80 molluscan genomes (Fig. 5). Our phylogenetic analysis of the mollusk sequences from CRY-PL placed all 440 sequences (excluding 4 likely contaminant sequences) into the following 6 classifications: CPD-II (phr), CRY-DASH, PCRY-like, DCRY (CRY-I), 6-4 PL, and MCRY (CRY-II) clades. Broadly, our CRY-PL ML phylogeny includes 10 major clades with high statistical support that also included sponge CRYs, AnthoCRYs, AnthoCRY-II, and CRY4 (Fig. 5). We did not resolve the PCRY/PCRY-like/CPD-III/CPD-I relationships (hence the collapsed clade in Fig. 5), as the phylogeny was constructed to classify the predicted sequences from BITACORA output and not meant to be an exhaustive reconstruction of the entire CRY-PL family. However, the PCRY-like clade does include a monophyletic mollusk group (supplementary fig. S11, Supplementary Material online), indicating strong support for our classification of those sequences. Broadly, the CRY-PL topology here (Fig. 5) recapitulates what has been reported from other studies (e.g. Oliveri et al. 2014; Deppisch et al. 2022; Vicedomini et al. 2022).

Most mollusk species have CRY-I, CRY-II, CRY-DASH, CPD-II (i.e. phr), and 6-4 PL (Fig. 3). We found that phr and the animal cryptochromes (CRY-I, CRY-II), were present across all molluscan classes and in nearly all species, while CRY-DASH and (6-4)PLs are absent in cephalopods and some terrestrial gastropods, with occasional losses from species in well-represented taxonomic classes (e.g. CRY-DASH absence in Dreissenidae, Bivalvia) (Fig. 3). The PCRY-like group is restricted to specific molluscan lineages, including 2 gastropod clades: Trochidae "top-snails" and Conidae "cone-snails," along with most bivalves, though it appears to have been lost in Pectinidae while maintained in other pteriomorphian families (PCRY-L in Fig. 3). This somewhat spotty recovery of PCRY-like in mollusks is similar the findings in Deppisch et al. (2022), who suggest that the presence of PCRY-like in the oyster *Crassostrea gigas* and other marine mollusks may be from horizontal gene transfer (HGT), as bacterial genes have been identified in marine metazoans, like the jellyfish *Nematostella vectensis* (Starcevic et al. 2008). However, our phylogeny recovers a monophyletic molluscan PCRY-like. Furthermore, the PCRY-like clade in Deppisch et al. (2022), too, has a monophyletic group of metazoan PCRY-like sequences, with a vertebrate clade and mollusk—annelid—arthropod clade sister to it. These results are not indicative of HGT unless it occurred in the last common ancestor of metazoan. Instead, the patterns of presence and absence of this sequence across mollusks and other metazoans likely reflect frequent loss, as proposed by Oliveri et al. (2014) in the first phylogenetic description of the widespread occurrence of PCRY-like sequences across bilaterians. Still, a greater survey of microbial genomes for PCRY-like sequences could help clarify the origin of this gene outside of plants. Other than evidence of rhythmic expression of this gene from zebrafish (Oliveri et al. 2014), virtually nothing else is known about the function of this protein.

Cryptochrome Abundance is Highly Restricted Relative to Opsins

We observed very few instances of retained cryptochrome duplication in the molluscan genomes we surveyed (Fig. 3). Second copies of each cryptochrome were identified spuriously across our species tree, but we did not observe any apparent shared duplication within taxonomic clades as we did with opsins. The only species that we recovered more than 2 copies for a given cryptochrome was in the zebra mussel, *Dreissena polymorpha*, in which we also found the highest number of opsins. In *D. polymorpha*, we found evidence of 3 (6-4)PLs and 5 CRY-II. In the congener species *D. rostriformis*, the quagga mussel, did not have the same CRY-II duplications, but the 2 species share a (6-4)PLs duplication. The annotations for the quagga and zebra mussel genome assemblies both reported

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msad263/7457512 by guest on 20 December 2023

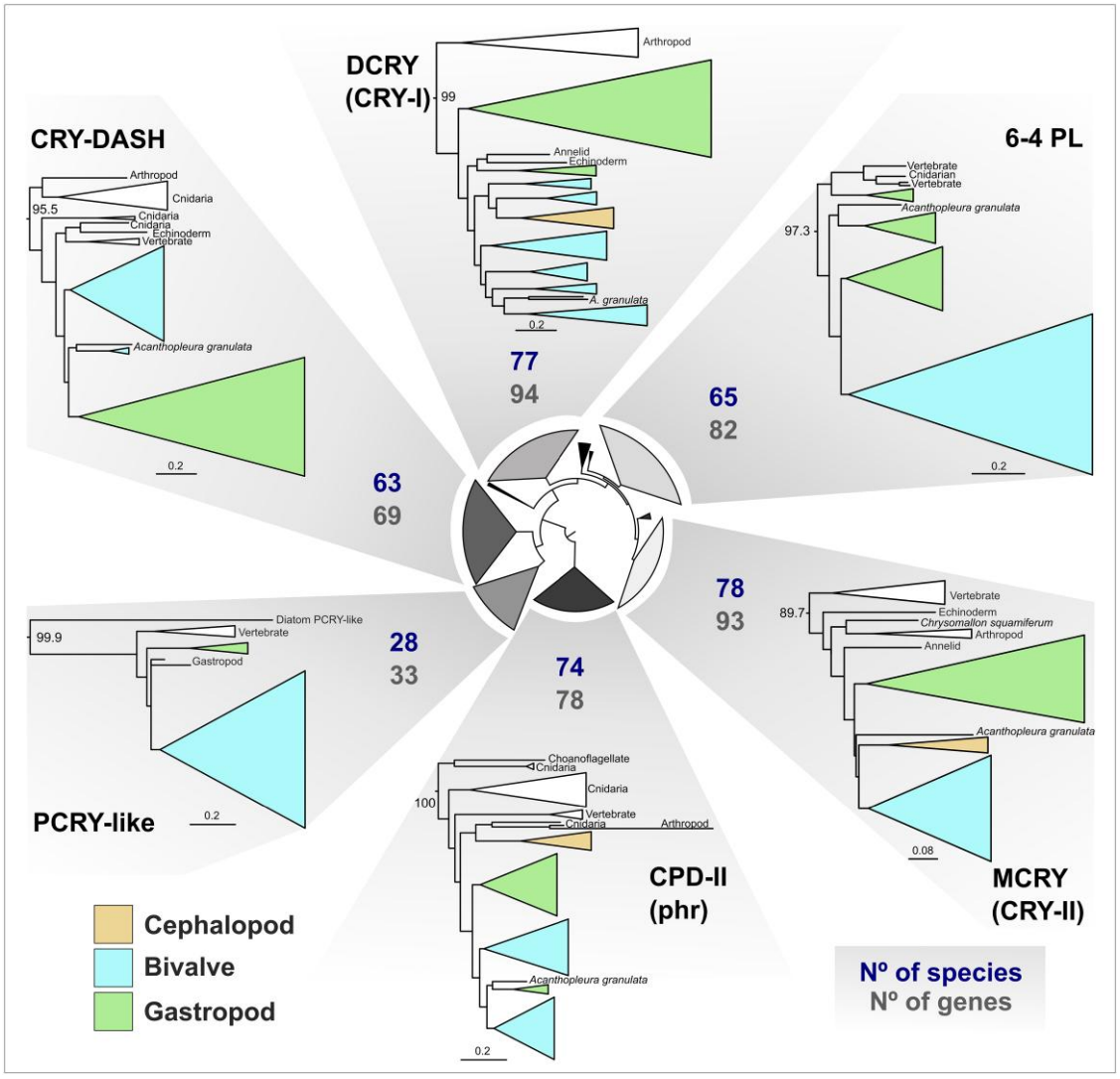

Fig. 5. CRY-PL clades are similar in total size and number of species represented. Subtrees of mollusk containing CRY-PL clades displayed from the IQ-TREE2 ML CRY-PL phylogeny (LG + R8 amino acid substitution model). For each of 6 CRY-PL types found in mollusks, the total number of sequences in the tree (N° of genes) and species (N° species) with that gene are displayed. Clades are collapsed according to taxonomic clades (e.g. bivalves) and color coded by mollusk class. CRY-PL names in parentheses (e.g. "CRY-I") reflect common nomenclature for animal proteins used in text. UF-bootstrap support values shown at base of named of CRY-PL clades. See supplementary fig. S12b, Supplementary Material online for branch support values for circular inner tree and mollusk-specific clades in supplementary fig. S11, Supplementary Material online.

exceptionally high numbers of predicted protein sequences for metazoans and numerous gene family expansions were described in the zebra mussel, though the authors acknowledged that their results are likely an overestimation (Calcino et al. 2019; McCartney et al. 2022).

Though cryptochromes do not diversify in a similar manner to opsins, there were a few notable parallels between opsin and cryptochrome loss in our results. Cephalopoda lost neuropsin and G_o-opsins and appears to have lost CRY-DASH and (6-4)PLs, 2 UV-induced DNA-damage repair PL in the CRY-PL. *Nautilus* also lacks phr while it is present in the other cephalopod genomes. Similarly, the group of land snails with reduced opsin repertoires, including loss of G_o-opsin, are also mostly missing CRY-DASH and (6-4) PLs. The exceptions being *Cepea nemoralis*, which has both genes, and *Arion vulgaris*, which has a DRY-DASH. The dreissenid clams also lost both CRY-DASH and G_o-opsins. Therefore, CRY-DASH and G_o-opsin have been jointly lost at least once in bivalves, cephalopods, and gastropods. Determining if a biological connection exists between CRY-DASH and G_o-opsins in mollusks will be important to our understanding of photopigment functions. Cryptochrome loss was also evident in the same deep-sea lineages lacking opsins (see below). The 2 deep-sea snails only had CRY-II, which, based on the mammalian circadian clock, functions independent of light (Griffin et al. 1999), and the deep-sea clam *Archivesica marissinica* lacked every CRY-PL sequence we searched for.

Opsin and Cryptochrome Loss in Deep-Sea Mollusks

The most apparent connection between opsin content and molluscan biology and ecology we found was

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msad263/7457512 by guest on 20 December 2023

complete opsin loss in 3 deep-sea mollusks. Two gastropods, the scaly foot snail, *Chrysomallon squamiferum* and *Gigantopelta aegis*, and a venerid clam, *Archivesica marissinica*, appear to have no opsin sequences in their genomes. None of the 3 opsinless species had particularly low numbers of BITACORA-predicted genes with opsins as input, indicating that this is not a failure of our approach. For *A. marissinica*, 177 genes were predicted (vs. 243 ± 108 in bivalves), *C. squamiferum* and *G. aegis* had 162 and 186, respectively (vs. 224 ± 117 in gastropods), indicating that the quality of the genomes and annotation approach were not biased. Furthermore, we detected many opsins in close relatives of these species, indicating that the phylogenetic distance to the reference opsin set was not preventing us from predicting opsins in these genomes. Additionally, we used tblastn to specifically query opsin amino acid sequences against these genomes and were unable to identify evidence of remnant opsin sequences (e.g. pseudogenes). Similarly, these 2 deep-sea snails have lost all members of the CRY/PL family except for the light-insensitive CRY-II, and the bivalve, *A. marissinica*, lacks any sequence from this protein family. Cryptochrome loss has also been observed in at least 1 deep-sea fish species, the coelacanth (Deppisch et al. 2022).

The 2 snail species belong to a single deep-sea transition in Peltospiridae, so our data only capture opsin loss in 2 distinct deep-sea mollusk lineages. Opsin loss has also been observed in deep-sea anthozoan cnidarians (Gornik et al. 2020). Loss of opsins in deep-sea mollusks and cnidarians is a stark contrast to some deep-sea teleost fish, where independent rhodopsin expansions—including up to 38 copies in the silver spinyfin—have been observed (Musilova et al. 2019). In the water beetle family Dytiscidae, independent transitions to subterranean life are associated with loss/decay of opsin and other phototransduction genes (Langille et al. 2022). Testing whether other proteins involved in phototransduction are missing or possibly nonfunctional from these opsinless deep-sea molluscan lineages will be a useful approach to exploring the genomics of trait decay.

From our taxonomic sampling, we cannot determine what factors lead to opsin loss in some deep-sea lineages and not others. Surprisingly, 2 other deep-sea mollusks in our dataset, the gastropod *Phymorhynchus buccinoides* and bivalve *Gigantidas platifrons* have similar numbers of opsin and CRY-PL genes when compared to their closest relatives. *Phymorhynchus buccinoides* occur in cold seeps versus the hydrothermal vents that support the scaly-foot snail. Both habitats should favor chemosensation over vision, and indeed, *C. squamiferum* has lost its eyes and numerous *Phymorhynchus* species, too, lack eyes (Zhang and Zhang 2017). In the deep-sea mussel *Bathymodiolus azoricus*, a close relative of *G. platifrons*, cryptochromes and other genes involved in circadian rhythm is expressed in a rhythmic fashion in the natural hydrothermal vent environment, apparently driven by tidal signal (Mat et al. 2020). Depth at which a species is found could be a parameter for opsin retention vs. loss. The scaly-foot snail is found at depths near 3,000 m while *P. buccinoides* occurs at 1,160–1,190 m. However, the deep-sea clam, *A. marissinica* is reported from 1,400 m and has completely lost its opsins. Another possible factor for loss of opsin in some species but not others could be the time since transitioning the aphotic deep-sea environments, as has been seen with repeated colonization of caves by teleost fish (Niemiller et al. 2013). Testing whether these species that lack opsins or cryptochromes can still perceive light will be critical to better understanding the degeneration of visual and light-sensing systems in deep-sea animals.

What Accounts for Differing Opsin Content in Mollusks?

From our genomic survey of photopigment genes in mollusk genomes, no apparent connection to eye or eye specialization emerges (Fig. 3), except for the loss of opsins and cryptochromes in some deep-sea lineages. We observed the highest numbers of opsins in eyeless bivalve species and the fewest number of opsins in cephalopods, a lineage with sophisticated eyes. Therefore, the degree of specialization in light-sensing organs appears to be independent of the diversity and abundance of opsins in mollusks, as was reported recently by De Vivo et al. (2023). The question is then, what other organismal and environmental factors drive opsin evolution for these species?

In addition to the increasing number of recorded extraocular functions for opsins, a growing body of literature points to numerous light-independent functions for opsins, suggesting that how we view opsins should change to describe them as a broad sensory-driven signaling molecule (Feuda et al. 2022). It has been argued that light-absorption may not even be the "original" function of opsins (Leung et al. 2020; Pisani et al. 2020; Feuda et al. 2022). Because opsins may be used to regulate a variety of physiological processes, taking in more than just light, we need to consider the possibility that the extensive diversification of opsins in species such as venerid clams and *Dreissina* species reflects selection on broad multisensory receptors.

Light, chemical, and other environmental cues may be especially important for animals with distinct life stages. The reduced sets of opsins in bryozoans and platyhelminths relative to mollusks and other lophotrochozoans may be connected to such life-stage distinction (De Vivo et al. 2023). The possibility that life cycle complexity influences opsin family evolution could extend to within-Mollusca differences in opsin diversity. Unlike most cephalopods and gastropods in freshwater or terrestrial environments that have direct development, most marine bivalves and gastropods have several morphologically distinct larval stages. These mobile larval stages are the main mechanism for dispersal for many marine mollusks, which after a "veliger" stage metamorphose and settle into a sessile adulthood. As a result, it is important for these species to be "choosy" when identifying appropriate substrates. Previous work has shown that bivalve species, like *Mytilus*, are highly selective of where to settle (e.g. Carl et al. 2012). Interestingly,

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msad263/7457512 by guest on 20 December 2023

chemical cues that likely induce abalone settlement and metamorphosis involve unidentified TM G-protein coupled receptors, the same superfamily of opsin (Baxter and Morse 1987, 1992). Our study uncovered large opsin repertoires in marine species, with abalone having the highest numbers among gastropods. We think that opsins, as TM receptors, are worthwhile proteins to explore in the context of life-stage triggers and "decisions" on settling in mollusks.

Among the most immediate next steps to better understand opsin use in mollusks is determining where and when these genes are expressed, e.g. different tissues and developmental stages. This question can be addressed with tissue-specific and single-cell RNA-seq, along with in situ RNA hybridization and immunohistochemistry. Whether these opsin sequences encode potential photopigments can be determined with heterologous expression assays in cell culture (e.g. Faggionato and Serb 2017; Smedley et al. 2022). There is the possibility that (some) of these molluscan opsins have the potential to detect light but serve other functions, i.e. multimodality, and can first be assessed from protein-ligand predictions based on protein models, such as the AlphaFill algorithm applied to Alphafold models (Hekkelman et al. 2023). Given the successful application of CRISPR gene editing in bivalves (Yu et al. 2019; Jin et al. 2021), cephalopods (Crawford et al. 2020; Ahuja et al. 2023), and gastropods (Perry and Henry 2015; Abe and Kuroda 2019), there is also potential to incorporate gene knock-out experiments into future explorations of molluscan opsin functions. Using molecular evolutionary analyses of opsin sequence to detect positively selected sites associated with photic and other environmental variables (e.g. Castiglione et al. 2017; Sondhi et al. 2021; Van Nynatten et al. 2021) will also be a powerful approach to characterizing molluscan opsins as more genomic data becomes available.

Conclusion

Our results reveal that mollusks vary greatly in the types and abundance of proteins involved in light-sense in their genomes. Our phylogenetic analysis of predicted opsin sequences from 80 molluscan genome assemblies supported at least 7 major opsin clades with species total opsin counts ranging from zero to 63 copies. We find that some types of opsins, like retinochrome, are evolutionarily static, characterized by rare cases of duplication, while other opsins repeatedly expand in lineage-specific manner, such as xenopsin. These results stand in stark contrast to another group of photopigments, cryptochromes, which have not diversified within mollusks. We found that some deep-sea species lack opsins and cryptochromes, indicating these proteins can be dispensable for mollusks in aphotic environments. Other than deep-sea-related loss of opsins, we see no clear connection between opsin diversity and photic environments but discuss possible connections to terrestriality. The bivalves in this study, most of which lack eyes, tended to have the greatest number of opsins, with repeated lineage-level expansions accounting for the observed diversity in opsin repertoire. The abundance of opsins in these eyeless species raises important questions about what function these genes play in how mollusks sense their environment.

Supplementary Material

Supplementary material is available at *Molecular Biology and Evolution* online.

Acknowledgments

We would like to thank other members of the Serb Lab for their helpful feedback on the manuscript. We would also like to thank the Iowa State Genome Informatics Facility for helpful discussions. We thank the reviewers and editors, whose feedback improved the quality of this work.

Author Contributions

K.E.M. and J.M.S. designed the study. K.E.M. conducted the analyses. J.A.A. collected biological and ecological trait information. K.E.M. and J.A.A. created the figures. K.E.M. and J.M.S. wrote the manuscript. K.E.M., J.A.A, and J.M.S. edited the manuscript.

Funding

This project was supported by the National Science Foundation under the Division of Environmental Biology (Grant No. 1754331 to JMS). The research reported in this paper is partially supported by the HPC@ISU equipment at Iowa State University, some of which has been purchased through funding provided by the National Science Foundation under the Major Research Instrumentation Program (Grant Nos. 1726447 and 2018594).

Conflict of interest statement. None declared.

Data Availability

All gene predictions from BITACORA, sequences used in opsin and CRY-PL phylogenetic analyses, alignments, and phylogenetic trees are available through FigShare (https://doi.org/10.6084/m9.figshare.23593641).

References

Abe M, Kuroda R. The development of CRISPR for a mollusc establishes the formin *Lsdia1* as the long-sought gene for snail dextral/sinistral coiling. *Development*. 2019:**146**(9):dev175976. https://doi.org/10.1242/dev.175976.

Ahuja N, Hwaun E, Pungor JR, Rafiq R, Nemes S, Sakmar T, Vogt MA, Grasse B, Diaz Quiroz J, Montague TG, *et al.* Creation of an albino squid line by CRISPR-Cas9 and its application for in vivo functional imaging of neural activity. *Curr Biol*. 2023:**33**(13):2774–2783.e5. https://doi.org/10.1016/j.cub.2023.05.066.

Anisimova M, Gil M, Dufayard J-F, Dessimoz C, Gascuel O. Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. *Syst Biol*. 2011:**60**(5):685–699. https://doi.org/10.1093/sysbio/syr041.

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msad263/7457512 by guest on 20 December 2023

Audino JA, Marian JEAR. On the evolutionary significance of the mantle margin in pteriomorphian bivalves. *Am Malacol Bull*. 2016:**34**(2):148–159. https://doi.org/10.4003/006.034.0212.

Audino JA, Marian JEAR. Detailed comparative anatomy of the pinnidae (Mollusca, Bivalvia) reveals further unusual mantle specializations. *Zool Anz*. 2020:**285**:27–36. https://doi.org/10.1016/j.jcz.2020.01.006.

Audino JA, Serb JM, Marian JEAR. Hard to get, easy to lose: evolution of mantle photoreceptor organs in bivalves (Bivalvia, Pteriomorphia). *Evolution*. 2020:**71**(9):2105–2120. https://doi.org/10.1111/evo.14050.

Ayers T, Tsukamoto H, Gühmann M, Veedin Rajan VB, Tessmar-Raible K. A go-type opsin mediates the shadow reflex in the annelid *Platynereis dumerilii*. *BMC Biol*. 2018:**16**(1):41. https://doi.org/10.1186/s12915-018-0505-8.

Baxter G, Morse DE. G protein and diacylglycerol regulate metamorphosis of planktonic molluscan larvae. *Proc Natl Acad Sci USA*. 1987:**84**(7):1867–1870. https://doi.org/10.1073/pnas.84.7.1867.

Baxter GT, Morse DE. Cilia from abalone larvae contain a receptor-dependent G protein transduction system similar to that in mammals. *Biol Bull*. 1992:**183**(1):147–154. https://doi.org/10.2307/1542416.

Beaudry FEG, Iwanicki TW, Mariluz BRZ, Darnet S, Brinkmann H, Schneider P, Taylor JS. The non-visual opsins: eighteen in the ancestor of vertebrates, astonishing increase in ray-finned fish, and loss in amniotes. *J Exp Zool B Mol Dev Evol*. 2017:**328**(7):685–696. https://doi.org/10.1002/jez.b.22773.

Birney E, Clamp M, Durbin R. GeneWise and genomewise. *Genome Res*. 2004:**14**(5):988–995. https://doi.org/10.1101/gr.1865504.

Borges R, Khan I, Johnson WE, Gilbert MTP, Zhang G, Jarvis ED, O'Brien SJ, Antunes A. Gene loss, adaptive evolution and the co-evolution of plumage coloration genes with opsins in birds. *BMC Genomics*. 2015:**16**(1):751. https://doi.org/10.1186/s12864-015-1924-3.

Bradshaw WE, Holzapfel CM. Evolution of animal photoperiodism. *Annu Rev Ecol Evol Syst*. 2007:**38**(1):1–25. https://doi.org/10.1146/annurev.ecolsys.37.091305.110115.

Brandon CS, Greenwold MJ, Dudycha JL. Ancient and recent duplications support functional diversity of daphnia opsins. *J Mol Evol*. 2017:**84**(1):12–28. https://doi.org/10.1007/s00239-016-9777-1.

Calcino AD, de Oliveira AL, Simakov O, Schwaha T, Zieger E, Wollesen T, Wanninger A. The quagga mussel genome and the evolution of freshwater tolerance. *DNA Res*. 2019:**26**(5):411–422. https://doi.org/10.1093/dnares/dsz019.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. *BMC Bioinformatics*. 2009:**10**(1):421. https://doi.org/10.1186/1471-2105-10-421.

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. Trimal: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. 2009:**25**(15):1972–1973. https://doi.org/10.1093/bioinformatics/btp348.

Carl C, Poole AJ, Williams MR, de Nys R. Where to settle—settlement preferences of Mytilus galloprovincialis and choice of habitat at a micro spatial scale. *PLoS One*. 2012:**7**(12):e52358. https://doi.org/10.1371/journal.pone.0052358.

Carter JG, Harries PJ, Malchus N, Sartori AF, Anderson LC, Bieler R, Bogan AE, Coan EV, Cope JCW, Cragg S, *et al*. Treatise online no. 48: Part N, Revised, Volume 1, Chapter 31: Illustrated glossary of the Bivalvia. *Treatise Online*. 2012:1–209. https://doi.org/10.17161/to.v0i0.4322.

Castiglione GM, Hauser FE, Liao BS, Lujan NK, Van Nynatten A, Morrow JM, Schott RK, Bhattacharyya N, Dungan SZ, Chang BSW. Evolution of nonspectral rhodopsin function at high altitudes. *Proc Natl Acad Sci U S A*. 2017:**114**(28):7385–7390. https://doi.org/10.1073/pnas.1705765114.

Chernomor O, von Haeseler A, Minh BQ. Terrace aware data structure for phylogenomic inference from supermatrices. *Syst Biol*. 2016:**65**(6):997–1008. https://doi.org/10.1093/sysbio/syw037.

Colbourne JK, Pfrender ME, Gilbert D, Thomas WK, Tucker A, Oakley TH, Tokishita S, Aerts A, Arnold GJ, Basu MK, *et al*. The ecoresponsive genome of *Daphnia pulex*. *Science*. 2011:**331**(6017):555–562. https://doi.org/10.1126/science.1197761.

Combosch DJ, Collins TM, Glover EA, Graf DL, Harper EM, Healy JM, Kawauchi GY, Lemer S, McIntyre E, Strong EE, *et al*. A family-level tree of life for bivalves based on a sanger-sequencing approach. *Mol Phylogenet Evol*. 2017:**107**:191–208. https://doi.org/10.1016/j.ympev.2016.11.003.

Crawford K, Quiroz JFD, Koenig KM, Ahuja N, Albertin CB, Rosenthal JJC. Highly efficient knockout of a squid pigmentation gene. *Curr Biol*. 2020:**30**(17):3484–3490.e4. https://doi.org/10.1016/j.cub.2020.06.099.

Cronin TW, Porter ML. *The evolution of invertebrate photopigments and photoreceptors BT - evolution of visual and non-visual pigments*. In: Hunt DM, Hankins MW, Collin SP, Marshall NJ, editors. Boston (MA): Springer US; 2014. p. 105–135.

Crooks GE, Hon G, Chandonia J-M, Brenner SE. WebLogo: a sequence logo generator. *Genome Res*. 2004:**14**(6):1188–1190. https://doi.org/10.1101/gr.849004.

Cunha TJ, Giribet G. A congruent topology for deep gastropod relationships. *Proc R Soc B Biol Sci*. 2019:**286**(1898):20182776. https://doi.org/10.1098/rspb.2018.2776.

Deppisch P, Helfrich-Förster C, Senthilan PR. The gain and loss of cryptochrome/photolyase family members during evolution. *Genes (Basel)*. 2022:**13**(9):1613. https://doi.org/10.3390/genes13091613.

De Vivo G, Crocetta F, Ferretti M, Feuda R, D'Aniello S. Duplication and losses of opsin genes in lophotrochozoan evolution. *Mol Biol Evol*. 2023:**40**(4):msad066. https://doi.org/10.1093/molbev/msad066.

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**:e55193. https://doi.org/10.7554/eLife.55193.

Faggionato D, Serb JM. Strategy to identify and test putative light-sensitive non-opsin G-protein-coupled receptors: a case study. *Biol Bull*. 2017:**233**(1):70–82. https://doi.org/10.1086/694842.

Feuda R, Goulty M, Zadra N, Gasparetti T, Rosato E, Pisani D, Rizzoli A, Segata N, Ometto L, Stabelli OR. Phylogenomics of opsin genes in Diptera reveals lineagespecific events and contrasting evolutionary dynamics in Anopheles and Drosophila. *Genome Biol Evol*. 2021:**13**:evab170. https://doi.org/10.1093/gbe/evab170.

Feuda R, Menon AK, Göpfert MC. Rethinking opsins. *Mol Biol Evol*. 2022:**39**(3):msac033. https://doi.org/10.1093/molbev/msac033.

Futahashi R, Kawahara-Miki R, Kinoshita M, Yoshitake K, Yajima S, Arikawa K, Fukatsu T. Extraordinary diversity of visual opsin genes in dragonflies. *Proc Natl Acad Sci U S A*. 2015:**112**:E1247–E1256. https://doi.org/10.1073/pnas.1424670112.

Gornik SG, Bergheim BG, Morel B, Stamatakis A, Foulkes NS, Guse A. Photoreceptor diversification accompanies the evolution of Anthozoa. *Mol Biol Evol*. 2020:**38**(5):1744–1760. https://doi.org/10.1093/molbev/msaa304.

Griffin EA, Staknis D, Weitz CJ. Light-Independent role of CRY1 and CRY2 in the mammalian circadian clock. *Science*. 1999:**286**(5440):768–771. https://doi.org/10.1126/science.286.5440.768.

Gühmann M, Porter ML, Bok MJ. The gluopsins: opsins without the retinal binding lysine. *Cells*. 2022:**11**(15):2441. https://doi.org/10.3390/cells11152441.

Hara T, Hara R. New photosensitive pigment found in the retina of the squid ommastrephes. *Nature*. 1965:**206**(4991):1331–1334. https://doi.org/10.1038/2061331a0.

Hara T, Hara R. Vision in Octopus and squid: rhodopsin and retinochrome in the squid retina. *Nature*. 1967:**214**(5088):573–575. https://doi.org/10.1038/214573a0.

Hekkelman ML, de Vries I, Joosten RP, Perrakis A. AlphaFill: enriching AlphaFold models with ligands and cofactors. *Nat Methods*. 2023:**20**(2):205–213. https://doi.org/10.1038/s41592-022-01685-y.

Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol*. 2018:**35**(2):518–522. https://doi.org/10.1093/molbev/msx281.

Hu J, Shi Y, Zhang J, Huang X, Wang Q, Zhao H, Shen J, Chen Z, Song W, Zheng P, *et al*. Melanopsin retinal ganglion cells mediate light-

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msad263/7457512 by guest on 20 December 2023

promoted brain development. *Cell*. 2022:**185**(17):3124–3137.e15. https://doi.org/10.1016/j.cell.2022.07.009.
Jin K, Zhang B, Jin Q, Cai Z, Wei L, Wang X, Zheng Y, Huang B, Zhang M, Qi Y, *et al.* CRISPR/cas9 system-mediated gene editing in the Fujian Oysters (*Crassostrea angulate*) by electroporation. *Front Mar Sci*. 2021:**8**:763470. https://doi.org/10.3389/fmars.2021.763470.
Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods*. 2017:**14**(6):587–589. https://doi.org/10.1038/nmeth.4285.
Katana R, Guan C, Zanini D, Larsen ME, Giraldo D, Geurten BRH, Schmidt CF, Britt SG, Göpfert MC. Chromophore-Independent roles of opsin apoproteins in *Drosophila* mechanoreceptors. *Curr Biol*. 2019:**29**(17):2961–2969.e4. https://doi.org/10.1016/j.cub.2019.07.036.
Keilwagen J, Hartung F, Paulini M, Twardziok SO, Grau J. Combining RNA-seq data and homology-based gene prediction for plants, animals and fungi. *BMC Bioinformatics*. 2018:**19**(1):189. https://doi.org/10.1186/s12859-018-2203-5.
Keilwagen J, Wenk M, Erickson JL, Schattat MH, Grau J, Hartung F. Using intron position conservation for homology-based gene prediction. *Nucleic Acids Res*. 2016:**44**(9):e89–e89. https://doi.org/10.1093/nar/gkw092.
Kennedy D. Neural photoreception in a lamellibranch mollusc. *J Gen Physiol*. 1960:**44**(2):277–299. https://doi.org/10.1085/jgp.44.2.277.
Kingston ACN, Kuzirian AM, Hanlon RT, Cronin TW. Visual phototransduction components in cephalopod chromatophores suggest dermal photoreception. *J Exp Biol*. 2015:**218**(10):1596–1602. https://doi.org/10.1242/jeb.117945.
Kocot KM, Cannon JT, Todt C, Citarella MR, Kohn AB, Meyer A, Santos SR, Schander C, Moroz LL, Lieb B, *et al.* Phylogenomics reveals deep molluscan relationships. *Nature*. 2011:**477**(7365):452–456. https://doi.org/10.1038/nature10382.
Kojima D, Terakita A, Ishikawa T, Tsukahara Y, Maeda A, Shichida Y. A novel go-mediated phototransduction cascade in scallop visual cells*. *J Biol Chem*. 1997:**272**(37):22979–22982. https://doi.org/10.1074/jbc.272.37.22979.
Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA x: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol*. 2018:**35**(6):1547–1549. https://doi.org/10.1093/molbev/msy096.
Kuraku S, Zmasek CM, Nishimura O, Katoh K. Aleaves facilitates on-demand exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity. *Nucleic Acids Res*. 2013:**41**(W1):W22–W28. https://doi.org/10.1093/nar/gkt389.
Langille BL, Tierney SM, Bertozzi T, Beasley-Hall PG, Bradford TM, Fagan-Jeffries EP, Hyde J, Leijs R, Richardson M, Saint KM, *et al.* Parallel decay of vision genes in subterranean water beetles. *Mol Phylogenet Evol*. 2022:**173**:107522. https://doi.org/10.1016/j.ympev.2022.107522.
Leung NY, Montell C. Unconventional roles of opsins. *Annu Rev Cell Dev Biol*. 2017:**33**(1):241–264. https://doi.org/10.1146/annurev-cellbio-100616-060432.
Leung NY, Thakur DP, Gurav AS, Kim SH, Di Pizio A, Niv MY, Montell C. Functions of opsins in *Drosophila* taste. *Curr Biol*. 2020:**30**(8):1367–1379.e6. https://doi.org/10.1016/j.cub.2020.01.068.
Li Y, Sun X, Hu X, Xun X, Zhang J, Guo X, Jiao W, Zhang L, Liu W, Wang J, *et al.* Scallop genome reveals molecular adaptations to semi-sessile life and neurotoxins. *Nat Commun*. 2017:**8**(1):1721. https://doi.org/10.1038/s41467-017-01927-0.
Lin C, Todo T. The cryptochromes. *Genome Biol*. 2005:**6**(5):220. https://doi.org/10.1186/gb-2005-6-5-220.
Madeira F, Madhusoodanan N, Lee J, Tivey ARN, Lopez R. Using EMBL-EBI services via web interface and programmatically via web services. *Curr Protoc Bioinformatics*. 2019:**66**(1):e74. https://doi.org/10.1002/cpbi.74.
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**(10):4647–4654. https://doi.org/10.1093/molbev/msab199.
Marin EP, Krishna AG, Zvyaga TA, Isele J, Siebert F, Sakmar TP. The amino terminus of the fourth cytoplasmic loop of rhodopsin modulates rhodopsin-transducin interaction*. *J Biol Chem*. 2000:**275**(3):1930–1936. https://doi.org/10.1074/jbc.275.3.1930.
Mat AM, Sarrazin J, Markov GV, Apremont V, Dubreuil C, Eché C, Fabioux C, Klopp C, Sarradin P-M, Tanguy A, *et al.* Biological rhythms in the deep-sea hydrothermal mussel bathymodiolus azoricus. *Nat Commun*. 2020:**11**(1):3454. https://doi.org/10.1038/s41467-020-17284-4.
McCartney MA, Auch B, Kono T, Mallez S, Zhang Y, Obille A, Becker A, Abrahante JE, Garbe J, Badalamenti JP, *et al.* The genome of the zebra mussel, Dreissena polymorpha: a resource for comparative genomics, invasion genetics, and biocontrol. *G3 (Bethesda)*. 2022:**12**:jkab423. https://doi.org/10.1093/g3journal/jkab423.
Mei Q, Dvornyk V. Evolutionary history of the photolyase/cryptochrome superfamily in eukaryotes. *PLOS ONE*. 2015:**10**:e0135940. https://doi.org/10.1371/journal.pone.0135940.
Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol*. 2020:**37**:1530–1534. https://doi.org/10.1093/molbev/msaa015.
Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, Tosatto SCE, Paladin L, Raj S, Richardson LJ, *et al.* Pfam: the protein families database in 2021. *Nucleic Acids Res*. 2021:**49**:D412–D419. https://doi.org/10.1093/nar/gkaa913.
Morel B, Kozlov AM, Stamatakis A, Szöllősi GJ. GeneRax: a tool for species-tree-aware maximum likelihood-based gene family tree inference under gene duplication, transfer, and loss. *Mol Biol Evol*. 2020:**37**:2763–2774. https://doi.org/10.1093/molbev/msaa141.
Musilova Z, Cortesi F, Matschiner M, Davies WI, Patel JS, Stieb SM, de Busserolles F, Malmstrøm M, Tørresen OK, Brown CJ, *et al.* Vision using multiple distinct rod opsins in deep-sea fishes. *Science*. 2019:**364**:588–592. https://doi.org/10.1126/science.aav4632.
Nagata T, Koyanagi M, Tsukamoto H, Saeki S, Isono K, Shichida Y, Tokunaga F, Kinoshita M, Arikawa K, Terakita A. Depth perception from image defocus in a jumping spider. *Science*. 2012:**335**:469–471. https://doi.org/10.1126/science.1211667.
Nakane Y, Ikegami K, Ono H, Yamamoto N, Yoshida S, Hirunagi K, Ebihara S, Kubo Y, Yoshimura T. A mammalian neural tissue opsin (Opsin 5) is a deep brain photoreceptor in birds. *Proc Natl Acad Sci U S A*. 2010:**107**:15264–15268. https://doi.org/10.1073/pnas.1006393107.
Niemiller ML, Fitzpatrick BM, Shah P, Schmitz L, Near TJ. Evidence for repeated loss of selective constraint in rhodopsin of amblyopsid cavefishes (teleostei: amblyopsidae). *Evolution*. 2013:**67**:732–748. https://doi.org/10.1111/j.1558-5646.2012.01822.x.
Nilsson D-E. The diversity of eyes and vision. *Annu Rev Vis Sci*. 2021:**7**:19–41. https://doi.org/10.1146/annurev-vision-121820-074736.
Nishiyama H, Nagata A, Matsuo Y, Matsuo R. Light avoidance by a non-ocular photosensing system in the terrestrial slug *Limax valentianus*. *J Exp Biol*. 2019:**222**:jeb208595. https://doi.org/10.1242/jeb.208595.
Ohuchi H, Yamashita T, Tomonari S, Fujita-Yanagibayashi S, Sakai K, Noji S, Shichida Y. A non-mammalian type opsin 5 functions dually in the photoreceptive and non-photoreceptive organs of birds. *PLoS One*. 2012:**7**:e31534. https://doi.org/10.1371/journal.pone.0031534.
Oliveri P, Fortunato AE, Petrone L, Ishikawa-Fujiwara T, Kobayashi Y, Todo T, Antonova O, Arboleda E, Zantke J, Tessmar-Raible K, *et al.* The cryptochrome/photolyase family in aquatic organisms. *Mar Genomics*. 2014:**14**:23–37. https://doi.org/10.1016/j.margen.2014.02.001.
Öztürk N. Phylogenetic and functional classification of the photolyase/cryptochrome family. *Photochem Photobiol*. 2017:**93**:104–111. https://doi.org/10.1111/php.12676.

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msad263/7457512 by guest on 20 December 2023

Öztürk N, Song S-H, Özgür S, Selby CP, Morrison L, Partch C, Zhong D, Sancar A. Structure and function of animal cryptochromes. *Cold Spring Harb Symp Quant Biol*. 2007:**72**:119–131. https://doi.org/10.1101/sqb.2007.72.015.

Penel S, Menet H, Tricou T, Daubin V, Tannier E. Thirdkind: displaying phylogenetic encounters beyond 2-level reconciliation. *Bioinformatics*. 2022:**38**:2350–2352. https://doi.org/10.1093/bioinformatics/btac062.

Perry KJ, Henry JQ. CRISPR/Cas9-mediated genome modification in the mollusc, *Crepidula fornicata*. *Genesis*. 2015:**53**:237–244. https://doi.org/10.1002/dvg.22843.

Pisani D, Rota-Stabelli O, Feuda R. Sensory neuroscience: a taste for light and the origin of animal vision. *Curr Biol*. 2020:**30**:R773–R775. https://doi.org/10.1016/j.cub.2020.05.009.

Plachetzki DC, Oakley TH. Key transitions during the evolution of animal phototransduction: novelty, "tree-thinking," co-option, and co-duplication. *Integr Comp Biol*. 2007:**47**:759–769. https://doi.org/10.1093/icb/icm050.

Porath-Krause AJ, Pairett AN, Faggionato D, Birla BS, Sankar K, Serb JM. Structural differences and differential expression among rhabdomeric opsins reveal functional change after gene duplication in the bay scallop, *Argopecten irradians* (Pectinidae). *BMC Evol Biol*. 2016:**16**:1–15. https://doi.org/10.1186/s12862-016-0823-9.

Porter ML, Awata H, Bok MJ, Cronin TW. Exceptional diversity of opsin expression patterns in *Neogonodactylus oerstedii* (Stomatopoda) retinas. *Proc Natl Acad Sci U S A*. 2020:**117**:8948–8957. https://doi.org/10.1073/pnas.1917303117.

Porter ML, Blasic JR, Bok MJ, Cameron EG, Pringle T, Cronin TW, Robinson PR. Shedding new light on opsin evolution. *Proc R Soc B Biol Sci*. 2011:**279**:3–14. https://doi.org/10.1098/rspb.2011.1819.

Ramirez MD, Oakley TH. Eye-independent, light-activated chromatophore expansion (LACE) and expression of phototransduction genes in the skin of *Octopus bimaculoides*. *J Exp Biol*. 2015:**218**:1513–1520. https://doi.org/10.1242/jeb.110908.

Ramirez MD, Pairett AN, Pankey MS, Serb JM, Speiser DI, Swafford AJ, Oakley TH. The last common ancestor of most bilaterian animals possessed at least nine opsins. *Genome Biol Evol*. 2016:**8**:3640–3652. https://doi.org/10.1093/gbe/evw248.

Raven CP. *The analysis of molluscan development*. Oxford: Pergamon Press; 1966.

Rawlinson KA, Lapraz F, Ballister ER, Terasaki M, Rodgers J, McDowell RJ, Girstmair J, Criswell KE, Boldogkoi M, Simpson F, *et al*. Extraocular, rod-like photoreceptors in a flatworm express xenopsin photopigment. *eLife*. 2019:**8**:1–28. https://doi.org/10.7554/eLife.45465.

Rivera AS, Ozturk N, Fahey B, Plachetzki DC, Degnan BM, Sancar A, Oakley TH. Blue-light-receptive cryptochrome is expressed in a sponge eye lacking neurons and opsin. *J Exp Biol*. 2012:**215**:1278–1286. https://doi.org/10.1242/jeb.067140.

Rovati GE, Capra V, Neubig RR. The highly conserved DRY motif of class A G protein-coupled receptors: beyond the ground state. *Mol Pharmacol*. 2007:**71**:959–964. https://doi.org/10.1124/mol.106.029470.

Sakai K, Ikeuchi H, Fujiyabu C, Imamoto Y, Yamashita T. Convergent evolutionary counterion displacement of bilaterian opsins in ciliary cells. *Cell Mol Life Sci*. 2022:**79**:493. https://doi.org/10.1007/s00018-022-04525-6.

Sancar A. Cryptochrome: the second photoactive pigment in the eye and its role in circadian photoreception. *Annu Rev Biochem*. 2000:**69**:31–67. https://doi.org/10.1146/annurev.biochem.69.1.31.

Schneider TD, Stephens RM. Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res*. 1990:**18**:6097–6100. https://doi.org/10.1093/nar/18.20.6097.

Senthilan PR, Piepenbrock D, Ovezmyradov G, Nadrowski B, Bechstedt S, Pauls S, Winkler M, Möbius W, Howard J, Göpfert MC. Drosophila auditory organ genes and genetic hearing defects. *Cell*. 2012:**150**:1042–1054. https://doi.org/10.1016/j.cell.2012.06.043.

Serb JM, Eernisse DJ. Charting evolution's trajectory: using molluscan eye diversity to understand parallel and convergent evolution. *Evol: Educ Outreach*. 2008:**1**:439–447. https://doi.org/10.1007/s12052-008-0084-1.

Serb JM, Porath-Krause AJ, Pairett AN. Uncovering a gene duplication of the photoreceptive protein, opsin, in scallops (bivalvia: pectinidae). *Integr Comp Biol*. 2013:**53**:68–77. https://doi.org/10.1093/icb/ict063.

Shen WL, Kwon Y, Adegbola AA, Luo J, Chess A, Montell C. Function of rhodopsin in temperature discrimination in *Drosophila*. *Science*. 2011:**331**:1333–1336. https://doi.org/10.1126/science.1198904.

Shichida Y, Matsuyama T. Evolution of opsins and phototransduction. *Philos Trans R Soc B Biol Sci*. 2009:**364**:2881–2895. https://doi.org/10.1098/rstb.2009.0051.

Sikka G, Hussmann GP, Pandey D, Cao S, Hori D, Park JT, Steppan J, Kim JH, Barodka V, Myers AC, *et al*. Melanopsin mediates light-dependent relaxation in blood vessels. *Proc Natl Acad Sci U S A*. 2014:**111**:17977–17982. https://doi.org/10.1073/pnas.1420258111.

Smedley GD, McElroy KE, Feller KD, Serb JM. Additive and epistatic effects influence spectral tuning in molluscan retinochrome opsin. *J Exp Biol*. 2022:**225**:jeb242929. https://doi.org/10.1242/jeb.242929.

Smith SO. Structure and activation of the visual pigment rhodopsin. *Annu Rev Biophys*. 2010:**39**:309–328. https://doi.org/10.1146/annurev-biophys-101209-104901.

Smith SA, Wilson NG, Goetz FE, Feehery C, Andrade SCS, Rouse GW, Giribet G, Dunn CW. Resolving the evolutionary relationships of molluscs with phylogenomic tools. *Nature*. 2011:**480**:364–367. https://doi.org/10.1038/nature10526.

Sondhi Y, Ellis EA, Bybee SM, Theobald JC, Kawahara AY. Light environment drives evolution of color vision genes in butterflies and moths. *Commun Biol*. 2021:**4**:177. https://doi.org/10.1038/s42003-021-01688-z.

Speiser DI, Gagnon YL, Chhetri RK, Oldenburg AL, Johnsen S. Examining the effects of chromatic aberration, object distance, and eye shape on image-formation in the mirror-based eyes of the bay scallop *Argopecten irradians*. *Integr Comp Biol*. 2016:**56**:796–808. https://doi.org/10.1093/icb/icw099.

Speiser DI, Pankey MS, Zaharoff AK, Battelle BA, Bracken-Grissom HD, Breinholt JW, Bybee SM, Cronin TW, Garm A, Lindgren AR, *et al*. Using phylogenetically-informed annotation (PIA) to search for light-interacting genes in transcriptomes from non-model organisms. *BMC Bioinformatics*. 2014:**15**:350. https://doi.org/10.1186/s12859-014-0350-x.

Starcevic A, Akthar S, Dunlap WC, Shick JM, Hranueli D, Cullum J, Long PF. Enzymes of the shikimic acid pathway encoded in the genome of a basal metazoan, *Nematostella vectensis*, have microbial origins. *Proc Natl Acad Sci USA*. 2008:**105**:2533–2537. https://doi.org/10.1073/pnas.0707388105.

Terakita A. The opsins. *Genome Biol*. 2005:**6**:1–9. https://doi.org/10.1186/gb-2005-6-3-213.

Terakita A, Nagata T. Functional properties of opsins and their contribution to light-sensing physiology. *Zool Sci*. 2014:**31**:653–659. https://doi.org/10.2108/zs140094.

Tong D, Rozas NS, Oakley TH, Mitchell J, Colley NJ, McFall-Ngai MJ. Evidence for light perception in a bioluminescent organ. *Proc Natl Acad Sci U S A*. 2009:**106**:9836–9841. https://doi.org/10.1073/pnas.0904571106.

Van Nynatten A, Castiglione GM, de A. Gutierrez E, Lovejoy NR, Chang BSW. Recreated ancestral opsin associated with marine to freshwater croaker invasion reveals kinetic and spectral adaptation. *Mol Biol Evol*. 2021:**38**:2076–2087. https://doi.org/10.1093/molbev/msab008.

Vermeij GJ, Watson-Zink VM. Terrestrialization in gastropods: lineages, ecological constraints and comparisons with other animals. *Biol J Linn Soc*. 2022:**136**:393–404. https://doi.org/10.1093/biolinnean/blac053.

Vicedomini R, Bouly JP, Laine E, Falciatore A, Carbone A. Multiple profile models extract features from protein sequence data and resolve functional diversity of very different protein families. *Mol Biol Evol*. 2022:**39**:msac070. https://doi.org/10.1093/molbev/msac070.

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msad263/7457512 by guest on 20 December 2023

Vizueta J, Sánchez-Gracia A, Rozas J. Bitacora: a comprehensive tool for the identification and annotation of gene families in genome assemblies. *Mol Ecol Resour*. 2020:**20**:1445–1452. https://doi.org/10.1111/1755-0998.13202.

Vöcking O, Kourtesis I, Tumu SC, Hausen H. Co-expression of xenopsin and rhabdomeric opsin in photoreceptors bearing microvilli and cilia. *eLife*. 2017:**6**:e23435. https://doi.org/10.7554/eLife.23435.

Vöcking O, Leclère L, Hausen H. The rhodopsin-retinochrome system for retinal re-isomerization predates the origin of cephalopod eyes. *BMC Ecol Evol*. 2021:**21**:215. https://doi.org/10.1186/s12862-021-01939-x.

Vöcking O, Macias-Muñoz A, Jaeger SJ, Oakley TH. Deep diversity: extensive variation in the components of complex visual systems across animals. *Cells*. 2022:**11**:3966. https://doi.org/10.3390/cells11243966.

Wang S, Zhang J, Jiao W, Li J, Xun X, Sun Y, Guo X, Huan P, Dong B, Zhang L, *et al*. Scallop genome provides insights into evolution of bilaterian karyotype and development. *Nat Ecol Evol*. 2017:**1**:1–12. https://doi.org/10.1038/s41559-017-0120.

Weisman CM, Murray AW, Eddy SR. Mixing genome annotation methods in a comparative analysis inflates the apparent number of lineage-specific genes. *Curr Biol*. 2022:**32**:2632–2639.e2. https://doi.org/10.1016/j.cub.2022.04.085.

Yaguchi J, Yaguchi S. Sea urchin larvae utilize light for regulating the pyloric opening. *BMC Biol*. 2021:**19**:64. https://doi.org/10.1186/s12915-021-00999-1.

Yoshida MA, Ogura A, Ikeo K, Shigeno S, Moritaki T, Winters GC, Kohn AB, Moroz LL. Molecular evidence for convergence and parallelism in evolution of Complex brains of cephalopod molluscs: insights from visual systems. *Integr Comp Biol*. 2015:**55**:1070–1083. https://doi.org/10.1093/icb/icv049.

Yu Z, Fischer R. Light sensing and responses in fungi. *Nat Rev Microbiol*. 2019:**17**:25–36. https://doi.org/10.1038/s41579-018-0109-x.

Yu H, Li H, Li Q, Xu R, Yue C, Du S. Targeted gene disruption in pacific oyster based on CRISPR/cas9 ribonucleoprotein complexes. *Mar Biotechnol*. 2019:**21**:301–309. https://doi.org/10.1007/s10126-019-09885-y.

Zhang S, Zhang S. A new species of the genus phymorhynchus (neogastropoda: raphitomidae) from a hydrothermal vent in the Manus back-arc basin. *Zootaxa*. 2017:**4300**:441. https://doi.org/10.11646/zootaxa.4300.3.7.

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msad263/7457512 by guest on 20 December 2023