

Chapter 3

The molecular evolution of animal
phototransduction and photoreceptor
cells

Abstract

The origin of vision has been a major novelty in animals, playing a fundamental role in the evolution of complex behaviours, such as mate choice and predator avoidance, that distinguish animals from other organisms. Vision starts with a light-triggered phototransduction cascade that occurs in specialised neurons, known as the photoreceptor cells (PRCs). The two main PRC types, ciliary and rhabdomeric, employ specific as well as common genes for phototransduction. While fundamental for vision, the origin and evolution of photoreceptor cells and their phototransduction pathways are still unclear.

Using gene-tree to species-tree reconciliation methods we studied the pattern of gene duplications for all phototransduction genes in more than 80 species, including non-bilaterian metazoans and other eukaryotes. Next, we investigated the expression of phototransduction genes in available single-cell RNA-sequencing data of various animals, including non-bilaterians. Using phototransduction genes as markers, we identified putative photoreceptor-like cells across animals and compared their regulatory genes toolkits.

We found that phototransduction gene families are generally very ancient, predating the origin of vision. Major family expansions and diversifications often occurred just prior to or at the base of animals, reflecting the ability of animals to broaden their responses to the environment. Moreover, we identified putative photoreceptor cells in non-bilaterians expressing some but not all components of the two well characterised phototransduction pathways, suggesting potential lineage-specific components involved in phototransduction. Finally, we found that most of the regulatory genes shared across animal PRCs are transcription factors, with the most predominant families including bZIP TF, zf C2H2 and homeobox. While several regulatory genes are recurrent throughout animals, exact same combinations of these genes rarely span all phyla.

Introduction

Animal evolution has gone hand in hand with an increasing refinement of the ability to sense and respond to the environment. One fundamental sense for animals is vision. While non-visual photoreception is present throughout the tree of life, visual photoreception is an animal-specific trait (Nilsson 2009; Nilsson 2013). At a molecular level, the visual process begins with the reception of light by a photosensitive molecule. This light-activated molecule in turn triggers a chain of molecular signalling within the cell that culminates into ion channel opening/closing resulting in electrical signalling. This phototransduction process occurs within specialised neurons called photoreceptor cells (Nilsson 2009).

The photosensitive molecule is composed of an opsin, a membrane bound G-protein coupled receptor (GPCR), and a light-sensitive chromophore bound to it (Terakita 2005). This chromophore, the retinal, derives from the metabolism of vitamin A. In dark, the retinal is in its 11-cis state. When hit by light photons, it isomerizes into its all-trans state, inducing the structural change in the opsin that in turn initiates the phototransduction pathway (Terakita 2005; Palczewski and Kiser 2020; Widjaja-Adhi and Golczak 2020).

Two alternative phototransduction cascades have been described in detail. In *Drosophila melanogaster* (Figure 3.1A), the opsin activates a Gq-type G protein. The alpha subunit detaches from the complex and activates phospholipase C beta that initiates a phosphoinositide cascade. This results in the opening of transient receptor potential (trp) and trp-like (trpl) channels with consequent depolarization of the cell (Wang and Montell 2007; Hardie and Juusola 2015). Whereas in vertebrates, as exemplified by *Homo sapiens* (Figure 3.1B), the opsin activates transducin (Gt) a G protein of the Gi/o-type that activates phosphodiesterase 6 (PDE6) that hydrolyses cyclic GMP. The drop in cGMP levels causes the cyclic nucleotide gated ion channels (CNGCs) to close, followed by a hyperpolarization of the cell (Lamb 2020). Some molecular components are shared between both pathways, whilst others are specific to either one or the other pathway (Figure 3.1 and Table 3.1). Reconstructing the evolutionary history of each phototransduction gene family is necessary to understand when the complete phototransduction pathways originated and may have started to acquire their visual function.

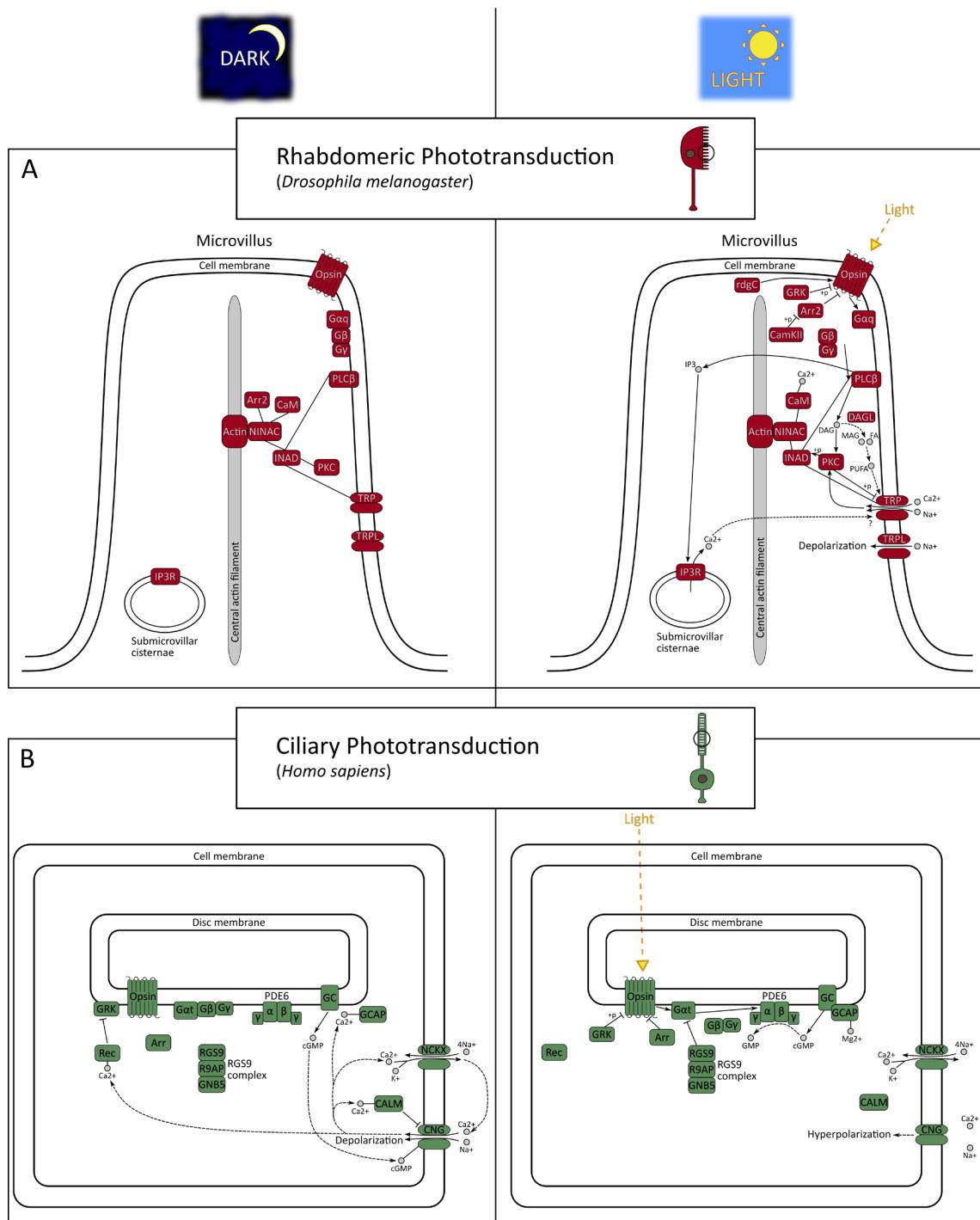


Figure 3.1. Schematics of rhabdomeric and ciliary phototransduction pathways. (A) Rhabdomeric phototransduction in *Drosophila melanogaster*. This cascade occurs in the microvilli of the rhabdomere, a structure within the cell body of the photoreceptor cell. The opsin interacts with a G alpha q that activates phospholipase C (PLC) initiating a phosphoinositide cascade that culminates in depolarisation of the photoreceptor cell. (B) Ciliary phototransduction in *Homo sapiens*. This cascade occurs in a specialised cilium of the photoreceptor cell. The opsin activates the G alpha of transducin that in turn activates phosphodiesterase 6 with consequent cascade that causes the hyperpolarization of the photoreceptor cell. In rod photoreceptors, the opsin and the other membrane proteins, with the exception of the ion channels, are in the membrane of the disk as depicted here. In cone photoreceptors, whilst the components and the cascade are the same, all membrane components are in the cell membrane (not depicted here). The pathways are based primarily on the Kegg maps ko04745 (rhabdomeric) and ko04744 (ciliary). Additional references were (Hardie and Juusola 2015) for *D. melanogaster* phototransduction and (Lamb 2020) for *H. sapiens*.

Protein components are coloured in red (rhabdomeric pathway) or green (ciliary pathway), ions and other non-protein molecules are represented by small grey circles. Lines between components indicate physical interaction, normal arrows between components indicate activation, normal arrows through channels indicate passage of ions, inhibitory arrows indicate inactivation, dotted arrows indicate movement/transition towards, +p indicates phosphorylation, -p indicates de-phosphorylation, ? indicates unclear mechanism.

The two phototransduction pathways occur in different subtypes of photoreceptor cells (PRCs). Rhabdomeric PRCs utilise the phosphoinositide pathway while the ciliary PRCs use the phosphodiesterase 6 pathway (Arendt 2003). Both cell types occur throughout Metazoa (Horridge 1964; Hattar et al. 2002; Arendt 2003; Nordström et al. 2003; Arendt et al. 2004; Kozmik et al. 2008; Passamaneck et al. 2011; Ullrich-Lüter et al. 2011; Jékely et al. 2015; Tamm 2016; von Döhren and Bartolomaeus 2018; Picciani et al. 2018; Valencia et al. 2021). The homology of the two photoreceptor cell types is still under debate, as is the question of the ancestral state in the ancestor to all animals (Arendt 2008; Arendt et al. 2016). While the identification of PRCs throughout animals has traditionally relied on morphological studies, now, the availability of single-cell RNA sequencing allows us to explore the presence of these cell-types in a growing number of organisms.

In this study we explored the evolutionary history of the molecular components essential for vision by first reconstructing the evolution of the genes involved in the two major phototransduction pathways; and then identifying PRC-like cell-types throughout animals and comparing their genetic profiles.

Table 3.1. All phototransduction components with respective gene and protein names. Common components are listed for both *Drosophila melanogaster* and *Homo sapiens*. Rhabdomeric components are listed for *D. melanogaster* and ciliary components are listed for *Homo sapiens*. The gene and protein names are based on FlyBase, GeneCards and UniProt.

	Component (Gene Family)	<i>Drosophila melanogaster</i>		<i>Homo sapiens</i>	
		Gene Name(s)	Protein Name(s)	Gene Name(s)	Protein Name(s)
Common	Opsin	ninaE	Opsin Rh1	OPN1LW	Long-wave-sensitive opsin 1
		Rh 2 to 7	Opsins Rh2 to 7	OPN1SW	Short-wave-sensitive opsin 1
	G beta	Gbeta76C	Gbeta76C (or Gbe)	OPN1MW	Medium-wave-sensitive opsin 1
		Ggamma30A	Ggamma(e)	RHO	Rhodopsin
	G gamma			GNB 1 to 4	G protein subunit beta 1 to 4
				GNGT1	G protein G(T) subunit gamma-T1
	Calmodulin	Cam	Calmodulin (or CaM)	GNGT2	G protein G(T) subunit gamma-T2
	Arrestin	Arr1	Phosrestin-2 (or Arrestin-1)	CALM 1 to 3	Calmodulin 1 to 3
		Arr2	Phosrestin-1 (or Arrestin-2)	ARR3	Arrestin-C
	GRK	Gprk1	GPCR kinase 1	SAG	S-arrestin
	GRK7			GRK1	Rhodopsin kinase GRK1
	GRK7			GRK7	Rhodopsin kinase GRK7
Rhabdomeric	Component (Gene Family)	<i>Drosophila melanogaster</i>			
		Gene Name(s)	Protein Name(s)		
	G alpha q	Galphiq	G protein alpha q subunit		
	PLC	norpA	Phosphoinositide phospholipase C-beta		
	PKC	inaC	Protein kinase C, eye isoform (or Eye-PKC)		
	INAD	inaD	Inactivation-no-after-potential D protein		
	MYO3	ninaC	Neither inactivation nor afterpotential protein C		
	Actin	Act5C	Actin-5C		
	TRP C	trp	Transient receptor potential protein		
		trpl	Transient-receptor-potential-like protein		
	IP3R-A	Itpr	Inositol 1,4,5-trisphosphate receptor (or IP3R)		
	CamKII	CaMKII	Calcium/calmodulin-dependent protein kinase		
	DAGL	inaE	Inactivation no afterpotential E		
	PPEF	rdgC	Serine/threonine-protein phosphatase rdgC (or Retinal degeneration C protein)		
Ciliary	Component (Gene Family)	<i>Homo sapiens</i>			
		Gene Name(s)	Protein Name(s)		
	G alpha i	GNAT1	G protein G(t) subunit alpha 1		
		GNAT2	G protein G(t) subunit alpha 2		
	PDE6 A/B/C	PDE6A	Rod cGMP-specific 3',5'-cyclic phosphodiesterase subunit alpha		
		PDE6B	Rod cGMP-specific 3',5'-cyclic phosphodiesterase subunit beta		
		PDE6C	Cone cGMP-specific 3',5'-cyclic phosphodiesterase subunit alpha'		
	PDE6 G/H	PDE6G	Retinal rod rhodopsin-sensitive cGMP 3',5'-cyclic phosphodiesterase subunit gamma		
		PDE6H	Retinal cone rhodopsin-sensitive cGMP 3',5'-cyclic phosphodiesterase subunit gamma		
	GC2	GUCY2D	Retinal guanylyl cyclase 1 (or GC2D)		
		GUCY2F	Retinal guanylyl cyclase 2 (or GC2F)		
	GCAP	GUCA1A	Guanylyl cyclase-activating protein 1 (GCAP1)		
		GUCA1B	Guanylyl cyclase-activating protein 2 (GCAP2)		
		GUCA1C	Guanylyl cyclase-activating protein 3 (GCAP3)		
	CNG	CNGA 1 to 4	cGMP-gated cation channel alpha 1 to 4		
		CNGB 1 and 3	Cyclic nucleotide-gated cation channel beta 1 and 3		
	NCKX	SLC24A 1, 2 and 4	Sodium/potassium/calcium exchanger 1, 2 and 4 (or NCKX1,2 and4)		
	Recoverin	RCVRN	Recoverin		
	RGS9	RGS9	Regulator of G-protein signaling 9		
	RGS9BP	RGS9BP	Regulator of G-protein signaling 9-binding protein (RGS9BP or RGBP) or RGS9-anchoring protein (R9AP)		
	GNB5	GNB5	G protein subunit beta 5		

Results and Discussion

Extended gene families of phototransduction components are generally broadly distributed throughout Eukarya.

To study the evolution of the phototransduction cascades, we first investigated the presence of each phototransduction component in 86 eukaryotic species (Table 3.2). We focused on early-branching animals and sister groups to animals, but also included in our search a balanced sampling of all major eukaryotic groups (see Supplementary Table S3.1 with source information). The phototransduction components examined were based primarily on the Kegg maps ko04745 (*D. melanogaster* rhabdomeric cascade) and ko04744 (*H. sapiens* ciliary cascade). The data mining was carried out with a combination of sequence similarity and protein motif analyses. We then constructed maximum likelihood phylogenetic trees and gene tree to species tree reconciliations for each gene family (see Methods for details). Gene tree to species tree reconciliations were performed both with ctenophore-first and sponge-first scenarios and comparison of total number of events (duplications and losses) revealed that overall, there were no major differences between the two scenarios (Supplementary Table S3.2).

Most gene families examined were broad, therefore, within each gene family we focused on identifying the sub-group containing the *D. melanogaster* and/or *H. sapiens* genes that are known to function in the phototransduction cascades. We were interested in understanding whether early-branching animals and non-animal species might possess genes that could perform in phototransduction, and in many cases this might include non-orthologous but related genes, therefore, we expanded our definition of the group of interest to include a broader set of genes within an orthogroup of interest. We found that while the specific orthogroup of interest is often present only within animals or in sister-groups to animals (Figure 3.2), closely related sub-groups were present in the next related species, and when considering the extended gene family as a whole, the distribution would often span Eukarya. This adds an extra layer of detail to our knowledge of when exactly the functional phototransduction pathways might have originated.

		Clades			Species	% Complete BUSCOs (tot) (eukaryota_odb10)	
Amorphea	Obazoa	Opisthokonta	Holozoa	Bilateria	Arthropoda	<i>Drosophila melanogaster</i> 100.00%	>90%
					Nematoda	<i>Caenorhabditis elegans</i> 97.70%	>80%
					Tardigrada	<i>Strigamia maritima</i> 92.60%	>70%
					Mollusca	<i>Daphnia pulex</i> 97.70%	>60%
					Brachiopoda	<i>Pristionchus pacificus</i> 80.40%	>50%
					Annelida	<i>Loa loa</i> 96.10%	<50%
					Bryozoa	<i>Ramazzottius varieornatus</i> 92.50%	
					Vertebrata	<i>Octopus bimaculoides</i> 92.90%	
					Urochordata	<i>Lottia gigantea</i> 96.50%	
					Cephalochordata	<i>Capitella teleta</i> 94.20%	
Diaphoreticks	CRuMs	Archaeplastida	Nucleomycota (Holomycota)	Metazoa	Echinodermata	<i>Helobdella robusta</i> 97.30%	
					Hemichordata	<i>Bugula neritina</i> 54.50%	
					Placozoa	<i>Homo sapiens</i> 100.00%	
					Cnidaria	<i>Mus musculus</i> 100.00%	
					Porifera	<i>Danio rerio</i> 99.60%	
					Ctenophora	<i>Epipatretus burgeri</i> 85.90%	
					Choanoflagellata	<i>Ciona intestinalis</i> 97.30%	
					Ichthyosporea	<i>Branchiostoma belcheri</i> 96.90%	
					Fungi	<i>Acanthaster planci</i> 91.40%	
					Rotsphaerida	<i>Strongylocentrotus purpuratus</i> 96.00%	
Diaphoreticks	Breviata	Amoebozoa	Apusomonadida		Apusomonadida	<i>Saccoglossus kowalevskii</i> 94.10%	
					Breviata	<i>Acropora digitifera</i> 42.80%	
					Amoebozoa	<i>Acropora tenuis</i> 27.80%	
					Amoebozoa	<i>Astreopora sp</i> 69.80%	
					Amoebozoa	<i>Porites australiensis</i> 81.60%	
					Amoebozoa	<i>Fungia scutaria</i> 84.40%	
					Amoebozoa	<i>Montastraea cavernosa</i> 80.40%	
					Amoebozoa	<i>Madracis auretenra</i> 68.30%	
					Amoebozoa	<i>Stylophora pistillata</i> 85.10%	
					Amoebozoa	<i>Anthopleura elegantissima</i> 62.00%	
Diaphoreticks	CRuMs	Archaeplastida	Nucleomycota (Holomycota)		Amoebozoa	<i>Nematostella vectensis</i> 93.40%	
					Amoebozoa	<i>Gorgia ventinala</i> 56.50%	
					Amoebozoa	<i>Clytia hemisphaerica</i> 84.70%	
					Amoebozoa	<i>Hydra magnapapillata</i> 70.90%	
					Amoebozoa	<i>Aurelia sp</i> 67.40%	
					Amoebozoa	<i>Trichoplax adhaerens</i> 96.10%	
					Amoebozoa	<i>Holitunga hongkongensis</i> 96.80%	
					Amoebozoa	<i>Amphimedon queenslandica</i> 92.50%	
					Amoebozoa	<i>Haliclona tubifera</i> 75.30%	
					Amoebozoa	<i>Ephydatia muelleri</i> 93.40%	
Diaphoreticks	CRuMs	Archaeplastida	Nucleomycota (Holomycota)		Amoebozoa	<i>Styliissa carteri</i> 43.60%	
					Amoebozoa	<i>Leucosolenia complicata</i> 94.90%	
					Amoebozoa	<i>Sycon ciliatum</i> 97.30%	
					Amoebozoa	<i>Oscarella pearsei</i> 94.90%	
					Amoebozoa	<i>Mnemiopsis leidyi</i> 83.60%	
					Amoebozoa	<i>Pleurobrachia bachei</i> 47.50%	
					Amoebozoa	<i>Acanthoeca spectabilis</i> 93.30%	
					Amoebozoa	<i>Helgoeca nana</i> 94.10%	
					Amoebozoa	<i>Diaphanoeca grandis</i> 92.20%	
					Amoebozoa	<i>Didymoea costata</i> 94.90%	
Diaphoreticks	CRuMs	Archaeplastida	Nucleomycota (Holomycota)		Amoebozoa	<i>Choanoeca perplexa</i> 92.10%	
					Amoebozoa	<i>Monosiga brevicollis</i> 78.80%	
					Amoebozoa	<i>Mylnosiga fluctuans</i> 93.30%	
					Amoebozoa	<i>Salpingoeca kvevrii</i> 94.90%	
					Amoebozoa	<i>Amoebiidium parasiticum</i> 89.90%	
					Amoebozoa	<i>Sphaeroforma arctica</i> 63.20%	
					Amoebozoa	<i>Sphaerothecum destruens</i> 68.60%	
					Amoebozoa	<i>Capsaspora owczarzaki</i> 93.70%	
					Amoebozoa	<i>Corallochytrium limacisporum</i> 90.20%	
					Fungi	<i>Fusarium oxysporum</i> 97.30%	
Diaphoreticks	CRuMs	Archaeplastida	Nucleomycota (Holomycota)		Fungi	<i>Saccharomyces cerevisiae</i> 93.30%	
					Fungi	<i>Ustilago maydis</i> 98.80%	
					Fungi	<i>Mortierella elongata</i> 98.40%	
					Fungi	<i>Rhizopus microsporus</i> 97.30%	
					Fungi	<i>Spizellomyces punctatus</i> 94.90%	
					Fungi	<i>Parvularia atlantis</i> 75.70%	
					Fungi	<i>Fonticula alba</i> 71.00%	
					Fungi	<i>Thecamonas trahens</i> 81.60%	
					Fungi	<i>Pygsia biforma</i> 62.40%	
					Fungi	<i>Dictyostelium discoideum</i> 94.10%	
Diaphoreticks	CRuMs	Archaeplastida	Nucleomycota (Holomycota)		Fungi	<i>Vermamoeba vermiciformis</i> 89.00%	
					Fungi	<i>Cunea sp</i> 76.50%	
					Fungi	<i>Diphylleia rotans</i> 90.60%	
					Fungi	<i>Rigifila ramosa</i> 86.30%	
					Rhodelphis	<i>Rhodelphis marinus</i> 89.80%	
					Rhodelphis	<i>Galdieria sulphuraria</i> 77.30%	
					Rhodelphis	<i>Chloroplatista thaliana</i> 99.60%	
					Cryptophysaceae	<i>Guillardia theta</i> 79.60%	
					Haptophyta	<i>Phaeocystis globosa</i> 75.30%	
Diaphoreticks	CRuMs	Archaeplastida	Nucleomycota (Holomycota)		Rhizaria	<i>Plasmiodiophora brassicae</i> 89.00%	
					Stramenopiles	<i>Phytophthora infestans</i> 90.20%	
					Alveolata	<i>Colponymida sp</i> 93.30%	
					Euglenozoa	<i>Euglena longa</i> 83.50%	
					Heterolobosea	<i>Pharyngomonas kirbyi</i> 85.10%	

Table 3.2. List of eukaryotic species used for the phylogenetic analysis of phototransduction gene families. The respective percentages of total complete BUSCO genes are indicated. BUSCO was conducted using species proteomes versus the eukaryota_odb10 database.

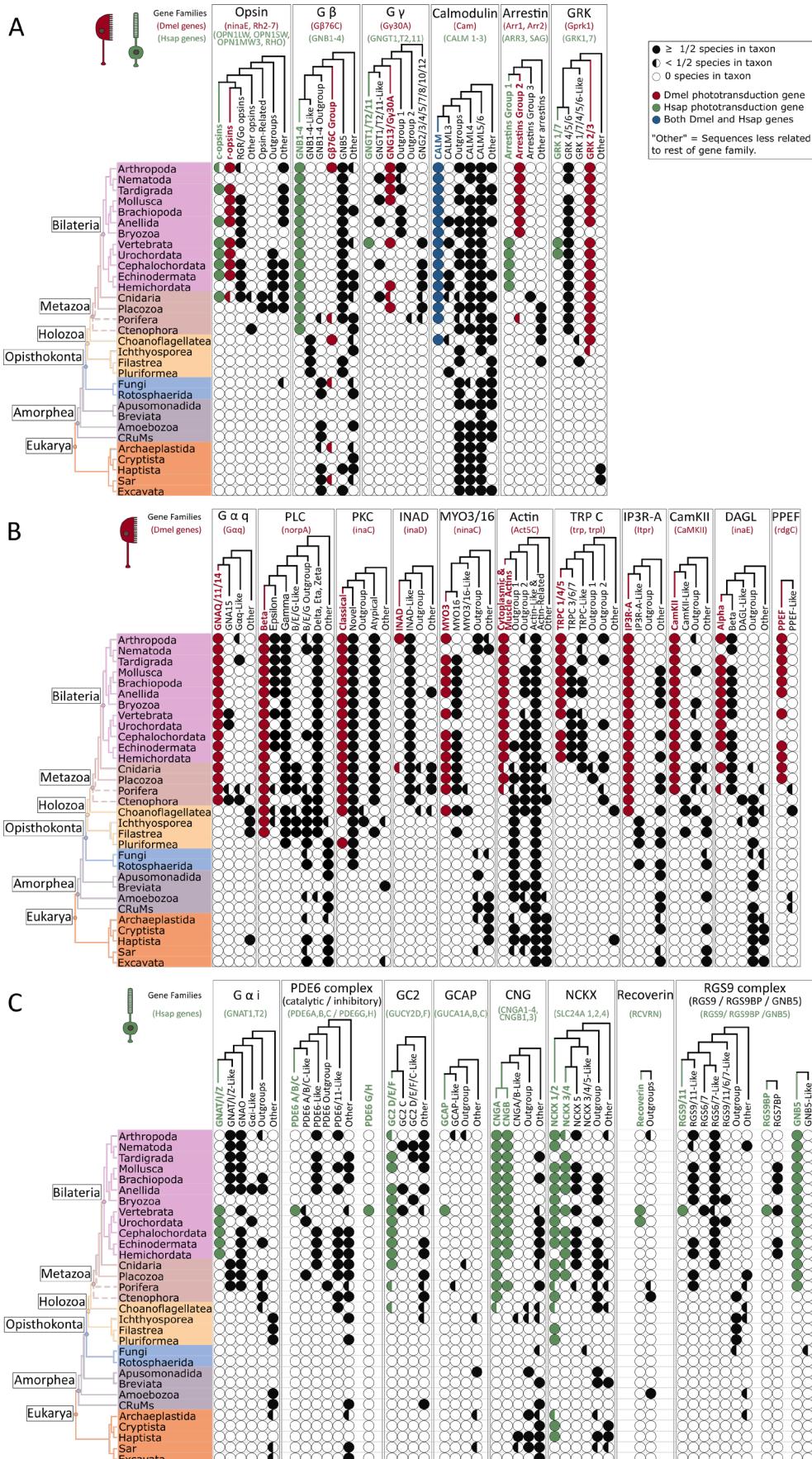


Figure 3.2. Evolutionary history of phototransduction components gene families and distribution across Eukarya. We reconstructed the evolution of each gene family for all common (A), rhabdomeric-specific (B) and ciliary-specific (C) components and we mapped their distribution across all major groups of Eukarya. For each gene family, we obtained a gene tree based on maximum likelihood phylogenetic

trees and gene tree to species tree reconciliations. Most gene families examined were broad, therefore, within each gene family tree we highlight the branch containing the *D. melanogaster* and/or *H. sapiens* gene that is known to function in the phototransduction pathway. When mapping the presence/absence of the phototransduction components throughout the tree of eukaryotes, we distinguish for each gene family whether the presence refers to the specific orthogroup of interest or to any of the other related sub-lineages within the broad gene family. While the specific-orthogroup of interest is often present only within animals or in sister-groups to animals, we detected numerous cases in which organisms more distantly related to animals possessed related genes within the broad gene family.

Common phototransduction components

Amongst the common components (Figure 3.2A), the orthogroup of interest is present either strictly in animals (opsins, G gamma and arrestin) or within Holozoa (G beta, calmodulin, GRK). However, if we consider the extended gene family, then G beta, calmodulin and GRK span all Eukarya, arrestin is present up to Holozoa, and only opsin and G gamma remain animal-specific. While for opsin this reflects the expected scenario (Fleming et al. 2020), for G gamma, an animal-oriented definition of the gene family (e.g. during protein motif filtering, see Methods) may have resulted in the exclusion of non-animal G gamma-types. Indeed G gamma is the least studied subunit of the G protein and its subtypes outside of animals are not well characterised (Krishnan et al. 2015). As the other two subunits of the G protein (G beta and G alpha) are present outside of animals, it is possible that a non-animal G gamma-type exists but was not detected here. Of note, in Figure 3.2, the G alpha family is divided into two subfamilies corresponding to those used in fly rhabdomeric (Gq family) (Hardie and Juusola 2015) and vertebrate ciliary (Gi/o family) (Lagman et al. 2012) phototransduction, as these are the two pathways that were used here as reference. However, we also conducted a comprehensive analysis of the full G alpha family that includes other important subfamilies such as Gs (see supplementary reconciliation files on GitHub).

Rhabdomeric-specific phototransduction components

Within the gene families of the rhabdomeric-specific components (Figure 3.2B), the orthogroup of interest is animal-specific for seven out of eleven gene families, with the remaining four families being holozoan-specific. It is therefore of striking contrast that the extended gene families are all present throughout Eukarya, except for INAD (inactivation no afterpotential D) that appears to be restricted to animals and choanoflagellates. The presence outside of Holozoa is perhaps questionable also for G

alpha q and TRP C, however, overall rhabdomeric extended gene families appear to be ancient.

Ciliary-specific phototransduction components

The majority of the ciliary-specific orthogroups of interest (Figure 3.2C), are also animal-specific (eight out of eleven). Two components are present also in Holozoa, while only NCKX, a sodium-calcium-potassium exchanger involved in numerous other pathways (Altimimi and Schnetkamp 2007), is present throughout Eukarya. The situation is dramatically different if you consider the extended gene families, as in this case nine families are present in Eukarya and only two remain animal-specific. Of further note for the ciliary components, in contrast to common and rhabdomeric components that within animals are more or less distributed in all or most phyla, the ciliary components often have a patchy presence also within animals, with vertebrates being the only group that contains all the gene families. This indicates that some of the components of the ciliary pathway we used as reference are likely vertebrate innovations, while other components are more ancient and represent the core part of the cascade. An example of this can be seen for the PDE6 complex. The alpha/beta subunits belong to the same protein family and constitute the essential catalytic subunits of the complex, while the gamma subunits are accessory inhibitory subunits that have been described only in vertebrate PDE6 (Lagman et al. 2016; Lamb 2020). Here, our result confirms the notion that PDE6 gamma subunits are a vertebrate novelty.

Patterns of major duplication, speciation and loss events clarify gene family expansions.

Our approach of reconciling the gene trees to the species tree not only allowed us to define the orthogroups of interest, but also revealed the specific patterns of duplication, speciation and loss events that characterise the lineage of the orthogroup of interest and all other lineages in the gene family. Here, we discuss the key findings for a few gene families of particular interest.

GPCR Kinases: an ancient family that expands in Metazoa

An interesting case amongst the common phototransduction components is that of the G-protein-coupled receptor kinases (GRK) (Figure 3.3A). This family has an ancient origin with presence in some distantly related eukaryotes, however, it is characterised by a series of key duplications just prior to and at the base of animals that gave rise to the various sub lineages of interest for either rhabdomeric or ciliary phototransduction (Figure 3.3A). In photoreceptor cells, the GRKs are essential for the inactivation phase of phototransduction. The light-activated visual pigment is capable of activating hundreds of G proteins (Shichida and Matsuyama 2009). To avoid the signal to continue long after the original light stimulus occurred, the visual pigment must be shut-off (Wang and Montell 2007; Lamb et al. 2018). After shut-off, photoreceptors have to recover their pre-illumination state and the quicker this occurs, the more they can adjust to rapidly changing lighting conditions (Orban and Palczewski 2016). GRKs, protein kinases of the serine/threonine protein kinases superfamily, phosphorylate target GPCRs facilitating the binding of arrestin to the GPCR (Mushegian et al. 2012; Orban and Palczewski 2016). The arrestin-capped GPCR is blocked from interacting with its G-protein. Therefore, GRKs initiate the desensitisation of GPCRs and deactivation of GPCR signalling (Gurevich and Gurevich 2016; Orban and Palczewski 2016).

In vertebrates there are seven GPCR kinases (GRK 1-7), and the ones involved in phototransduction shut-off are GRK1 (rods) and GRK7 (cones) (Lamb et al. 2018; Lamb 2020). The fruit fly *Drosophila melanogaster* possesses two GRK genes, Gprk1 and Gprk2. The one that is involved in phototransduction shut-off is Gprk1, which is more closely related to GRK 2/3 (Lee et al. 2004; Wang and Montell 2007). Overall within Metazoa, the GRK family is split into two major clades: one clade includes GRK 2 and 3; while the other contains all other GRKs and in turn is composed of two subgroups, one with GRK 1 and 7 and the other with GRK 4, 5, and 6 (Mushegian et al. 2012). An extensive phylogenetic analysis of GRKs (Mushegian et al. 2012) previously found that GRKs are an ancient family that arose well before Metazoa. In that study, the authors concluded that the GRK family underwent a first split into GRK 2/3 type and GRK 1/7+4/5/6 type at some point before the advent of animals within the history of opisthokonts. Further expansions occurred later within animals, likely to reflect the greater need for rapid signalling to adapt to the surrounding environment (Mushegian et al. 2012).

With our much broader set of eukaryotic lineages examined, our focus on early-branching animals and sister groups of animals, and our gene tree to species tree reconciliation, we were able to expand our knowledge of the evolution of the GRK family adding further details compared to (Mushegian et al. 2012). In accordance with previous results, the duplication that gives rise to the GRK 1 and 7 sub-groups is at the split between urochordates and vertebrates (see supplementary files with the full reconciliation for GRK on GitHub). The GRK 4/5/6 sub-groups all derive from two subsequent duplications at the base of jawed vertebrates. Interestingly, the split between GRK 1/7 and GRK 4/5/6 appears much more ancient than expected as it derives from a gene duplication at the base of Metazoa. This holds true in both ctenophore-first and sponge-first scenarios (Figure 3.3A). Whilst the GRK 4/5/6 lineage is widespread throughout animals, the GRK 1/7 lineage seems to have been lost in all animal groups except in Olfactores (urochordates and vertebrates) and potentially in ctenophores, according to the sponge-first scenario only (Figure 3.3A).

The duplication that gave rise to the split between GRK 2/3 and GRK 1/7+4/5/6 occurred at the base of Holozoa (Figure 3.3A). Therefore, the closest relatives to Metazoa inherited both lineages, as previously proposed (Mushegian et al. 2012). Although we too see that several holozoans lost either one or the other lineage as described previously, our larger taxonomic sampling allowed us to clarify that at least within choanoflagellates, both lineages were originally present, contrary to what previously thought (Mushegian et al. 2012).

Finally, outside of Holozoa, GRKs are not present in other opisthokonts (e.g., fungi) nor in any other Amorphea group (e.g., Amoebozoa). An orthologous lineage to the GRKs 1-7 is instead present in the other major eukaryotic branch, the Diaphoretickes (Figure 3.3A). However, the presence is limited to a small subset of groups, namely the SAR and Haptophyta (see supplementary reconciliation files).

Gene tree to species tree reconciliations under either ctenophore-first or sponge-first scenarios provided the same overall results, with one minor exception: in ctenophore-first scenario, the GRK 1/7 lineage is present only in Olfactores and the ctenophore branch includes only GRK 2/3 and GRK 4/5/6; instead in the sponge-first scenario, the GRK 1/7 lineage is present also in the ctenophore branch, that has lost GRK 4/5/6 (but retained GRK 2/3) (Figure 3.3A).

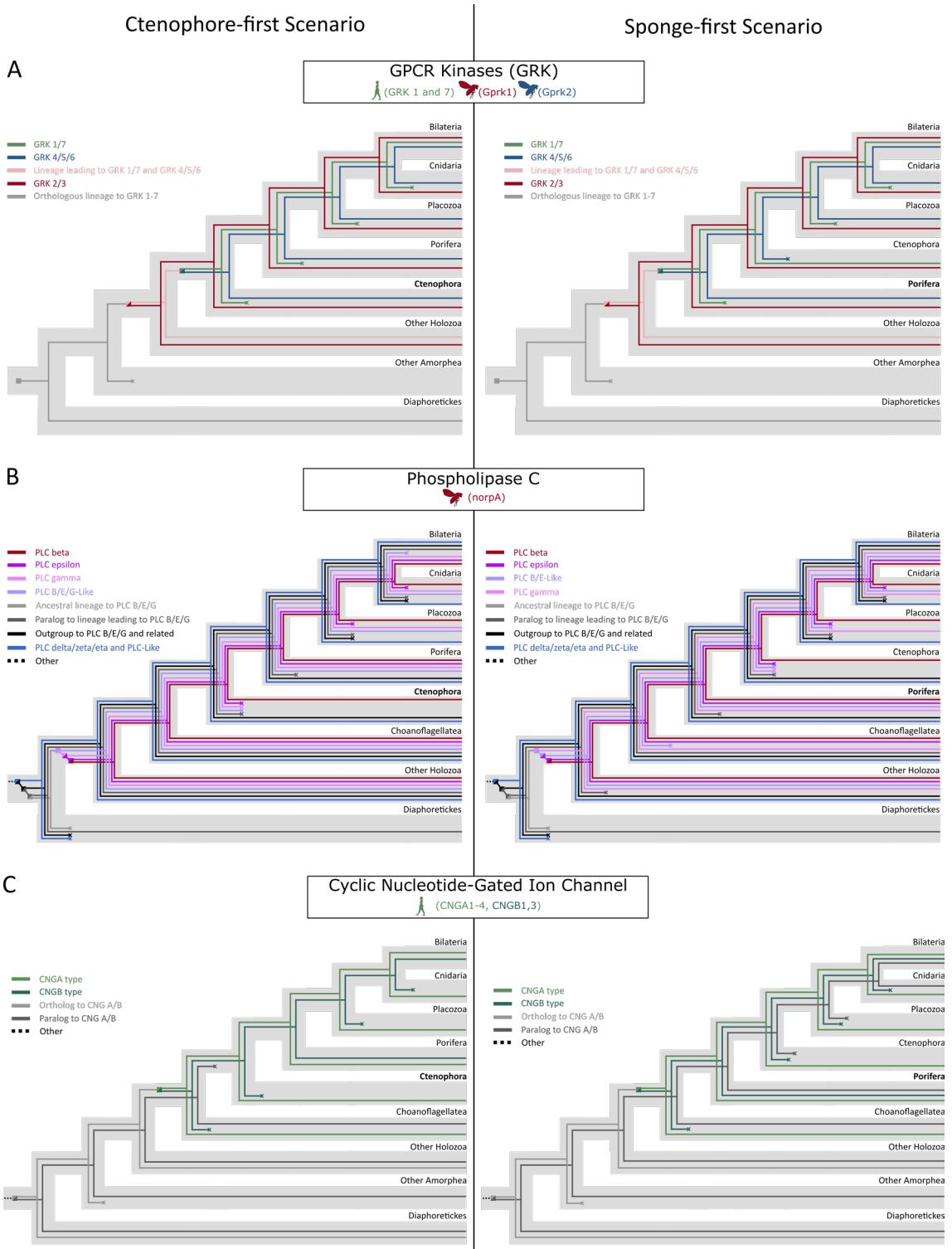


Figure 3.3. Major events of duplication, speciation, and losses for three phototransduction gene families of interest. Reconciliations were constructed under both ctenophore-first and sponge-first scenarios and no major differences were found. **(A)** GPCR Kinases (GRK) are important for the shut-off of light response in both rhabdomeric and ciliary phototransduction. The gene family has an ancient eukaryotic origin, however the key duplication events that gave rise to the diversity of the family present in animals, occurred just prior to animals and at the base of animals. The lineage that gives rise to the *Drosophila melanogaster* gene Gprk1 that is used in rhabdomeric phototransduction derives from a duplication at the base of Holozoa. While a duplication at the base of animals gave rise to the lineage that includes the human GRK1 and GRK7 involved in ciliary phototransduction. **(B)** The phospholipase C

(PLC) is important for the initial steps of the rhabdomeric phototransduction. It is a very broad family of enzymes that includes many subgroups. The *Drosophila* gene NorpA involved in phototransduction is a PLC type beta. This lineage, like most others in the family, derived from a duplication at the base of Holozoa. (C) The cyclic nucleotide gated ion channels (CNG) are responsible for the hyperpolarization of vertebrate photoreceptor cells at the end of the signal cascade. It is again a very ancient family and the two subunits, alpha and beta, that compose vertebrate CNG channels originated from a duplication at the split between choanoflagellates and animals.

Phospholipase C: Holozoan origin of the beta subgroup from an ancient eukaryotic family

As mentioned, most rhabdomeric gene families have an ancient origin. An example of a gene family with an extensive repertoire of sub lineages deriving from very ancient gene duplications is the family of phospholipases of type C (PLC) (Figure 3.3B). PLCs are a broad family of enzymes that catalyse the hydrolysis of the phospholipid PIP2 into DAG and IP3, that both function as second messengers (Suh et al. 2008). In *Drosophila* a PLC of type beta is the one used in phototransduction. During the phototransduction cascade, IP3 interacts with its receptor (IP3R) on the endoplasmic reticulum, causing the release of calcium, and DAG goes on to activate the eye-specific protein kinase C (PKC) that is involved in the deactivation of the visual cascade (Wang and Montell 2007; Hardie and Juusola 2015).

Although PLCs have been described throughout Eukarya (Tsutsui et al. 1995; Koyanagi et al. 1998; Rebecchi and Pentyala 2000; Mikami 2014; Wang et al. 2020), few studies have looked into their evolution. In mammals there are 6 subgroups of PLCs: beta; gamma; delta; epsilon; zeta; and eta (Suh et al. 2008). Candidate beta-type and gamma-type PLCs have been cloned in the sponge *Ephydatia fluviatilis* and a delta-like in the cnidarian *Hydra magnipapillata* (Koyanagi et al. 1998). While the PLCs in fungi (and plants) have been described as similar to delta-type (Rebecchi and Pentyala 2000). A comprehensive phylogenetic analysis of the family is lacking.

Our data mining recovered for Human, the 13 known PLCs belonging to the 6 subgroups plus two inactive PLC-Like sequences; and for *Drosophila*, the PLC beta used in phototransduction, encoded by NorpA, plus two other PLCs: PLC21C and small wing (sl). Gene tree to species tree reconciliation revealed that *Drosophila* NorpA arises from a duplication at the base of Cnidaria+Bilateria and that from the same duplication arises Human PLCbeta4 (see supplementary files with the full reconciliation for PLC on GitHub). Instead, *Drosophila* PLC21C is more related to Human PLCbeta1/2/3, and their

lineage originates with a duplication at the base of Metazoa. A prior duplication at the same species node is the one that separates the PLC21C + PLCbeta1/2/3 on the one hand from the NorpA + PLCbeta4 on the other. These duplication patterns are consistent between ctenophore-first and sponge-first scenarios. Several additional duplications for the PLC beta lineage also occur at the base of Metazoa (in both ctenophore-first and sponge-first scenarios), indicating that PLC beta underwent a great expansion at the base of Metazoa (see supplementary reconciliation files). The origin of the PLC beta lineage is from a duplication at the base of Holozoa where its direct paralog lineage is the PLC epsilon (Figure 3.3B). At the same species node, a previous duplication gave rise to the PLC beta/epsilon lineage on the one hand and the PLC gamma on the other. The position of PLC epsilon as sister group to PLC beta is recovered with both ctenophore-first and sponge-first scenarios. This is a novel insight into the evolution of PLC subfamilies, as PLC beta has been considered to be related to gamma and delta (Rebecchi and Pentyala 2000), while here we show that its closest relative seems to be epsilon. Our data shows that PLC beta/epsilon/gamma are more related to each other than to the other PLCs including PLC delta (Figure 3.3B). This clarification can be crucial, especially when trying to identify possible candidate genes involved in a putative rhabdomeric-like phototransduction pathway in non-model organisms such as non-bilateria. Tracing backwards the lineage of PLC beta/epsilon/gamma, uncovers that it originates from a duplication at the base of Eukaryotes (Figure 3.3B). Here at this species node, there are multiple other duplications, including the one that gives rise to the lineage of all the other subgroups of PLCs (delta, zeta, eta) known in mammals. These major subgroup relationships remain consistent between ctenophore-first and sponge-first scenarios (Figure 3.3B).

Cyclic Nucleotide Gated Ion Channels: ancient origin of alpha and beta subtypes

Amongst the ciliary phototransduction components the cyclic nucleotide gated ion channels (CNGs) gene family is one of the ones with the broadest distribution across Eukarya (Figure 3.2C and Figure 3.3C). CNGs belong to the broad family of voltage-gated ion channels (Anderson and Grenberg 2001) and function in response to the binding of cyclic nucleotides. They are non-selective cation channels through which the passage of Ca²⁺ ions in particular is of importance for the excitation of sensory cells (Kaupp and Seifert 2002).

During phototransduction the drop of cyclic GMP, caused by its hydrolysis by phosphodiesterase (PDE), induces the closure of CNG channels which in turn causes the hyperpolarization of the photoreceptor cell. Apart from this role in the activation of phototransduction, CNG channels are also involved in the Ca²⁺-feedback regulation of the cascade and thus in photoreceptor light adaptation (Kaupp and Seifert 2002).

The ion channel complex is composed of two groups of subunits, alpha and beta. Jawed vertebrates possess six genes encoding for CNG subunits: CNGA1-4 encode for four alpha subunits while CNGB1 and CNGB3 encode for beta subunits (Kaupp and Seifert 2002, Lamb 2020). The ion channel complex consists in the combination of four subunits around a pore. Native rod channels consist of three alpha1 (CNGA1) and one beta1 (CNGB1) subunits, while cone channels comprise two alpha3 (CNGA3) and two beta3 (CNGB3) subunits. Subunits alpha2 (CNGA2) and alpha4 (CNGA4) together with beta1 (CNGB1) are instead used in CNG channels of olfactory receptor neurons. Phylogenetic and gene synteny analyses led (Lamb 2020) to the reconstruction that the gene lineages of alpha and beta subunits derived from a duplication that occurred before the split of protostomes and deuterostomes (Lamb 2020). Likewise, CNGA4 split from the other branch of CNGA that later gave rise to CNGA1-3, prior to the protostome-deuterostome split. The authors speculate that the ancestral CNG channel was composed of two alpha and two beta subunits (Lamb 2020).

Outside of vertebrates, homologs to the CNG genes have been found in the nematode *Caenorhabditis elegans*, the fruit fly *Drosophila melanogaster* and the horseshoe crab *Limulus polyphemus*, where likely they are involved in chemosensation (Kaupp and Seifert 2002). Amongst early branching animals, CNGs have been found in the cnidarian *Hydra magnipapillata* where it is implicated in phototransduction (Plachetzki et al. 2010). CNG channels are in fact not confined to animals as they are present also in plants (Saand et al 2015) and prokaryotes (Brams et al 2014, Napolitano et al 2021). However, while much attention has been given to the evolution of the CNG genes within and at the base of vertebrates, not much is known about the ancient evolutionary history of this gene family and the relationship between animal and non-animal CNG lineages.

Our phylogenetic analysis and gene tree to species tree reconciliation of the CNG family revealed that the alpha and beta gene lineages derive from a gene duplication at the split between choanoflagellates and animals (Figure 3.3C). This remains constant whether the species tree used for reconciliation is ctenophore-first or sponge-first. Although it was

already hypothesised that this gene duplication was ancient (Lamb 2020), it had not yet been clarified when it had occurred precisely.

According to our reconstructions, while the alpha lineage seems to be present in all major animal groups and choanoflagellates, the beta lineage seems to be present only in Bilateria and sponges (Figure 3.3C).

The orthologous lineage to the CNG alpha/beta lineage is present in other holozoan species and in some Diaphoretickes. More distantly related CNG genes are present throughout Eukarya, but not in animals, according to the ctenophore-first reconciliation. While in the sponge-first reconciliation, this group of less related CNGs appears to be present also in Porifera, Cnidaria and Bilateria.

Identification of putative photoreceptor cells throughout animals.

To understand the origin and early evolution of vision, we must understand not only when a functional phototransduction pathway evolved, but also in which cell type it started to function.

Our detailed analysis of phototransduction gene family evolution with clarifications of the relationships amongst sub lineages, allowed us to compile a list of best candidate phototransduction genes for every species. These were used as markers to identify candidate photoreceptor cells (PRCs) from the available single-cell RNA sequencing data of a variety of animal species. We focused our investigation on twelve species that spanned Metazoa with particular emphasis on early-branching animals. *Drosophila melanogaster* was used as representative of rhabdomeric PRCs; *Homo sapiens* and *Mus musculus* as representatives of ciliary PRCs. The urochordate *Ciona intestinalis* and the sea urchin *Strongylocentrotus purpuratus* were used as bridge species between protostomes and vertebrates. Finally, amongst non-bilaterians we investigated the cnidarians *Hydra vulgaris*, *Clytia hemisphaerica*, *Stylophora pistillata* and *Nematostella vectensis*; the placozoan *Trichoplax adhaerens*, the sponge *Amphimedon queenslandica* and the ctenophore *Mnemiopsis leidyi*. A comprehensive list of scRNAseq data sources and sample details for each species are in Table 3.3.

Bilateria	<i>D. mel</i>		Adult optic lobe	Ozel et al. 2020. Nature.
	<i>H. sap</i>		Adult retina	Lukowski et al. 2019. The EMBO Journal.
	<i>M. mus</i>		Juvenile retina	Macosko et al 2015. Cell.
	<i>C. int</i>		Late larvae brain	Sharma et al. 2019. Developmental Biology.
	<i>S. pur</i>		3-day whole larvae	Paganos et al. 2021. Elife.
Cnidaria	<i>N. vec</i>		Adult whole organism	Sebe-Pedros et al. 2018. Cell.
	<i>S. pis</i>		Adult whole organism	Levy et al. 2021. Cell.
	<i>C. hem</i>		Adult whole organism	Chari et al. 2021. Science Advances.
	<i>H. vul</i>		Adult whole organism	Siebert et al 2019. Science.
Placozoa	<i>T. adh</i>		Adult whole organism	Sebe-Pedros et al. 2018. Nat Ecol Evol.
Porifera	<i>A. que</i>		Adult whole organism	Sebe-Pedros et al. 2018. Nat Ecol Evol.
Ctenophora	<i>M. lei</i>		Adult whole organism	Sebe-Pedros et al. 2018. Nat Ecol Evol.

Table 3.3. Datasets used for the single cell analyses. Single cell RNA sequencing data from 12 species were used for the single cell analyses. Phylogenetic relationships, as well as developmental stage, tissue type and source are indicated for each species. Focus was given on non-bilaterian species. Species silhouettes are images with CC0 1.0 Universal Public Domain Dedication licences obtained from <https://www.phylopic.org/>. Abbreviations: *D. mel*: *Drosophila melanogaster*; *H. sap*: *Homo sapiens*; *M. mus*: *Mus musculus*; *C. int*: *Ciona intestinalis*; *S. pur*: *Strongylocentrotus purpuratus*; *N. vec*: *Nematostella vectensis*; *S. pis*: *Stylophora pistillata*; *C. hem*: *Clytia hemisphaerica*; *H. vul*: *Hydra vulgaris*; *T. adh*: *Trichoplax adhaerens*; *A. que*: *Amphimedon queenslandica*; *M. lei*: *Mnemiopsis leidyi*.

While for *D. melanogaster*, *H. sapiens* and *M. musculus* photoreceptor cells are well characterised, and for *C. intestinalis*, *S. purpuratus* and some species of cnidaria photoreceptors have at least been reported, for other species the presence of photoreceptors is unknown. Moreover, when searching for putative homologous cell types to the PRCs in these species, it is uncertain whether they might possess a more rhabdomic-like or ciliary-like profile. Therefore, we developed a pipeline, described in detail in the Methods section, to identify PRC-like “metacells” or cell states based on phototransduction gene expression.

The presence/absence of phototransduction genes, whether belonging to the best orthogroup or to another related lineage, provides some form of evidence to understand the diversity of PRC-like profiles amongst animals (Figure 3.4).

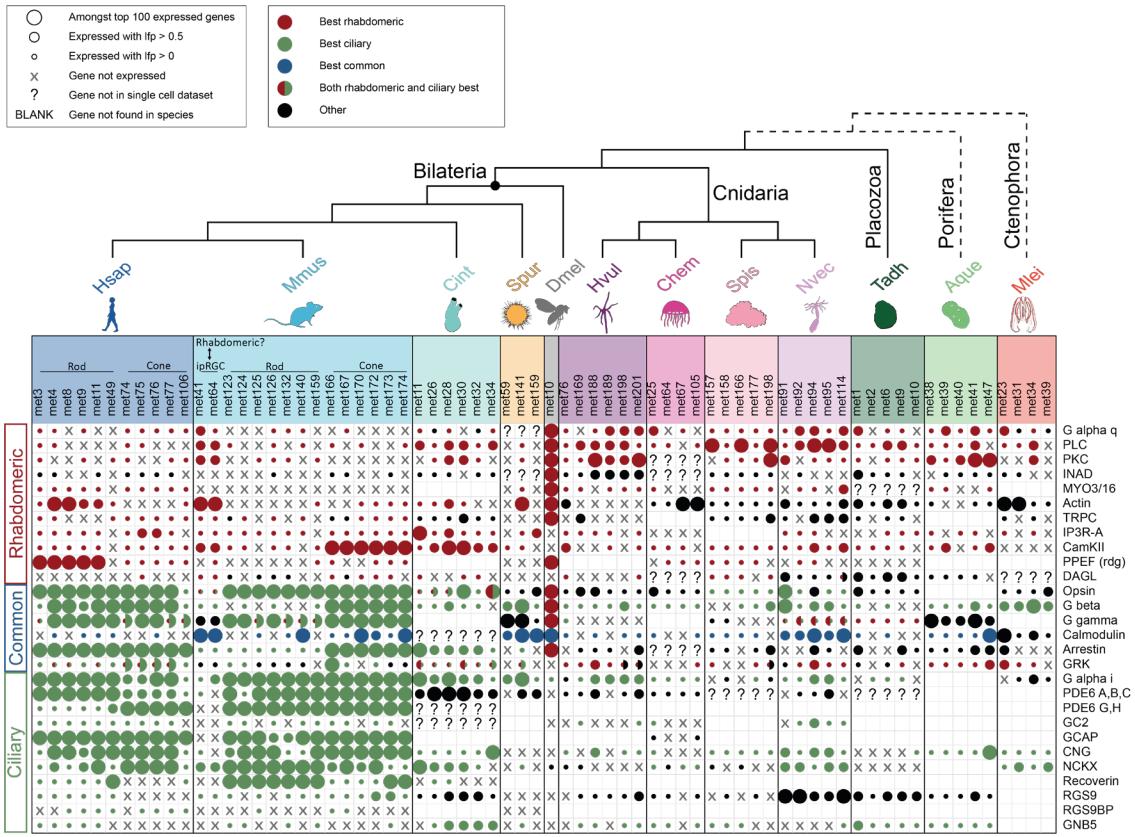


Figure 3.4. Expression of phototransduction genes in photoreceptor-like cells across animals. The single-cell RNA sequencing analysis identified putative PRC-like metacells across all the species examined, including all non-bilaterian phyla. Human and mouse ciliary PRCs express mainly ciliary type genes, but also some rhabdomeric type ones. Instead, *Drosophila* PRC expresses almost exclusively rhabdomeric type genes. Two candidate ipRGC (intrinsically photosensitive retinal ganglion cells) in mouse display a rhabdomeric-type profile. *Ciona intestinalis* metacells appear to have ciliary-like profiles. Outside of chordates, a large amount of phototransduction genes is either not present in the genome or not detected in the scRNAseq data, and, overall, most species have a mixture of rhabdomeric and ciliary genes expressed. Species silhouettes were modified from images with CC0 1.0 Universal Public Domain Dedication licences obtained from <https://www.phylopic.org/>. Abbreviations: *D. mel*: *Drosophila melanogaster*; *H. sap*: *Homo sapiens*; *M. mus*: *Mus musculus*; *C. int*: *Ciona intestinalis*; *S. pur*: *Strongylocentrotus purpuratus*; *N. vec*: *Nematostella vectensis*; *S. pis*: *Stylophora pistillata*; *C. hem*: *Clytia hemisphaerica*; *H. vul*: *Hydra vulgaris*; *T. adh*: *Trichoplax adhaerens*; *A. que*: *Amphimedon queenslandica*; *M. lei*: *Mnemiopsis leidyi*.

Ciliary and Rhabdomeric PRCs in model organisms

As expected, *D. melanogaster* PRC type expresses rhabdomeric phototransduction genes while *H. sapiens* and *M. musculus* express ciliary genes (Figure 3.4). The *Drosophila* PRC metacell possesses a rhabdomeric-exclusive profile, with the only ciliary gene expressed being the G alpha of type i/o. All other ciliary genes were either not detected in the *Drosophila* genome or are not expressed in its PRC metacell. In contrast, *H. sapiens* and *M. musculus* metacells, while having a clear ciliary-oriented profile, still have a significant amount of rhabdomeric genes expressed, albeit at a lower level compared to ciliary genes. This could reflect the possibility that ciliary photoreceptor cells, whilst

using the ciliary pathway for phototransduction, may contemporarily employ rhabdomeric-like signalling either to modulate the phototransduction or, alternatively, to perform unrelated tasks, as previously proposed (Yau and Hardie 2009).

Furthermore, it has been proposed that melanopsin (OPN4) expressing cells in the vertebrate retina are the homologous cell type to rhabdomeric photoreceptor cells (Provencio et al. 2000; Hattar et al. 2002; Arendt 2003; Rollag et al. 2003; Fu et al. 2005). Specifically, a subclass of retinal ganglion cells, called intrinsically photosensitive retinal ganglion cells (ipRGC), are known to use OPN4 to mediate non-visual forming light sensing functions, e.g., circadian entrainment and pupillary light reflex (Hahn et al. 2023). The human OPN4 was not detected in the human retina single cell dataset used in this study, so no candidate rhabdomeric profile could be identified. Instead in the mouse dataset, two metacells were found to express the mouse OPN4 (metacells 41 and 64) (Figure 3.4). Both also expressed another ipRGC marker, the transcription factor EOMES (Hahn et al. 2023) (the list of all genes with their respective lfp values is available on GitHub). Interestingly, these two mouse metacells express a lot less ciliary genes compared to other mouse metacells. Metacell 64 in particular is missing some of the key genes involved in the ciliary pathway, including all the PDE6 subunits and the CNG channel. Additionally, metacells 41 and 64 are the mouse metacells that express the highest number of rhabdomeric genes, with metacell 64 expressing all except two rhabdomeric genes. These results suggest that based on phototransduction genes these two ipRGC metacells have a rhabdomeric profile.

***C. intestinalis* and *S. purpuratus* PRC metacells**

The two deuterostome invertebrates examined here have both been reported to possess photoreceptor cells. The sea squirt *Ciona intestinalis* is known to possess a ciliary-type PRC (Eakin and Kuda 1970; Ryan et al. 2016). The sea urchin *Strongylocentrotus purpuratus* has been reported to have both rhabdomeric-type (Ullrich-Lüter et al. 2011) and ciliary-type (Valencia et al. 2021) PRCs. For both species the expression of phototransduction genes provided somewhat mixed results (Figure 3.4).

C. intestinalis metacells express both some rhabdomeric and some ciliary genes, with the common components being predominantly of ciliary type. However, many genes were either not found in the genome or not detected in the single cell data, so cannot be assessed. If focusing on the opsins, then the majority of the metacells express only c-opsins, while some express contemporarily c-opsins and r-opsins. In this sense our results

are consistent with the literature that has described a ciliary type PRC based on morphology (Eakin and Kuda 1970; Ryan et al. 2016). Whereas we are unable to exclude or to suggest the possibility of the presence of a rhabdomeric-type PRC profile.

Similarly, in *S. purpuratus* several genes are missing either from the genome or from the single cell data (Figure 3.4). However, compared to *Ciona*, in the sea urchin there are also many genes that are present in the genome and the single cell data but that are not expressed in the PRC-like metacells. Of note we were only able to identify 3 PRC-like metacells in the sea urchin, likely due to the fact that of all the opsins expressed in the genome, only two opsins were detected in the single cell data. For example, neither Sp-Opsin-4, an r-opsin described to be expressed in candidate rhabdomeric cells (Ullrich-Lüter et al. 2011), nor Sp-Opsin-3.2, a Go-opsin expressed in candidate ciliary cells (Valencia et al. 2021), were detected in the single cell dataset. The opsins that are in the single cell dataset (Sp-Opsin2 and Sp-Opn5L) are echinoderm-specific echinopsins (D'Aniello et al. 2015) that according to our phylogenetic analysis fall in the broad lineage of RGR/Go opsins (see supplementary files with the full reconciliation for opsins on GitHub). While they likely initiate a functioning phototransduction cascade, it is not certain whether it could be a rhabdomeric or ciliary pathway.

Photoreceptor-like metacells in non-bilateria

PRC-like in Cnidaria

Amongst all the non-bilaterian phyla, the Cnidaria are the only group in which there is clear evidence of the presence of photoreceptor cells (Piatigorsky and Kozmik 2004; Kozmik et al. 2008; Vöcking et al. 2022) and of which some components of the phototransduction cascade have been described (Plachetzki et al. 2010; Gornik et al. 2021). The results from our analysis revealed that although several phototransduction genes were missing in the genomes/transcriptomes and/or in the single cell data of cnidarian species, overall, this phylum seems to have the most complete repertoire of phototransduction components compared to other non-bilateria (Figure 3.4). Furthermore, having examined four species, we were able in part to compensate for absences in single species. In general, there is no clear-cut distinction between rhabdomeric profile or ciliary profile. *Stylophora pistillata* and *Nematostella vectensis* both express ciliary type opsins, while *Hydra vulgaris* and *Clytia hemisphaerica* express opsins that are RGR/Go type according to our phylogenetic analysis (see supplementary files with the full reconciliation). The opsin expression may suggest a potentially more

ciliary-like profile as has been suggested (Plachetzki et al. 2010). However, the overall difficulty in distinguishing between rhabdomeric and ciliary profile may reflect a growing view that cnidaria possess a different pathway, that while sharing some components with the two traditional cascades, also includes cnidaria-specific elements yet to be characterised (Vöcking et al. 2022).

PRC-like in Placozoa

The placozoan *Trichoplax adhaerens* has a very simple body plan in which only a handful of cell types have been described morphologically (Smith et al. 2014), although molecular studies have uncovered a broader diversity (Sebé-Pedrós, Chomsky, et al. 2018; Varoqueaux et al. 2018). While *Trichoplax* seems to have at least some basic response to light (Heyland et al. 2014), there is no morphological evidence of the presence of photoreceptor cells, furthermore, it does not possess *bona fide* opsins, but rather phylogenetically related placopsins (Feuda et al. 2012). Bearing this in mind, here our goal was to test whether we could at least find any PRC-like profile that could be further explored as candidate homologous cell type to PRCs, whether or not it may indeed have a role in light response. Our analysis of single cell data (see methods) highlighted 5 candidate metacells (Figure 3.4). Interestingly, from the *Trichoplax* genome we identified all rhabdomeric genes and these were all detected in the single cell data except one. This is in contrast to the ciliary genes, of which only a handful were present in the genome. Although this asymmetry complicates the comparison between potential rhabdomeric versus ciliary profiles, it is important to note that most rhabdomeric components are expressed in the *Trichoplax* PRC-like metacells. Further functional exploration of this cascade could therefore be of relevance in the future.

PRC-like in Porifera

Although sponges lack opsins and, like placozoans, do not possess neurons, they are known to be receptive to light (Leys and Degnan 2001; Maldonado et al. 2003; Elliott and Leys 2004; Wong et al. 2022). It has been proposed that sponges may utilise the light sensitive cryptochrome (Rivera et al. 2012; Müller et al. 2013), or even other GPCRs such as glutamate receptors (Wong et al. 2022), instead of opsins for photoreception. Furthermore, in *Amphimedon queenslandica* two rhabdomeric phototransduction genes have been implicated in phototactic behaviour of the larvae (Wong et al. 2022), further suggesting the existence of a phototransduction pathway and potentially a photoreceptor

cell type in these animals. From our phylogenetic analysis, we found that a couple of rhabdomeric genes and most ciliary genes were missing from the *Amphimedon* genome. Overall, this species, together with the ctenophore (see below), is the one with fewest phototransduction genes recovered in the genome. In the PRC-like metacells that we recovered from the single cell analysis, the few ciliary genes found in the genome are all expressed, as are all the common genes, and most of the rhabdomeric genes. Due to the paucity of ciliary genes in comparison to the rhabdomeric genes, we may be inclined to suggest that a rhabdomeric-like profile is predominant. However, like for cnidaria, it could be that sponges utilise some components of the classic phototransduction cascades alongside more lineage-specific components.

PRC-like in Ctenophora

In the ctenophore *Mnemiopsis leidyi*, a morphologically ciliary-type photoreceptor cell (Horridge 1964; Tamm 2016) has been reported. Although PRCs are not entirely characterised, ctenophores are generally considered to be more likely to possess PRCs compared to placozoans and sponges, as they have neurons and complex behaviours that include predation (Jékely et al. 2015). Importantly, *M. leidyi* has opsins ((Schnitzler et al. 2012) and this study) which is another clue that there might be functional PRCs, although we do not know if the phototransduction pathway used might be similar to one of the already described ones or could be independent. Here we find 4 candidate metacells (Figure 3.4). As many phototransduction genes, especially ciliary ones, were missing from the genome, it is difficult to make strong conclusions. Although more rhabdomeric genes were present in the genome compared to ciliary genes, in the PRC-like metacells, the few ciliary genes are almost entirely expressed, in contrast to the rhabdomeric genes that are expressed in less metacells. The most extreme case is in metacell 39 that expresses all three ciliary genes it has in the genome but only two of the eight rhabdomeric genes available in the genome. A previous study (Schnitzler et al. 2012) reported to have found many ciliary phototransduction genes in *Mnemiopsis leidyi*, in contrast to only a handful of rhabdomeric genes. Overall that study reported more phototransduction genes than the ones we report here, however, their data mining was exclusively based on BLAST, with phylogenetic analysis dedicated only to the opsin gene, so likely some of those genes were filtered out in our more rigorous phylogenetic analysis. In any case, their conclusion that *M. leidyi* PRCs have a ciliary type phototransduction is compatible with our results,

although we caution that possibly ctenophores have some alternative specific components in their cascade.

Shared regulatory toolkit of PRC-Like metacells throughout animals.

The putative PRCs we have identified throughout animals were based on the expression of phototransduction genes. This helped us to identify cells that may have the molecular machinery to perform phototransduction and are therefore similar to known PRCs at least from a potentially functional perspective. However, to explore the potential homology amongst cell types across species, we must focus on the core regulatory complex of the cells, namely the set of genes, such as transcription factors, that regulate the expression of other genes and determine the cell identity (Arendt et al. 2016).

We collected a list of orthogroups of regulatory genes that are differentially expressed in each of the PRC-like metacells (see Methods) and used this information to further understand relationships amongst metacells across species.

Orthogroups of regulatory genes

A total of 806 EggNog orthogroups (see Methods) for regulatory genes were identified amongst the highly expressed genes of the PRC-like metacells of the 12 species examined. On average, each species possessed more orthogroups that were shared with at least one other species rather than species-specific ones (Figure 3.5), suggesting some degree of communal regulatory profile amongst PRCs across animals. Orthogroups that were species-specific were often metacell-specific within that species as well (Figure 3.6). Metacell-specific orthogroups are unlikely to be indicative of a universal core PRC cell profile, therefore, these were discarded from further analyses. This left us with 421 orthogroups that were shared across at least two metacells (Figure 3.7). Of these, 286 orthogroups were present in at least two species (Extended Figure 3.8A, available on GitHub), 219 were present in at least 2 phyla and 69 are in 3 or more phyla (Figure 3.8A).

Structure of relationships amongst PRC-like metacells

As a first step in comparing the regulatory toolkit of animal PRC-like metacells, we investigated how many orthogroups were shared and by which metacells (Figure 3.7A). We found that whilst the highest number of shared genes is between PRCs of the same or

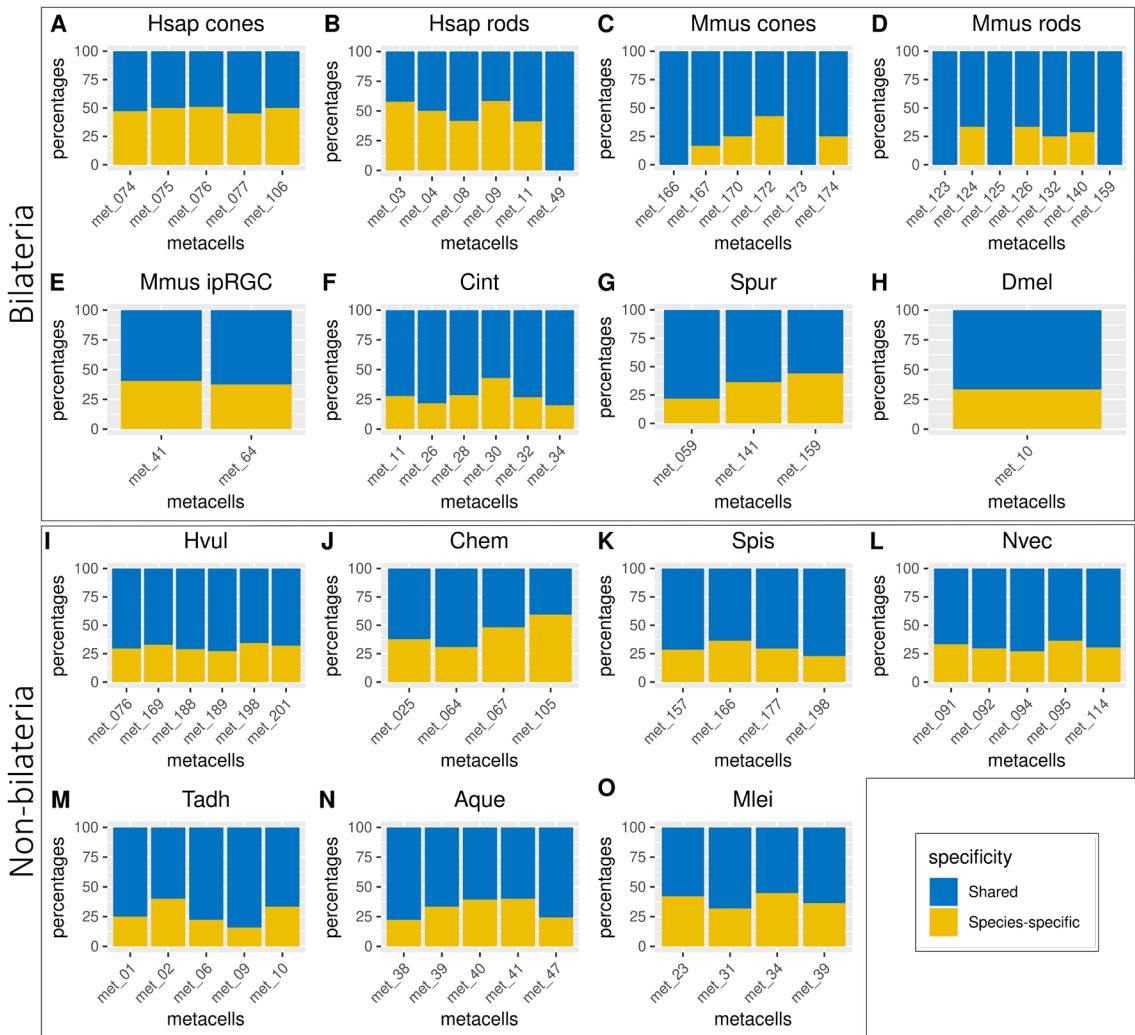


Figure 3.5. Shared versus species-specific orthogroups of regulatory genes in PRC-like metacells throughout animals. For each PRC-like metacell of each species a list of the regulatory genes expressed was compiled. Regulatory genes orthogroups were defined through EggNog (see Methods). Cross-species comparison revealed that the majority of the orthogroups examined are shared with at least one other species. This was true both in bilaterian and in non-bilaterian animals. Abbreviations: *D. mel*: *Drosophila melanogaster*; *H. sap*: *Homo sapiens*; *M. mus*: *Mus musculus*; *C. int*: *Ciona intestinalis*; *S. pur*: *Strongylocentrotus purpuratus*; *N. vec*: *Nematostella vectensis*; *S. pis*: *Stylophora pistillata*; *C. hem*: *Clytia hemisphaerica*; *H. vul*: *Hydra vulgaris*; *T. adh*: *Trichoplax adhaerens*; *A. que*: *Amphimedon queenslandica*; *M. lei*: *Mnemiopsis leidyi*.

closely related species, there are still several shared genes also amongst distantly related species. To have a broader understanding of the relationships amongst PRCs of different species, we constructed a network to visualise connections amongst metacells based on the number of shared transcriptional genes (Figure 3.7B). We built this network using only the genes that were amongst the top 100 highly expressed genes for each metacells, as this would provide us with higher confidence connections (see Methods). The strength of the network approach is both to obtain an overview of metacell relationships and to identify indirect connections that are otherwise difficult to spot. The network of all

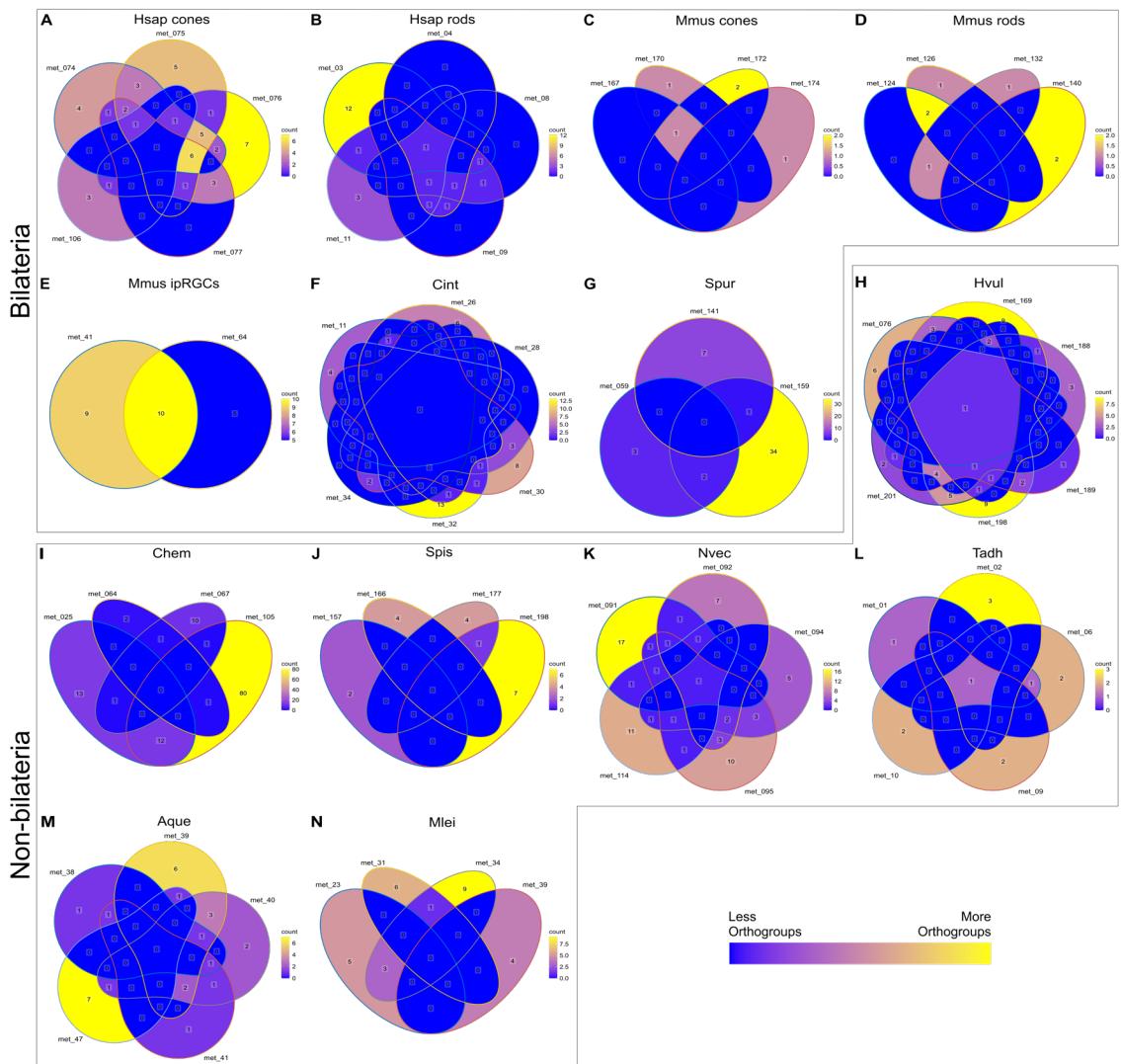


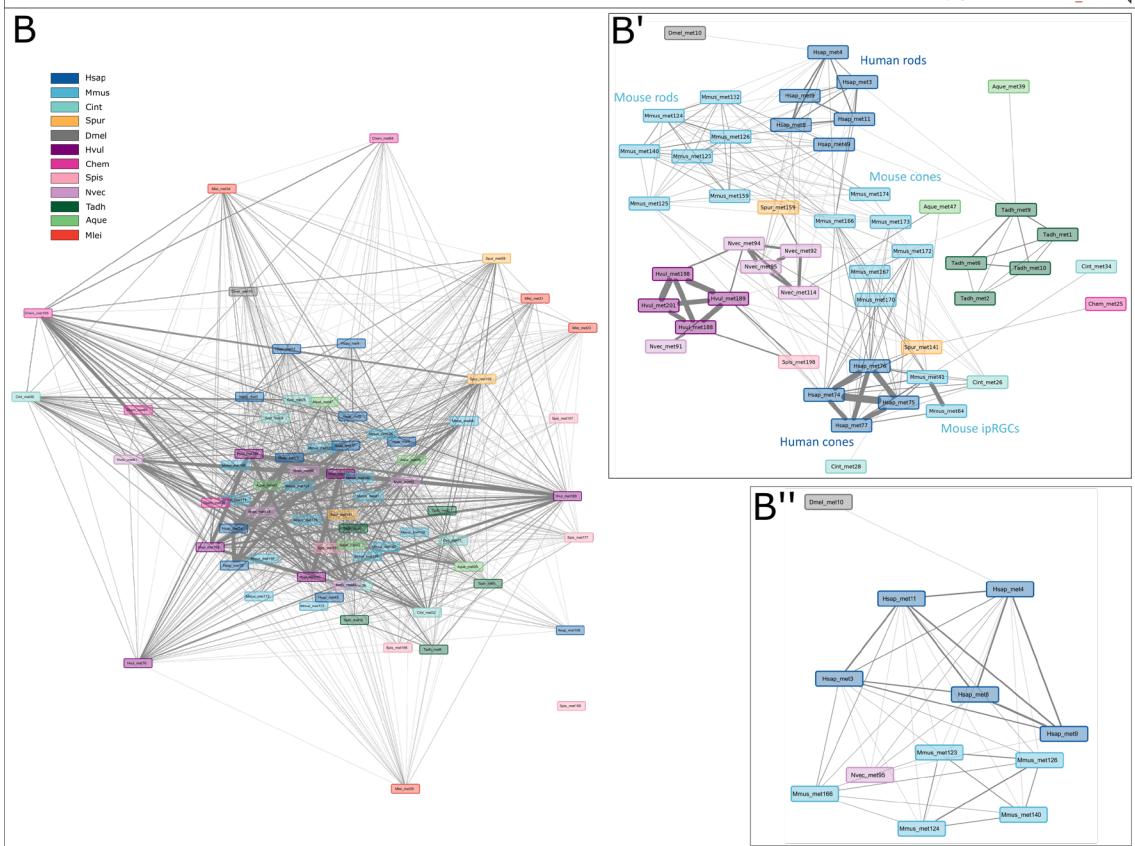
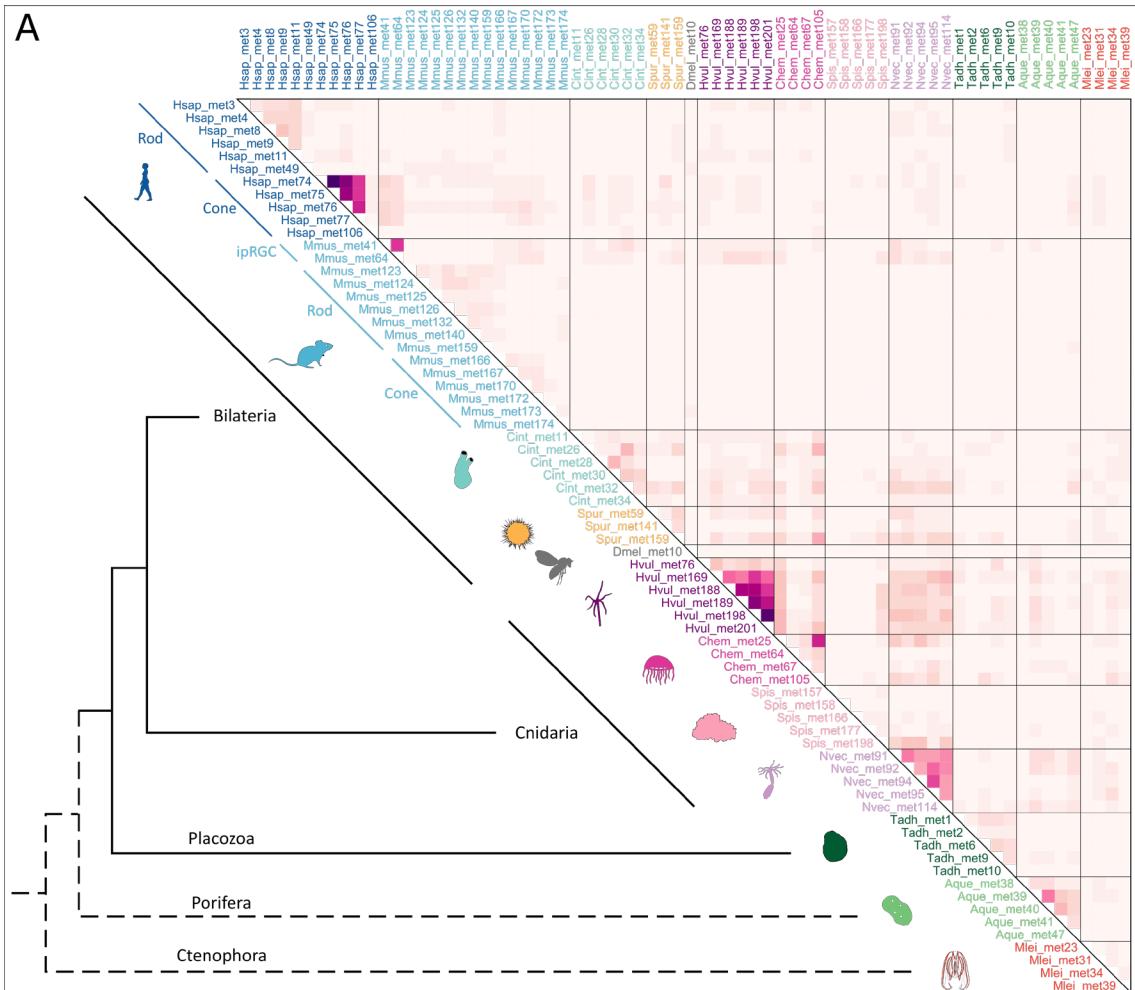
Figure 3.6. Species-specific orthogroups of regulatory genes across PRC-like metacells for each species. Venn diagrams were made to visualise how many orthogroups were shared amongst metacells of the same species. In many cases, species-specific orthogroups of regulatory genes are also metacell-specific or at least shared by only a small subset of PRC-like metacells within a given species. The most striking exception is within the ipRGCs of mouse (E), where the majority of the species-specific orthogroups are shared between the two ipRGC metacells of mouse. Human (A-B) and mouse (C-E) metacells are divided here by PRC type (cones, rods or ipRGCs) to maintain the graphic representation human-readable; in any case orthogroups that are not shared with same type PRCs are generally not shared with different PRC types. The fly (*D. melanogaster*) is missing here, as it has only one PRC-like metacell. Abbreviations: *H. sap*: *Homo sapiens*; *M. mus*: *Mus musculus*; *C. int*: *Ciona intestinalis*; *S. pur*: *Strongylocentrotus purpuratus*; *N. vec*: *Nematostella vectensis*; *S. pis*: *Stylophora pistillata*; *C. hem*: *Clytia hemisphaerica*; *H. vul*: *Hydra vulgaris*; *T. adh*: *Trichoplax adhaerens*; *A. que*: *Amphimedon queenslandica*; *M. lei*: *Mnemiopsis leidyi*.

metacells revealed a vast number of connections linking metacells either directly or indirectly. To better discern the relationships amongst a subset of metacells, we extracted subnetworks of the metacells most closely related to human PRC metacells (Figure 3.7B') and to the *Drosophila melanogaster* PRC metacell (Figure 3.7B'').

From the human PRCs subnetwork (Figure 3.7B’), we observe that human rods cluster together and are closely connected to mouse rods. Similarly, human cones are strongly clustered together and connect to mouse cones. While the connection between mouse rods and cones is solid, the direct connection between human rods and cones is weaker. Three *Ciona intestinalis* metacells are directly connected to various combination of metacells of the broad cluster containing human and mouse cones as well as mouse ipsRGC and a sea urchin metacell. Curiously, these urochordate metacells are not directly connected to each other. Two sea urchin metacells are also quite related to human and mouse PRCs, one to the cone type and the other to the rod type. A cluster of cnidarian metacells appears connected to both rod and cone clusters. A *Trichoplax adhaerens* cluster has a few connections with the rod cluster. The sponge metacell 39 is loosely connected to rod type PRCs via the *Trichoplax* cluster. Curiously, the *Drosophila* metacell has a connection with one human metacell of the rods cluster but no connection to the two mouse ipRGCs (41 and 64), that are candidate homologs to rhabdomeric PRCs, that instead cluster more closely to the cones cluster. Therefore, while these rhabdomeric-like mouse PRCs may utilise a rhabdomeric-like cascade, from a regulatory perspective they do not appear to share a similar identity to the classic rhabdomeric cell type of *Drosophila*.

Drosophila melanogaster metacell 10 is the only representative of the rhabdomeric type PRC. The subnetwork of this metacell with its closest related metacells (Figure 3.7B’’) confirms the connection to the rod ciliary cluster through direct connection to only one human metacell. No other relationship with the rest of the dataset is detected. Whilst this suggests a unique transcription factor profile for this PRC type, it is important to note that this network analysis was conducted considering only the top 100 highest expressed genes of each metacell, therefore, connections based on shared regulatory genes with lower expression do not appear.

Figure 3.7. Cross-species comparison of orthogroups of regulatory genes expressed in PRC-like metacells. Orthogroups of regulatory genes were identified with EggNog. (A) Heatmap showing the number of orthogroups in common amongst PRC-like metacells across species. While the majority of the shared orthogroups are amongst metacells of the same or closely related species, there is still some degree of shared orthogroups amongst distantly related species. (B) A network analysis of the orthogroups in common highlighted many indirect connections, indicating some level of relationship across all PRC-like metacells. B’ A subnetwork of human PRC metacells and their most closely related metacells (first two neighbours) reveals details about the relationships between potentially ciliary type metacells. Of note, the mouse ipRGCs (candidate rhabdomeric PRCs) appear more similar to human cone PRCs rather than to the *Drosophila* rhabdomeric metacell. B’’) A subnetwork with the *Drosophila* PRC and its closest relatives (first two neighbours) does not provide any evidence of rhabdomeric type metacells in other species, as the *Drosophila* metacell primarily connects to a human rod PRC.



Species-specific combinations of regulatory genes across Metazoan PRC-like metacells

Next, we examined which genes were responsible for the above network connections. Interestingly, from a vast list of hundreds of orthogroups of regulatory genes, only 69 were expressed in 3 or more phyla (Figure 3.8A), while the majority were expressed in 2 or 1 phyla (Extended Figure 3.8A). Of the regulatory genes in common between 3 or more phyla, some are already known to be involved in photoreceptor identity and/or specification, for example Six6/3, Meis2 and Tbx2 (Zuber et al. 2003; Alvarez-Delfin et al. 2009; Vopalensky and Kozmik 2009), while for others there is no known connection. Furthermore, some transcription factors that are well known to be involved in photoreceptor identity/specification, for example Otx or Rx (Arendt 2003; Vopalensky and Kozmik 2009), did not pass the threshold of 3 or more phyla in our dataset (Extended Figure 3.8A). Curiously, while there seems to be some conserved pattern of combinations of regulatory genes expressed within PRC-like metacells of the same species, across different species there seems to be little conservation. This explains all the indirect connections that we found in the network. Ultimately all metacells are “related” to each other to some degree because they share one or few genes with a metacell that in turn shares another set of few genes with a different metacell and so on. The results of this comparative analysis suggest that the core regulatory complex of PRC-like metacells in different species comprises a set of species-specific genes. Although some regulatory genes make a recurrent presence across species, the exact combination of genes is often different.

Transcription factors amongst the most abundant regulatory genes in PRC-like metacells throughout animals

Finally, we classified the orthogroups of regulatory genes into finer categories by performing BLASTP versus the Animal Transcription Factor Database (ATFDB v4) (Shen et al. 2023). This allowed us to clarify that amongst the 69 orthogroups that are shared across 3 or more phyla, around 60% belonged to transcription factor families (Figure 3.8B). The remainder of the orthogroups were either transcription cofactors (~23%), genes that interact with transcription factors but do not directly bind DNA, or other regulatory genes (~17%) like RNA polymerases and proteins that interact with the chromatin structure. The transcription factors belonged to multiple different families, with bZIP transcription factors, zinc finger C2H2 and homeoboxes being amongst the

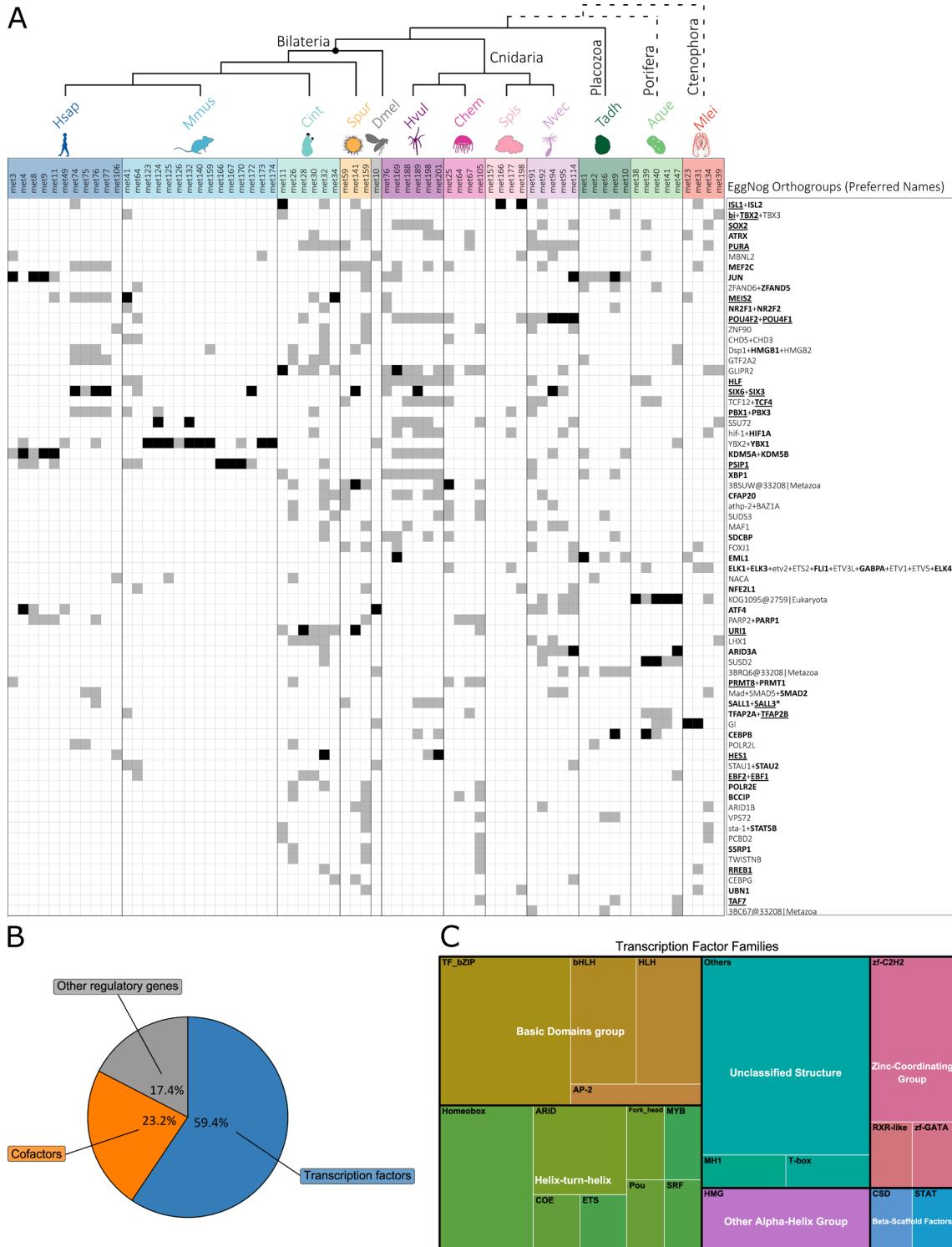


Figure 3.8. Most common orthogroups of regulatory genes shared across PRC-like metacells throughout animals. (A) 69 orthogroups are present in 3 or more phyla. Orthogroups are ordered by most frequent (with the hierarchy: present in most phyla, present in most species, present in most metacells). While some orthogroups are frequently expressed throughout animals, their exact combination of co-expression varies in the different species. Genes that were amongst the top 100 differentially expressed genes of a metacell are indicated with a black square; a grey square indicates expression with log-fold change ($\text{lfp} > 0.5$). Therefore, black squares indicate strong markers for a given metacell, while grey squares indicate that the gene is expressed in the metacell but differential expression level is not as high. Names derive from the Preferred_names of the respective EggNog orthogroup, where present, or the EggNog orthogroup itself. GeneCards/Flybase were used to

characterise the human/*Drosophila* representatives. Genes are highlighted: in bold if there is some evidence of involvement in vision and/or eye/photoreceptor development; in bold and underlined if there is strong evidence of involvement in vision and/or are expressed in the retina, although not necessarily in cones and rods; in bold, underlined and with asterisk if they are specifically expressed in photoreceptor cells. Species silhouettes were modified from images with CC0 1.0 Universal Public Domain Dedication licences obtained from <https://www.phylopic.org/>. Abbreviations: *D. mel*: *Drosophila melanogaster*; *H. sap*: *Homo sapiens*; *M. mus*: *Mus musculus*; *C. int*: *Ciona intestinalis*; *S. pur*: *Strongylocentrotus purpuratus*; *N. vec*: *Nematostella vectensis*; *S. pis*: *Stylophora pistillata*; *C. hem*: *Clytia hemisphaerica*; *H. vul*: *Hydra vulgaris*; *T. adh*: *Trichoplax adhaerens*; *A. que*: *Amphimedon queenslandica*; *M. lei*: *Mnemiopsis leidyi*. **(B)** The majority of the orthogroups of regulatory genes are transcription factor families (59.4%). Transcription cofactors are also abundant (23.2%). The remaining orthogroups include a mixture of other genes that are involved in transcription, such as polymerases and genes involved in chromatin conformation. **(C)** Treemap of the most abundant families of transcription factors shared across PRC-like metacells, organised by broad groups based on the type of DNA-binding domain. Basic domains and helix-turn-helix are the most abundant.

most predominant (Figure 3.8C). Overall, the most abundant transcription factor families possess varied types of DNA binding domains, but the most popular are the basic domains and the helix-turn-helix, although many others have an unclassified structure. When we examined all 421 orthogroups that were present in at least 2 metacells (Supplementary Table S3.3) the percentages of these categories were very similar: ~59% transcription factors; ~29% cofactors; ~12% other regulatory genes. With zinc finger C2H2 and homeoboxes being again amongst the most frequent transcription factor families. bZIP were still very abundant, although marginally surpassed by HLH and bHLH.

Conclusions

Our comprehensive analysis of the evolution of phototransduction genes revealed that their broad families mostly originated anciently in eukaryotes. Notably, even the sub-lineages containing the precise genes specialised in phototransduction functions often trace back to pre-Metazoan/Holozoan times, with few exceptions primarily amongst ciliary components. This in turn has important implications for understanding the evolution of the photoreceptor cell type in which phototransduction is employed.

Using the phototransduction genes we found in non-model organisms, including all non-bilaterian phyla, we were able to detect photoreceptor cell-like profiles in their single cell dataset. In early branching animals a mixed situation in the expression of the core components of either one or both the classical rhabdomeric/ciliary pathways, suggests that some shared components were likely employed early on in phototransduction, but then different animal lineages recruited a specific set of other components. Future research should therefore focus on uncovering these species-specific phototransduction variants in early branching animals (Vöcking et al. 2022).

Furthermore, our analysis of regulatory genes differentially expressed in these photoreceptor-like cells uncovered that the most common category of regulatory genes shared across animal PRCs are transcription factors. Although our results suggest that the exact combinations of these genes are species-specific, there are some transcription factor families that are more abundant than others. For example, bZIP transcription factors, zinc fingers C2H2 and homeoboxes are amongst the most frequent. These transcription factor families in turn deploy varied DNA-binding-domains, suggesting a broad spectrum of mechanisms through which transcription can be regulated in photoreceptor-like cells throughout animals.

Finally, in this work we have compiled an extensive list of molecular components that could be involved in phototransduction and photoreceptor-like cell identity in non-bilaterians. This can be used as a valuable resource in future research on the functional characterisation of visual systems in these organisms.

Methods

Reconstruction of the Evolution of Phototransduction Components.

Species List and Species Tree

To investigate the deep origin of the gene families of the phototransduction components, the search was broadened to all eukaryotes. 86 species representatives of Eukarya were chosen based on proteome completeness and taxonomic sampling (Table 3.2). Focus was given to sister taxa of Metazoa (8 choanoflagellates and 5 other holozoans) and non-bilaterian Metazoa (25 species), since functional visual processes must have originated at an early stage of animal evolution. The proteome completeness was assessed with BUSCO (v4.0.6) (Simão et al. 2015; Waterhouse et al. 2018) using the eukaryota_odb10 database of 255 BUSCO genes (See Supplementary Table S3.1 with proteome details). Prior knowledge of species relationships can provide a backbone for species-tree-aware gene tree construction (Boussau and Scornavacca 2020). Therefore, we used the BUSCO genes from each species for the construction of a species tree. Briefly, BUSCO genes were extracted and aligned with MAFFT v7.470 (--auto) (Katoh et al. 2002; Katoh and Standley 2013) and trimmed with Trimal v1.4.rev22 (-automated1) (Capella-Gutiérrez et al. 2009). Trimmed alignments of all BUSCO genes were concatenated with FASconCAT v1.11 (Kück and Meusemann 2010) into a super-matrix. The super-matrix was used as input for species tree construction with IQTREE v2.0.6 (Hoang et al. 2018; Minh et al. 2020), after running Model Finder (Kalyaanamoorthy et al. 2017) for best-fitting model. The resulting species tree was inspected to confirm that known species and phyla relationships were recovered. Our species tree places Ctenophores as the most basal animal phylum. As this is one of the currently accepted scenarios (Whelan et al. 2017; Schultz et al. 2023), this topology was kept. The alternative topology (Sponges as sister-group to all other animals) (Feuda et al. 2017) was obtained by manually swapping branches with Mesquite v3.6.1 (Maddison and Maddison 2008). Both species topologies were kept for downstream applications (and are available on GitHub).

Data Mining

Molecular components of interest were based on *Drosophila melanogaster* and *Homo sapiens* pathways as representative of rhabdomeric and ciliary phototransduction

respectively. Some elements of the pathways are composed of multiple subunits encoded by different genes. In total 28 gene families were identified based primarily on the KEGG maps ko04745 (rhabdomeric) and ko04744 (ciliary) (Kanehisa et al. 2021). Two additional genes, the RGS9BP and GNB5 subunits of the RGS9 complex, were added based on updated references of vertebrate phototransduction (Lamb et al. 2018) (Figure 3.1 and Table 3.1). Queries were collected from the KEGG Orthology lists (Kanehisa 2019) for each component present in the KEGG pathways and from (Lamb et al. 2018) for the two additional gene families. BLASTP (Camacho et al. 2009) was conducted (e-value cut-off of 1e-5) for each query versus the species database. Potential duplicates were removed with cd-hit (Li et al. 2001; Fu et al. 2012) with identity threshold of 100%. Outputs were used for another BLASTP versus the SwissProt database (Poux et al. 2017). Sequences were kept only if the gene family of interest was within the top five hits and parsing was carried out with gene family-specific keywords (See Supplementary Table S3.4 with list of keywords per component). This provided a first level of similarity-based filtering. A second round of filtering was conducted based on the presence of gene family-specific protein motifs. The filtered dataset was scanned with InterProScan (Quevillon et al. 2005; Jones et al. 2014) and sequences were kept only if they contained the combination of motifs characteristic to their gene family (See Supplementary Table S3.4 with list of protein motifs). To provide an annotation to the final collections of sequences, we used the top hit from BLASTP versus SwissProt.

Phylogenetic Trees

Gene trees constructed for each gene family followed a standard pipeline: alignment of sequences with MAFFT (--auto) (Katoh et al. 2002; Katoh and Standley 2013); trimming of sequences to eliminate columns with more than 70% gaps (Trimal with -gt 0.3) (Capella-Gutiérrez et al. 2009); tree construction after running Model Finder in IQTREE2 (Kalyaanamoorthy et al. 2017). The list of models tested and the best-fit models for each family, chosen based on the Bayesian Information Criterion (Schwarz 1978), can be found in Supplementary Table 3.5. Any polytomy in the gene trees was randomly resolved with ETE3 (Huerta-Cepas et al. 2016) so as to obtain fully bifurcating gene trees necessary as inputs for the gene tree to species tree reconciliations (see below).

Gene tree to species tree reconciliation

The resulting gene trees were used as starting trees for a gene tree to species tree reconciliation using Generax (v1.2.3) (Morel et al. 2020). The model used to compute the reconciliation was set to account for duplication and loss, but not transfer events. Both alternative species trees (ctenophore-first and sponge-first) were tested. The number of duplications and losses were extracted for each gene family and compared between ctenophore-first and sponge-first scenarios (see Supplementary Table S3.2).

Resulting reconciled trees were manually examined to trace the evolution of the genes of interest. The *D. melanogaster* and *H. sapiens* genes known to function in phototransduction were used to identify the orthogroups of interest and the duplication and loss events that characterised their lineages. Other subgroups within the gene families and their relationship with the orthogroups of interest were also identified. Comparison between the two alternative reconciliations with ctenophore-first versus sponge-first species tree provided a more comprehensive picture for the reconstruction of the evolutionary history of the gene families.

Collection of phototransduction marker genes for photoreceptor cells in non-model organisms

By tracing the presence of the orthogroups of interest, as identified with the reconciliations, throughout all the species examined, we were able to collect a list of candidate marker genes for phototransduction also in non-model organisms. Where the orthogroup of interest was not present, closely related lineages were used as potential markers. These marker genes were used for identifying candidate photoreceptor cell types in non-model organisms, including several non-bilaterians, for which single-cell RNA sequencing data was available. See more details below.

Identification of putative photoreceptor cell types from single-cell RNA-sequencing data.

Species datasets

To obtain a sample of photoreceptor cell diversity throughout Metazoa, we focused the single-cell analysis on twelve species based on scRNAseq data availability and phylogenetic representation. *Drosophila melanogaster* (Özel et al. 2021) served as an example for rhabdomeric-type PRCs, while *Homo sapiens* (Lukowski et al. 2019) and

Mus musculus (Macosko et al. 2015) were representative for ciliary-type PRCs. Two additional deuterostomes (the urochordate *Ciona intestinalis* (Sharma et al. 2019) and the sea urchin *Strongylocentrotus purpuratus* (Paganos et al. 2021) served as bridge species between vertebrate PRCs and protostome PRCs as represented by *Drosophila*. Finally, of particular interest for this project are non-bilaterian animals: we therefore included four cnidarian species (*Hydra vulgaris* (Siebert et al. 2019), *Clytia hemisphaerica* (Chari et al. 2021), *Stylophora pistillata* (Levy et al. 2021) and *Nematostella vectensis* (Sebé-Pedrós, Saudemont, et al. 2018)), the placozoan *Trichoplax adhaerens* (Sebé-Pedrós, Chomsky, et al. 2018), the sponge *Amphimedon queenslandica* (Sebé-Pedrós, Chomsky, et al. 2018), and the ctenophore *Mnemiopsis leidyi* (Sebé-Pedrós, Chomsky, et al. 2018). The details of these scRNAseq datasets are summarised in Table 3.3.

MetaCell pipeline for clustering cells

For the search of photoreceptor-like cells in the species of interest, we used the approach of identifying “metacells” or cell states to account for potential low depth of sequencing in non-model organisms, especially when the dataset is of the whole body.

Unique Molecular Identifiers (UMI) count matrices for each species were used as input for an established pipeline using the MetaCell v0.3.6 (Baran et al. 2019) R package, as described on MetaCell GitHub ([Analyzing whole-organism scRNA-seq data with metacell • metacell \(tanaylab.github.io\)](https://github.com/tanaylab/metacell)). Once the metacells were computed, heatmaps for all the species-specific phototransduction markers were generated to visualise which metacells were overexpressing them and indeed whether they were co-expressed in the same metacell. To better visualise the situation for single genes, we also generated bar plots with the log fold change values (lfp) of each gene in each metacell and 2D graphs with the expression of single genes mapped into the metacells 2D graph (see supplementary figures on GitHub). Finally, complete lists of lfp values for all genes in all metacells for each species were extracted for downstream analysis. See metacell scripts on GitHub for each species to reproduce these gene lists as well as the figures.

Identification of photoreceptor metacells in the model organisms *D. melanogaster*, *H. sapiens* and *M. musculus*

As a first step, we tested our pipeline on model organisms to determine whether photoreceptor cells (PRCs) could reliably be identified. *D. melanogaster* rhabdomeric

phototransduction genes were used to pinpoint a rhabdomeric PRC profile; and ciliary phototransduction genes of *H. sapiens* and *M. musculus* were used to identify ciliary-type PRCs. In the case of human and mouse, since it has been proposed that OPN4 (melanopsin) expressing cells, such as retinal ganglion cells, of vertebrates are homologous to rhabdomeric PRCs (Provencio et al. 2000; Hattar et al. 2002; Arendt 2003; Rollag et al. 2003; Fu et al. 2005), we searched also for candidate rhabdomeric PRC profiles. For this we used OPN4 (an r-opsin) together with the other rhabdomeric genes that were found in human and mouse as markers.

In the case of *D. melanogaster*, the identification of a rhabdomeric PRC profile was extremely straightforward. It was possible to spot a candidate metacell already with the heatmap of phototransduction marker expression. This metacell was kept as an example for rhabdomeric PRC-type for comparison with non-model organisms (see below). Conversely, in human and mouse datasets, multiple metacells were good candidate PRCs. This was likely due to the fact that both datasets used were from retinal samples and it is indeed expected that we identify multiple PRC profiles, especially rods that are known to be more abundant than cones. Instead, the *Drosophila* dataset came from an entire optic lobe, where we do expect a more diverse set of cell types. Although it is sensible to keep in consideration several metacells per species as PRC candidates (as effectively each metacell is a cell state so there could be several PRC cell states in the dataset), we still needed to discriminate between PRC cells and non-PRC cells present in the retina. Therefore, further steps to decide which metacells to keep were carried out for human and mouse. In order to be consistent with the non-model organisms, the same pipeline was used and is described below.

Identification of candidate photoreceptor metacells in non-model organisms

As identifying photoreceptors in non-model organisms is not straightforward, particularly for some non-bilaterians for which we do not even have any evidence that there may be photoreceptors at all (e.g. placozoa (Smith et al. 2014)), we developed a pipeline to pick-up the metacells that could be most likely a PRC-type. By this we mean that there was sufficient evidence based on the expression of combinations of phototransduction genes, to say that they have at least a PRC-like profile.

First, we filtered out metacells in which opsin log fold change (lfp) was below 0.2. This is because the opsin is the strongest marker for a photoreceptor cell, so we expect it to be

at least slightly overexpressed. The exception was *Amphimedon queenslandica*, as sponges do not possess opsins ((Feuda et al. 2012) and our results). To detect potential photoreceptor cell homologs in the sponge we had to rely only on the other phototransduction genes. We also ranked all metacells based on highest differential expression (lfp) of an opsin.

Next, we assessed the level of phototransduction gene expression in the metacells. For this we checked both the percentage of phototransduction genes co-expressed in the same metacell and their level of differential expression within the metacell. Specifically, we calculated the percentage of phototransduction genes expressed and their average lfp for: all genes; all common genes; all rhabdomeric genes; and all ciliary genes. Between the latter two, we kept the highest value as we assume that metacells lean more towards either a rhabdomeric or a ciliary profile.

Therefore, to classify metacells into best PRC candidates we had available all of the following evidences: 1) lfp of highest expressed opsin in metacell; 2) average lfp of all phototransduction genes; 3) average lfp of common phototransduction genes; 4) average lfp of either ciliary or rhabdomeric genes (whichever is highest); 5) highest percentage of all phototransduction genes; 6) highest percentage of common phototransduction genes; 7) highest percentage of either ciliary or rhabdomeric phototransduction genes (whichever is highest).

For each of these categories of evidence, we ranked the metacells from best (1st) to worst (nth). We then summed the ranking values for all the metacells to obtain a final ranking. For all rankings, if metacells tied, they got the same ranking value. We decided to keep as best candidate PRCs to be used for further analyses the PRCs that are in the top 5 of the final ranking. As a result, we have circa 5 metacells for each species. Some species have less because less than 5 metacells passed the initial threshold of opsin >0.2. Other species have slightly more than 5 metacells because some metacells tied in the final ranking. In Supplementary Table S3.6 we show these ranking calculations, and we also show an alternative ranking system. In the latter case, metacells were ranked based on how many times they appeared in the top 5 of each of the separate categories. In most cases final best metacells correspond between the two methods. We show the alternative method for completeness.

Note that in the case of mouse and human ciliary PRCs, this procedure was done separately for rod and cone metacells and the top 5 were collected for both types as indeed

their genetic profile can be a bit different and in this way we have full representation of the ciliary type.

Exploration of the regulatory genes' toolkit of candidate PRCs and comparison across species.

After having identified PRC-like metacells based on the expression of phototransduction genes as markers (see previous sections), we then moved on to further characterise the genetic profile of these candidate PRCs. We focused our analysis on regulatory genes, such as transcription factors, as these genes influence the rest of the genetic profile of the cell and are considered the core regulatory complex that defines cell identity (Arendt et al. 2016).

For all candidate PRCs of all species, we collected: i) the top 100 most highly expressed genes, these should be considered as additional markers for the metacell; and ii) all genes that have lfp above 0.5, these represent genes that are mildly overexpressed in the given metacell.

Identifying regulatory genes in PRC-like metacells

To identify genes involved in transcription, we used two tools. First, we annotated all the collected genes with Eggnog mapper (Cantalapiedra et al. 2021). We filtered out only the genes that fell into the COG category K, as that indicates that they are involved in regulating transcription. Contemporarily, we scanned our sequences for Pfam profiles of known transcription factors (see Supplementary Table S3.7 with list of profiles searched). These two approaches are complementary, as the first selects genes based on whether they may have a regulatory role in transcription, and the second focuses on collecting genes based on the presence of protein domains known to be present in certain transcription factors. Combining these two approaches, we collected a list of transcription factors and genes involved in transcription for all metacells.

For comparison across species, we used the Eggnog Orthogroup (Eggnog_OG) of the genes. As we are comparing amongst distantly related animals, we chose to compare preferably the Metazoa level of the Eggnog_OG, and only when the Eggnog_OG did not reach Metazoa level, did we collect the most stringent level available (often either Eukarya or Opisthokonta).

Cross species comparison of PRC-like metacells based on shared regulatory genes

To understand the extent to which regulatory genes were shared across species we first made an all-against-all comparison (Figure 3.7A) with all metacells of all species. This was done using genes collected with the lfp cut-off of 0.5.

To better visualise the relationships amongst metacells and species based on shared regulatory genes, we created a network graph using Cytoscape v3.9.1 (Shannon et al. 2003) (Figure 3.7B). In order to avoid over clustering and to focus on the highest confidence connections, the network was constructed using only the genes that were amongst the top 100 highly expressed for each metacell. (Figure 3.7B). As the network of all metacells from all species still contained too many connections to easily focus on relationships amongst specific subsets of metacells, we extracted subsets of the networks to identify more meaningful connections. So, we extracted the subnetwork containing Human PRCs and the first two neighbouring metacells (directly connecting metacells, and metacells connecting to the directly connected metacells) to explore connections amongst candidate ciliary PRCs (Figure 3.7B'). For candidate rhabdomeric PRCs, we made a subnetwork with *Drosophila* metacell and its next two neighbours (Figure 3.7B'').

Uncovering what type of regulatory genes are most common in PRC-like metacells

The network graphs provided broad information about how many connections are shared amongst metacells, however, we also wanted to understand which genes were behind the connections. For that we mapped the presence/absence of all regulatory genes orthogroups across all species and ordered them by most frequent. (Figure 3.8A and Extended Figure 3.8A).

Furthermore, to understand how many of these orthogroups were transcription factors as opposed to other regulatory genes (e.g., transcription cofactors), we performed a BLASTP versus the Animal Transcription Factor Database (ATFDB version 4) (Shen et al. 2023) database collecting first hits (Figure 3.8B). For the transcription factor orthogroups, we also used this database to categorise them into transcription factor families, in turn distributed across broader groups based on the DNA-binding-domain (Figure 3.8C).

Data Availability

Additional supplementary material and raw output files are available at the GitHub repository: [put link](#).

Acknowledgements

I am extremely grateful to Maryam Ghaffari Saadat and Julien Devilliers for their help in coding and automatising steps used in the methodologies described in this chapter.

References

- Altimimi HF, Schnetkamp PPM. 2007. Na⁺/Ca²⁺-K⁺ Exchangers (NCKX):Functional Properties and Physiological Roles. *Channels* 1:62–69.
- Alvarez-Delfin K, Morris AC, Snelson CD, Gamse JT, Gupta T, Marlow FL, Mullins MC, Burgess HA, Granato M, Fadool JM. 2009. Tbx2b is required for ultraviolet photoreceptor cell specification during zebrafish retinal development. *Proc. Natl. Acad. Sci.* 106:2023–2028.
- Arendt D. 2003. Evolution of eyes and photoreceptor cell types. *Int. J. Dev. Biol.* 47:563–571.
- Arendt D. 2008. The evolution of cell types in animals: emerging principles from molecular studies. *Nat. Rev. Genet.* 9:868–882.
- Arendt D, Musser JM, Baker CVH, Bergman A, Cepko C, Erwin DH, Pavlicev M, Schlosser G, Widder S, Laubichler MD, et al. 2016. The origin and evolution of cell types. *Nat. Rev. Genet.* 17:744–757.
- Arendt D, Tessmar-Raible K, Snyman H, Dorresteijn AW, Wittbrodt J. 2004. Ciliary photoreceptors with a vertebrate-type opsin in an invertebrate brain. *Science* 306:869–871.
- Baran Y, Bercovich A, Sebe-Pedros A, Lubling Y, Giladi A, Chomsky E, Meir Z, Hoichman M, Lifshitz A, Tanay A. 2019. MetaCell: analysis of single-cell RNA-seq data using K-nn graph partitions. *Genome Biol.* 20:206.
- Boussau B, Scornavacca C. 2020. Reconciling Gene trees with Species Trees. In: Scornavacca C, Delsuc F, Galtier N, editors. *Phylogenetics in the Genomic Era*. No commercial publisher | Authors open access book. p. 3.2:1-3.2:23. Available from: <https://hal.science/hal-02535529>
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:421.

- Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. 2021. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. Available from: <https://www.biorxiv.org/content/10.1101/2021.06.03.446934v2>
- Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAI: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25:1972–1973.
- Chari T, Weissbourd B, Gehring J, Ferraioli A, Leclère L, Herl M, Gao F, Chevalier S, Copley RR, Houlston E, et al. 2021. Whole-animal multiplexed single-cell RNA-seq reveals transcriptional shifts across *Clytia medusa* cell types. *Sci. Adv.* 7:eabh1683.
- D'Aniello S, Delroisse J, Valero-Gracia A, Lowe EK, Byrne M, Cannon JT, Halanych KM, Elphick MR, Mallefet J, Kaul-Strehlow S, et al. 2015. Opsin evolution in the Ambulacraria. *Mar. Genomics* 24:177–183.
- von Döhren J, Bartolomaeus T. 2018. Unexpected ultrastructure of an eye in Spiralia: the larval ocelli of *Procephalothrix oestrymnicus* (Nemertea). *Zoomorphology* 137:241–248.
- Eakin RM, Kuda A. 1970. Ultrastructure of sensory receptors in ascidian tadpoles. *Z. Für Zellforsch. Mikrosk. Anat.* 112:287–312.
- Elliott GRD, Leys SP. 2004. SPONGE LARVAL PHOTOTAXIS: A COMPARATIVE STUDY. *BMIB - Boll. Dei Musei E Degli Ist. Biol.* [Internet] 68. Available from: <https://riviste.unige.it/index.php/BMIB/article/view/625>
- Feuda R, Dohrmann M, Pett W, Philippe H, Rota-Stabelli O, Lartillot N, Wörheide G, Pisani D. 2017. Improved Modeling of Compositional Heterogeneity Supports Sponges as Sister to All Other Animals. *Curr. Biol.* 27:3864–3870.e4.
- Feuda R, Hamilton SC, McInerney JO, Pisani D. 2012. Metazoan opsin evolution reveals a simple route to animal vision. *Proc. Natl. Acad. Sci.* 109:18868–18872.
- Fleming JF, Feuda R, Roberts NW, Pisani D. 2020. A Novel Approach to Investigate the Effect of Tree Reconstruction Artifacts in Single-Gene Analysis Clarifies Opsin Evolution in Nonbilaterian Metazoans. *Genome Biol. Evol.* 12:3906–3916.
- Fu L, Niu B, Zhu Z, Wu S, Li W. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28:3150–3152.
- Fu Y, Liao H-W, Do MTH, Yau K-W. 2005. Non-image-forming ocular photoreception in vertebrates. *Curr. Opin. Neurobiol.* 15:415–422.
- Gornik SG, Bergheim BG, Morel B, Stamatakis A, Foulkes NS, Guse A. 2021. Photoreceptor Diversification Accompanies the Evolution of Anthozoa. *Mol. Biol. Evol.* 38:1744–1760.
- Gurevich VV, Gurevich EV. 2016. G Protein-Coupled Receptor Kinases (GRKs) History: Evolution and Discovery. In: Gurevich VV, Gurevich EV, Tesmer JJG,

- editors. G Protein-Coupled Receptor Kinases. New York, NY: Springer New York. p. 3–22. Available from: https://doi.org/10.1007/978-1-4939-3798-1_1
- Hahn J, Monavarfeshani A, Qiao M, Kao A, Kölsch Y, Kumar A, Kunze VP, Rasys AM, Richardson R, Baier H, et al. 2023. Evolution of neuronal cell classes and types in the vertebrate retina. :2023.04.07.536039. Available from: <https://www.biorxiv.org/content/10.1101/2023.04.07.536039v1>
- Hardie RC, Juusola M. 2015. Phototransduction in Drosophila. *Curr. Opin. Neurobiol.* 34:37–45.
- Hattar S, Liao HW, Takao M, Berson DM, Yau KW. 2002. Melanopsin-containing retinal ganglion cells: architecture, projections, and intrinsic photosensitivity. *Science* 295:1065–1070.
- Heyland A, Croll R, Goodall S, Kranyak J, Wyeth R. 2014. Trichoplax adhaerens, an Enigmatic Basal Metazoan with Potential. In: Carroll DJ, Stricker SA, editors. Developmental Biology of the Sea Urchin and Other Marine Invertebrates: Methods and Protocols. Methods in Molecular Biology. Totowa, NJ: Humana Press. p. 45–61. Available from: https://doi.org/10.1007/978-1-62703-974-1_4
- Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol. Biol. Evol.* 35:518–522.
- Horridge GA. 1964. Presumed photoreceptive cilia in a ctenophore. *Q. J. Microsc. Sci.* [Internet]. Available from: <https://openresearch-repository.anu.edu.au/handle/1885/167542>
- Huerta-Cepas J, Serra F, Bork P. 2016. ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. *Mol. Biol. Evol.* 33:1635–1638.
- Jékely G, Paps J, Nielsen C. 2015. The phylogenetic position of ctenophores and the origin(s) of nervous systems. *EvoDevo* 6:1.
- Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, et al. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30:1236–1240.
- Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermini LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14:587–589.
- Kanehisa M. 2019. Toward understanding the origin and evolution of cellular organisms. *Protein Sci.* 28:1947–1951.
- Kanehisa M, Sato Y, Kawashima M. 2021. KEGG mapping tools for uncovering hidden features in biological data. *Protein Sci.* [Internet] n/a. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1002/pro.4172>
- Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30:3059–3066.
- Katoh K, Standley DM. 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* 30:772–780.

- Koyanagi M, Ono K, Suga H, Iwabe N, Miyata T. 1998. Phospholipase C cDNAs from sponge and hydra: antiquity of genes involved in the inositol phospholipid signaling pathway¹The nucleotide sequence data reported in this paper will appear in the DDBJ, EMBL and GenBank nucleotide sequence databases.1. *FEBS Lett.* 439:66–70.
- Kozmík Z, Ruzickova J, Jonasova K, Matsumoto Y, Vopalensky P, Kozmíkova I, Strnad H, Kawamura S, Piatigorsky J, Paces V, et al. 2008. Assembly of the cnidarian camera-type eye from vertebrate-like components. *Proc. Natl. Acad. Sci.* 105:8989–8993.
- Krishnan A, Mustafa A, Almén MS, Fredriksson R, Williams MJ, Schiöth HB. 2015. Evolutionary hierarchy of vertebrate-like heterotrimeric G protein families. *Mol. Phylogenet. Evol.* 91:27–40.
- Kück P, Meusemann K. 2010. FASconCAT, Version 1.0, Zool. Forschungsmuseum A. Koenig, Germany, 2010.
- Lagman D, Franzén IE, Eggert J, Larhammar D, Abalo XM. 2016. Evolution and expression of the phosphodiesterase 6 genes unveils vertebrate novelty to control photosensitivity. *BMC Evol. Biol.* 16:124.
- Lagman D, Sundström G, Ocampo Daza D, Abalo XM, Larhammar D. 2012. Expansion of transducin subunit gene families in early vertebrate tetraploidizations. *Genomics* 100:203–211.
- Lamb TD. 2020. Evolution of the genes mediating phototransduction in rod and cone photoreceptors. *Prog. Retin. Eye Res.* 76:100823.
- Lamb TD, Patel HR, Chuah A, Hunt DM. 2018. Evolution of the shut-off steps of vertebrate phototransduction. *Open Biol.* 8:170232.
- Lee S-J, Xu H, Montell C. 2004. Rhodopsin kinase activity modulates the amplitude of the visual response in *Drosophila*. *Proc. Natl. Acad. Sci.* 101:11874–11879.
- Levy S, Elek A, Grau-Bové X, Menéndez-Bravo S, Iglesias M, Tanay A, Mass T, Sebé-Pedrós A. 2021. A stony coral cell atlas illuminates the molecular and cellular basis of coral symbiosis, calcification, and immunity. *Cell* 184:2973–2987.e18.
- Leys SP, Degnan BM. 2001. Cytological Basis of Photoresponsive Behavior in a Sponge Larva. *Biol. Bull.* [Internet]. Available from: <https://www.journals.uchicago.edu/doi/10.2307/1543611>
- Li W, Jaroszewski L, Godzik A. 2001. Clustering of highly homologous sequences to reduce the size of large protein databases. *Bioinformatics* 17:282–283.
- Lukowski SW, Lo CY, Sharov AA, Nguyen Q, Fang L, Hung SS, Zhu L, Zhang T, Grünert U, Nguyen T, et al. 2019. A single-cell transcriptome atlas of the adult human retina. *EMBO J.* 38:e100811.
- Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas AR, Kamitaki N, Martersteck EM, et al. 2015. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* 161:1202–1214.

- Maddison W, Maddison D. 2008. Mesquite: A modular system for evolutionary analysis. *Evolution* 62:1103–1118.
- Maldonado M, Durfort M, McCarthy DA, Young CM. 2003. The cellular basis of photobehavior in the tufted parenchymella larva of demosponges. *Mar. Biol.* 143:427–441.
- Mikami K. 2014. Structural divergence and loss of phosphoinositide-specific phospholipase C signaling components during the evolution of the green plant lineage: implications from structural characteristics of algal components. *Front. Plant Sci.* 5:380.
- Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* 37:1530–1534.
- Morel B, Kozlov AM, Stamatakis A, Szöllősi GJ. 2020. GeneRax: A Tool for Species-Tree-Aware Maximum Likelihood-Based Gene Family Tree Inference under Gene Duplication, Transfer, and Loss. *Mol. Biol. Evol.* 37:2763–2774.
- Müller WEG, Schröder HC, Markl JS, Grebenjuk VA, Korzhev M, Steffen R, Wang X. 2013. Cryptochrome in Sponges: A Key Molecule Linking Photoreception with Phototransduction. *J. Histochem. Cytochem.* 61:814–832.
- Mushegian A, Gurevich VV, Gurevich EV. 2012. The Origin and Evolution of G Protein-Coupled Receptor Kinases. *PLOS ONE* 7:e33806.
- Nilsson D-E. 2009. The evolution of eyes and visually guided behaviour. *Philos. Trans. R. Soc. B Biol. Sci.* 364:2833–2847.
- Nilsson D-E. 2013. Eye evolution and its functional basis. *Vis. Neurosci.* 30:5–20.
- Nordström K, Wallén null, Seymour J, Nilsson D. 2003. A simple visual system without neurons in jellyfish larvae. *Proc. R. Soc. Lond. B Biol. Sci.* 270:2349–2354.
- Orban T, Palczewski K. 2016. Structure and Function of G-Protein-Coupled Receptor Kinases 1 and 7. In: Gurevich VV, Gurevich EV, Tesmer JJG, editors. *G Protein-Coupled Receptor Kinases*. New York, NY: Springer New York. p. 25–43. Available from: https://doi.org/10.1007/978-1-4939-3798-1_2
- Özel MN, Simon F, Jafari S, Holguera I, Chen Y-C, Benhra N, El-Danaf RN, Kapuralin K, Malin JA, Konstantinides N, et al. 2021. Neuronal diversity and convergence in a visual system developmental atlas. *Nature* 589:88–95.
- Paganos P, Voronov D, Musser JM, Arendt D, Arnone MI. 2021. Single-cell RNA sequencing of the *Strongylocentrotus purpuratus* larva reveals the blueprint of major cell types and nervous system of a non-chordate deuterostome. Tessmar-Raible K, Bronner ME, Martinez Serra P, Revilla-i-Domingo R, Hinman V, editors. *eLife* 10:e70416.
- Palczewski K, Kiser PD. 2020. Shedding new light on the generation of the visual chromophore. *Proc. Natl. Acad. Sci. U. S. A.* 117:19629–19638.
- Passamaneck YJ, Furchheim N, Hejnol A, Martindale MQ, Lüter C. 2011. Ciliary photoreceptors in the cerebral eyes of a protostome larva. *EvoDevo* 2:6.

- Piatigorsky J, Kozmik Z. 2004. Cubozoan jellyfish: an Evo/Devo model for eyes and other sensory systems. *Int. J. Dev. Biol.* 48:719–729.
- Picciani N, Kerlin JR, Sierra N, Swafford AJM, Ramirez MD, Roberts NG, Cannon JT, Daly M, Oakley TH. 2018. Prolific Origination of Eyes in Cnidaria with Co-option of Non-visual Opsins. *Curr. Biol.* 28:2413-2419.e4.
- Plachetzki DC, Fong CR, Oakley TH. 2010. The evolution of phototransduction from an ancestral cyclic nucleotide gated pathway. *Proc. R. Soc. B Biol. Sci.* 277:1963–1969.
- Poux S, Arighi CN, Magrane M, Bateman A, Wei C-H, Lu Z, Boutet E, Bye-A-Jee H, Famiglietti ML, Roechert B, et al. 2017. On expert curation and scalability: UniProtKB/Swiss-Prot as a case study. *Bioinformatics* 33:3454–3460.
- Provencio I, Rodriguez IR, Jiang G, Hayes WP, Moreira EF, Rollag MD. 2000. A Novel Human Opsin in the Inner Retina. *J. Neurosci.* 20:600–605.
- Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R. 2005. InterProScan: protein domains identifier. *Nucleic Acids Res.* 33:W116–W120.
- Rebecchi MJ, Pentyala SN. 2000. Structure, Function, and Control of Phosphoinositide-Specific Phospholipase C. *Physiol. Rev.* 80:1291–1335.
- Rivera AS, Ozturk N, Fahey B, Plachetzki DC, Degnan BM, Sancar A, Oakley TH. 2012. Blue-light-receptive cryptochrome is expressed in a sponge eye lacking neurons and opsin. *J. Exp. Biol.* 215:1278–1286.
- Rollag MD, Berson DM, Provencio I. 2003. Melanopsin, Ganglion-Cell Photoreceptors, and Mammalian Photoentrainment. *J. Biol. Rhythms* 18:227–234.
- Ryan K, Lu Z, Meinertzhagen IA. 2016. The CNS connectome of a tadpole larva of *Ciona intestinalis* (L.) highlights sidedness in the brain of a chordate sibling. Marder E, editor. *eLife* 5:e16962.
- Schnitzler CE, Pang K, Powers ML, Reitzel AM, Ryan JF, Simmons D, Tada T, Park M, Gupta J, Brooks SY, et al. 2012. Genomic organization, evolution, and expression of photoprotein and opsin genes in *Mnemiopsis leidyi*: a new view of ctenophore photocytess. *BMC Biol.* 10:107.
- Schultz DT, Haddock SHD, Bredeson JV, Green RE, Simakov O, Rokhsar DS. 2023. Ancient gene linkages support ctenophores as sister to other animals. *Nature*:1–8.
- Schwarz G. 1978. Estimating the Dimension of a Model. *Ann. Stat.* 6:461–464.
- Sebé-Pedrós A, Chomsky E, Pang K, Lara-Astiaso D, Gaiti F, Mukamel Z, Amit I, Hejnol A, Degnan BM, Tanay A. 2018. Early metazoan cell type diversity and the evolution of multicellular gene regulation. *Nat. Ecol. Evol.* 2:1176–1188.
- Sebé-Pedrós A, Saudemont B, Chomsky E, Plessier F, Mailhé M-P, Renno J, Loe-Mie Y, Lifshitz A, Mukamel Z, Schmutz S, et al. 2018. Cnidarian Cell Type Diversity and Regulation Revealed by Whole-Organism Single-Cell RNA-Seq. *Cell* 173:1520-1534.e20.

- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* 13:2498–2504.
- Sharma S, Wang W, Stolfi A. 2019. Single-cell transcriptome profiling of the *Ciona* larval brain. *Dev. Biol.* 448:226–236.
- Shen W-K, Chen S-Y, Gan Z-Q, Zhang Y-Z, Yue T, Chen M-M, Xue Y, Hu H, Guo A-Y. 2023. AnimalTFDB 4.0: a comprehensive animal transcription factor database updated with variation and expression annotations. *Nucleic Acids Res.* 51:D39–D45.
- Shichida Y, Matsuyama T. 2009. Evolution of opsins and phototransduction. *Philos. Trans. R. Soc. B Biol. Sci.* 364:2881–2895.
- Siebert S, Farrell JA, Cazet JF, Abeykoon Y, Primack AS, Schnitzler CE, Juliano CE. 2019. Stem cell differentiation trajectories in *Hydra* resolved at single-cell resolution. *Science* 365:eaav9314.
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210–3212.
- Smith CL, Varoqueaux F, Kittelmann M, Azzam RN, Cooper B, Winters CA, Eitel M, Fasshauer D, Reese TS. 2014. Novel Cell Types, Neurosecretory Cells, and Body Plan of the Early-Diverging Metazoan *Trichoplax adhaerens*. *Curr. Biol.* 24:1565–1572.
- Suh P-G, Park J-I, Manzoli L, Cocco L, Peak JC, Katan M, Fukami K, Kataoka T, Yun S, Ryu SH. 2008. Multiple roles of phosphoinositide-specific phospholipase C isozymes. *BMB Rep.* 41:415–434.
- Tamm SL. 2016. Novel Structures Associated with Presumed Photoreceptors in the Aboral Sense Organ of Ctenophores. *Biol. Bull.* 231:97–102.
- Terakita A. 2005. The opsins. *Genome Biol.* 6:213.
- Tsutsui K, Minami J, Matsushita O, Katayama S, Taniguchi Y, Nakamura S, Nishioka M, Okabe A. 1995. Phylogenetic analysis of phospholipase C genes from *Clostridium perfringens* types A to E and *Clostridium novyi*. *J. Bacteriol.* 177:7164–7170.
- Ullrich-Lüter EM, Dupont S, Arboleda E, Hausen H, Arnone MI. 2011. Unique system of photoreceptors in sea urchin tube feet. *Proc. Natl. Acad. Sci. U. S. A.* 108:8367–8372.
- Valencia JE, Feuda R, Mellott DO, Burke RD, Peter IS. 2021. Ciliary photoreceptors in sea urchin larvae indicate pan-deuterostome cell type conservation. *BMC Biol.* 19:257.
- Varoqueaux F, Williams EA, Grandemange S, Truscello L, Kamm K, Schierwater B, Jékely G, Fasshauer D. 2018. High Cell Diversity and Complex Peptidergic Signaling Underlie Placozoan Behavior. *Curr. Biol.* 28:3495-3501.e2.

- Vöcking O, Macias-Muñoz A, Jaeger S, Oakley TH. 2022. Deep Diversity: Extensive Variation in the Components of Complex Visual Systems across Animals. Available from: <https://www.preprints.org/manuscript/202209.0432/v1>
- Vopalensky P, Kozmík Z. 2009. Eye evolution: common use and independent recruitment of genetic components. *Philos. Trans. R. Soc. B Biol. Sci.* 364:2819–2832.
- Wang T, Montell C. 2007. Phototransduction and retinal degeneration in Drosophila. *Pflüg. Arch. - Eur. J. Physiol.* 454:821–847.
- Wang X, Liu Y, Li Z, Gao X, Dong J, Yang M. 2020. Expression and evolution of the phospholipase C gene family in Brachypodium distachyon. *Genes Genomics* 42:1041–1053.
- Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, Klioutchnikov G, Kriventseva EV, Zdobnov EM. 2018. BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics. *Mol. Biol. Evol.* 35:543–548.
- Whelan NV, Kocot KM, Moroz TP, Mukherjee K, Williams P, Paulay G, Moroz LL, Halanych KM. 2017. Ctenophore relationships and their placement as the sister group to all other animals. *Nat. Ecol. Evol.* 1:1737–1746.
- Widjaja-Adhi MAK, Golczak M. 2020. The molecular aspects of absorption and metabolism of carotenoids and retinoids in vertebrates. *Biochim. Biophys. Acta Mol. Cell Biol. Lipids* 1865:158571.
- Wong E, Anggono V, Williams SR, Degnan SM, Degnan BM. 2022. Phototransduction in a marine sponge provides insights into the origin of animal vision. *iScience* 25:104436.
- Yau K-W, Hardie RC. 2009. Phototransduction Motifs and Variations. *Cell* 139:246–264.
- Zuber ME, Gestri G, Viczán AS, Barsacchi G, Harris WA. 2003. Specification of the vertebrate eye by a network of eye field transcription factors. *Development* 130:5155–5167.