A deep dive into the sablefish (*Anoplopoma fimbria)* opsin repertoire: insight into melanopsin expression, localization and function in an unlikely demersal model.

by

Hayley V. Barnes
B.Sc., University of Victoria, 2019

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Biology

University of Victoria

We acknowledge and respect the lək̓ʷəŋən peoples on whose traditional territory the university stands and the Songhees, Esquimalt and W̱SÁNEĆ peoples whose historical relationships with the land continue to this day.

A deep dive into the sablefish (*Anoplopoma fimbria)* opsin repertoire: insight into melanopsin expression and localization in an unlikely demersal model.

by

Hayley V. Barnes
B.Sc., University of Victoria, 2019

Supervisory Committee

Dr. John Taylor, Supervisor
Department of Biology

Dr. Robert Chow, Departmental Member
Department of Biology

Dr. Ben Koop, Departmental Member
Department of Biology

Dr. Steve Perlman, Departmental Member
Department of Biology

Abstract

Light regulates many biological processes through light-sensitive proteins called opsins. Opsins are involved in vision, but they are also expressed in extraretinal tissue, where their roles are far less clear. Fish have large opsin repertoires, derived from a long history of gene duplication and divergence, making them useful models to study opsin diversity and function. I introduce the deep-sea sablefish (*Anoplopoma fimbria*) as a model for opsin research for three main reasons: i) the availability of a draft genome and transcriptome, simplifying the characterization of this species' opsin repertoire, ii) the proximity of the only sablefish aquaculture facility in the world, providing exclusive access to a large number of individuals at all developmental stages, iii) the observation that sablefish occupy very different light environments during the course of development, ranging from well-lit shallow waters to the aphotic zone, which provides a light environment context for opsin gene expression data.

My survey of the genome showed that sablefish have 36 distinct opsin genes (7 visual and 29 non-visual), even though they spend most of their lives in the dark. The sablefish opsin sequences and repertoire are similar to those of other teleost fish. To test the hypothesis that the sablefish opsin repertoire is being expressed/transcribed during the comparatively brief period of time when this species is exposed to light (the free-swimming larval stage through to the juvenile stage), I quantified the expression of five paralogous genes from a well-studied non-visual opsin family (OPN4's) in the brain across life stages. Data show statistically stable expression of *Opn4m1* and *Opn4m3* among life stages, a rough association of *Opn4x1* and *Opn4m2* expression with age and light environment, and little-to-no expression of *Opn4x2.* I localized proteins encoded by the most highly expressed class of OPN4 genes in the brain, the Opn4m genes, to the surface of the optic tectum just below a cranial 'window'; a zone that has been shown to express dozens of opsins in zebrafish (a distant relative, with their ancestor diverging more than 230 million years ago). Thus, in some cases, expression appears to be correlated with light exposure not only temporally, but also spatially. By studying non-visual opsins in sablefish, I have challenged and broadened the current understanding of opsin evolution and function in fish and provided the foundation for future studies to test brain regions for light-sensitivity, perform opsin gene knock-outs, and explore potential light-independent processes.

Table of Contents

List of Tables

List of Figures

Glossary

AC: amacrine cell

cDNA: complementary deoxyribonucleic acid

Cq: quantification cycle (often interchanged for Ct)

DAPI: 4’,6-diamidino-2-phenylindole

DNA: deoxyribonucleic acid

EST: expressed sequence tags

gDNA: genomic deoxyribonucleic acid

GPCR: G-protein coupled receptor

HC: horizontal cell

HMM: Hidden Markov Model

IPL: inner plexiform layer

ipRGC: intrinsically photosensitive retinal ganglion cell

mRNA: messenger ribonucleic acid

OPL: outer plexiform layer

PCR: polymerase chain reaction

PGZ: periventricular grey zone

qPCR: quantitative polymerase chain reaction

RGC: retinal ganglion cell

RNA: ribonucleic acid

RQ: relative quantification

RT-: reverse transcriptase negative control

RT-PCR: reverse transcription polymerase chain reaction

SAC: stratum album centrale

SFGS: stratum fibrosum et griseum superficiale

SGC: stratum griseum centrale

SO: stratum opticum

SRA: sequence read archive

TGD: teleost genome duplication

TSA: transcriptome shotgun assembly

Acknowledgements

I owe so much of my success as a graduate student to my supervisor, John Taylor. John took every opportunity to share his insight and knowledge with me, while still allowing me the space to learn and grown on my own. By encouraging curiosity and exploration, he has shaped me into a better scientist. More than anything, I am grateful for John's endless support, patience and kindness. John – thank you for your encouragement during both the highs and lows of this degree. I am so lucky to have had you as a mentor.

I would also like to thank my committee members Steve Perlman, Bob Chow and Ben Koop for their valuable guidance, feedback and insight.

Without collaboration, the completion of this project would not have been possible. I am eternally grateful to everyone who contributed. A sincere thank you to:

My lab mate Nilou Mokariasl, for helping me troubleshoot, supplementing my work, and for being a great friend to me. Past lab members Tricia Rubi and Alana McPherson for their contributions in the early days of this research, as well as the various work study students who have lent a hand.

Our collaborators at Golden Eagle Sablefish, especially Briony Campbell, for being so accommodating and allowing me to pursue this exciting research opportunity.

Inigo Novales Flamarique, Kennedy Bolstad, Brent Gowen, Stephen Hughes and Wayne Davies, and the Chow and Swayne Labs for their immunohistochemistry guidance, reagents and equipment. The Koop Lab for their work on the sablefish genome and transcriptome – thanks for the head start.

I appreciate your belief in me and my research.

Dedication

I dedicate this thesis to the people I could not have completed it without, my friends and family. Dayton, Mom, Dad, Mitch, Lauren and Dora: thank you for your unwavering love and encouragement. I am forever grateful for everything you have done for me, the hard times you have gotten me through and the moments we have celebrated together. Your support has meant the world to me.

CHAPTER 1 - Characterization of 36 sablefish opsins, evaluation of repertoire and sequence divergence, and evidence of transcription.

INTRODUCTION

The world of opsins

Light is one of the most reliable signals on earth. Though light intensity and wavelength can vary on any given day, the planet's constant rotation provides dependable day and night cycles to almost all inhabitants. Not surprisingly, light sensitivity is an attribute of all three domains of life and is mediated by a diversity of proteins, including flavoproteins (e.g., phototropins and cryptochromes), phytochromes, xanthopsin and two 'types' of rhodopsins (Ernst et al. 2014).

Microbial rhodopsins (type I rhodopsins) occur in archaea, bacteria and eukarya. They were discovered in haloarchaea and act as light-driven proton pumps in some species and light sensors necessary for phototaxis and photoadaptation in others (Ordterhelt & Stoeckenius, 1973). While the taxonomic distribution of type I rhodopsins suggests these membrane-bound chromophore-binding proteins were present in the common ancestor of all extant species, those in eukaryotes appear to be products of horizontal gene transfer (Sharma et al., 2007). Light-sensitivity in animals is mediated by type II rhodopsins. Thus, despite having a similar structure and utilizing a similar chromophore, type II and type I rhodopsins are not homologous (i.e. not derived from a common ancestor/origination event).

Type II rhodopsins, commonly called opsins, are sorted into five main lineages and have been detected in the genomes of all groups of Metazoans, from Placozoa to Mammalia (Clarke, 2020). These genes encode seven transmembrane G-protein coupled receptors (GPCR) that bind a vitamin A-derived chromophore (e.g. 11-*cis* retinal or another retinal variant). When exposed to light, the chromophore undergoes a series of conformational changes (to all-*trans* retinal) in the opsin binding pocket and this results in the activation of intracellular G-proteins. Duplication and divergence events have produced large opsin repertoires in many species and resulted in great sequence diversity among members of this gene superfamily (Dulai et al.,1999; Rennison et al., 2012; Beaudry et al., 2017).

Different opsins vary in wavelength sensitivity; consequently, cells that express different opsins are maximally sensitive to different wavelengths of light. Opsin proteins have diversified

in other ways as well, including in their interactions with the chromophore. Some are monostable (i.e., forming a stable interaction with 11-*cis* retinal only) and others are bistable (stable when complexed with 11-*cis* or all-*trans* retinal) (Zhong et al., 2012). Additionally, G-protein binding varies among opsin groups; some signal through a G_q or G_s pathway, through which activation leads to a depolarization in membrane potential, and others (e.g., those in rod and cone cells of the retina) bind G_t or $G_{i/o}$, which lead to cell hyperpolarization (Yau & Hardie, 2009). Despite the variability described above, the ability of opsins to relay light-mediated chromophore isomerization to intracellular G-proteins depends upon a structure determined by several highly conserved amino acid residues, making it possible to align the amino acid sequences of distantly related species and align distantly related opsins within a species.

Opsins are important in a diversity of processes, with vision being the most obvious. Opsins involved in image formation are termed 'visual opsins', and most opsin research is focused on this group; however, phylogenetic analyses show that opsin gene diversity evolved well before vision (Beaudry et al., 2017) and that vision evolved independently along different branches of the opsin gene tree (Arendt, 2003). Many invertebrates see using rhabdomeric opsins (OPN4 family) in their compound or ocelli-like eyes, while the vertebrate visual opsins are ciliary opsins (OPN1 family) (Lamb, 2013). The anatomical units and morphology of the various animal eyes point toward a lack of homology, and recently, it was confirmed that no single gene regulates the development of eyes across metazoans (Koenig & Gross, 2020). While understanding a key sensory system (the eye) has its merit, in the grand scheme of opsin diversity and function, it covers only the tip of the iceberg. In recent years, the non-visual opsins (opsins playing roles beyond image formation) have become better characterized and are beginning to become more prominent in research. While still fairly enigmatic, light-sensitive non-visual opsins are known to play roles in circadian rhythm entrainment (Basili et al., 2021; Cavallari et al., 2021), camouflage (Kelley & Davies, 2016), development (Nguyen et al., 2019), and pupillary constriction (Lucas et al., 2001; Woelders et al., 2018).

Fish in opsin research

Ray-finned fish are of particular interest and value in the field of opsin research, as they have especially large opsin repertoires, some with more than 40 distinct genes (Davies et al., 2015; Beaudry et al., 2017). This larger repertoire relative to other vertebrates is due to a

combination of the retention of opsins present in the ancestor of all vertebrates, whole genome duplication in the ancestor of teleosts (TGD), and lineage-specific duplication events (Cortesi et al., 2015), the latter leading to great variation in visual opsin repertoires among fishes (Rennison et al., 2012). The reason why fish have retained so many of these genes is not clear, though their persistence in almost all lineages suggests they contribute to highly conserved functions. Fish provide an opportunity to better study the fate of opsin duplicates through their opsin domain diversity, and to tie together genetics and organismal biology.

Some of the earliest visual opsin research used fish as models: Kottgen and Abelsdorf (1896) identified a difference in retinal pigment color when comparing freshwater fish to mammals and frogs, Wald (1939) isolated the vitamin A chromophore from a fish retina, Dartnell (1952) observed a greater diversity of visual pigments in fish compared to other vertebrates, and Denton and Warren (1956) showed an increased blue visual sensitivity in deep sea fish, to name a few examples.

With more recent technological advances (namely sequencing, gene transcript quantification and knock-out techniques), fish have remained at the forefront of opsin research. Expression and repertoire-level observations have guided what we know about the function of this gene family. For example, Li et al. (2005) showed that expression of the long-wave sensitive cone opsin in zebrafish increased during the day and decreased during the night, consequently making them more sensitive to red objects during the day, and Smith et al. (2012) also showed that long-wave opsin expression is tuned to environmental light and mediates behavior in cichlids. Visual opsin repertoire diversity has been studied in species among diverse photic environments; some deep-sea fish, for example, have expanded rhodopsin repertoires (Musilova et al., 2019), while other species show repertoires adapted to turbid water habitats (Musilova et al., 2021). The visual opsin repertoires of cichlids have even been linked to color preference in mate choice (Wright et al., 2017). Fish have also been important models for studying spatial patterns of expression. The four-eyed fish and archer fish have interesting intraretinal variation in opsin expression associated with their morphological and behavioral adaptations (Temple et al., 2010; Owens et al., 2009).

As aforementioned, opsin diversity evolved before vision. Early extraocular light-sensitivity research also included fish. Before the term ‘opsin’ was coined, Frisch (1911), working with minnows, discovered that a color changing response to light occurred in blind fish.

After the removal of their pineal glands, the minnows' ability to change color significantly diminished (though not entirely), pointing to the presence of light receptors in both the pineal and the deep brain. As non-visual opsin transcripts were uncovered in the early 1990's, fish such as salmon (Philp et al., 2000), lamprey (Foster et al., 1994) and catfish (Blackshaw et al., 1997) were used to study this phenomenon further.

The zebrafish (*Danio rerio*) is a model organism in many scientific fields and currently one of the best for understanding opsin gene family evolution and opsin function (both visual and non-visual). Its genome and transcriptomes from a diversity of tissues and developmental stages are well characterized: zebrafish have 42 opsins, of which 32 are non-visual, and these are expressed in nearly every tissue in the body (Davies et al., 2015). Zebrafish opsin research has consequently helped connect opsin sequence variation to wavelength sensitivity (Chinen et al., 2005), differential expression of duplicates (Kojima et al., 2008), opsin function through transgenic experiments (Crespo et al., 2018) and expression differences among tissues and through development (Takechi et al., 2005). Though zebrafish have contributed much to the field, no species alone can uncover the whole opsin story; the teleost class is incredibly species-rich and the life histories and inhabited photic environments are also highly variable.

Sablefish as a new model

Here, I introduce sablefish (*Anoplopoma fimbria)*, also called black cod, as a new model for opsin research. It is a distant relative of the zebrafish and occupies very different light environments. Adult sablefish live and spawn in the bathypelagic/bathydemersal zone, up to thousands of meters deep, where light is severely restricted. As larvae deplete their yolk-sacs (at approximately 40 days), they move from the depths of the ocean to the surface to feed, where there is an abundance of down-welling light. The juvenile stage is spent inshore, and after two to three years, they migrate back to the deep, where they reach sexual maturity and remain for decades (Figure 1). The deep-sea adults also have vertical migration patterns, moving up and down the water column (typically at night), to forage for food (Sigler & Echave, 2019). Protocols have recently been developed for raising this species in large numbers in aquaculture, making all tissues available at all stages of development and in large numbers. These features give sablefish great potential to be useful in the study of opsin function.

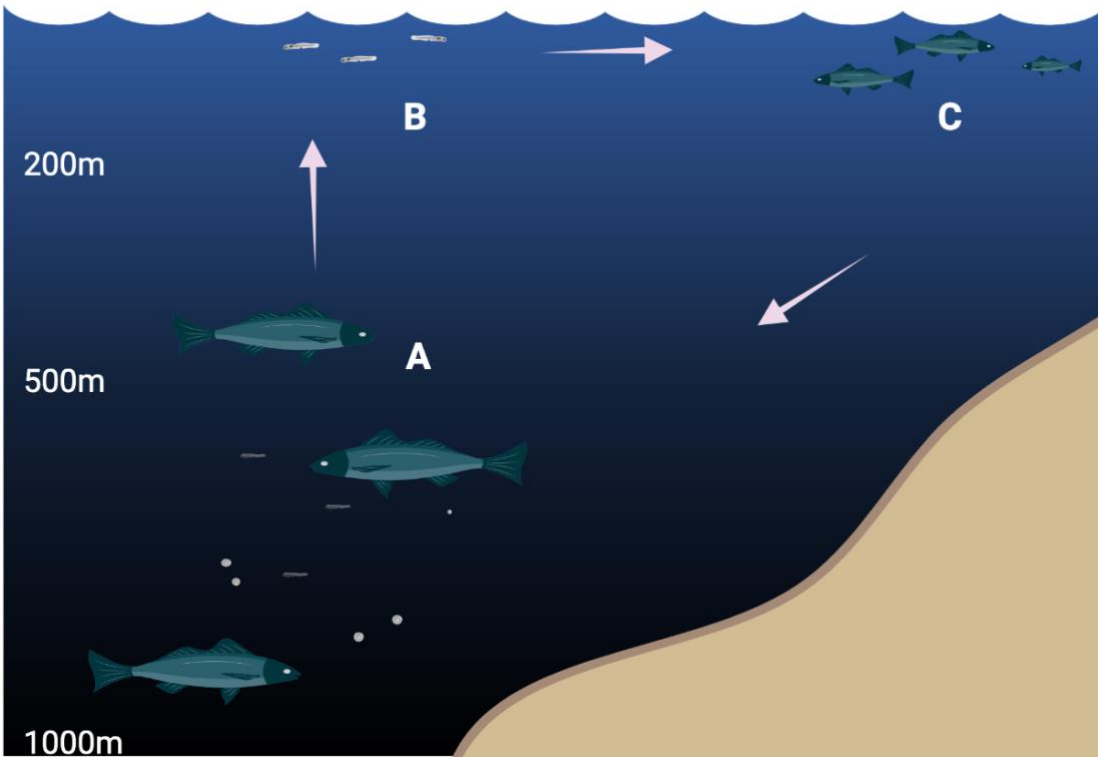


Figure 1. Illustration of wild sablefish movement in the ocean throughout development, from depths up to 1000m or deeper, to the surface, inshore and back. A) Mature adults, eggs and yolk-sac dependent larvae. B) Free-feeding larvae. C) Juveniles. Ocean depths are approximate.

By studying sablefish, I can investigate connections between opsin expression, stages in development, and light exposure. Much of what is known about environmentally regulated opsin expression comes from comparing species living in different conditions. Plasticity can be studied within a single species as well by observing changes associated with life history and light exposure, but this has only been done for a few species, including flatfish (Wang et al., 2021), sharks (Fasick et al., 2020) and select deep-sea teleosts (Lupse et al., 2021). Sablefish provide this same opportunity, but over a longer temporal and more defined developmental span than has previously been done with other species. A fully characterized opsin repertoire is the first step and is the focus of this chapter of my thesis.

I used BLAST to identify opsin-bearing contigs and scaffolds in the sablefish genome. Two gene prediction tools were used to produce full-length coding sequences (i.e., spliced exons) and gene identification was confirmed using phylogenetic analyses. The sablefish opsin sequences are first compared to those of other fish species and then to other Metazoans. Reverse transcription polymerase chain reaction (RT-PCR) and a search of opsins in the sablefish

transcriptome provided evidence that most of these opsins are transcribed. My characterization of the sablefish opsin repertoire provides a foundation for research on expression in visual and non-visual tissues and at all stages of development, as well as the raw data necessary for future gene transcript quantification, localization and knock out experiments.

METHODS

Extracting opsins from draft genomes

At the beginning of this study, there was one draft sablefish genome in NCBI, a collection of 208,506 contigs (BioProject PRJNA202249, assembly GCA_000499045.1, N50=5,156). The 42 zebrafish opsins reported by Davies et al. (2015) were used as query sequences in a BLASTn search (Altschul et al., 1990) of the database. While there are teleost species far more closely related to sablefish, zebrafish opsins were used because the entire repertoire is characterized and available in NCBI. This approach identified 48 sablefish contigs that contained sequences with significant similarity to the zebrafish opsins. Each contig was surveyed for genes using gene prediction tools Augustus (Stanke & Morgenstern, 2005) and GENSCAN (Burge & Karlin, 1997). These programs use models generated by other species (including zebrafish) and gene structure to determine splice sites, start and stop codons. The two programs were used in conjunction to best predict the correct coding sequence. As several genes were identified by these programs in any given contig, the output gene sequences were used in a BLASTx search of the NCBI non-redundant protein sequence database to confirm their similarity to other identified opsins or G-protein coupled receptors.

I aligned the predicted opsin sequences to orthologs (opsins in other fish species), previously characterized by Beaudry et al. (2017). This exposed gaps (e.g., missing exons or exons truncated by the prediction tools) suggesting that some Augustus and GENSCAN-derived opsin were incomplete. The contigs from this first genome assembly were short and there were many instances in which a single opsin gene was split up amongst contigs. In these cases, I used a phylogenetic tree approach to confirm that the stretches of sequence belonged to the same gene; I truncated the alignment file comprised of the Beaudry et al. (2017) derived sequences to span only the region of interest, generated a gene tree, and checked the contig's position in the tree relative to the other genes. This was repeated for the other regions and if they were supported in the same clade, then the sequences were concatenated to form a full opsin sequence.

In instances where small regions were missing from the sablefish opsins, it was difficult to identify them in the genome using the same approach as above, due to the zebrafish query being distantly related and limited by word size constraints in BLAST. When necessary, I used opsins from more closely related fish, such as *Oreochromis aureus* (a cichlid) and *Gasterosteus aculeatus* (a stickleback) in BLAST searches to fill these gaps. Many non-model fish species do not have their non-visual opsins characterized in NCBI and the Beaudry et al. (2017) alignment is not comprehensive for each species; thus, in instances without any other solution, I characterized cichlid and stickleback opsin genes using the same zebrafish BLAST approach. These sequences, when aligned to sablefish opsins and used in subsequent searches of the sablefish genome, improved my ability to uncover short opsin sequence fragments.

Conveniently, during this study, a second sablefish genome draft was released, comprised of 26,624 scaffolds and 50,159 contigs (BioProject PRJNA6567280, assembly GCA_000499045.2, contig N50=29,750, scaffold N50=288,526). Re-doing the zebrafish opsin BLAST searches yielded more complete sequences, as scaffolds were much longer, but with many more errors (denoted as 'N'). There were hits to 34 scaffolds, some of which contained multiple opsin sequences. The same gene predictions tools and related steps were employed.

The output from both sablefish databases and both gene prediction programs were compared to opsin sequences of other fish species (Beaudry et al., 2017) and final 'consensus' sablefish opsin sequences were manually generated. In some cases, this required omitting stretches of sequence identified by the prediction tools or concatenating sequences from the two genome databases.

Phylogenetic analyses

A Neighbor-Joining tree (JTT amino acid model, 800 bootstrap replicates) was made using the peptide sequences of the newly characterized sablefish opsins to show the evolutionary relationships among them. To show conservation among sablefish opsins, the sequence fingerprint feature of TEXshade (Beitz, 2000) was used. This depicts each opsin sequence in a horizontal line, with vertical lines across it symbolizing each residue. Two shading modes were used: one that depicts sequence conservation, and the other shade residues according to the chemical properties of the amino acids.

An additional tree (using the same parameters as above) was generated using an alignment of opsin sequences from sablefish and three other fish species, zebrafish (*Danio rerio*), starry flounder (*Platichthys stellatus*) and Asian arowana (*Scleropages formosus*). These species, whose sequences were retrieved from Beaudry et al. (2017), span the ray-finned fish lineage (each belonging to a different order) and the resulting tree can thus be considered somewhat representative of teleosts as a whole, but many others can be considered in addition.

In total, 137 opsin sequences were aligned from sablefish, zebrafish, starry flounder and Asian arowana. This alignment was truncated due to the high degree of sequence variation in the 5' and 3' ends; consequently, position one in the alignment corresponded to position 55 in *Bos taurus* rhodopsin, the standard for amino acid numbering in opsins, and ended at the position equivalent of 312 in *Bos taurus* rhodopsin. With insertions and gaps introduced from aligning all of the fish species, the final alignment was 382 residues long.

A search for sablefish-specific amino acid motifs

To determine if sablefish opsins had any distinct sequence features or motifs (i.e., differ from the consensus of the other species), which may potentially be correlated to them spending most of their lives in the dark, the subfamily logo feature of TEXshade (Beitz, 2000) was used. Subfamily logos draw attention to residues in a chosen subfamily that deviate significantly from those of all other sequences in a full alignment at a given position. The program calculates the difference between the frequency of a residue in the subfamily and all other sequences, then weights it by information content, taking into account general amino acid variability at a given position. Here, I used this feature in a slightly unorthodox way (though I use it for its traditional purpose later in this chapter as well); I set the whole sablefish opsin repertoire as the 'subfamily' to be compared to the other three teleost opsin repertoires in our four-species alignment. This was done for the complete alignment first, then broken down by the following opsin families: OPN1, OPN3, OPN4, OPN5 and Rgr/Peropsin (the photoisomerases).

I repeated these steps using an alignment of opsins from many more teleost fish species, obtained from the Beaudry et al. (2017) supplemental dataset. The alignment originally contained mammals and other non-fish vertebrates, which I removed, leaving 458 sequences. This dataset showed far better sequence diversity than the four-species alignment, however, the file exceeded the memory capacity of the TEXshade program. To reduce the size of the

alignment, the EMBOSS skipredundant (Rice & Bleasby, 2000) program was used to filter out sequences that were more than 75% identical to others in the original alignment. This approach got rid of many very similar genes within a subfamily, and thus some valuable information (i.e., the residues that were not redundant) was lost. It is important to note that this filtering is done in the order that the sequences appear in the alignment; that is, if a zebrafish rhodopsin sequence appears first, then subsequent rhodopsin sequences with more than 75% sequence similarity would be deleted. Because of this, I added the sablefish repertoire to the alignment after the filtering step so that they were not removed. Despite the caveats of this approach, I compared the results obtained with this set of 72 sequences to those of the four-species survey.

A search for well-conserved residues among fish species

I also aimed to identify residues that were well-conserved across all opsins in ray-finned fish. Because of the lack of species diversity in the Beaudry et al. (2017) alignment after being filtered for redundancy, I only focused on the opsin sequences from the four representative teleost species for this survey. A sequence logo was made from all opsins characterized in the four species using TEXshade (Beitz, 2000) to show sequence conservation. The program was also used to shade residues that were between 90% and 100% identical across every sequence in the alignment (the lower limit due to the fact that gaps in the alignment are considered non-identical). These positions were screened manually to ensure perfect conservation, ignoring gaps, as the gaps in this alignment are more likely to be due to incomplete gene characterization.

Residues characteristic of opsin subfamilies

I identified residues or motifs that were characteristic of each of the different opsin subfamilies when the four fish species were considered. To do this, two different approaches were taken. First, consensus sequences were generated for each of the following opsin subfamilies: OPN1, OPN3, OPN4, OPN5 and Rgr/Peropsins. Residues that were previously identified as being common to all teleost opsins were filtered out. To identify residues that are specific to each opsin family, I used the consensus sequences as a guide and manually screened the alignment, checking that the subfamily residues did not occur in any of the other subfamilies at that position.

Second, the subfamily logo feature in TEXshade was used, this time for its typical function of comparing different gene subfamilies. The same five major opsin subfamilies were each compared to the others one by one (e.g., OPN1 vs the four other subfamilies, etc.), and residues were flagged as statistically significant by the program. Because of the nature of the subfamily logos and their statistical basis, this analysis is quite different from the previous approach; TEXshade provides evolutionary context for the observed conservation but misses the *extent* of conservation shown by the subfamily consensus sequences.

The analyses above show conservation at the level of ray-finned fish. To investigate whether these subfamily specific/characteristic residues (as well as the sites conserved in all opsins) extend to all vertebrate orthologs, and even to invertebrates, TEXshade subfamily logos were also generated using clustered opsin sequences across Metazoa. These sequences were retrieved from Dr. Neil Clarke (personal communication, 2020); they are derived from UniProt (Bateman et al., 2021) and have been clustered based on their how well they score against different opsin Hidden Markov Models (HMM; Baum & Petrie, 1966). For the purpose of this study, I used only the clusters defining the same five main vertebrate opsin subfamilies we focused on for the fish – OPN1, OPN3, OPN4, OPN5 and Rgr/Peropsins – though these include sequences from non-bilateria to mammalia. The sequences were aligned using ClustalW Multiple Sequence Alignment (Thompson, Higgins & Gibson, 1994).

With more than 2000 highly diverse opsin sequences total and again being faced with limited TEXshade memory capacity, the sequences >60% redundant were filtered from each cluster alignment separately using EMBOSS skipredundant (Rice & Bleasby, 2000). Here, skipredundant was useful in that it mediated biases caused by the taxonomic diversity of the sequences derived from UniProt (i.e. with the majority of the sequences coming from vertebrates). The remaining sequences from each cluster were aligned and run in TEXshade to once again identify residues characteristic of each subfamily, this time on a much larger evolutionary scale. The results were compared to those of the four fish species mentioned above.

Sablefish opsin expression

i) Transcriptome survey

There are currently four transcript datasets for sablefish at NCBI: RNA-Seq data (raw reads) from juvenile testes (accession: SRX3346431) and ovaries (accession: SRX3346430) and a transcriptome shotgun assembly (TSA), which combines expressed sequence tag (EST) and RNA-Seq data. The EST sequences (34,080, assembled into 12,060 contigs) were derived from sablefish mixed brain, kidney and gill tissue samples. The RNA-Seq transcripts (96,733,584 reads assembled into 91,888 Unigene sequences) come from mRNA isolated from kidney, gonad and liver tissue of hatchery-raised adults. Combined, the EST and RNA-seq data produced 34,728 contigs greater than 400 nucleotides long (Rondeau et al., 2013). The raw reads from that same RNA-Seq experiment (again, from kidney, gonadal and liver tissue) exist in a separate sequence read archive (SRA) database (accession: SRX278009).

All sablefish opsin sequences (obtained using the methods described above) were used as query sequences in BLAST searches of the four transcript databases. The TSA database serves as the best indicator for true expression and functionality, as hundreds to thousands of reads must be assembled beyond a specific threshold (200 minimum assembly score, 0.96 stringency) to form a complete transcript. The raw read databases are useful for identifying evidence of low transcription of opsin genes, which fail to appear in the TSA database, as well as revealing the read count of the high expressors which were ultimately assembled into transcripts. Only perfect or near-perfect hits in these databases (% identity >90, e-val<0.01) were considered relevant and reported in the results. The results were constrained by read length, so the percent query cover was not factored in.

Worth noting, there is a second TSA assembly based solely on Illumina (no EST data). It is made from the same raw RNA-Seq data as above but using a different assembly method; this is not described in a publication. The second assembly produced more contigs (47,327 greater than 400 nucleotides long). Because of likely differences in filtering and assembly parameters, I reported BLAST hits to both assemblies.

ii) RT-PCR

In addition to the bioinformatics approach, RT-PCR was done in order to study opsin transcription – this is an ongoing process. Primers were designed to overlap exon boundaries (when possible) in PrimerExpress (Thermo Fisher) and checked for specificity using BLAST and

self- and cross-complementarity using BeaconDesigner (Premier BioSoft) (Table 1; see Chapter 2 for detailed primer design considerations). To cover the most likely centers of opsin-expression and various life stages, whole brains and eyes were isolated from wild adult and hatchery juvenile sablefish brains and eyes (3/4 years and 147 days old, respectively), as well as whole yolk-sac larvae (30 days old). These fish were from the same cohorts as those sampled in Chapter 3; tissue dissection and light exposure is covered in more detail there. Total RNA was extracted using the RNeasy Lipid Mini Kit (Qiazol) and reverse transcribed to cDNA (standardized to 100ng/μL) using SuperScript VILO (Thermo Fisher). The brain, eye and larval cDNA was pooled for subsequent PCR runs. Opsin genes were amplified from the pooled cDNA using Platinum Superfi II DNA Polymerase protocols (Thermo Fisher), which uses a universal annealing temperature of 60°C. PCR products were run and separated in a 3% SYBR Green (Thermo Fisher) agarose gel at 70V for 40 minutes and visualized under UV light for correct DNA band sizes.

Table 1. Opsin primers used for sablefish RT-PCR, designed using PrimerExpress (Thermo Fisher). Included are the 21 assays designed and tested thus far.

Opsin gene	Forward Primer (5'-3')	Reverse Primer (5'-3')	Amplicon size
Rh1	CGTAGGCTTCCCCGTCAAC	GGGTTCGCAGCTTCTTGTG	70
Rh2	TGTTAACGGCTATTTCATTCTTGGA	GAGAGCAACCTCACCTCCTAGTG	82
Sws1	CAAAGACTATCGACTCGTCACCAT	GGCATTAAACTGTTTGTTCATGAA	100
Sws2a1	CGGCTTCTGCTTTGCTGTTC	CGCTGCCATTTTCAGTGTAATTA	79
Sws2a2	ATCCTGGTGAACTTGGCCG	GGCTCACCATACCGCCAAG	151
Lws	GGCCACAAGTGGAATAGTGTTCT	GGCCACAAGTGGAATAGTGTTCT	81
VA1	TGGCATTCATGGTGGGTTG	TGGATTAGGCACTTCCTGAACTGT	180
Parapinopsin 1	TGGACTCTACGCCAGGTGACC	CGAGGGCCATGACAGCATAC	144
Parietopsin	CTACCTCATCCTCTACACGCTGCT	GCTCCACACTCCTGTTCAGCTTATT	110
Opn3	CATGATCTACTGCTACGGGAACAT	CCGTCTGGAGGTCCTGGAT	71
Opn4x1	AAGGCAGATGTTTGGAGTGATGT	GCTTCCCTTTTCATTGGTCTTCT	135
Opn4x2	CCCTGACATGGTGTTAAAGGATGT	TCCGAAATCTCACTGGTCAGACT	300
Opn4m1	GACCTCAATTGGCGTGATGTC	GGCACTCCAGCCGAAGAAC	106
Opn4m2	CACGCCCACTACATCATTGG	GCGTCAGAGACATCAGGAAGTCT	166
Opn4m3	ATGTCTCGCAGGAAAGCCC	CATAAGCACTCCAGCCAAAGAAG	94
Opn6a1	GGAGCTACAAAGATCGTGGGTATG	CACCGGGACGAAGAAGCAG	188
Opn6a2	CTCCACCATCCACAAGTCCTACA	ACAATGGAAATCCTTGTGACATCTC	118
Opn7a	GCTGTGGTCTACTCTGTTTTTGGAGT	TGATGATGAAGTACTCGGCTGGT	112
Opn7b(cd)	TCATCTTTCAAACACAGGTGGCT	GGTAAGCAGACTTTAAGGCAGCAC	137
Opn5	GATGAAACTCACAAAGGTGGCAA	GTTGTACATGGCTGAGGACTTGG	165
Rgr1	CTCGCCAGCATCCACTTCAT	GCTTTGTCCTGGTGCAGTACTG	70

RESULTS

BLAST searches of the sablefish genome (GCA_000499045.1) using zebrafish opsins (Davies et al., 2015) as query sequences uncovered many opsin-bearing contigs (Appendix B). With the use of two draft genomes (one with longer scaffolds and the other with less sequence gaps) and two gene prediction tools (which tended to differentially predict exon boundaries), I assembled complete opsin sequences (Appendix A). There are 36 opsins in the sablefish genome. Phylogenetic analyses were used to show that 7 are visual opsins (*Rh1, Rh2, Sws1, Sws2A1, Sws2A2, Sws2B, Lws*) and 29 are non-visual opsins (*Exorh, Va1, Parapinopsin, Parapinopsin2, Parietopsin, Opn3, Tmt1a, Tmt1b, Tmt2a, Tmt2b, Tmt3a, Tmt3b, Rgr1, Rgr2, Peropsin, Opn6a1, Opn6a2, Opn8a, Opn8b, Opn8c, Opn5, Opn9, Opn7a, Opn7b(cd), Opn4m1, Opn4m2, Opn4m3, Opn4x1, Opn4x2*). The phylogeny of only sablefish opsins is shown in Figure 2.

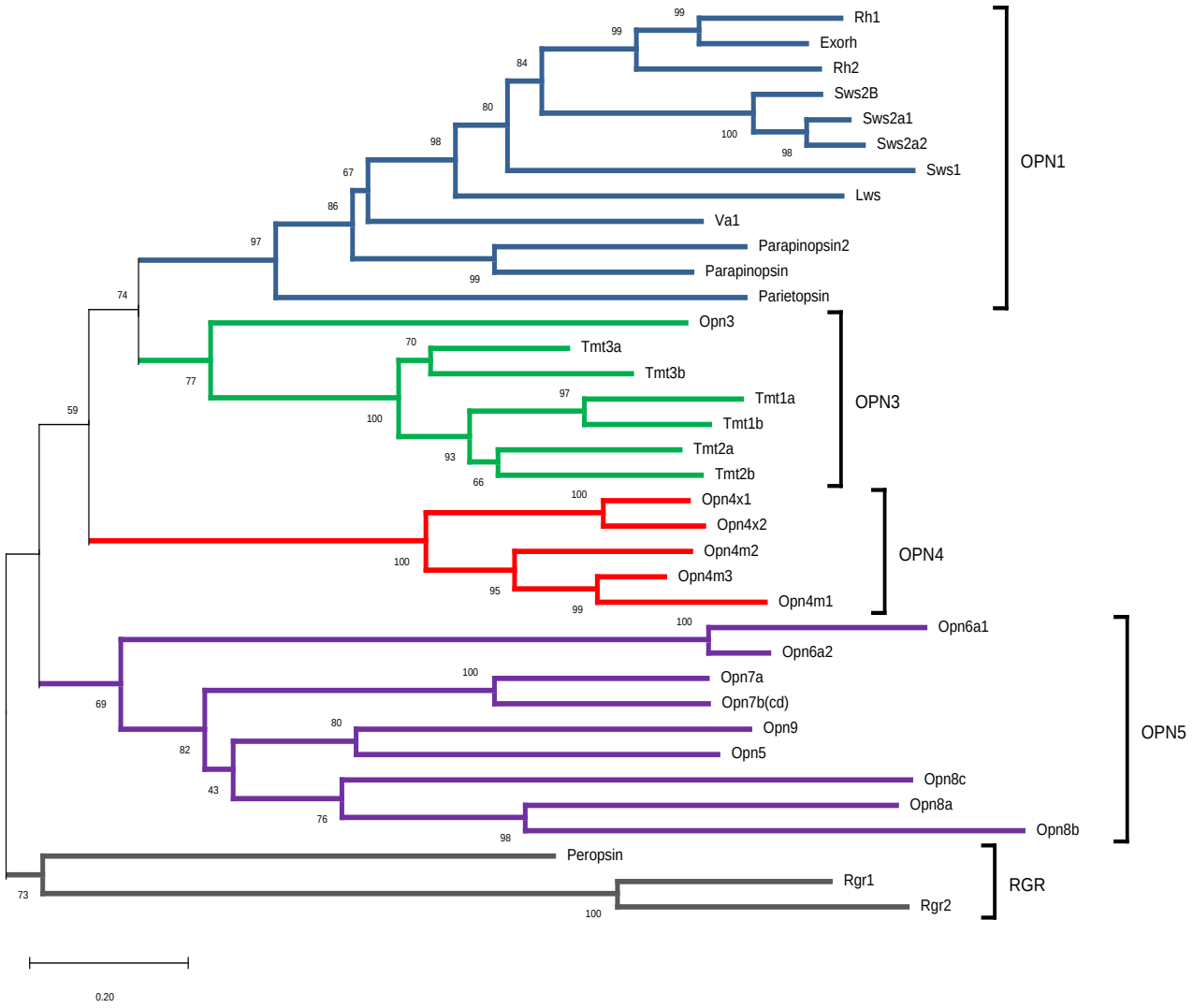


Figure 2. Evolutionary relationships among 36 sablefish opsins. Genetic distance was computed using the Jones-Taylor-Thornton amino acid substitution model and was reconstructed from these distance data using the Neighbor-Joining method. The root was placed at the base of the Rgr/Peropsins. Evolutionary analyses were conducted in MEGA. The percentage of 800 replicate trees in which the associated taxa clustered together in the bootstrap test are shown next to the nodes. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree.

Visual opsins

Sablefish have fewer visual opsins than zebrafish, seven compared to ten (Table 2). This observation reflects the fact that there have been relatively recent *Rh1, Rh2* and *Lws* duplication events in zebrafish (Rennison et al., 2012); the difference is not due to opsin loss in sablefish.

Both species have one *Sws1* gene. Sablefish have three *Sws2* genes (two more than zebrafish). The phylogenetic analysis that includes Asian arowana, starry flounder and zebrafish (taken with findings from Beaudry et al., 2017) provided no evidence of sablefish-specific visual opsin gene duplication (Appendix C). A tandem duplication at the base of the Acanthopterygii, and a percomorph-specific duplication of *Sws2a* have left sablefish *Sws2a1, Sws2a2 and Sws2B* (Rennison et al., 2012).

Non-visual opsins

Sablefish and zebrafish have similar non-visual opsin repertoires as a result of shared duplication events and retention of the same duplicates. That said, some differences were observed. For example, a *VA2* duplicate retained in zebrafish has been lost in Euteleosteomorpha (Beaudry et al., 2017), leaving sablefish with only *VA1*. Furthermore, zebrafish have four Opn7 genes (*Opn7a, b, c* and *d*), and sablefish have only *Opn7a* and the pro-ortholog of the zebrafish *Opn7b, c* and *d* genes. While both *Opn6a* and *Opn6b* occur in zebrafish, sablefish appear to have lost *Opn6b* but retained *Opn6a1* and *6a2* paralogs (from TGD), which are also found in Asian arowana and starry flounder, among other species (Appendix C). In general, the variation between the zebrafish and sablefish non-visual opsin repertoires comes largely from zebrafish-specific duplications (i.e. the Opn7 paralogs). There do not appear to be any opsin duplication events exclusive to sablefish. Thus, the large opsin repertoire in sablefish is not generated by species-specific pressures, but rather they have retained and appear to continue to use the opsins generated in a distant ancestor.

Table 2. Comparison of sablefish and zebrafish opsin repertoire sizes. The number refers to the number of genes in that particular subfamily.

Opsin Subfamilies	Visual	Exorh	Pinopsin	VA	Parietopsin	Parapinopsin	Opn3	Tmt1	Tmt2	Tmt3	Opn4x	Opn4m	Opn5/9	Opn8	Opn6a	Opn6b	Opn7a	Opn7b(cd)	Peropsin	Rgr
Sablefish	7	1	0	1	1	2	1	2	2	2	2	3	2	3	2	0	1	1	1	2
Zebrafish	10	1	0	2	1	2	1	2	2	2	2	3	2	3	1	1	1	3	1	2

Transcriptome database searches

The sablefish transcriptome included only five of the 36 sablefish opsins: *Rgr1*, *Opn8c*, *Tmt1b*, *Tmt2b* and *Opn9*, all of which are non-visual (Table 3). Opsin genes aligned to transcripts from both assembly versions (RNA-Seq and EST, and RNA-Seq only). These transcripts were derived from mixed brain, gill, kidney, liver and gonad tissue, therefore, each opsin may be expressed in all five tissues or just a subset of them. Interestingly, the assembly version without EST data had hits to more opsins, though this is more likely due to differences in filtering and other alignment parameters.

For the five genes mentioned above, the genome-derived query sequence found complete or partial hits in the TSA database. For example, the *Opn8c* transcript was complete and the *Rgr1* transcript aligned to 95% of the query; however, the *Tmt1b* transcript had only 16% query cover. In such cases, where there is low coverage of the whole transcript, this may not represent real transcription, but rather fragments (reads) from regions of other opsins, GPCRs or unrelated genes that BLAST is recognizing as the query sequence. That said, in these cases, the hits shared 92-100% identity to the opsin query sequences over the covered region and e-values for the hits were all well under 0.01, suggesting that they are significant.

Table 3. Opsin hits (partial and complete) to the sablefish TSA database. Transcripts from both assembly versions were included: the combined RNA-Seq and EST version and the one RNA-Seq data only.

Database	Opsin	Amount of gene represented by transcript (%)	Transcript accession number (NCBI)
Combined EST and RNA-Seq TSA (brain, gill kidney, gonad, liver)	*Tmt1b*	17	JO668938.1
	Opn8c	100	JO660947.1
	Rgr1	51	JO672341.1
RNA-Seq only (kidney, liver, gonad)	*Tmt1b*	16	GAJJ01028902.1
	Tmt2b	30	GAJJ01003272.1
		22	GAJJ01046995.1
		(52 total)	
	Opn9	38	GAJJ01003484.1
	Opn8c	100	GAJJ01034577.1
	Rgr1	95	GAJJ01008900.1

The RNA-Seq raw reads (SRX278009) which make up the TSA unsurprisingly have hundreds to thousands of perfect matches to *Opn8c* and *Rgr1* (Table 4). Many of the other opsins, which only had partial hits to transcripts or did not appear in the TSA search at all, had several significant hits in the raw read database. Each individual hit/read covered only a small region of the query sequence, as the reads are relatively short (~100 nucleotides), but together, they often spanned much of the query opsin sequence. Among the opsins with the next greatest number of strong hits (e-value <0.01, >90% identity) were *Parietopsin, Opn3, Tmt1a, Tmt1b, Tmt2b, Opn9* and *Peropsin*, though there had between 50 and 100 reads per gene, compared to the thousands for the opsin genes with high enough expression to be assembled into full transcripts in the TSA database. The genes with a higher read count also tended to have greater coverage of the query sequence (i.e., close to the full transcript).

Very few opsins had hits to reads in a BLAST search of juvenile testes and ovary RNA-Seq raw read databases (SRX3346431 and SRX3346430, respectively). Those that did typically only had one significant match below the e-value<0.01 threshold. In cases where there was more

than one match, the reads were not distinct (i.e., they were overlapping in the same region of the transcript). In the juvenile testes, reads aligned to 9 opsins; in ovaries, 6. *Rh2*, *Tmt1a*, and *Opn9* occurred in both tissues (Table 4).

RT-PCR

Using RT-PCR, I provided additional evidence of expression. As I was targeting the most likely centres of opsin expression here (brain and eye), the tissues used in the making of the transcriptomes (above) and those pooled for my RT-PCR research do not overlap, apart from adult brain tissue. The aim of this experiment was not to quantify or localize expression, but rather to look for evidence of transcription for each of the opsin genes in any tissue type; that is, to confirm that the genes identified in the draft genomes are transcribed. 21 opsins have been included in this RT-PCR survey so far and all were shown to be expressed (Table 4).

Table 4. Summary of evidence of opsin transcription in sablefish from BLAST searches across three raw read sequence collections (shown as: number of reads hits/transcript coverage) and from RT-PCR (amplification in any sample). Only hits with >90% identity and e-value<0.01 were considered here. Both the number of reads and % of the query transcript covered were reported, as many reads were not distinct (i.e., were overlapping).

Gene	Juvenile testes	Juvenile ovaries	Adult gonad, liver, kidney	RT-PCR
Rh1	**1 / 16%**	0	**2 / 19%**	**+**
Rh2	**3 / 46%**	**3 / 38%**	0	**+**
Exorh	0	0	0	TBD
Sws1	0	0	**2 / 20%**	+
Sws2a1	0	0	**5 / 31%**	+
Sws2a2	0	0	**5 / 40%**	+
Sws2b	0	0	**3 / 19%**	TBD
Lws	0	0	0	+
VA1	0	0	**13 / 57%**	+
Parapin1	0	0	**6 / 23%**	+
Parapin2	0	0	0	TBD
Pariet	0	0	**50 / 56%**	+
Opn3	**2 / 48%**	0	**67 / 95%**	+

Tmt1a	**1 / 10%**	**2 / 21%**	**53 / 55%**	TBD
Tmt1b	0	**2 / 33%**	**94 / 41%**	TBD
Tmt2a	0	0	**13 / 45%**	TBD
Tmt2b	**1 / 19%**	0	**84 / 63%**	TBD
Tmt3a	0	0	**16 / 44%**	TBD
Tmt3b	**1 / 35%**	0	**25 / 54%**	TBD
Opn4x1	0	0	**2 / 5%**	+
Opn4x2	0	0	0	+
Opn4m1	0	0	**8 / 25%**	+
Opn4m2	0	0	**3 / 16%**	+
Opn4m3	0	0	**2 / 11%**	+
Opn6a1	0	0	**4 / 17%**	+
Opn6a2	0	0	**14 / 48%**	+
Opn7a	0	0	**2 / 11%**	+
Opn7b(cd)	0	0	**2 / 13%**	+
Opn8a	0	0	0	TBD
Opn8b	0	**1 / 29%**	**19 / 56%**	TBD
Opn8c	0	**25 / 78%**	**1597 / 88%**	TBD
Opn5	**1 / 12%**	0	**12 / 36%**	+
Opn9	**1 / 20%**	**2 / 10%**	**86 / 92%**	TBD
Rgr1	**6 / 64%**	0	**300 / 100%**	+
Rgr2	0	0	**1 / 11%**	TBD
Peropsin	0	0	**12 / 49%**	TBD

Sablefish-specific opsin signatures

Sequence conservation and divergence among the 36 sablefish opsins is represented by a TEXshade fingerprint (Figure 3), which ultimately forms the basis of the relationships depicted in the phylogenetic tree (Figure 2). Many of the conserved sites shown here are not specific to this species, but shared amongst many. For example, the retinal-binding lysine and the NPxxY motif known to be important for G-protein activation in the seventh transmembrane domain occur in all or most opsins. The image also illustrates regions that are variable, such as the intracellular loop between helices five and six, which does not directly interact with the

chromophore, nor does it contain a disulfide bridge. These regions tend to have more variability in other species for the same reasons.

To determine if there are any key differences between sablefish opsins and those of other fish species (e.g., conserved sites represented in Figure 3 that others do not have), additional TEXshade analyses were conducted. The two all-opsin comparisons, one using filtered fish opsin sequences from Beaudry et al. (2017) (Appendix E) and the other using the entire suite of Asian arowana, starry flounder and zebrafish opsins (Figure 4), yielded the same results. Sablefish opsin sequences, taken as an entire set, are similar to all other fish opsins; there were no residues flagged as significantly characteristic or specific to sablefish opsins as a whole, despite them being the only deep-sea fish of the four teleost-representative species. When the alignments were further separated by subfamily and the same analyses were run, there were still no signs of significant differences between sablefish and the other species (example in Appendix G). As this method (the subfamily logo) is not explicitly designed to find residues specific to a single species in the alignment, it is possible that there are key signatures in sablefish that were not identified here; the important takeaway is that there are no key residues shared by other teleosts that are *missing* in sablefish.

Figure 3. TEXshade fingerprint showing all sablefish opsins. Each row depicts an opsin sequence and each vertical line along it represents one amino acid. The shading shows the degree of sequence conservation among all of the 36 sablefish opsins. The transmembrane regions depicted represent those of *Bos taurus* rhodopsin. An alternative chemical fingerprint (depicting amino acid properties) is available in Appendix D.

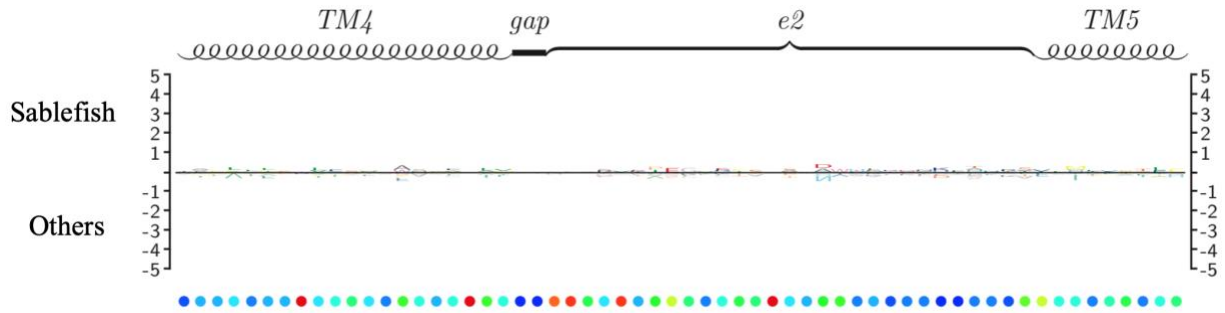


Figure 4. All sablefish opsins compared to the opsins of starry flounder, Asian arowana and zebrafish using TEXshade subfamily logo in an atypical way (see Methods). This is a cropped region of the full alignment (available in the Appendix F, showing the same results). Residues characteristic of sablefish opsins (taken as a whole) appear above the mid-line and residues characteristic of the three other fish species opsins appear below the mid-line. Proximity to the mid-line (as is shown at every position across the full alignment) indicates that there are no meaningful deviations in sablefish opsins and those of the other species. The Y-axis is measured in bits (a unit of information). The structural images overhead correspond to the positions in bovine Rho. Conservation is shown as bullets on a hot-cold scale, with red meaning high conservation in the entire alignment at that position.

Sequence conservation

To search for conserved sites amongst all opsins of the four chosen fish species, the same all-opsin alignment was used to generate a TEXshade sequence logo (Figure 5) and a consensus sequence (>90% conservation). This revealed 18 well-conserved amino acid residues; 5 were perfectly conserved in each of the four different species (Table 5). These residues are largely known to be involved in chromophore binding, G-protein activation and other basic GPCR and opsin roles.

The remaining 13 residues of the initial 18 identified were generally well-conserved (again, over 90% conservation among the opsins in the alignment), but not perfectly conserved like the five sites aforementioned in Table 5. In some instances, there was only one sequence that differed from all others at that position, however, since these four species were meant to be representative of all ray-finned fish, a difference in one of the four selected species could certainly suggest that other ray-finned fish differ at that position as well. It is also possible that some of these differences are due to uncertainties in the alignment. Under any circumstance, these sites are reported separately in Table 6.

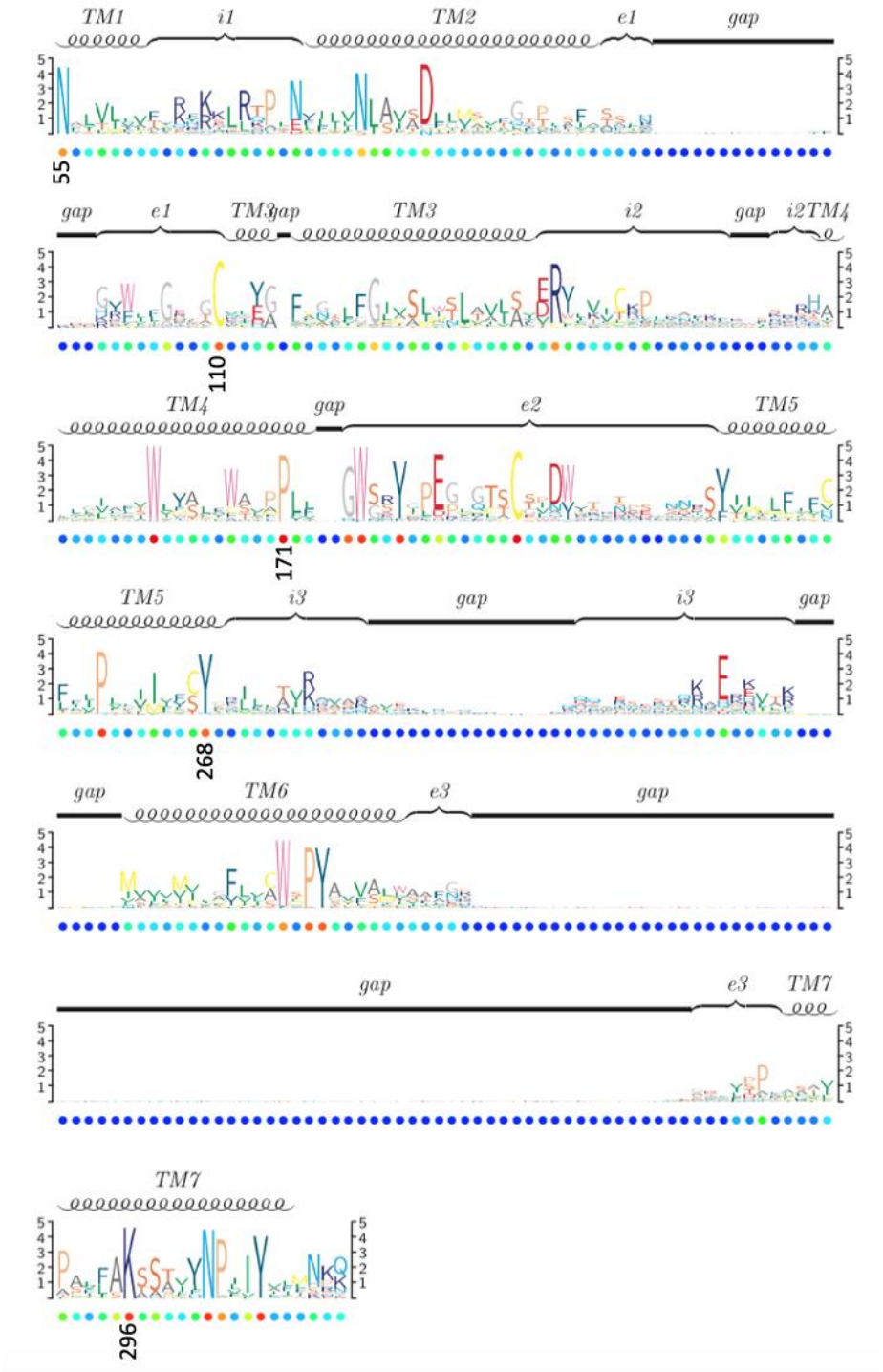


Figure 5. A sequence logo created by TEXshade showing conserved amino acid positions in an all-opsin alignment of four teleost fish: sablefish, zebrafish, Asian arowana and starry flounder. The symbol height is measured in bit (information content), and the amino acids are colored according to chemical composition. Perfectly conserved residues show their position (according to *Bos taurus* Rho) below. The helical and loop structures are displayed are also according to their positions in bovine Rho. Conservation is shown as hot-cold bullet points, with red being a highly conserved position.

Table 5. Perfectly conserved amino acid residues across all opsins in the four-species alignment and their proposed roles. Gaps at these positions were treated as missing data rather than deviations from the consensus. [1]van Hazel et al., 2016; [2]Giraldo-Calderon et al., 2017; [3]Kozmik et al., 2008; [4]Morra et al., 2018.

Amino Acid	Position according to *Bos taurus* Rho	Notes from literature
N	55	For stability/folding requirement, chromophore regeneration[1]
C	110	Part of disulfide bridge[2]
P	171	Opsin trace residue, retinal binding[3]
Y	268	Conformational changes during isomerization/formation of binding pocket[2], lipid translocation[4]
K	296	Chromophore binding site (Schiff base binds to retinal)[2]

Table 6. Amino acid residues that are largely conserved across all opsins in the four species (>90%), but with some deviations, as mentioned in column three. [1]Giraldo-Calderon et al., 2017; [2]Adamian & Liang, 2002; [3]Gross et al., 2004; [4]Kimata et al., 2016; [5]Deupi, Standfuss & Schertler, 2012; [6]Audet & Bouvier, 2012.

Amino Acid	Position according to *Bos taurus* Rho	Not conserved in	Notes from literature
R	135	*Opn7*	R in the (E)DRY motif, Gt activation[1]
W	161	Asian arowana *Opn8c*	Polar clamp, stabilizing[2]
G	174	*Opn7*	Edge of helix 4
W	175	*Opn9*	"Opsin-specific trace residue", chromophore regeneration[3]
Y	178	Zebrafish *Opn5*, Asian arowana *Lws*	In extracellular loop
C	187	Asian arowana *Lws*	Disulfide bridge, protein folding[1]
P	215	*Rgr1*	Hinge that guides configuration changes and stabilizes[4]
Y	223	*Opn8c*	Protein stabilization[5]
W	265	*Sws1*	Spectral tuning site, conformational changes in binding pocket[1]
P	267	Asian arowana *Opn8c*	Gt activation[1]

N	302	*Rgr2*	NPxxY motif, G protein activiation (maintain open state)[6]
P	303	*Rgr*	NpxxY motif, G protein activation
Y	306	Sablefish *Sws1*	NPxxY motif, Gt activation

Specific opsin sequences and subfamilies that introduced variation at the well-conserved sites were identified (Table 6). Largely, they were opsins belonging to the OPN5 family (i.e., Opn5, Opn7, Opn8 and Opn9) or Rgr photoisomerase subfamily. Asian arowana (and notably *not* sablefish) stood out as a species in which more deviations occurred.

Of the 18 positions that were initially identified as perfectly or well-conserved, 12 of them occur in a transmembrane helix (assuming the protein structure follows that of bovine rhodopsin) and 6 occur in either the intracellular or extracellular loops.

Subfamily characteristic residues

While the residues that are shared amongst all teleost opsins are of great importance, the molecular differences that allow us to classify each sequence into a subfamily are also key. These subfamily specific characteristics form the basis of phylogenetic analyses and they also lead to functional diversity in the opsin gene family, ultimately explaining why there are so many opsins.

As part of my endeavor to determine whether or not sablefish opsins differed from those in other fish, I looked for residues that were specific to subfamilies, with the goal of seeing if these occurred in sablefish subfamilies too. The TEXshade logos above compared the full repertoire of opsin sequences in sablefish (or full suite of genes in a subfamily) to those of the other three representative teleost species; thus, if there were marks of pseudogenization or important residues missing in a single sablefish gene, they could have been masked by the numerous other opsin sequences in the analysis.

The first effort to identify opsin subfamily-specific residues used consensus sequences generated from the four-species subfamily alignments. This showed me well-conserved sites within individual subfamilies, but I had to visually scan the all-opsin alignment to see if the same amino acids occurred in any other subfamilies at the same position (i.e., if they were specific or not). Certain subfamilies had many more conserved residues in general (not necessarily specific) compared to others. For example, OPN4's had a much higher degree of conservation amongst

paralogs and orthologs than OPN5's when the same species are considered. Also worth noting is that some subfamilies contain many more sequences/paralogs across the four teleost species used. For instance, there are 38 sequences that belong to the OPN5 family, but only 19 OPN4 sequences, thus this plays a role in the degree of conservation observed. Even groups with high sequence conservation at many positions had relatively few amino acids that were well-conserved *and* specific only to that group at a particular position (i.e., do not occur in any other subfamilies). An example is the OPN3 subfamily, which had 30 sites at which amino acids were more than 90% conserved, but had no residues that were exclusive to OPN3's at that site. Only the OPN4 and Rgr/Peropsin (photoisomerase) subfamilies had specific, well-conserved residues; six and two, respectively (Table 7). This analysis revealed to me that there were few perfect subfamily 'signatures' that could be used to evaluate sablefish opsin conservation and function.

Table 7. Well-conserved amino acid residues *specific* to selected monophyletic groups/subfamilies, according to the 4 species all-opsin alignment. The number of sequences refers to the number of sequences in that selected group from all four species. Their positions are given according to bovine Rho, rather than their position in the alignment.

Subfamily/Group	Number of sequences in the group	Group-specific amino acid @ position according to bovine Rho
OPN1	44	N/A
OPN3	24	N/A
OPN4	19	Q90, W189, V211, W248, Y286, I294
OPN5	38	N/A
Rgr/Per	12	E67, P292

Despite this, sablefish opsins nest among known opsin subfamilies the phylogenetic tree, so it is clear that other characteristic (but not necessarily perfectly conserved and specific) residues among subfamilies should also be taken into account when evaluating the sablefish opsins. To find residues characteristic of the various fish opsin sub-families, I made use of the TEXshade subfamily logo feature. This selects only relevant residues, based on a stronger computational/statistical basis (i.e. calculates the difference between the frequency of a residue in the subfamily and all other sequences, then weights it by the information content). Figure 6, cropped from a larger subfamily logo of the Opn4's, is used below to highlight differences between the two approaches and their results.

As an example, the visual screen of the OPN4 subfamily's most well conserved residues revealed that a tryptophan (W) at position 189 (numbering according to *Bos taurus* rhodopsin) occurs in all OPN4's and no other sequences; however, this residue is not flagged as relevant/characteristic of the OPN4's by the TEXshade subfamily logo. Instead, tyrosine (Y) at position 191 is deemed more characteristic of that group. Y191 is not specific to the OPN4 subfamily; many of the visual and Rgr opsins share this amino acid at this position. In fact, 23 non-Opn4 sequences have this Y191 along with of the 19 OPN4's in the four-species alignment, but this is a position where little variation occurs in the entire all-opsin alignment. Thus, despite more non-OPN4 opsins having this tyrosine than the OPN4 subfamily itself, due to the fact that nearly every other non-OPN4 opsin has a tryptophan (W) 191 (95 sequences), the frequency of Y191 ends up being higher in the OPN4 subfamily.

TEXshade subfamily logos were generated for the five main opsin subfamilies. These subfamilies were further subdivided into smaller clades as well. As these logos are space-consuming, the findings from this analysis have been summarized in Table 8. In some cases, positions flagged as significant/characteristic of a subfamily were not attributed to a conserved amino acid, but rather to a lack of a conservation for a particular amino acid that most other subfamilies have at that position.

The consensus residues, the subsequently identified subfamily-specific residues and the significant TEXshade subfamily logo residues have contrasting results (Figure 7), however, they are all informative in their respective ways. That said, the TEXshade subfamily logo approach is strong in its ability to identify well-conserved sequences and specific sequences, but only in situations in which variability at a certain position is low for all other sequences in the alignment, as well as in its ability to identify a lack of conservation in subfamilies, as mentioned above. Some of the residues identified in this set of analyses have previously been identified (mostly those differentiating the visual opsins from the others), but the majority have not been identified or functionally characterized.

Regardless of the method used, these surveys once provide no evidence that sablefish opsins are different than those of other fish, as their subfamilies are comprised of the same characteristic amino acid motifs as the other species included in the analyses. Therefore, I believe my research in later chapters using sablefish as a model can be extrapolated to gain insight to other fish species as well.

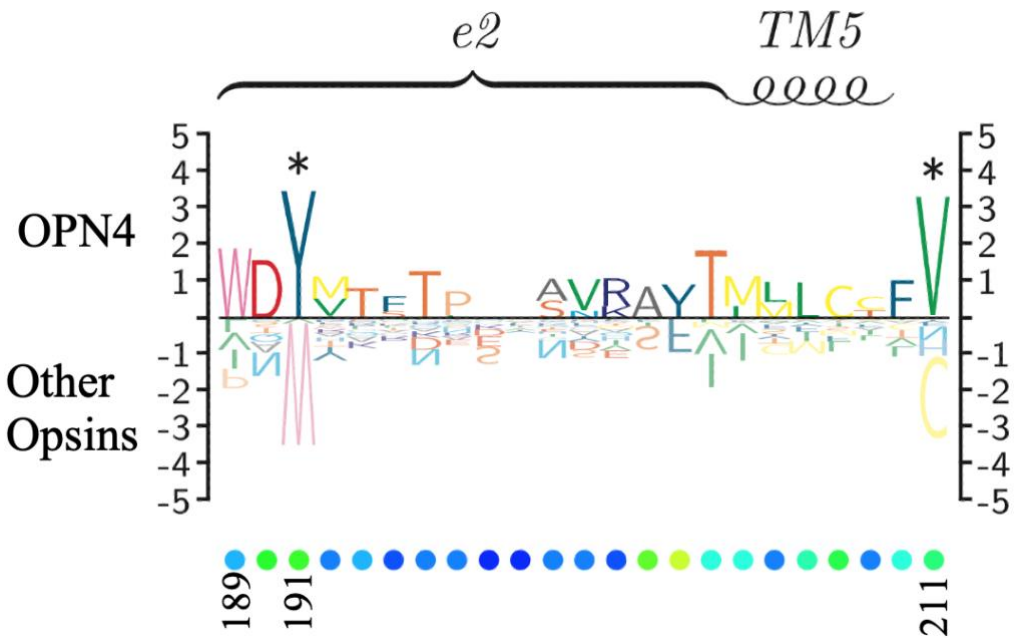


Figure 6. A subfamily logo generated with TEXshade. This is a cropped image from a comparison of the four-species OPN4's vs. the rest of the opsins in the alignment. All letters appearing on the upper side of the mid-line correspond to residues that are characteristic of the subfamily of interest (Opn4's), which meaningfully deviate from the rest of the opsins at that position. The letters upside down under the mid-line are characteristic of all other opsins. The asterisks are placed above the most relevant conserved residues (bit over 2.321). The Y-axis is the measured in bits. The structural images overhead correspond to the equivalent positions in bovine Rho. The conservation is shown on a hot-cold scale, with red meaning high conservation in the entire alignment at the position. The numbers below the logo show positions according to bovine Rho.

Table 8. Residues flagged as significant by the TEXshade subfamily alignment feature. The main opsins families (bolded) were further subdivided into smaller subfamilies. The number of sequences in the subfamily may influence the degree of conservation and therefore the number of subfamily characteristic residues. Cases where the subfamily is notably missing a residue that all other sequences seem to have is denoted by (X); for example, G/N(XD) at position 83 shows that the subfamily of interest has both G and N at that position, but the site was flagged as significant due to the fact that it deviates from the D that most other opsins have at position 83.

Subfamily	**# Sequences in subfamily**	**Amino acid and position according to *Bos taurus* rhodopsin**
OPN1	44	G/N(XD)83, F103, E113, E134, Q312
Visual and Exorh OPN1	31	G/N(XD)83, F/M(XD)103, E113, G/C(XF)120, E249, Q312
Non-visual OPN1	13	F103
OPN3	43	I79, S80, M/V/L(XC)140
Opn3	4	N/S(XG)106, D113, V140, A142, N176, L180, F203, M219, S310
Tmt	20	I79, M/L(XC)140, Q/T/S190, Y/G261, I301
OPN4	19	Q90, K101, A114, D134, T140, Y191, V211, S264, H309
OPN5	38	L69, E/D(XN)73, H101, V/T134, G176, P/R182
Opn5/9	6	E73, H101, D134
Opn6	8	L69, D73, S/W(XG)106, D113, T/S134, D/A181, R182, Y184, G185, T291, A298, S310
Opn7	11	L69, E73, L90, H101, A114, V134, C135, F167, G/H(XG)174, H/Q247, L253, K310, N312
Opn8	12	L/R(XR)69, E73, H101, A19
Rgr/Per	12	A/V(XG)121, D134, M/V(XI)219, T/Q(XE)247, A/L(XP)291, A/V(XP)303

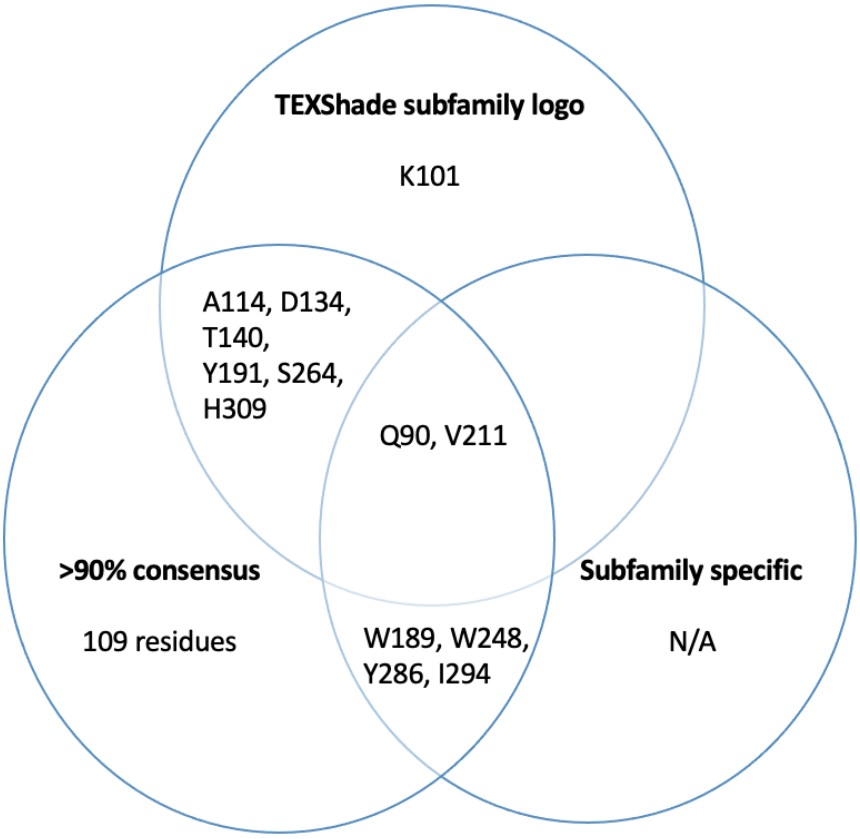


Figure 7. A Venn diagram demonstrating shared and contrasting results from different methods used to identify important amino acid residues in teleost opsin subfamilies – this example shows the OPN4 family. Consensus residues refer to highly conserved (>90%) amino acids among all OPN4's (including residues shared by all opsins/other subfamilies). Subfamily specific residues are those that occur in all OPN4 sequences and in no other opsins at that particular site. TEXshade residues are those flagged as statistically significant or *characteristic* of OPN4's using the program's default statistics. In this example, there are only two amino acid sites that all three approaches agree upon.

Extension to other Metazoans

With a collection of sites that appear to be important to each of the teleost opsin subfamilies, I aimed to determine if these trends extend to other Metazoans. TEXshade subfamily logos were generated using opsin sequences from a greater diversity of Metazoans (identified by HMM's derived from vertebrate opsin families) including echinoderms, insects and mammals. Given the impressive species and sequence diversity, the results of this survey differed immensely from those reported for teleost opsins alone. Namely, there were very few residues identified as characteristic of the broader Metazoan opsin clusters; unsurprising as HMM's cluster sequences based on a match to a profile, rather than a pairwise alignment of

amino acids. The sequence diversity within a particular cluster is striking. From my observations, many of the subfamily characteristic sites identified in teleosts extend to other vertebrates, but rarely further (though I did not identify precisely where the conservation is lost). Therefore, sablefish research may not help to speculate function of very distantly related organisms, but it can be applied to and aid in the understanding of other vertebrates to some extent.

DISCUSSION

Opsins are excellent models of molecular evolution. The ancestor of vertebrates had 23 opsins, of which 18 were non-visual (Beaudry et al., 2017), and this repertoire expanded significantly to a diversity of life histories, morphologies and habitats. Fish have especially large opsin repertoires, making them good candidates to advance the understanding of opsin gene function and diversity. Here, I have added the deep-sea sablefish to the short list of fish species whose full opsin repertoires have been characterized and analysed.

Sablefish opsins – a means for comparison

My research shows that sablefish have a repertoire of 36 opsins (7 visual and 29 non-visual). Despite the fact that the genome available at the time of my research was fragmented, full coding sequences were identified (in many cases assembled from exons distributed among multiple contigs) for each opsin gene. With future updates to the genome (i.e., the chromosome-level assembly currently in progress), these sequences should be further solidified. Additionally, their chromosomal location, and consequently their linkage and origin through various means of duplication, will also be better understood.

The large opsin repertoire in the sablefish genome is in line with previous findings that show a remarkable increase of opsins in teleosts. Whole genome duplication events specific to ray-finned fish and subsequent tandem duplications are the basis of this expansion (Taylor et al., 2003; Rennison et al., 2012; Braasch et al., 2016; Beaudry et al., 2017). Duplicated genes have many fates, with the most common being degeneration or loss. However, as purifying selection relaxes briefly following a duplication event, there is opportunity for newly duplicated opsins to gain new function and persist (Prince & Pickett, 2002). With respect to opsins, these new functions can include changes in their sensitivity to light, their lambda max (Cortesi et al., 2015), their interactions with chromophores and G-proteins (Terakita, 2005), and their expression

patterns (levels of expression, timing of expression, or tissue specificity) (Kojima et al., 2008). The persistence of duplicates over evolutionary time suggest that they are advantageous to the host organism.

Sablefish have a very similar opsin repertoire to zebrafish, with orthologs for almost all zebrafish opsins. Both repertoires are impressively large, with 36 and 42 genes, respectively. The majority of the duplicates generated from whole genome duplication have been preserved in both, though the two species have experienced their own history of opsin gene loss over evolutionary time (e.g. the loss of sablefish *Opn6b*). In the case of zebrafish, there are also clear examples of species-specific duplication events leading to more functional opsin genes (particularly within the visual opsins, e.g. *Lws* and *Rh2*). There is no such evidence in sablefish, though it is unclear whether from a lack of chance mutations resulting in duplications, or if any duplicates that did arise were selected against.

The theory that large opsin repertoires are maintained in teleosts because they are advantageous in highly variable photic environments still stands, but the breadth of each species' repertoire is not well correlated with life history complexity and environmental light exposure. That is, sablefish are exposed to a much more diverse range of photic environments than zebrafish, from the deep aphotic zone to more shallow waters, yet zebrafish have the larger repertoire of the two. On the other hand, though sablefish spend most of their lives in the dark, their opsin repertoire has not been significantly reduced, likely due to heavy selective pressures and the need for these opsins in the stages that occur in the light. A more extreme example of this lack of correlation between repertoire diversity and photic habitat is the Mexican cavefish, which lives in perpetual darkness yet has 33 opsin genes (Simon et al., 2019). It is generally surprising that such distant and different species share such similar opsin repertoires. Opsin gene loss/pseudogenization associated with life history has been observed in several species, but this is typically for visual opsins; for example, the loss of UV cone opsins in blind mole rats (David-Gray et al., 2002) and long-wave sensitive opsins in many deep-sea whale species (Meredith et al., 2013). Most of the variation between teleost species lies among the visual opsin repertoires, despite the understanding that all opsins use light as a stimulus, which may support the finding that non-visual opsin repertoires do not appear to be as evolutionarily plastic as visuals opsin repertoire (Beaudry et al., 2017).

While we do not know what conserved role non-visual opsins play in this diversity of species, we can now strive to find out using the sablefish as a new model species. The first step toward this goal will be to further investigate expression. The same widespread non-visual opsin expression that is seen in zebrafish (Davies et al., 2015) is not well reflected in the survey of the sablefish transcriptome presented in this chapter, as only a few sablefish opsins aligned completely to transcripts. The transcriptome is limited to a small number of ESTs and reads, so genes with low expression levels would be less likely to show up. On the other hand, the RT-PCR analysis using eye and brain tissue (the most opsin-rich tissues, according to work done in zebrafish; Davies et al., 2015), shows expression of all 21 opsins tested to date.

We are left to wonder if sablefish have retained this large opsin repertoire for use only during the stages exposed to light (after the loss of their yolk-sacs in the larval stage until they are around two years of age). Chapter 3 tests this hypothesis. Other deep-sea fish occupying similar depths in the ocean as sablefish (which also spend early developmental stages in shallow waters) express green opsins exclusively during larval stages, when they are “needed”, for example (Lupse et al., 2021). Non-visual opsins have not been studied in this way, and it will be interesting to see, using sablefish, whether this same trend is mirrored. While this thesis explores a hypothesis based on environmental light, one can postulate that the highly conserved non-visual opsin repertoire among so many fish species suggests an alternative, light-independent, role of non-visual opsins. Given repertoire similarity to other fish, the findings uncovered in sablefish may help us to understand how opsins are being used in other species too.

It is difficult to speculate on functional adaptations, advantages and differences between opsins without first studying the protein sequences. Sequence analyses in conjunction with knowledge of orthologous genes provide a strong foundation for understanding gene family function in newer model organisms, such as sablefish.

The comparison between Asian arowana, zebrafish, starry flounder and sablefish revealed not only that the sablefish opsin repertoire is similar to those of other species, but also that this similarity extends also to the opsin sequences. That is, despite their differing life histories and their occupation of many photic environments, on a molecular basis, sablefish opsin sequences are not unlike those of other ray-finned fish (i.e. there is no evidence of species-specific signatures). A more comprehensive study of several deep-sea fish in the future may reveal sites that were not identified here using only four species, such as spectral tuning sites for adaptive

photosensitivity in the deep (which has been done for some visual opsins; Musilova et al., 2019), or we may come to learn that adaptation depends most on the regulation of expression of these large opsin repertoires.

Opsin evolution in general

As part of this sablefish opsin sequence analysis (comparisons with other teleosts), I identified residues that are perfectly conserved and nearly perfectly conserved across the four teleost-representative species. These sites are mostly well-documented in the literature as having some known necessary opsin function, such as chromophore binding or G-protein activation. The conserved amino acids in this list more commonly occur within the helical structures of the opsin GPCR, which is also in line with the functional requirement of chromophore binding, as this occurs within the binding pocket. In terms of the residues that were not perfectly conserved across all four species, but well-conserved nonetheless, variation was often a result of subfamily-level differences (e.g., the photoisomerases and OPN5 opsins) or single gene deviations, with the latter having potential to be results of sequence misalignment. There is no evidence that any one of the four species studied was responsible for introducing variation at sites largely conserved in the others. As their name suggests, the photoisomerase opsins do not activate G-proteins, but rather reisomerize retinal from the all-*trans* form to the 11-*cis* form (Hao & Fong, 1999). As such, it was not surprising to discover that this group differed from the other subfamilies in terms of the make-up of the protein at positions important for G-protein activation. The OPN5 family, composed of Opn5, Opn6, Opn7, Opn8 and Opn9, has only been fully characterized in fish within the past decade and most of the genes encoding these proteins (with the exception of *Opn5*) have yet to be studied intensively. The finding that the translated Opn7 and Opn9 sequences lack residues that are perfectly conserved across all other teleost opsins suggests that they may deviate from typical opsin mechanisms (e.g., chromophore handling, G-protein binding, etc.). Studying them further may reveal more about opsin functional diversity.

Most opsin studies do not characterize the entire opsin repertoire of the study organism, nor do they offer a thorough sequence analysis. Often, only the most well-conserved motifs and known functional domains are compared between opsins, but there is far more that unites the opsin subfamilies. This is made clear through their phylogenetic relationships, but it can be difficult to identify significant differences between the groups.

Here, I identified amino acid residues specific to the major opsin subfamilies and used subfamily logos to further identify residues characteristic of each subfamily compared to all others. These approaches yielded different results, both of which have potential to be important in the functional division of opsin groups; however, the subfamily logos effectively used the frequency of amino acids as well as overall conservation at particular site, which is more likely to be biologically significant. In cases where these two approaches converge on a residue, it may be of particular interest. Many of the subfamily characteristic residues have not been explicitly identified or surveyed before, thus whether they contribute to functional diversity remains unknown for the time being, but these findings are an excellent launching point for future studies.

This survey highlights the importance of a comparative approach to opsin research; without the context of the repertoires from multiple fish species, it is very difficult to discern important differences between the subfamilies. The approach taken here will aid in understanding why fish have retained so many distinct opsin genes. It is clear that the suite of opsins identified in sablefish must be carrying out conserved functions in an incredible diversity of light environments, given the photic zones they inhabit through development. Since sablefish opsin sequences are similar to those of the current fish model, zebrafish, and to the two other species studied here (Asian arowana and starry flounder), the logical null hypothesis moving forward will be that the opsins shared between these species have similar expression patterns and function; this hypothesis can be tested in the future by investigating expression domains in various species.

While the most well-conserved sites identified among all opsins in fish are also present in other Metazoan subphyla, the residues flagged as characteristic of the various ray-finned fish opsin subfamilies are not always extended beyond vertebrates. That is, invertebrate opsins from the Dr. Neil Clarke's HMM clusters show enough similarity to be grouped with vertebrate opsins into subfamilies, but this similarity is not reflected in *specific* amino acid residues, and rather a weighted matrix of likely substitutions. The observation that invertebrate Metazoans have homologs of almost all of the opsin subfamilies highlights the long evolutionary history of opsins and light-sensitivity, culminating in diverse repertoires within modern day species, such as sablefish.

Continuing sablefish opsin research

With a repertoire of 36 opsins and access to all stages of life, sablefish are not only amenable to experimental manipulation, but they can also provide insight into opsin expression throughout development. The connection to development is particularly pertinent in the case of sablefish, as their life stages coincide with distinct photic environments and may therefore have different light-sensing requirements. The tracking of opsin expression through development has been done for some opsin families in zebrafish (Schmitt, Hyatt & Dowling, 1999; Takechi, 2005), but zebrafish develop entirely in the light; sablefish, on the other hand, remain in the deep until they are swimming larval fish, and thus provide an interesting basis for comparison. Sablefish also develop at a much slower rate than zebrafish, which presents an opportunity to study opsin expression on a much finer scale.

Currently, little is known about opsins in organisms inhabiting light-restricted or dark environments. Some research has characterized opsins in the genome of dark-dwelling fish; some deep-sea species have evolved multiple Rh1 genes (the dim-light rod opsin) to aid their vision while losing cone opsins (Musilova et al., 2019), and as aforementioned, cavefish have been shown to have an impressive suite opsin genes (Simon et al., 2019; Cavallari et al., 2011). These dark environments are not within the Sun's reach, meaning light stimuli are relatively weak (for example from bioluminescence), and it is not clear if and how effectively opsins are being 'turned on' in these environments, especially in internal tissues. Thus, beyond a need to navigate or source food visually, having light mediate other biological processes in the dark environments is not intuitive and may not be productive. Recent studies have uncovered light-independent (chemo- or thermosensory) roles of opsins as well (Leung et al., 2020; Perez-Cerezales et al., 2015), which may contribute to the persistence of many opsin genes in sablefish which spends most of their lives in the dark. The compelling questions moving forward: when and where are all 36 of these sablefish opsins expressed, and what are they being used for?

Until recently, the visual opsins have been the main focus of opsin research. While they are certainly fundamental in understanding how and what animals can see, there exists a world beneath the conscious visual stream that is underexplored. The discovery of non-visual opsins and the subsequent discovery that most vertebrates, especially fish, have very large and very similar non-visual opsins repertoires suggest that light is a far more important biological regulator than previously appreciated (or the alternative: that we know very little about what

opsins do and how they are regulated). The characterization of opsin repertoires from different species has been instrumental in not only supporting this suggestion, but also in understanding the evolution of the opsin gene family as basally non-visual photoreceptive proteins. Every species studied contributes new insight and understanding into the fairly enigmatic topic of opsins, each bringing along their unique life histories, morphologies, behaviors, etc. which all have the potential to influence opsin expression and light-sensitivity.

CONCLUSION

This study provides the first evidence of opsins in sablefish. I characterized a repertoire of 36 opsins, compared these to one another and to opsins from other species. I also identified opsin transcripts in sablefish. Despite occupying different light environments than the other teleosts considered here, the sablefish repertoire and opsin sequences appear to be typical. Though these do not appear to be well correlated with life history and light exposure, the expression of these genes may be. Thus, the next step is to determine when opsins are expressed because this information will, on its own, provide insight into function and will tell us what life stages are most suitable for experiments to further study function.

CHAPTER 2 – Optimization and best practices for accurate qPCR, from primer design to data analysis.

INTRODUCTION

Quantitative PCR (qPCR) is a powerful technique used to quantify gene expression. It is a technique I used to test the hypothesis that sablefish opsins are expressed at higher levels during stages of development that coincide with greater light exposure. To be accurate, qPCR experiments must be optimized, and various factors must be taken into account. Ahead of what is considered my 'main' qPCR experiment, which is outlined in Chapter 3, was a long process of troubleshooting to provide confidence in the final results.

In this chapter, I present experiments that helped me understand the strengths and weaknesses of the technique I used to quantify opsin gene expression in sablefish. These experiments include those performed in most qPCR studies (e.g., primer efficiency experiments) and experiments not found in many other studies, which I added in order to evaluate the impact of genomic DNA contamination. By sorting these components of my project into a distinct chapter, I have been able to focus entirely on the sablefish-specific data in the next chapter.

Almost all cells contain a species' complete genome. Cells, tissues and periods of development differ with respect to the use of the gene repertoire. Genes are 'used' when they are transcribed from double-stranded DNA to single-stranded mRNA, which is then further translated into a protein. qPCR is used to quantify mRNA after it has been artificially reverse-transcribed into complementary DNA (cDNA). The more mRNA transcripts of a gene there is in a sample (i.e., the more it is expressed), the more cDNA copies there will be. Like traditional PCR, primers and DNA polymerase work to amplify a specific transcript (or gene) of interest. Unlike traditional PCR, qPCR quantifies relative transcript abundance by tracking the exponential amplification that occurs. Genes that are expressed at higher levels reach the linear phase of amplification sooner than genes expressed at lower levels.

A common alternative to qPCR is RNA-seq. RNA-seq is a next generation sequencing approach to quantifying RNA transcripts; it is strong in its ability to discover new gene transcripts and variants, but it has low scalability and low quantification limits. As discussed in Chapter 1, very few opsins were well-represented in the sablefish transcriptome. Considering this

finding (and relative cost) qPCR is a clear choice for further quantifying the expression of sablefish opsin transcripts among different tissue types and life stages.

There are several commonly used qPCR methods, including digital PCR, which is an array-based method for carrying out thousands of PCR reactions at once, Sybr-based detection of cDNA amplicons, and the method I used, TaqMan probe-based qPCR. Each of these methods is based on fluorescence, but the ways in which fluorescence is produced varies. For example, Sybr green dye universally binds all double stranded DNA. When bound, the molecule releases energy in the form of fluorescence (Ponchel et al., 2003). On the other hand, TaqMan probes are specifically designed oligos with a reporter dye and quencher. As DNA polymerase extends the target, it cleaves the probe, separating the reporter dye from the quencher and allowing it to fluoresce. Both methods quantify gene expression using this fluorescence that accumulates as templates are amplified. Specifically, they identify Cq (quantification cycle) values; this is the PCR cycle number at which the fluorescence intensity in the reaction reaches a particular threshold, exceeding the background fluorescent signal. A lower Cq signifies higher expression. This is because the greater the number of transcripts in the sample, the sooner the amplicon concentration increases and crosses this threshold.

The draw to TaqMan qPCR, and the reason I chose it over Sybr qPCR, is that it has higher target specificity because of the additional probe oligo which must hybridize to the target to fluoresce; this is particularly important for opsin work, as there can be high sequence similarity among opsin paralogs. TaqMan qPCR is also more sensitive to low numbers of DNA copies, and multiple targets can be amplified in a single reaction by using different reporter dyes.

As previously mentioned, qPCR is a highly sensitive method that can easily be skewed or biased by a variety of factors; thus, it requires careful set up and optimization in order to give biologically meaningful Cq readings. Optimization includes specific primer design, estimating primer efficiencies, controlling for contamination and validating reference genes, among other criteria.

Primer design

It is important that typical primer design parameters are taken into account for qPCR. These considerations include primer and amplicon lengths, GC content and clamp, and hairpin and dimer formation. Primer design programs, such as PrimerExpress (Thermo Fisher) can aid in

fitting these parameters, but it is not always possible to meet all criteria when sequences are short and when there are other constraints.

At least one primer (or probe) in a gene quantification assay should cross an exon-intron junction, when possible. This step can eliminate, or at least reduce, the impact of genomic DNA contamination in the transcript-derived cDNA sample. In theory, a primer that crosses a junction should work only on spliced cDNA; contaminating genomic DNA will have the primer-complementary sequence fragmented by an intron, ultimately preventing it from binding and being amplified in the PCR reaction.

Given that many opsin genes share sequence similarities, primers and probes must be also checked for specificity to the desired target. A melt curve, which a technique that is only possible with Sybr-based qPCR, can confirm whether one or multiple products have been amplified; a single peak on a graph of temperature vs. fluorescence demonstrates specificity. Sequencing of the PCR product can further confirm that the product is the gene of interest. One can also use a bioinformatic approach to predict specificity to the target, by BLAST searching the genome for off-target sequences that match the primer.

Primer efficiencies

Most PCR methods assume that there is a perfect doubling of the target product/amplicon with each cycle; however, suboptimal primer and probe design or PCR inhibitors in the samples can influence this. A perfect doubling of product corresponds to a PCR efficiency of 100%. As it is difficult to achieve this level of amplification, the acceptable range of primer efficiencies is considered to be between 90-110% (Taylor et al., 2019).

One can determine the efficiency of a given assay using a standard curve; a serial dilution of cDNA of known concentration to ensure that amplification is reflective of the number of target gene molecules. For example, a sample that has been diluted by a factor of 2 should have a Cq value exactly one cycle later than the undiluted sample. The slope of the standard curve is used to calculate % primer amplification efficiency across the given range of cDNA concentrations.

The calculated efficiency is incorporated into the relative quantification calculations, as certain primer sets may be more efficient than others, which could lead to inaccurate comparisons among genes. This can be done either directly in some data analysis software

(where Cq values are directly altered by input efficiencies) or by using a calculation (such as the Pfaffl method; Pfaffl, 2001) that takes the slope of the standard curve into account.

Multiplexing assays

One advantage of the TaqMan probe-based qPCR approach is the ability to multiplex targets; that is, amplify and distinguish between multiple targets within a single well. This significantly reduces the amounts of reagents used and time spent pipetting. In my experiments, I amplified up to two opsins and one housekeeping gene in a single reaction. This is done by designing probes with different fluorophores for different genes. Some optimization is required; the efficiencies of each assay on their own should not change when run together. There can be competition between primers and inhibition when two assays (and their respective reagents) interact, so this must be tested.

The best practice for multiplexing is to have the most abundant gene paired with the least fluorescent/powerful reporter dye, and the lower expressed gene with the stronger dye so that the signal from the high expressor is not overpowering (Taylor et al., 2019). Thus, preliminary qPCR experiments are undertaken to estimate relative expression levels.

Validation of housekeeping/reference genes

Housekeeping genes (commonly called reference or control genes in this context) are crucial to qPCR. They are used to control for experimental variation among biological replicates through normalization. This involves subtracting the Cq of the target gene from the Cq of the housekeeping gene and using this value (known as ΔCq) in subsequent quantification calculations. A suitable normalizer is expressed at the same level across samples. It is important to validate the stability of the housekeeping genes across the biological treatments in an experiment to confirm that they will properly normalize the samples. It is best to use a minimum of two housekeeping genes in a given experiment to increase reliability of the results.

Non-biological variation can be introduced from inconsistencies in mRNA isolation, cDNA conversion and concentration readings, despite efforts to standardize these. If a sample had a higher concentration of cDNA compared to others, for example, the Cq values for both the housekeeping genes and the target gene would be elevated. When normalized, however, the

values would be comparable to other samples. An example of this is shown in the Results section below.

Genomic DNA contamination

qPCR results can be skewed by genomic DNA contamination in the mRNA sample, because the cells that the mRNA is derived from also contain nuclei with copies of the target gene. Efforts are made to reduce the impact of genomic DNA contamination, such as a Trizol based RNA extraction methods (which separates RNA from DNA), DNA washing steps in RNA column kits, DNase treatments to the mRNA sample prior to reverse transcription and primer design over exon junctions (see above); however, each of these efforts are imperfect (Huang et al., 1996). Combining several of these methods, and aiming to prevent contamination in the first place, are the most reliable ways to control for contamination.

Non-detects

In some instances, target transcripts may not amplify within the set number of cycles in the PCR reaction (40); these are known as 'non-detects' (McCall et al., 2014). When all or most of the biological replicates (samples with the same conditions – life stage and tissue type in this case) show a lack of amplification, the simplest explanation is that there is no expression of the target gene. However, biological replicates do not always corroborate in this way. For example, half of the biological replicates may fail to amplify in 40 cycles and others may amplify just before the 40^{th} cycle. In more extreme cases, some may not amplify in 40 cycles, and others may amplify much sooner (e.g., around 25 cycles). I encountered both of these scenarios with my sablefish opsin experiments.

Different approaches are taken for non-detects. Some researchers will assign a Cq of 40 (the upper cycle limit) to samples that did not show amplification, so as to not decrease the sample size. For many, a Cq of 35 and beyond is considered a negligible level of expression. Others assign the non-detects the average value of the biological group (i.e., of the other replicates), again to maintain sample size, though both of these approaches can artificially influence the variance of the data and skew statistical significance (McCall et al., 2014). On the other hand, if one were to exclude the non-detect, then the reduced sample size may lead to a

failure to support gene expression differences. Each of these techniques introduce bias and the choice should rely on context given by the detects.

Other considerations and guides

Beyond the considerations aforementioned, which required special attention for my work, there are many additional best practices to follow for a reliable qPCR experiment. For example, one must include at least three biological replicates per treatment (or more, depending on the statistical analyses to be performed and effect size) and two to three technical replicates per biological sample. One must also ensure that the RNA used in the experiment is sufficiently pure using spectrophotometry and the ratio of absorbance at various wavelengths (OD 260/280 > 1.8).

With so many factors influencing the reliability of qPCR experiments, and given how widely used the technique is, others have written comprehensively on the topic of optimization. Bustin et al. (2009) put forward a set of guidelines for the minimum information required for publication, and others have followed suit by describing the major sources of error in qPCR work (Taylor et al., 2019).

Each of the qPCR considerations described above can have major impacts on the data and its interpretation, but they are sometimes overlooked, or performed without being reported in publications. By thoughtfully troubleshooting, optimizing and validating these parameters, I aim to strengthen my arguments in the next chapter.

METHODS

Primer design and validation

For my sablefish qPCR experiments, primers and probes were designed for ten genes using Primer Express software (Thermo Fisher). Typical primer design considerations were taken into account, including ~50% GC content, a GC cap on the 3’ end, primer length of 18-24bp, amplicon length of ~100-150bp, and a melting temperature around 60°C.

For every gene assay, at least one of the primers (or the probe) was designed over an exon junction (identified in Chapter 1), except those used to amplify single exon genes. Due to the amplicon length limit, GC content/cap and the relatively few numbers of exons, the placement of the exon-spanning oligo was restricted.

It is important that primers and probes do not bind to one another or to themselves because this lowers their target-binding efficiencies and therefore their PCR efficiency: I used the program Beacon Designer (Premier Biosoft) to ensure that no hairpins or self- and cross-dimers would form, paying particular attention to the 3' ends. I checked primer-target specificity with BLASTn searches of the sablefish whole genome shotgun (WGS) database. I set a cut off of 80% identity and query cover and manually scanned the best hits for similarity to the 3' end of the primers. I ensured that the top hit was in the target-containing contig and that no other locus (i.e., genome contig) in the sablefish genome was a good target for my primer-probe combinations.

Due to the short length of the primer and probe sequences and nucleotide redundancy, there were sometimes non-specific BLAST hits outside of the contig of interest. I ran another set of BLASTn searches using both the forward and reverse primers of each target gene concatenated together to see if I could identify hits to both within a single contig. If not, I assumed that the primers would likely be too distant from one another and well over the suggested 150bp amplicon length to work effectively and cause an issue. For a non-specific target to influence the results, both primers and the probe would all have to bind within close proximity to amplify.

To determine whether or not my primers were amplifying the intended target, I used them in RT-PCR reactions (according to the same protocol outlined in Chapter 1) with cDNA from mixed sablefish brains and eyes as templates. I then ran the PCR products in 3% agarose Sybr-stained gel and visualized the bands under UV light to see if the product was the correct size.

Calculating primer efficiencies

To determine the PCR efficiencies of the primer/probe assays, I performed a 1:2 5-point serial dilution on a pooled sample of cDNA (derived from brain tissue across several life stages, three technical replicates) with a starting concentration of 50000pg. I chose non-visual tissue, as that is the focus of the following chapter. I tested each opsin gene assay (primer-probe sets) to be used in my qPCR experiment across all of the different cDNA dilutions in singleplex first. I removed outliers from the triplicates, graphed the standard curve and, from the slope, calculated the efficiencies using the formula: $\%efficiency=((-1/10^{slope}-1)x100)$. I ensured that they fell within the acceptable range of 90-110%; if they did not, I removed either the most concentrated

or most dilute points of the curve until I could determine a range of concentrations over which the primers amplified their target efficiently.

Multiplexing assays

Upon determining that each gene assay had acceptable PCR efficiencies in singleplex, I developed multiplex assays. As my work is the first to quantify opsins in extraretinal tissue in sablefish, it was difficult to know beforehand which genes would be best run together, i.e., which genes should be surveyed together in a way that optimizes amplicon concentration and probe brightness (see Introduction). I used zebrafish expression levels (reported in Davies et al., 2015) as a guide. I ran the same serial dilution experiment with the multiplexed assays and compared both the PCR efficiencies and the general Cq values at given concentrations to the singleplex assays to ensure that they did not give vastly different results.

Housekeeping gene choice and validation

I selected two housekeeping genes for normalization in my qPCR experiments. The first, encoding ribosomal protein subunit 18 (*Rps18*), was shown to be stable during development and across different stages in zebrafish (McCurley & Callard et al., 2008) and suitable reference genes in other species, such as carp and cavefish (Stahl & Gross, 2017; Ye et al., 2010). The second gene, encoding elongation factor 1 alpha (*Eef1a),* had previously been used in sablefish as a reference gene and was shown not to vary between the sexes (Fairgrieve et al., 2016). Though primers had been designed for the gene in the past, the *Eef1a* sequence was not available in NCBI. Thus, to design *Eef1a* and *Rps18* primers and probes suitable for TaqMan qPCR, I used the same approach detailed in Chapter 1 to uncover opsins: a tBLASTn search of the sablefish WGS database using zebrafish *Rps18* and *Eef1a* as query sequences. All subsequent steps, i.e., gene prediction, confirming similarity to orthologs, etc. were also followed according to the same protocol as Chapter 1. The *Rps18* and *Eef1a* contigs have accession numbers AWGY02001014.1 and AWGY02002662.1, respectively.

To ensure that the housekeeping genes I chose were stable across the different biological groups, I quantified their expression using qPCR across all life stages (five stages, N=6 for each) in tissues of interest (brain and eye). These data are also reported in Chapter 3, where sampling is explained more thoroughly. As these data could not be normalized (since they are the

normalizers themselves), raw Cq values were reported. Given that some of the mRNA extraction methods for the different life stages varied (by necessity), I expected the Cq values to vary slightly between them. To confirm that the variation in housekeeping gene expression is not biological (i.e., influenced by development), I compared the difference between the Cq's of the two housekeeping genes amongst life stages. As the two chosen genes play different functions, one would not expect them to vary in the same way if the variation was biological. A one-way ANOVA was conducted to see if the mean difference between the housekeeping genes varied across the life stages to be used in Chapter 3.

Controlling for contamination

The RNA extraction method used (RNeasy Lipid Kit by Qiagen) is Trizol-based, which is fairly effective at separating RNA from DNA and proteins. The protocol also includes a silica membrane column and several washing steps to remove contaminants. Any genomic DNA (gDNA) that 'survives' the mRNA extraction protocol should fail to amplify during qPCR because the primers were designed such that at least one crosses an exon-intron boundary. Nonetheless, in early experiments, I observed bands in my no reverse transcriptase (RT-) controls and the qPCR machine detected fluorescence (indicative of target amplification) after 30 cycles.

To see if this signal was coming from genomic DNA and whether the exon-spanning primer sets could still work on genomic DNA, I extracted DNA from adult sablefish fin clips using DNeasy Blood & Tissue Kits (Qiagen), re-ran my RT-PCR experiments (protocol described in Chapter 1) and visualized the product on a 3% agarose gel.

I also performed a series of experiments in order to identify the step of the protocol at which the contamination was being introduced: during RNA extraction or cDNA conversion. I did this using DNase treated and untreated RNA, and cDNA (both derived from the same pooled eye and brain sample) as templates; DNase treatment protocols are outlined in Chapter 3. RT- controls (for the cDNA samples) were also included. To gain quantitative insight, i.e., how DNase treatment alters the raw Cq values of these samples, I ran these reactions in the quantitative PCR machine with their fluorescent probes. I compared amplification patterns (Cq values) between the different conditions in order to determine whether or not DNase treatment is effective at removing contamination and necessary to implement.

Handling non-detects

The presence of non-detects in my data was unanticipated, thus it was something I had to troubleshoot. Three genes that I chose to explore in Chapter 3 had non-detects across five life stages and biological replicates (N=6 at each life stage): *Opn4m2, Opn4x1 and Opn4x2*. Two of these genes were low expressors and the other was expressed at fairly high levels. Of the three genes, some also had a higher proportion of non-detects than others.

After the non-detects were re-run and continued to yield the same results multiple times, I aimed to determine the best approach to mitigate their effect (and whether the same approach can be applied in all scenarios). I compared the results of setting the Cq of non-detects to 40, setting the Cq to the mean of the biological replicates (i.e., life stage replicates) and excluding the non-detects from the data set. As an example here, I focused on one life stage and tissue type: juvenile brains. I checked to see if and how statistical significance (one-way ANOVA and Tukey Multiple Comparison) differed based on the three approaches.

RESULTS AND DISCUSSION

Satisfying primer parameters and specificity

Primers were successfully designed for the genes of interest (Table 9), though with relatively short sequence length and positional constraints, not every parameter was always met. Yet, with pooled brain and eye cDNA as a template in RT-PCR, all amplicons were the predicted size, confirming proper design and target specificity. According to a series of BLAST searches of the sablefish genome, these primer/probe sets are unlikely to work on other opsins and non-target genes, due to imperfect complementarity of all three oligos. Primer dimers were very minimal, if present at all.

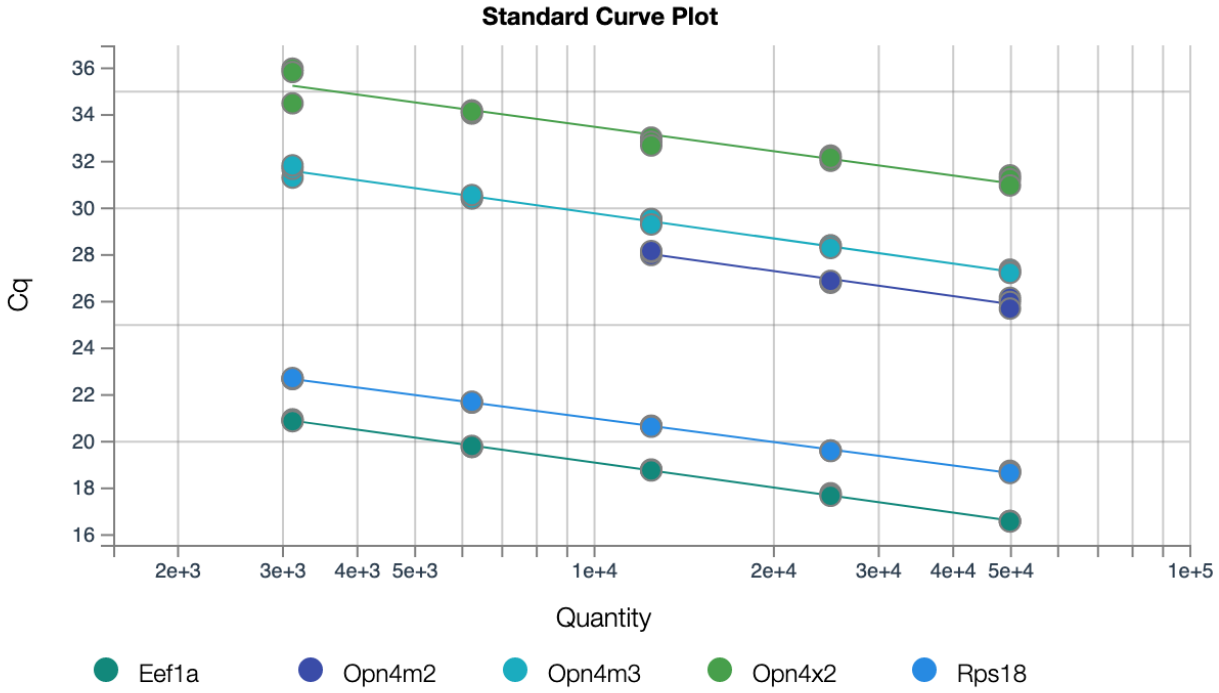
Table 9. Examples of custom primers and TaqMan QSY probes designed to amplify opsins and housekeeping genes in sablefish, designed using Primer Express software (Thermo Fisher). This shows the primers (F, forward; R, reverse) and probes (P) used specifically in Chapter 3, which were also primarily used for this Chapter's experiments. The targets are paired to reflect the different multiplex assays. PCR efficiency was calculated for each multiplex set.

Target	Oligo	Sequence (5'-3')	Amplicon size	% Efficiency
Opn4m1	F	GACCTCAATTGGCGTGATGTC	106	92.7
	R	GGCACTCCAGCCGAAGAAC		
	P	VIC-CTACTCCCTGGGATGGAGCCTGCC-QSY		
Opn4x1	F	AAGGCAGATGTTTGGAGTGATGT	135	95.7
	R	GCTTCCCTTTTCATTGGTCTTCT		
	P	ABY-CCCTTGGAGCTCTGAAGGACAGGGATC-QSY		
Opn4m2	F	CACGCCCACTACATCATTGG	166	90.7
	R	GCGTCAGAGACATCAGGAAGTCT		
	P	FAM-TGCGGTCGCCGAGCAACCTT-QSY		
Rps18	F	GCTGACTGATGATGAGGTTGAG	77	99.1
	R	AACCAGTCTGGGATTTTGTACTG		
	P	ABY-TGGTCACCATCATGCAGAACCCTCG-QSY		
Opn4m3	F	ATGTCTCGCAGGAAAGCCC	94	90.5
	R	CATAAGCACTCCAGCCAAAGAAG		
	P	ABY-TCCGCAGGATGGAGCCTTCCAC-QSY		
Eef1a	F	GGAAATCCGTCGTGGATACG	82	91.2
	R	GATGACCTGGGCGTTGAAAT		
	P	VIC-TGGAGACAGCAAGAACGACCCACCC-QSY		
Opn4x2	F	CCCTGACATGGTGTTAAAGGATGT	300	100.2
	R	TCCGAAATCTCACTGGTCAGACT		
	P	ABY-CGCCTCTCTCCCCTTACTGGTTCTTACGA-QSY		
Sws1	F	CAAAGACTATCGACTCGTCACCAT	100	104.6
	R	GGCATTAAACTGTTTGTTCATGAA		
	P	FAM-CTCCAAAAGCTCGTGTGTCTACAACCCG-QSY		
Rgr1	F	CTCGCCAGCATCCACTTCAT	70	108.2
	R	GCTTTGTCCTGGTGCAGTACTG		
	P	VIC-CCGCCATCGAATGGGACAGA-QSY		
Rh1	F	GCAAGCCCATCAGCAACTTC	82	94.0
	R	GCCAGGTGAGTCCCAAACC		
	P	FAM-CCCTCACCCTCTACGTCACCCTCGA-QSY		

Primer efficiencies in single- and multiplex

All tested primer probe assays were validated as efficient (between 90-110% efficiency, Table 9) over a certain range of cDNA concentrations. For many, this was the full range of the serial dilution (3,125pg-50,000pg). For others, such as the *Opn4m2* gene, the range had to be narrowed by removing the most dilute points (Figure 8), meaning that the primers did not work efficiently at those low concentrations. This seems to be unrelated to gene expression levels, as

assays targeting lower expressed genes (such as *Opn4x2*, Figure 8) are efficient across the full range of concentrations.



Target: Rps18 Slope: -3.345 R^2: 0.999 Y-Inter: 34.334 Eff%: 99.057 Error: 0.026

Target: Opn4m3 Slope: -3.574 R^2: 0.994 Y-Inter: 44.042 Eff%: 90.471 Error: 0.079

Target: Opn4m2 Slope: -3.566 R^2: 0.974 Y-Inter: 42.608 Eff%: 90.725 Error: 0.222

Target: Opn4x2 Slope: -3.466 R^2: 0.942 Y-Inter: 47.318 Eff%: 94.309 Error: 0.238

Target: Eef1a Slope: -3.554 R^2: 0.999 Y-Inter: 33.281 Eff%: 91.150 Error: 0.028

Figure 8. Example of a standard curve to test primer efficiencies across various cDNA concentrations (a 5 point 1:2 serial dilution beginning at 50,000pg). The slope of the line and subsequent calculated efficiencies (Eff%) are shown below the graph.

After determining that each assay was functional, amplifying the correct target and efficient in singleplex, I moved on to test the assays in multiplex. I combined assays with compatible reporter dyes and compared the multiplexed Cq values to those of the singleplexed assays (of the same sample at the same concentration) to ensure that they did not change significantly. Those that did not change were classified as a valid and accurate pair and were used together in all future experiments. Figure 3 shows an example of two assays, *Opn4m2* and *Rps18*, each run singleplex and together in multiplex; it is evident that the amplification curves overlap and Cq values are within 0.2 cycles. Multiplexing proved to be an easy optimization;

most pairs were successful in maintaining the same Cq readings, even in cases in which the higher expressed gene had the brighter reporter dye.

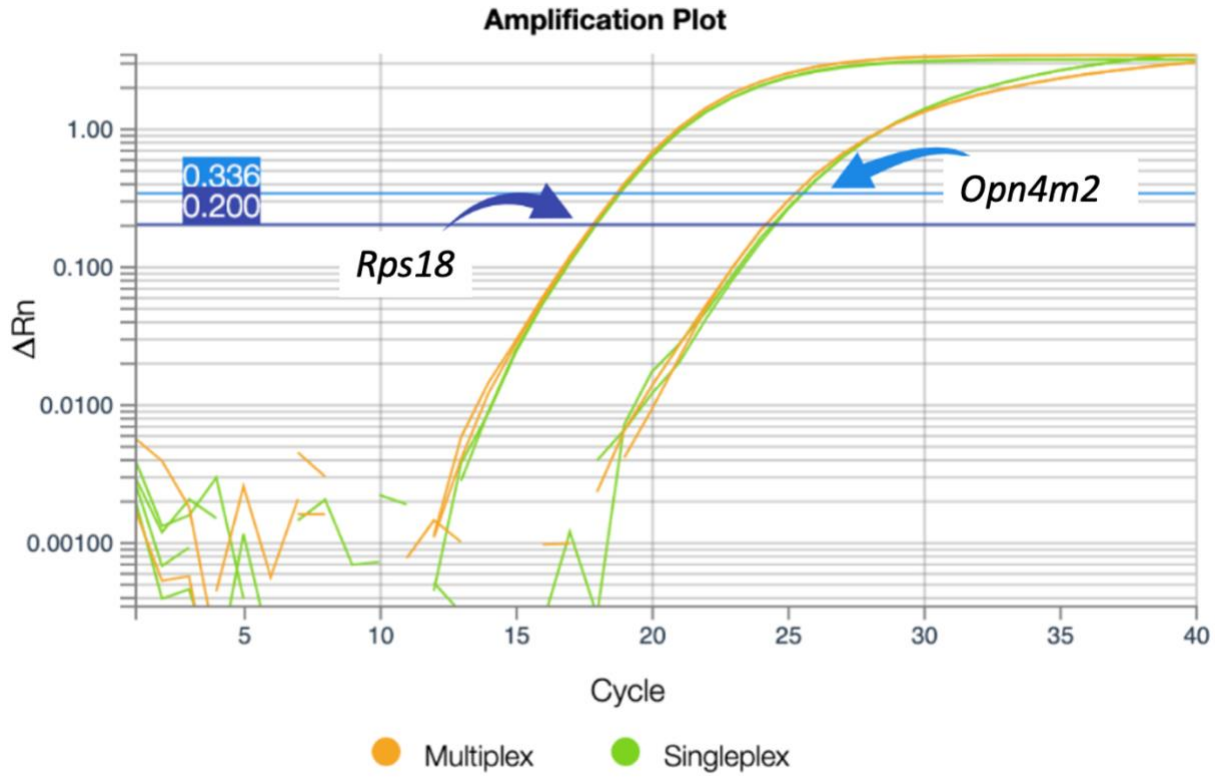


Figure 9. Amplification plot of two example genes (one opsin gene, *Opn4m2,* and one housekeeping gene, *Rps18*) showing near-identical Cq values when run separately in singleplex (green) and together in multiplex (orange). Units of the Y axis, ΔRn, represent the magnitude of the signal from the reporter dye, normalized to the passive reference signal and subtracted by the background signal.

Housekeeping gene stability

When testing the chosen housekeeping genes across different life stages, I was first surprised to see that the Cq's were not stable. For example, expression of *Rps18* differed as much as 4 cycles (a 16-fold difference) across some of the life stages. This seems more troubling than it is; all targets in qPCR are normalized to housekeeping genes to account for variation among biological replicates and different treatments. If a given treatment, in this case a life stage, has a less successful RNA extraction or cDNA conversion, then you would expect both the housekeeping and target genes to be expressed at lower levels. Since the housekeeping gene normalizes the target gene, any discrepancy between treatments is controlled for, as the relative difference between the housekeeping gene and target remains the same.

For this reason, it can be difficult to validate a single housekeeping gene; it may be difficult to know whether the change in Cq is because of true gene-specific biological differences or simply because of experimental differences. To remediate this, a second housekeeping gene (with a different function) is added. If experimental differences (e.g., variation in quality or quantity of RNA or cDNA) are causing the variation in Cq's across different treatments, then the second housekeeping gene would show the same trend. If it was gene-specific variation, the trend would differ. The housekeeping genes I've chosen demonstrate the former (Figure 10), with both sharing the same pattern (decreasing or increasing together). Thus, they are reliable to use for normalization in qPCR experiments. The difference in Cq between the two housekeeping genes (ΔCq) remains consistent (one-way ANOVA, p-val=0.1951) across replicates and life stages. In essence, Figure 11 demonstrates that they will do their job properly, not only for methodological variation among replicates of a particular life stage, but also between the different life treatments.

In my experiments, different life stages unfortunately required different extraction methods. For example, juvenile brains could be homogenized whole in the lysis buffer, whereas adult brains exceeded the tissue weight limit for the kit, which led me to grind them up prior to adding them to the lysis reagent. The younger larval stages also required additional treatment prior to dissection. These details will be discussed further in Chapter 3, but it is ultimately not surprising that there are methodological/experimental differences in reference gene Cq among the life stages. While it is best practice to have all of the preparations consistent, it does seem as

though the introduced differences are being controlled for and normalized by the reference genes.

Worth noting is that, since the experiments presented in this thesis, another member of the Taylor Lab has added a third housekeeping gene (encoding TATA box binding protein), which is expressed in the same patterns as the two housekeeping genes presented in this chapter and the next. Additionally, the range of life stages has also since been extended to include the earliest egg and hatching stages, and *Rps18* and *Eef1a* continued to be stably expressed and work well as normalizers.

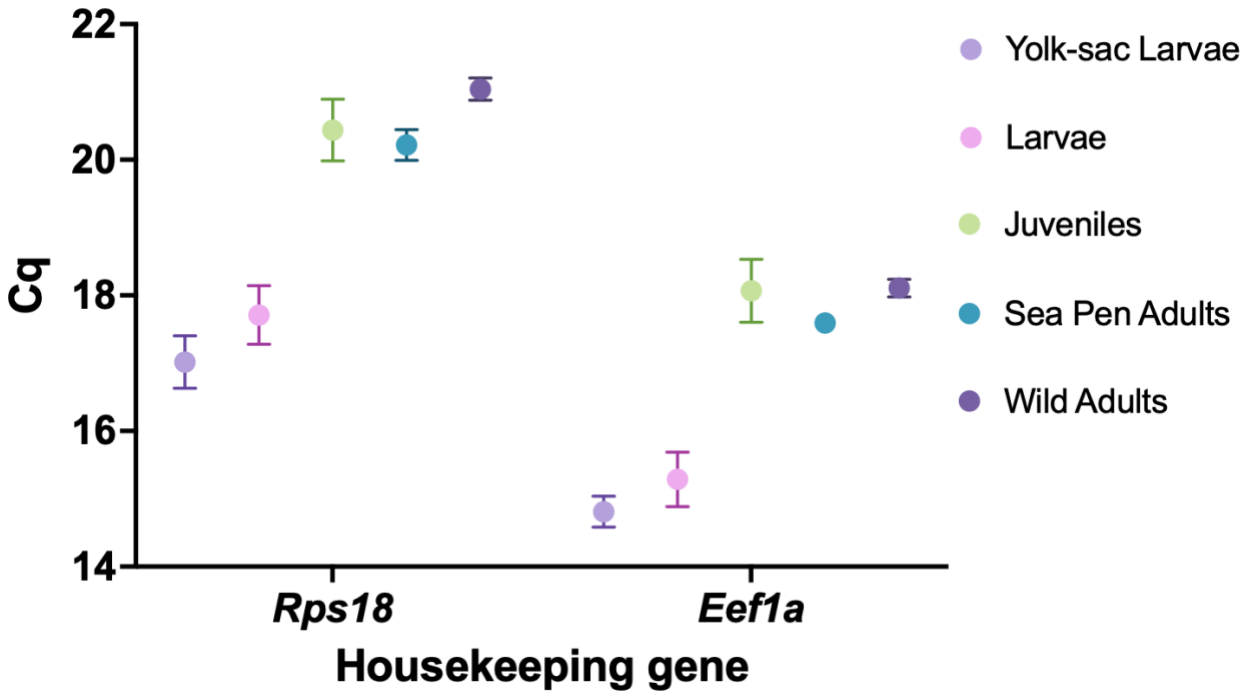


Figure 10. Mean Cq values (+/- SD) of the two housekeeping genes, *Rps18* and *Eef1a*, across five life stages to check for stability and trends mirrored across both genes (N=6).

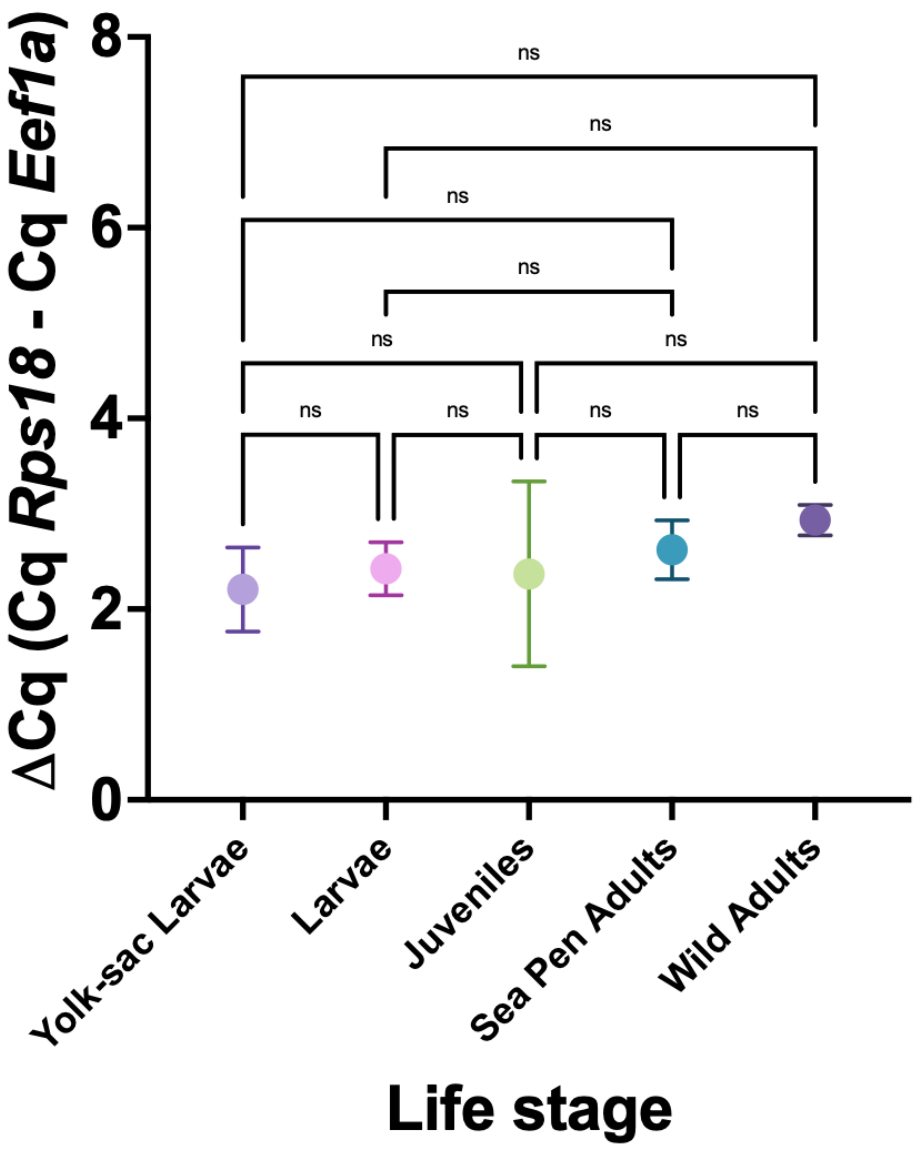


Figure 11. The difference in mean Cq (ΔCq+/-SD) between the housekeeping genes at each of the life stages studied in Chapter 3 (N=6). Significance reported from Tukey Multiple Comparison comparing the mean ΔCq's of the various life stages (P-values all >0.1951).

Removing genomic contamination

Contamination in qPCR, the most common form being genomic DNA contamination, can pose great risks to data interpretation and the accuracy of results. Because of this, I designed at least one primer for each target gene over an exon junction (with the exception of single exon genes), in hopes that cDNA would be the only template that can be successfully amplified, negating the worry of genomic contamination. Despite my best efforts, this was not always

successful. The first sign of genomic contamination was amplification in the no reverse transcriptase (RT-) controls; in theory, this control should only contain RNA, which is not compatible with the designed primers. The amplification occurring was sometimes the same size as the predicted cDNA amplicon, and other times larger. In subsequent tests using genomic DNA as a template in PCR, it became evident that many of the primers could amplify from this source (Figure 12).

There was little flexibility in terms of positioning the primers that spanned exon junctions (because of GC content and other primer design considerations). As such, junction-spanning primers were split to varying extents across the boundaries. In some cases, the majority of the primer was complementary to the 3' exon (Figure 12A). In these scenarios, the likelihood that the primer would also work on genomic DNA was quite high. The 3' end of a primer is most crucial for DNA polymerase extension, and it often appeared as though the 3' part of the primer could effectively bind and amplify without complementary binding of the 5' end of the primer. Evidence of this was my observation that gDNA could produce an amplicon that was the same size as the one expected from the amplification of cDNA. Because there are only four nucleotides in the genetic code, it is also possible that by chance, some of the 5' end of the primer sequence is complementary to intronic sequence, allowing it to easily bind gDNA too.

Primers that were split more or less half and half between the exons, with a slightly longer portion on the 5' exon (Figure 12B) performed best, failing to amplify from genomic DNA. This suggests that if there was contaminating gDNA in the extracts for qPCR, the gDNA would not amplify and therefore would not contribute to the signal.

Primers weighted *too* heavily in the 5' exon amplified both genomic and cDNA, but with different sized bands (Figure 12C). That is, the primer worked well to bind to cDNA, but was also able to bind gDNA, and the intronic region of the contaminating DNA was amplified. This leads to a smaller amplicon with cDNA as a template and a larger one with gDNA as a template. While one can tell whether gDNA is contributing to the signal in this circumstance by running the PCR product on a gel, it is impossible to distinguish the contamination signal from the true target signal in qPCR.

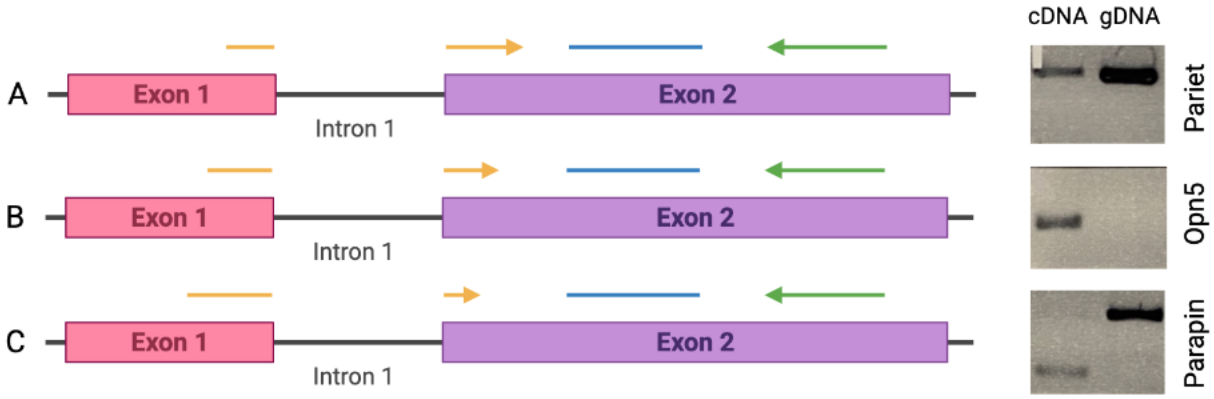


Figure 12. Visualization of primer design over exon junctions to prevent primers from binding to contaminating genomic DNA. The yellow arrow symbolizes the region to which the forward primer binds on the opposite strand, the blue line represents the probe and the green arrow represents the reverse primer. Different primer placements over the exon junction and their resulting ability to amplify genomic DNA, shown by amplicon bands on a gel: **A)** a primer with most of its sequence complementary to the 3' exon, exon 2 (*Parietopsin*); **B)** a primer weighted slightly more in the 5' exon, exon 1 (*Opn5);* **C)** a primer mostly complementary to the 5' exon (*Parapinopsin)*.

These results demonstrate that, despite designing primers to span exon boundaries, genomic contamination can sometimes successfully amplify. This is a cause for concern in qPCR, as amplification from genomic DNA will contribute to the overall signal. This reiterates the importance of contamination control protocols, including DNase treatment.

The experiment conducted to determine the extent to which genomic DNA is a concern yielded interesting results. It became evident that the RNA itself was contaminated. Using untreated RNA as a template in qPCR, amplification occurred quite late, around 30 cycles. This amplification should only be due to genomic contamination as DNA polymerase does not recognize RNA. The same RNA that had been treated with DNase (which cleaves double-stranded DNA) showed a significant improvement in that the Cq was pushed back by five or more cycles for any given target (Figure 13).

The same treated and untreated RNA was reverse transcribed to cDNA, and as expected, the cDNA sample derived from untreated RNA was skewed to have lower Cq values because of contamination compared to the cDNA derived from treated RNA. Consequently, the no reverse transcriptase controls (RT-) showed the same trend (Figure 13). This is a strong argument for implementing a DNase step in the RNA protocol.

Even after treatment, some contamination can remain in the RNA and negative controls; however, these are very low levels, and beyond 35 cycles is typically considered the no-expression zone. The remaining contamination could be from excess genomic DNA or, albeit unlikely, from first strand cDNA (a DNA:RNA hybrid which is not targeted by DNase). On the other hand, sometimes treatment with DNase can effectively get rid of all contamination, as shown by failure to amplify from the no reverse transcriptase controls (Figure 14).

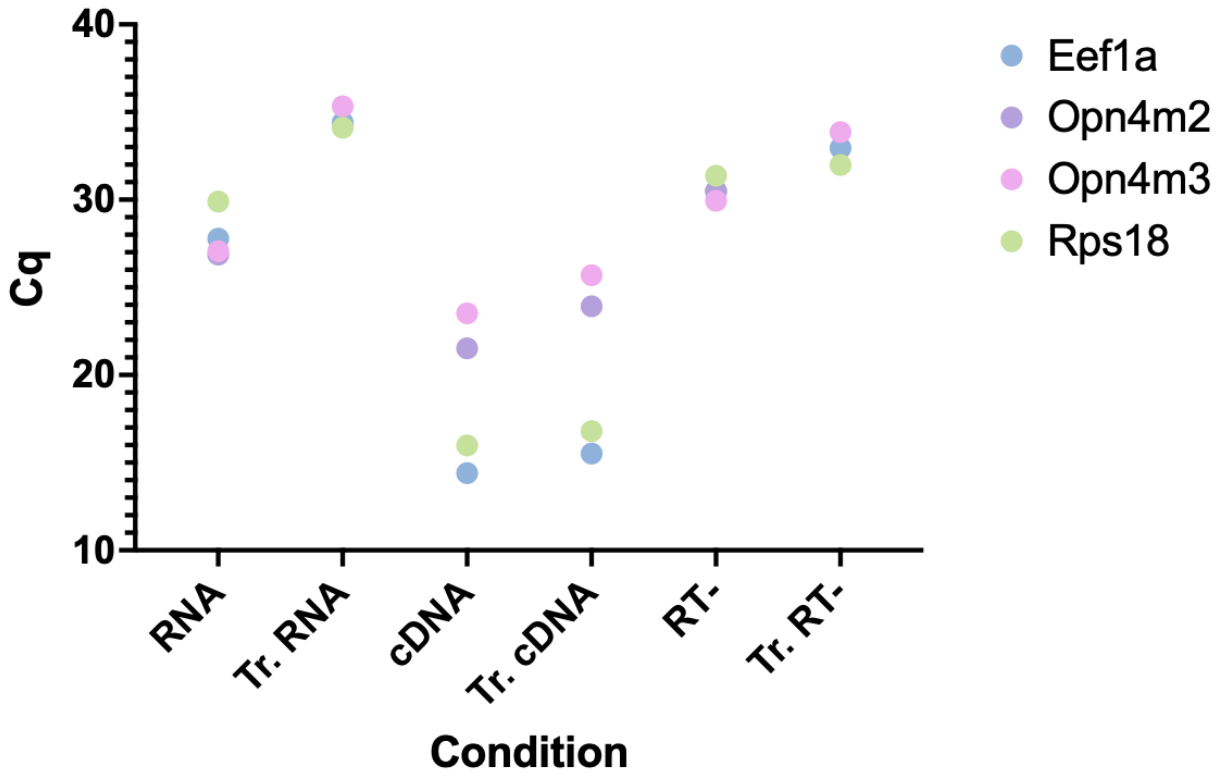


Figure 13. Plot of raw Cq values of a subset of opsin genes and housekeeping genes under various conditions. Conditions are treated (Tr.) and untreated RNA, cDNA and reverse transcriptase negative controls (RT-) derived from the same sample.

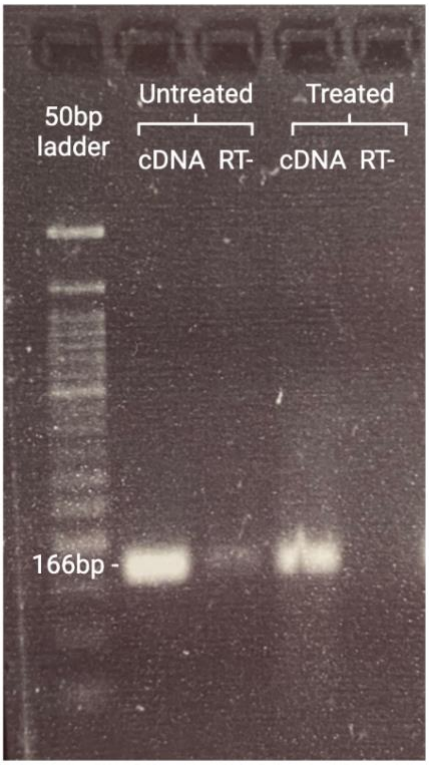


Figure 14. *Opn4m2* RT-PCR products generated from both DNase-treated and untreated cDNA templates, run on a 3% agarose gel with a 50bp ladder showing the effectiveness of this technique at removing genomic DNA contamination. This is evidenced by amplification or lack thereof from the reverse transcriptase negative (RT-) controls; these results are mirrored in Figure 13. Smearing in the treated cDNA lane may represent digested DNA or overly concentrated cDNA in the well; a melt curve, while not performed here, can reveal whether a single, specific product was produced from this reaction.

A general suggestion in qPCR studies is that the no reverse transcriptase controls should have a Cq at least seven cycles away from the target of interest (personal correspondence with Thermo Fisher representatives). This value is somewhat arbitrary but stems from the fact that a difference of seven cycles in qPCR represents a 128-fold difference in expression. Thus, any contamination present would be contributing only ~1% to the overall signal.

Housekeeping genes, and in fact many genes in clinical research, are often expressed at high levels in the chosen tissue. As such, many researchers likely avoid worry of contamination having a great effect on the target's signal, even without taking extra precautions or applying DNase treatments. GPCRs are generally expressed at low levels in all organisms, despite the superfamily being very abundant in the genome. Certainly, some exceptions exist, such as odorant receptors in the nasal cavity and visual opsins in the retina; however, GPCRs outside of

the typical sensory organs account for a very small proportion of genes expressed in a given transcriptome (Fredriksson & Schioth, 2005). With low expression, they are often much closer in Cq to the signal from contamination in the negative controls. Extra treatments can mitigate this to an extent, but opsin gene quantification outside of the eye appears to be far more heavily impacted by contaminating sources than the housekeeping genes are.

This experiment certifies the need to treat RNA samples with DNase, especially when dealing with low expressed genes. Furthermore, it hints at a trouble area for non-visual opsin gene expression work. Those that are expressed at very low levels may not always exceed a 7-cycle difference from the negative control. Even if the targets are consistently peaking 3 or 4 cycles before the negative control, a large proportion of that signal is simply due to the remaining contaminants and it is therefore not a reliable reading. One must then consider this as no expression or background expression.

Non-detects

The data from Chapter 3 showed three genes with non-detects across conditions and biological replicates (25 in total): *Opn4m2* had 5, *Opn4x1* had 3, and *Opn4x2* had 19 non-detects. *Opn4m2* detects were on average amplifying at cycle 25 across all samples (range: 24-27), *Opn4x1* detects had an average Cq of 30 (range: 28-36) and *Opn4x2* had an average of 33 (range: 29-36). Worth noting is that for each sample with a non-detected gene, their housekeeping genes (and often all other opsin genes tested) were detected at comparable levels to other replicates.

The various approaches to non-detects yielded very different results (Figure 7, Table 1). Assigning the non-detects with the average of the replicates was the least conservative approach. In doing so, the variance of the replicates was decreased, and differences between groups became more significant (artificially) (Table 10).

Assigning the Cq of non-detects as 40 influences the three opsin genes in different ways (Table 10). This approach seems most suitable for genes with low expression (close to 40) such as *Opn4x2*, as the non-detects in this scenario are less likely to be random; that is, as the average Cq of the replicates increases toward 40, the more likely it is that some replicates will not amplify under 40 cycles. In the case of a gene like *Opn4m2*, with detects amplifying around 25 cycles and very few non-detects, it makes much less sense to assign a Cq of 40 to the non-detects, as these are more likely to be missing at random.

Excluding non-detects entirely seems to be a common ground between the other two alternatives. That is, it is not overly conservative, nor is it too radical. The issue with this approach is a reduction in sample size/replicates. Not only does this introduce the possibility of masking real differences in gene expression, but it also affects the normality of the data. Additionally, for genes with many non-detects, such as *Opn4x2*, relying on the very few detects to draw conclusions is not wise. Such genes should be clearly represented as having little-to-no expression.

Given that non-detects are not very common in my dataset (apart from the *Opn4x2*), I chose to exclude non-detects in Chapter 3, as it appears to be the most reasonable approach. New programs (i.e., DirEst; Sherina et al., 2020) have been developed to impute values for non-detects without reducing the variance of the replicates. That said, this software is largely incompatible with my experimental set up at this time.

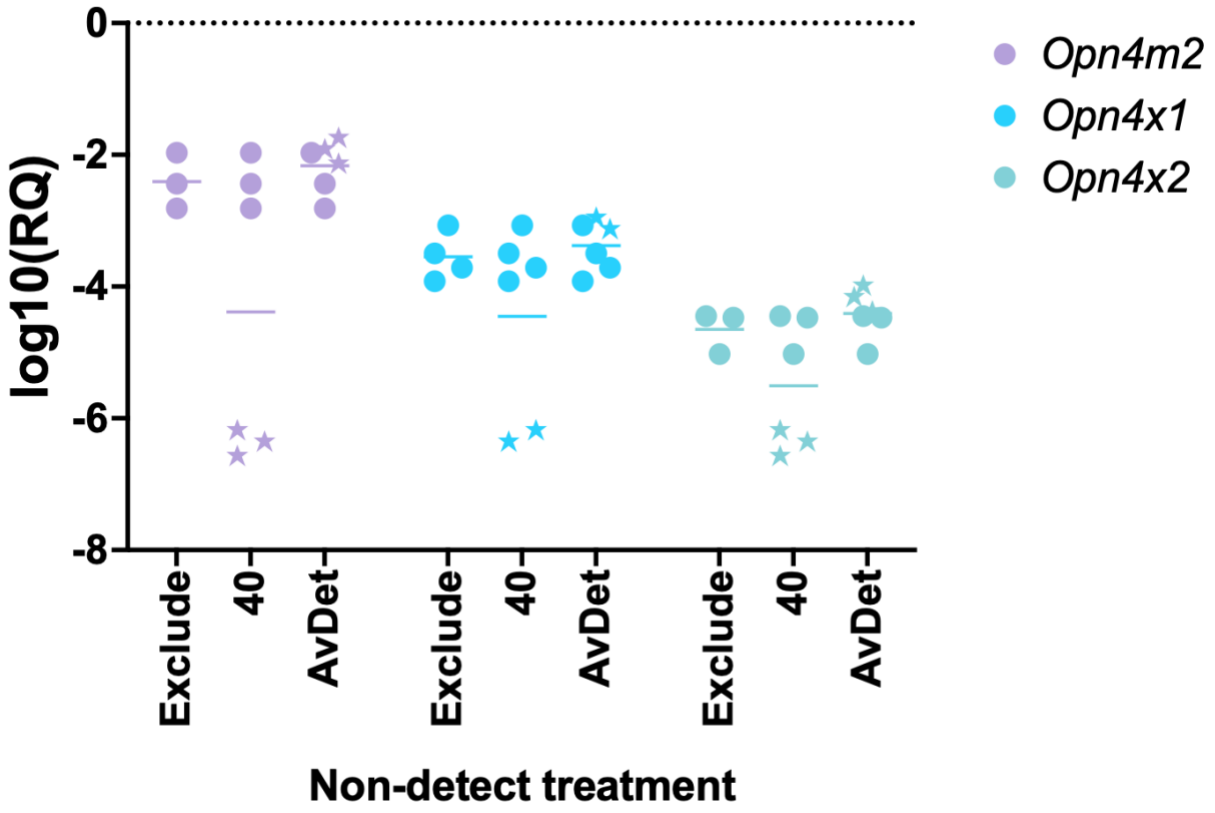


Figure 15. An example (using only juvenile brain data) of the outcomes using the three approaches to non-detects: excluding the non-detects entirely, setting their Cq to 40 and setting their Cq to the average of the detects among replicates (AvDet). This example shows the relative expression levels of three different genes (log10 transformed); the more negative the value, the lower the expression. Each point represents a biological replicate, stars represent non-detects, and horizontal lines depict the mean of the replicates.

Table 10. Overview of non-detects from the entire Chapter 3 qPCR dataset (5 life stages x 6 replicates each). A series of one-way ANOVAs for each opsin (with the null hypothesis that the mean expression level is the same across all life stages) was run to show how the outcome/significance varies using the different approaches to non-detects. Statistical significance was set to 0.05. P-value: >0.05 (ns), <0.05(*), <0.01(**), <0.001 (***), <0.0001.

			ANOVA (difference between life stages)		
Gene	**Average Cq of detects**	**# non-detects across all samples**	**Exclude non-detects**	**Non-detects = Cq 40**	**Non-detects = Cq average of group**
Opn4m2	25	5	**	ns	****
Opn4x1	30	3	****	*	****
Opn4x2	33	19	*	**	****

The cause of non-detects in the data is not precisely known. As previously mentioned, for genes such as *Opn4x2*, with detects that have very low expression, normal gene expression variation can cause replicates to go undetected, in a way that is non-random (i.e., true low/no expression).

This explanation is far less likely for genes that are detected with higher expression levels in other replicates. Many possible causes, such as multiplex and treatment issues, are ruled out because the same multiplex assays work perfectly for other replicates/samples, and samples with non-detects for some genes show normal amplification of other genes (including housekeeping genes). Other possible causes of non-detects include primer mismatches (individual-level variation at the primer binding site) and the possibility that the RNA sampled may not be representative of the whole tissue (i.e. regions that express a gene at high levels may have been inadvertently excluded), leading to the seemingly random pattern of non-detects. While most qPCR studies use a small sample size, as small as N=4, it may be worthwhile to increase the sample size in future experiments to account for non-detection.

CONCLUSION

qPCR is a powerful tool for quantifying and comparing gene expression levels; however, this chapter makes it clear that thorough optimization is required to get accurate and reliable results. Many qPCR studies fail to report these measures (most importantly, housekeeping gene validation), lowering the credibility of the data reported. The troubleshooting outlined in this chapter also indicate that low expressed genes, such as some of the non-visual opsins in extraretinal tissue, may require additional steps to remove contamination, as lower expressed genes are disproportionately affected by outside contamination. Each of the measures mentioned in this chapter were employed in the Chapter 3's qPCR experiments; as such, I can have great confidence in the results.

CHAPTER 3 – Investigating the temporal and spatial correlation of melanopsin expression and light exposure

INTRODUCTION

The enigmatic non-visual opsins

Large opsin repertoires in fish are the products of tandem gene duplication, retroduplication and whole genome duplication (Rennison et al., 2012; Beaudry et al., 2017). These repertoires have been sorted in a diversity of ways; the 'visual' (expressed in the vertebrate retinal photoreceptors and contributing to image formation) and 'non-visual' (not involved in image formation) subdivision is especially common.

Visual opsin repertoires tend to vary among species, and this variation appears to be driven by adaptation to different photic habitats. Expanding a repertoire can improve color vision (wavelength discrimination) in broad spectrum zones (Musilova et at., 2021) or improve light detection in light-restricted zones (Musilova et al., 2019). Many visual opsin genes, for example *Rh1* and *Rh2* in Pacific salmon, have experienced strong positive selective pressures early in their evolutionary histories, as the encoded proteins conferred advantageous new visual phenotypes important for species survival and evolution (Dann et al., 2004). Given how dynamic the evolution of visual opsins has been with respect to the light environment and visual advantages, it is surprising that non-visual opsin repertoires are i) so large and ii) so well-conserved across distantly related fish species in varying aquatic environments. Lineage-specific duplication and losses are far less common among teleost non-visual opsins, and repertoires have generally stabilized following the teleost genome duplication (Beaudry et al., 2017).

The differences between visual and non-visual opsin repertoires are likely associated with their distinct roles in image formation or lack thereof, respectively; however, these differences may be caused by something deeper, such as roles that do not involve phototransduction. Evidence that opsins are 'moonlighting' proteins (i.e., capable of performing multiple functions, for example through mutations or via secondary functional sites; Jeffrey, 2014; Feuda et al., 2022) has been accumulating in the past decade, mostly in research on fly models. Opsins can act as chemosensors (bitter taste receptors in flies; Leung et al., 2020), thermosensors (used in sperm thermotaxis; Perez-Cerezales et al., 2015) and mechanosensors (neuromasts of the literal line;

Backfisch et al., 2013). These light-independent functions vary in their use of a chromophore and the opsin biding pocket.

Though the fact that sablefish spend a short part of their lives in the light may, on its own, explain the conservation of 29 non-visual opsins, a non-light sensing role may allow these opsins to be transcribed and used in the aphotic zone (where they spend the vast majority of their lives) as well. That said, almost all opsins have key conserved amino acid residues, such as the chromophore-binding lysine, counterions and G-protein binding motifs (see Chapter 1), suggesting that thermo-sensitivity, for example, is a function that might have been 'appended' to the core photo-sensing opsins. Indeed, many non-visual opsins have had their wavelength sensitivities characterized and have been shown to mediate light sensitivity through a variety of techniques, including in cell culture (Sugihara et al, 2017; Matos-Cruz et al., 2011; Sato et al., 2016; Koyanagi et al., 2002; Fischer et al., 2013) and mRNA-injected *Xenopus* oocytes (Panda et al., 2005). Tissues that express non-visual opsins (such as the brain and pineal) have also shown to be light sensitive in isolation (Kuznetsova et al., 2021). Whole organisms with their eyes (and in some cases pineal) removed have demonstrated responses to light (von Frisch, 1911; Foster, et al., 1991; Fernandes et al., 2012). This might be due to non-visual opsins, though in very few instances have these opsins been proven to mediate this photosensitivity. For the vast majority of non-visual opsins, no clear connection between expression and light-sensitivity has been made.

Sablefish provide an opportunity to study opsin expression in a fish that is both exposed to and not exposed to light at different stages of its life. This allows me to show whether or not non-visual opsins are utilized only when they have the potential to detect light. Using sablefish, I can also determine precisely where non-visual opsins are expressed. If these genes encode light sensors, I would expect them to occur in tissue positionally optimized to capture light. In addition, by identifying opsin expressing cells, my work sets up experiments that go well beyond the general and ambiguous observations that only hint at a connection between non-visual opsins and light sensitivity, such as those in mentioned above. Sablefish are amenable to calcium imaging and even to CRISPR knock-out experiments that could get to the bottom of non-visual opsin function.

A focus on melanopsins

To quantify and localize opsin expression in sablefish, I first needed to narrow the scope of my research. With so many opsins in the sablefish genome and the possibility of expression across many organs (as in zebrafish; Davies et al., 2015), I chose to focus primarily on one tissue and one subfamily of opsins. With a particular interest in the highly-conserved non-visual opsins, I aimed to study the brain, given that it is second only to the eye in terms of opsin expression in fish (Simon et al., 2019; Davies et al., 2015). Knowing this, I chose OPN4's (also referred to as melanopsins) as my subfamily of interest; this is the most well-studied group of non-visual opsins in animals, and they are expressed in a variety of regions in fish, mammal, bird and reptile brains (Perez et al., 2019).

The first melanopsin gene was identified in African clawed frog (*Xenopus laevis*) by Provencio et al. (1998). The researchers were in search of the basis of dermal melanophore photosensitivity, hence the name it was given. Upon its discovery, it was recognized as a non-visual opsin, despite also being expressed in the retina. Orthologs were quickly characterized in other vertebrates, which ultimately led to the discovery of different classes of melanopsin or Opn4 genes (Bellingham et al., 2006). The two classes are named after the species in which they were first discovered: mammalian-like (*Opn4m*) and *Xenopus*-like (*Opn4x*). These two ancestral genes arose from a genome duplication event at approximately the time of the origin of vertebrates (Borges et al., 2012). Thus, despite this naming convention, non-mammalian vertebrates have *Opn4m*, and *Opn4x* is not restricted to frogs.

In addition to the two broad classes of Opn4 genes, teleosts have a diversification of the ancestral *Opn4m and Opn4x* genes through retroduplication and whole genome duplication. A retroduplication in the ancestral Opn4m led to the intronless *Opn4m2* gene (Matos-Cruz et al., 2011). This was followed by a teleost genome duplication event, creating mammalian-like *Opn4m1* and *Opn4m3* and *Xenopus*-like *Opn4x1* and *Opn4x2* genes (Davies et al., 2011; Beaudry et al., 2017). The *Opn4m2* retroduplicate must have also been duplicated in during this event, but only one copy has been found in all fish surveyed (Figure 16, Figure 17).

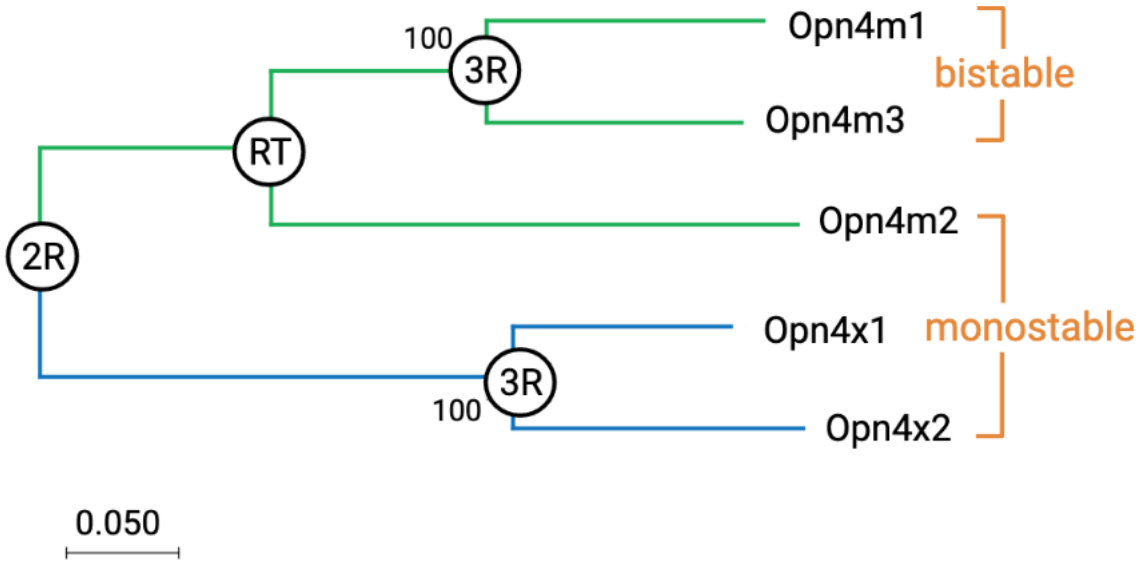


Figure 16. Sablefish OPN4 phylogeny. Genetic distance was computed using the Jones-Taylor-Thornton amino acid substitution model and the phylogeny was reconstructed from these distance data using the Neighbor-Joining method. Evolutionary analyses were conducted in MEGA. The percentage of 500 replicate trees in which the associated taxa clustered together in the bootstrap test are shown next to the nodes. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. Duplication nodes are annotated: 2R: two rounds of whole genome duplication in common ancestor of vertebrates, RT: retroduplication, 3R: teleost genome duplication. Chromophore valency (monostablility or bistability) is also noted.

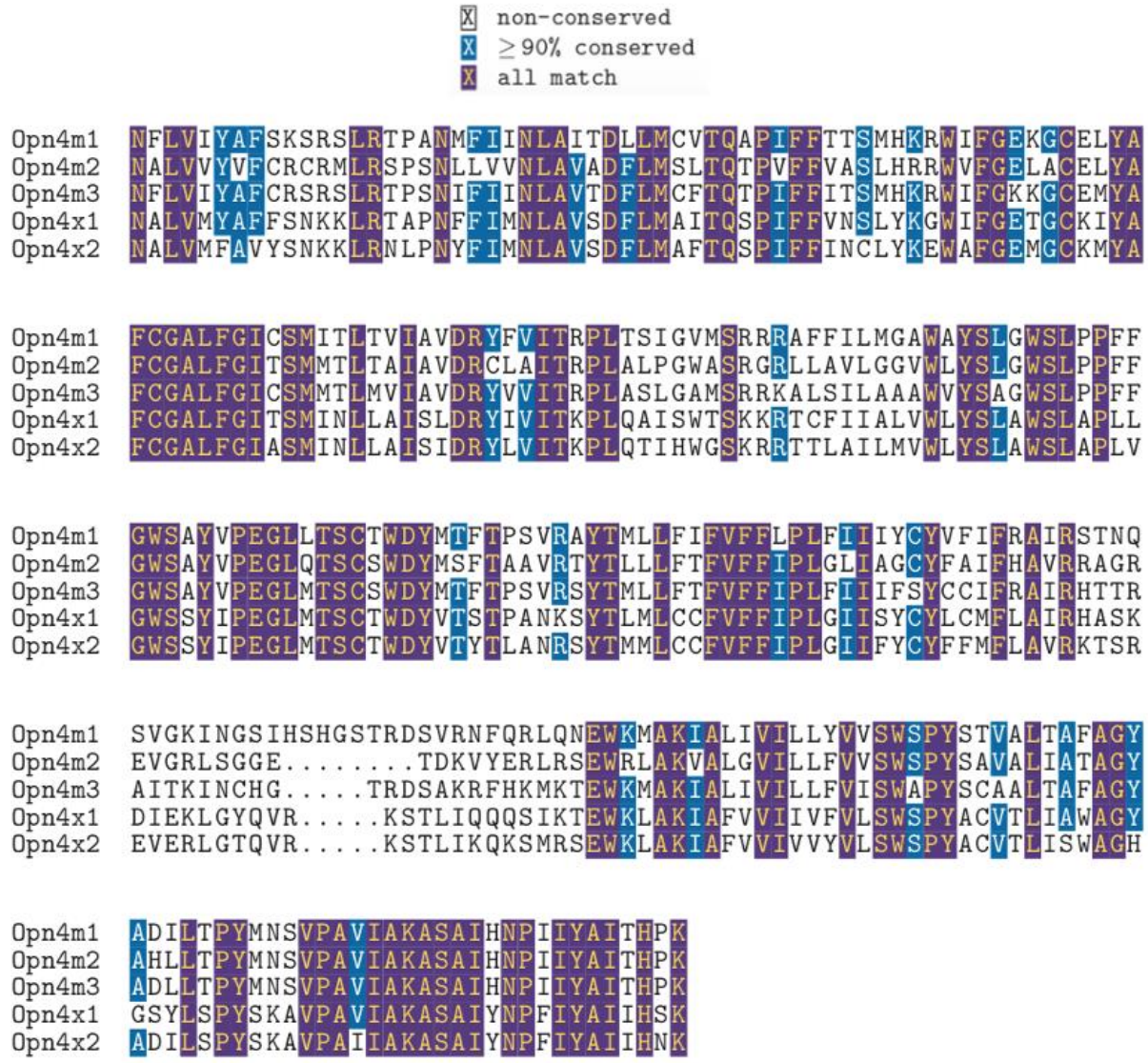


Figure 17. Aligned sablefish OPN4 peptide sequences, truncated to remove variable 5' and 3' regions. Purple shading shows residues that are 100% conserved amongst the five paralogs and blue shows >90% conservation.

Melanopsin sequences more closely resemble invertebrate rhabdomeric photoreceptors (the primary visual opsins in insects) than vertebrate ciliary opsins (Ignacio et al., 1998). The same is true about their signalling pathway (Koyanagi and Terakita, 2008). Melanopsins (maximally sensitive to light in the 403-483nm range; Perez et al., 2019) depolarize the cell by activating G_q proteins, causing a transduction cascade that induces subsequent activation of phospholipase-C and transient receptor protein (TRP) channels (Hardie, 2011). These observations, coupled with phylogenetic analysis, confirmed that invertebrate rhodopsins (the

opsin protostomes, e.g., insects and cephalopods, use for vision) are orthologs of vertebrate melanopsins.

Melanopsins are expressed in intrinsically-photosensitive retinal ganglion cells (ipRGCs), causing the depolarization of cells in response to light, independently of rods and cones (Hattar et al., 2002; Berson et al., 2002). Through melanopsin GPCRs, the ipRGCs feed light information to regions of the brain (the superchiasmatic nuclei in mammals and the optic tectum in non-mammalian vertebrates) responsible for maintaining a diversity of biological rhythms (Gooley, 2001), such as neuroendocrine regulation (Bertolesi et al., 2016). In addition to sending signals from the eye to the brain, Opn4 transcripts and proteins have been found in several different brain regions across vertebrates (Perez et al., 2019), with suggested roles in dark photo-kinesis (Fernandes, 2012) and seasonal breeding (Kang et al., 2010), among others. There is recent evidence that melanopsins in the zebrafish pineal play roles in melatonin production, the regulation of clock gene rhythmicity and locomotor activity (Dekens et al., 2022). Melanopsins also play a role in the pupillary light response (Chen et al., 2011) and in the representation of visual images through brightness and contrast discrimination (Allen et al., 2017). Recently, targeted melanopsin stimulation has even been shown to elicit a conscious perceptual experience, marked by unpleasant brightness, in humans (Spitschan et al., 2017).

Quantifying melanopsin expression in the brain

As mentioned in Chapter 2, one way to quantify levels of Opn4 expression in sablefish brains is using quantitative PCR (qPCR). qPCR quantifies transcripts using cDNA (reverse transcribed from isolated RNA) as a template. The greater the number of transcripts in a sample (i.e., the higher the expression of a gene), the sooner it will amplify in a PCR beyond the level of background noise. This technique is useful for comparing relative expression levels between paralogs, different tissues, life stages and treatments. Determining the conditions in which genes are upregulated can help to decode the role of their protein products. Given that the main (known) role of opsins is to act as light sensors, I predicted an upregulation of the genes that encode them in the light – in the case of sablefish, during the free-feeding larvae to juvenile stages.

Localizing transcripts and proteins is another important step toward understanding the function of a gene product. Though qPCR can generally compare expression between different

tissues, the results are only as fine scale as the tissue from which the RNA was extracted. For example, one could attempt to dissect the various lobes of a fish brain and isolate RNA from each individually, but there would be no way of discriminating where (within that lobe/region) expression might be localized. Complex and sensory organs, such as the eye and brain, have well-studied and highly specialized functional regions, thus pinpointing expression to the exact cell type or layer within a tissue can be highly informative. Additionally, if sablefish opsins are light-sensitive, I expect to find expression localized to peripheral (dorsal) regions of the brain rather than the deep brain, or regions in which the optic nerves terminate.

To localize gene products to sub-regions and cell types, two approaches are typically taken: in-situ hybridization (ISH) and immunohistochemistry (IHC). ISH localizes RNA in tissue sections using specific probes, which hybridize to a target of interest. The advantage to this approach is increased specificity toward particular products. ISH has been used in opsin research to identify centres of gene expression across aquatic and aerial fields of the four-eyed fish retina (Owens et al., 2012) and in regions of the brain and eye in zebrafish (Davies et al., 2011 and 2015), to compare *Opn5* expression among medaka, zebrafish and gar brains (Sato et al., 2016) and to observe co-expression of *Tmt* and *VA* opsins in medaka neurons (Fisher et al., 2013), to give a few examples. IHC, on the other hand, can localize the protein product of a gene of interest using a primary antibody raised against a small stretch of amino acids, followed by a secondary antibody (with a fluorophore) which reacts with the primary antibody for visualization. Given that many studies begin with either observing the presence/absence or quantification of gene transcripts through qPCR, IHC is a valuable second step to provide information about post-translational products and ultimately function. This is particularly important as many studies have come forward to show the discrepancy between gene transcript and functional protein levels, for example, because of post-translation modifications or protein degradation/half lives (Greenbaum et al., 2003). IHC is also commonly used in opsin research to show protein organization in tissue sections, for example, in deep brain regions of birds (Nakane et al., 2010), and in zebrafish retinal sections (Davies et al., 2011). Both of these techniques are semi-quantitative, in that it they can be used to quantify positive cells, and relative signal levels can be compared within tissue types. As opsins are transmembrane proteins, IHC has the added benefit of (potentially) labeling neural projections, which is useful in cell type differentiation, since axons can terminate in distant regions.

Here, I take advantage of the aquaculture availability of sablefish to determine whether melanopsin expression in the brain varies according to the light environments associated with particular stages of development. I hypothesize that Opn4 genes will be more highly expressed during the stages that naturally occur in photic zones (free-feeding larvae to juvenile). This qPCR-based research will provide insight as to why a fish that spends most of its life in the deep maintains so many opsin genes and whether the expression of melanopsins is regulated tightly by the light environment. I examine the general relationships of Opn4 paralogs across and within life stages to determine which of them appear to be more important in the brain and if they are differentially expressed. Finally, I also localize a subset of melanopsin proteins in the sablefish brain using IHC to confirm that the transcripts are translated into proteins, to see where they are localized in the brain, and whether these regions are optimized to collect light information (for example, in peripheral regions). While opsins have been studied and compared among fish inhabiting different environments, this set of experiments on sablefish is novel to non-visual opsins research in that this single species inhabits diverse photic zones in its lifetime.

METHODS

Sampling for qPCR

I aimed to quantify Opn4 expression in brains from as broad of an age range as possible. Golden Eagle Sablefish rears sablefish from egg until the young adult stage, approximately two years old (Figure 18).

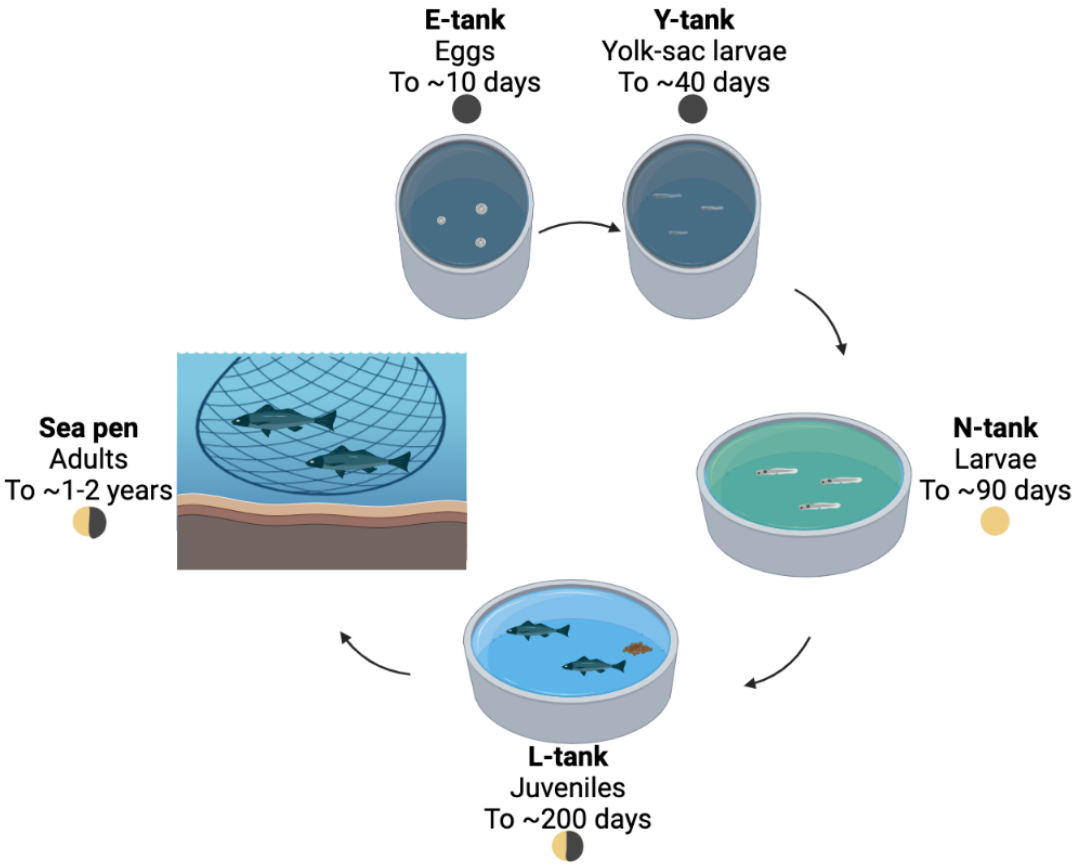


Figure 18. Illustration of Golden Eagle Sablefish hatchery and sea pens, showing the cycling of sablefish through tanks with their approximate ages and light conditions (black circle for total darkness, yellow for 24-hour light exposure, split color for light-dark cycle). The only stage not included in the qPCR experiments was E-tank/egg.

i) Y-tanks

The youngest samples used for this opsin expression analysis were larvae from yolk-tanks (also called Y-tanks), hereafter referred to as yolk-sac larvae. These samples were obtained by Dr. Tricia Rubi, another member of the Taylor Lab. They were sampled one week before ponding, approximately 30 days old. Ponding takes place when the fish are approximately 40 days old; facility workers draw fish to the surface of the Y-tank for collection and to transfer to N-tanks using broad spectrum light. Ponding is the first significant light exposure for the captive fish, as they are kept in dark, sealed tanks until this point. The larvae I surveyed were siphoned from different regions of the funnel-shaped Y-tanks and flash frozen whole in liquid nitrogen. Given that all other life stages include only brain tissue, a subset of the frozen yolk-sac larvae were later treated and thawed in RNAlater ICE (Invitrogen) according to the manufacturer's

protocol and enucleated. I used sterilized insect pins as scalpels under a dissecting microscope to separate the eyes from the rest of the body, providing a better early comparison for the later brain-only samples.

ii) N-tanks

Older larvae (hereafter referred to simply as larvae) were collected from the next set of tanks (N-tanks). They were sampled at 42 days post-ponding (approximately 82 days old) and had been under broad-spectrum 24-hour light and in green algae-tinted water since ponding. Individuals were collected from different regions of the tank. They were euthanized in tricane methanesulfonate (TMS)-treated water (100mg/L) and flash frozen whole in liquid nitrogen. RNAlater ICE (Invitrogen) was later used to thaw the fish for dissection. Horizontal incisions were made between the eyes and behind the gill slits, and these incisions were connected by vertical cuts on either side of the fish. The skin and delicate skull covering the brain were removed. The brain was disconnected from the spinal cord and optic nerves, isolated, kept whole, and immediately used for RNA isolation.

iii) L-tanks

L-tanks house juvenile sablefish, the oldest at the hatchery on Salt Spring Island. Brains from juveniles (sampled 107 days post-ponding, approximately 147 days old) were isolated fresh at the hatchery and flash frozen. These fish were under 12hr:12hr light-dark cycles and were sampled between 3-5 hours after the lights were turned on, so that observations could be compared to those from similar experiments in zebrafish (Davies et al., 2015). They were netted one at a time from the L-tank and tissues were sampled immediately after euthanasia in TMS-treated water (100mg/L). Tissue surrounding the brain was removed and the brains were flash frozen whole. Eyes were also isolated from juvenile fish because of their agreeable size to be used in a subset of experiments discussed later.

For all life stages aforementioned, samples were retrieved from tanks in small batches to ensure that the time between TMS treatment and the freezing of dissected tissue was minimized.

iv) Sea pens

After graduating from the L-tanks, sablefish are transported to open ocean pens in Kyuquot, which are approximately 70 feet deep, where they are kept until they are roughly two years old. These young adults (hereafter referred to as sea pen adults) were sampled by Golden Eagle Sablefish employees and ranged from 1.5-2 years of age. The fish were euthanized in TMS-treated water (100mg/L) around 5:00pm (roughly 8 hours after sunrise) – this was the only opportunity for sampling for logistical reasons. To isolate the brains in the larger fish, their heads were removed near the heart and brains were accessed from the ventral side with a deep lengthwise incision starting at the jaw. Organs were removed until there was a clear view of the brain. After cutting the optic nerves and spinal cord, the brains were isolated and flash frozen whole.

v) Wild caught

Given that the sea pen adults are in relatively shallow ocean pens and are only approximately two years of age (around the age that they begin to descend to the bathypelagic zone in the wild), I also included wild-caught adults in this survey. Alana McPherson, an honours student in the lab, purchased wild adult sablefish heads from Finest at Sea, a local seafood vendor. The sablefish were collected by long line fishermen between the Alaskan border and Barkley Canyon ridge, between depths of 500-1000m. The fish heads were separated from the rest of the body and flash frozen. Alana dissected the brains while the fish were still frozen; a shallow horizontal incision was made behind the eyes, continuing back to the point at which the head was initially removed from the rest of the body. This removed the top of the skull, allowing the brain to be isolated and used immediately in the RNA extraction process. Otoliths were also extracted from these wild fish and aged by the Pacific Biological Station, as age was much less certain in this population than any of the hatchery/sea pen samples. The ages ranged from 3 to 22 years, but for this experiment, I used 3-4 year-old fish.

In all brain isolations, it was difficult to identify the pineal, suggesting that it had been severed/discarded and not included in these brain samples. I performed several experiments on a separate set of formaldehyde-preserved juvenile heads and was unsuccessful in identifying a clear pineal structure, though I could often identify small gelatinous connections between the brain and the inner part of the skull.

The specific ages, light conditions, tissues sampled and sample size at each life stage are summarized in Table 11.

Table 11. Summary of the sablefish life stages used for the qPCR gene quantification assay, including information about light conditions, tissue sampled and sample size. Ages of both adult stages are approximate.

Life stage	Age	Light conditions	Tissue used	Sample size
Yolk-sac larvae	30 days	24-hour darkness, pre-ponding	Whole larvae and whole enucleated larvae (pooled 3 per sample)	N=5
Larvae	82 days	24-hour broad spectrum light, green algae-tinted water	Whole brain	N=6
Juveniles	147 days	12hr:12hr light dark cycle, broad spectrum light	Whole brain, eye	N=6
Sea pen adults	1-2 years	Natural outdoor light, depth 0-70 feet	Whole brain	N=6
Wild adults	3-4 years	Natural outdoor light, depth 500-1000 meters	Whole brain	N=6

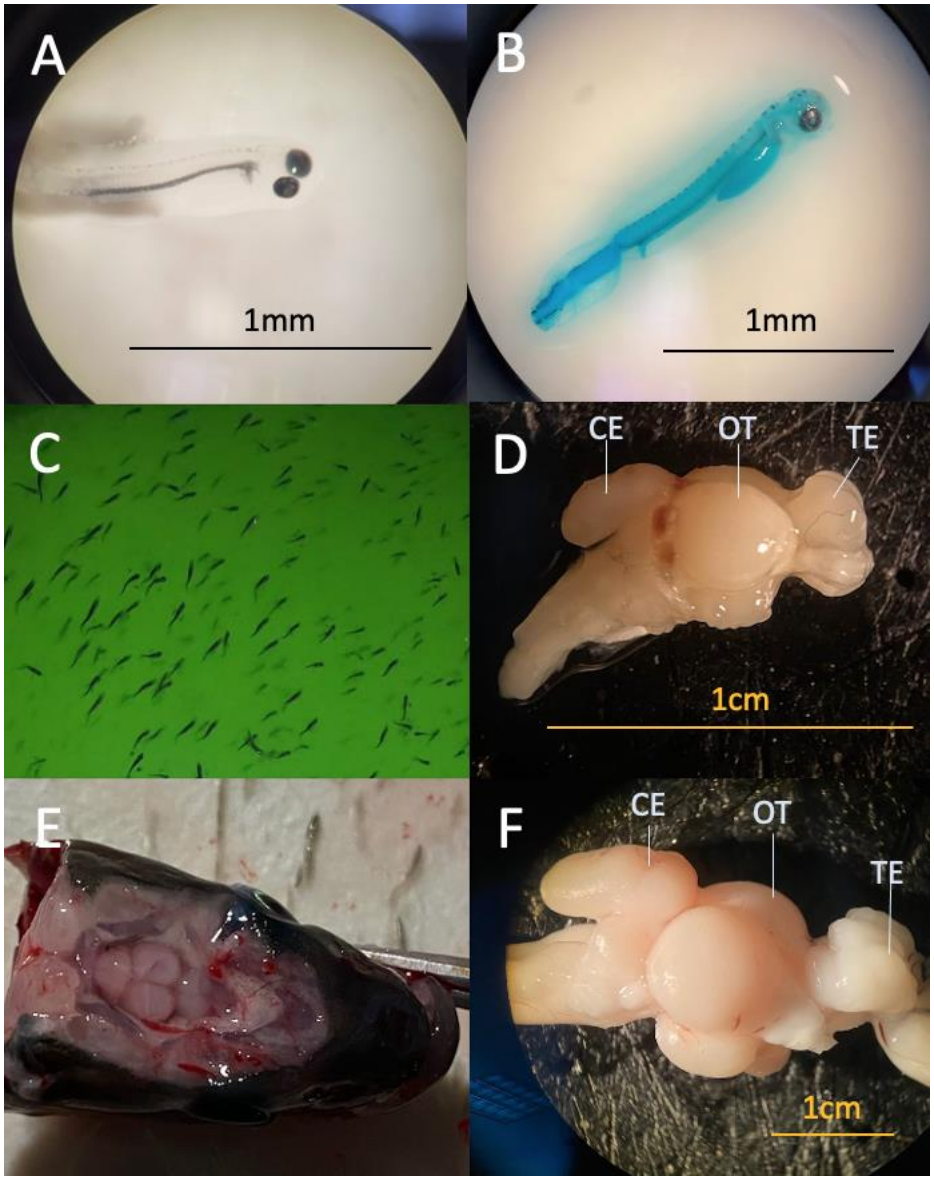


Figure 19. Photos taken during sampling. **A)** A yolk-sac larvae under the dissecting microscope, roughly the size of an eyelash. At this stage, larvae are not exposed to light and are transparent. The lack of a developed brain is also apparent here. **B)** A yolk-sac larvae from the same group as A, stained blue after RNALater-ICE (Invitrogen) treatment. This staining more clearly shows the depleting yolk-sac and what appear to be melanophores developing. **C)** N-tank larvae swimming in their algae-tinted tank. **D)** An isolated juvenile brain with clear fore-, mid- and hind-brain structures. Labeled for reference are the telencephalon (TE), optic tectum (OT) and cerebellum (CE). **E)** A juvenile head with a thin layer of skin and skull removed, showing the brain's position under the pineal window and its proximity to external environment. **F)** A sea pen adult brain with the same structures labeled in D).

RNA extraction

Total RNA was extracted using the RNeasy Lipid Mini Kit (Qiagen) according to the manufacturer's protocol. Briefly, the steps include lysis/homogenization, chloroform phase separation and ethanol precipitation, followed by column binding, washing and eluting. As previously mentioned, yolk-sac larvae were kept whole (or whole and enucleated), as brains were not possible to isolate. Two to three yolk-sac larvae were pooled and treated as one biological replicate to ensure that the RNA yield was high enough for cDNA concentrations of all life stages to be standardized in subsequent steps. Older larvae and juvenile brains were kept whole and added one at a time directly to the lysis reagent. Much larger sea pen and wild adult brains were macerated in a chilled mortar and pestle into a powder. The powder was mixed (in order to form a representative sample of the whole brain), and a fraction of it was added to the lysis reagent. After this point, all samples followed the same protocols through to qPCR.

RNA concentrations were measured using a NanoDrop, and OD 260/280 and 260/230 ratios were used to determine the quality of the RNA extracted. All samples used in subsequent analyses showed 260/280 > 2.0.

DNase treatment

Experiments presented in Chapter 2 showed a need for RNA samples to be treated with DNase to reduce genomic DNA contamination in subsequent qPCR amplification, as this impacted my ability to accurately evaluate expression levels. Each RNA sample was treated with DNAse I Amplification Grade (Thermo Fisher), with the standard protocol (incubation and deactivation with EDTA and heat), scaled up to accommodate and treat 2000ng of RNA total. RNA concentrations were measured using a NanoDrop once again after treatment. These final concentrations were used for standardizing subsequent steps.

Reverse transcription

All DNAse-treated RNA samples were reverse transcribed to cDNA using SuperScript VILO Master Mix (Thermo Fisher). All cDNA reactions were standardized to 25ng/uL. For each conversion, two reverse transcriptase negative (RT-) controls were included per life stage. These controls have had their reverse transcriptase enzyme/reagent denatured (rendered non-functional) at a high temperature prior to the subsequent conversion steps. With no RNA to cDNA

conversion, genomic DNA (if present) is the only template available for subsequent qPCR amplifications in these samples.

Primers and probes

Primers and probes were designed (from sequence data reported in Chapter 1) for all OPN4-class opsins: *Opn4m1, Opn4m2, Opn4m3, Opn4x1* and *Opn4x2,* along with the reference genes introduced and validated in Chapter 2, *Eef1a and Rps18*.

Additional primers and probes were designed for *Sws1* and *Rh1*, only to be used on certain life stages/conditions (discussed further in below). These were used to put melanopsin expression levels into perspective. *Sws1* and *Rh1* are classical visual opsin genes, typically expressed at high levels in cone and rod photoreceptors, respectively.

Each assay was designed according to the protocol in Chapter 2. Most importantly, in each assay, one of the primers spanned an exon junction, with the exception of the single exon *Opn4m2* gene. Testing the primers on genomic DNA (using RT-PCR) revealed that each of the Opn4 primers can amplify genomic DNA, but after additional DNase treatment, the contamination problem becomes negligible. As mentioned in Chapter 2, targets were run in duplex with one other assay with a compatible TaqMan QSY probe dye. PCR efficiencies (in duplex) were determined with a cDNA serial dilution. PCR oligo sequences, amplicon sizes and efficiencies are shown in detail in Chapter 1 (Table 9).

qPCR gene quantification

The samples used for the main qPCR experiment (with all Opn4 genes) were whole yolk-sac larvae, enucleated yolk-sac larvae (both 30 days old), larval brains (82 days old), juvenile brains (147 days old), sea pen adult brains (~1-2 years old) and wild caught adult brains (~3-4 years old). The sample size for each stage was 6, with the exception of yolk-sac larvae, which only had 5 biological replicates. Each biological replicate was run in triplicate (i.e., 3 technical replicates).

The additional experiment, quantifying other opsins to orient the brain Opn4 expression data, was run separately using select tissues only. I quantified *Rh1* and *Sws1* expression in the same subset of juvenile brains mentioned above, as well as whole juvenile eyes (N=6), as eyes are easiest to isolate and homogenize at this age due to their size. The Opn4 genes were also

quantified in juvenile eyes for comparison. Visual opsin and melanopsin genes are known to encode functional protein products in various regions of the eye, thus this experiment was included to see if the same genes were expressed at comparable and potentially functional levels in the brain. Importantly, the *Eef1a* and *Rps18* housekeeping genes were validated as stable across the two tissue types at the juvenile life stage, with less than a one cycle difference on average.

Target gene master mixes were made for each opsin and reference gene, which included forward and reverse primers (18uM), the probe (5uM) and 1X Tris-EDTA. A TaqMan master mix was made for each duplex assay that contained both target gene mixes, water and TaqMan FAST Advanced Master Mix (Thermo Fisher). TaqMan FAST Advanced Master Mix contains DNA polymerase, dNTPs, buffer components, and importantly, ROX dye as a passive reference which controls for variation (in pipetted volume) between wells. The various duplex TaqMan master mixes were aliquoted for each sample, the appropriate cDNA was added, and triplicates were plated with a final concentration of 25ng cDNA per well. No template controls were included for each assay. The PCR reaction was carried out according to the typical 'Comparative Ct FAST' protocol on the QuantStudio 5 (Thermo Fisher) machine, which includes a 20 second polymerase activation stage (95°C) and 40 short denaturation (1 second, 95°C) and annealing-extension (20 seconds, 65°C) cycles.

qPCR data analysis

The qPCR results were first analyzed using the Design and Analysis software (Thermo Fisher). The software automatically sets the critical thresholds, i.e., the point at which the signal exceeds the background fluorescent signal (this is slightly different for each assay). I performed a primary quality check, ensuring that the ROX passive reference appeared to have a stable (flat) signal across the whole plate. I input the %efficiency for each of my primer sets to account for deviations from the assumed perfect doubling of product with every cycle; as noted in Chapter 2, these ranged between 90-110% efficiency. I manually omitted triplicate outliers if they differed by more than one cycle from the others. I checked for amplification in RT- and no template controls to ensure that the components of the reaction are free of contamination.

I computed relative gene quantification manually in Excel following the ΔΔCq method (Table 12). This is similar to the Pfaffl method, which accounts for PCR efficiency. While the

Pfaffl method is considered most accurate (Pfaffl, 2001), I did not follow it exactly, since the Design and Analysis software already incorporated primer efficiencies into the output Cq values.

Table 12 outlines the ΔΔCq method. Briefly, Cq values of technical replicates are averaged for the target genes and housekeeping genes; this is referred to as the EqCq mean. For each biological replicate, the EqCq values for both housekeeping genes are averaged. The averaged housekeeping gene EqCq is then subtracted from the EqCq of the target genes, giving a value referred to as ΔCq (column 5); this is the normalization step. That is, if one replicate has a high target gene Cq because of some external factor (e.g., cDNA concentration, etc.), then the housekeeping gene (which is stably expressed) will also be higher. Taking the difference between the target and the housekeeping genes accounts for that external factor. Since I used the relative quantification qPCR approach (as opposed to absolute quantification), I then chose a calibrator in order to calculate fold change. The calibrator can either be a life stage or a particular gene that the expression of everything else is compared to. Since I aimed to compare expression of several opsin paralogs across many different life stages all at once, I chose one of my reference genes, *Eef1a,* as a calibrator. To calculate fold change, first the ΔΔCq is determined by calculating the difference between the ΔCq of the target gene and the calibrator. The fold change (or RQ, for relative quantification) is calculated by the equation 2^-ΔΔCq, where the 2 signifies the doubling of PCR product with each cycle. The RQ is log10 transformed for linearity of genes expressed at higher and lower levels than the housekeeping genes. After this point, the log transformed values of each biological replicate are averaged to get the mean log10(RQ) of the target relative to the calibrator *Eef1a* for a biological group or life stage.

To determine if the mean expression of each Opn4 gene differed across life stages, five separate one way ANOVAs were conducted, followed by Tukey Multiple Comparisons (both with alpha=0.05). The same tests were also done to compare all five Opn4 paralogs at a particular life stage. Despite having a small sample size, my data were normal (with very few deviations), as shown by Shapiro-Wilk tests. As ANOVAs are fairly robust to deviations in normality and non-parametric tests can be very underpowered, I decided that ANOVA would be best suited for my data.

Table 12. Data analysis for qPCR relative melanopsin quantification. This example shows the normalization of *Opn4m1* in six larval brain replicates using two housekeeping genes *(Eef1a* and *Rps18*), as well as the fold change calculation, relative to the chosen calibrator, *Eef1a*. Cq values (which already account for PCR efficiencies) of technical replicates are averaged for both the target *Opn4m1* gene and the housekeeping genes; this is referred to as the EqCq mean (column 3). For each biological replicate, the EqCq values for both housekeeping genes are averaged (column 4). The averaged housekeeping gene EqCq is then subtracted from the EqCq of the target *Opn4m1* gene, giving a value referred to as ΔCq (column 5); this is the normalization step. To calculate fold change, first the ΔΔCq is determined by calculating the difference between the ΔCq of the target gene and the calibrator, which is *Eef1a* in this case (column 6). The fold change or RQ is calculated by the equation 2^-ΔΔCq, where the 2 signifies the doubling of PCR product with each cycle (column 7). The RQ is log10 transformed for linearity of genes expressed at higher and lower levels than the housekeeping genes (column 8). After this point, the log transformed values of each biological replicate are averaged to get the mean log10(RQ) of *Opn4m1* relative to the calibrator *Eef1a* at the larval stage of development.

1 **Sample**	**2** **Target**	**3** **EqCq Mean** (mean of triplicates)	**4** **HKEqCq** 3(ref1)+3(ref2)/2	**5** **ΔCq** 3-4	**6** **ΔΔCq** 5(target)-5(calibrator)	**7** **RQ** 2^-(6)	**8** **log10** log10(7)
Brain 3	*Eef1a*	14.6598172	15.8649546	-1.2051374			
Brain 4	*Eef1a*	14.8780338	16.3180585	-1.4400247			
Brain 5	*Eef1a*	14.9424332	15.9556108	-1.0131776			
Brain 6	*Eef1a*	15.3483408	16.5704115	-1.2220707			
Brain 9	*Eef1a*	17.2311355	18.4766605	-1.245525			
Brain 10	*Eef1a*	14.6630991	15.8090788	-1.1459797			
Brain 3	*Rps18*	17.0700919	15.8649546				
Brain 4	*Rps18*	17.7580831	16.3180585				
Brain 5	*Rps18*	16.9687884	15.9556108				
Brain 6	*Rps18*	17.7924822	16.5704115				
Brain 9	*Rps18*	19.7221855	18.4766605				
Brain 10	*Rps18*	16.9550585	15.8090788				
Brain 3	*Opn4m1*	25.0746458	15.8649546	9.20969123	10.4148286	0.00073253	-3.1351758
Brain 4	*Opn4m1*	23.6664279	16.3180585	7.34836943	8.78839413	0.00226167	-2.6455702
Brain 5	*Opn4m1*	25.1568971	15.9556108	9.20128632	10.2144639	0.00084167	-3.07486
Brain 6	*Opn4m1*	25.204209	16.5704115	8.63379753	9.85586823	0.00107917	-2.966912
Brain 9	*Opn4m1*	34.492157	18.4766605	16.0154965	17.2610215	6.3667E-06	-5.1960852
Brain 10	*Opn4m1*	25.1503822	15.8090788	9.34130344	10.4872831	0.00069665	-3.1569868

Immunohistochemistry

After uncovering clear evidence of Opn4 transcription in the brain, I characterized protein-level expression using immunohistochemistry. There were no perfectly suitable commercially available melanopsin antibodies; those raised against the single mouse Opn4 lacked sequence similarity to the sablefish Opn4's, and the only zebrafish anti-Opn4 antibody commercially available was designed to label only one of Opn4 and it has not been used in any publications to date. The most suitable antibody in the literature was a custom anti-Opn4 antibody, called pas350, designed by Davies et al. (2011). Though not available to order, Wayne Davies and Stephen Hughes, authors of the aforementioned paper, generously sent me this antibody to use. Pas350 is a rabbit polyclonal antibody raised against a conserved stretch of amino acids among the zebrafish Opn4m's. The sequence, CVPFPTVDVPDHA, is a good match to sablefish Opn4m opsins as well (Figure 20A). Sablefish Opn4m1 and Opn4m3 share 11/13 amino acids and Opn4m2 shares 10/13. As such, the protein labeling is not specific to a particular opsin and cannot divulge differential localization of these Opn4 paralogs, but can provide insight on the Opn4m group as a whole.

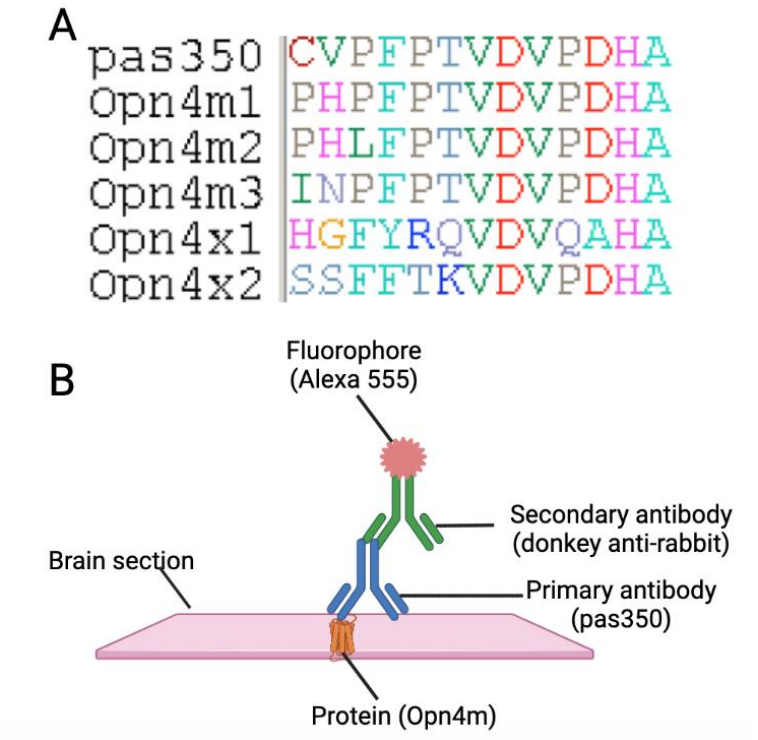


Figure 20. Opn4m immunohistochemistry. **A)** Alignment of the zebrafish pas350 antibody target epitope to the five sablefish Opn4 paralogs, showing a high degree of sequence similarity, particularly for Opn4m's. **B)** Diagram of Opn4m immunohistochemistry, showing the binding of the primary antibody (pas350) to the opsin epitope and the secondary antibody (donkey anti-rabbit) binding to the primary antibody, with its attached fluorophore (Alexa 555).

Prior to using the decade-old pas350 antibody on sablefish, it was tested on zebrafish eyes, the model for which it was originally designed, and the tissue in which Davies et al. got results. This original pas350 experiment showed that Opn4m proteins are expressed globally across the retinal cell types in zebrafish, including photoreceptors, horizontal cells, bipolar cells, amacrine cells and a subset of retinal ganglion cells. Gene-specific in-situ hybridization showed that mRNA of the five melanopsin paralogs was organized in a largely non-overlapping manner among these cell types (Davies et al., 2011).

Samples of adult zebrafish were obtained from collaborators for this step (N=2). They had been briefly fixed in 4% paraformaldehyde (PFA) and transported in phosphate-buffered saline (PBS). The tissues were placed in a cryoprotective solution (0.1 M PBS containing 25% sucrose and 20% O.C.T. compound) for approximately 24 hours, until the tissues sank to the bottom of the vial. Protected eyes were embedded whole in 100% O.C.T. and flash frozen in

liquid nitrogen. I took radial cryosections of the eye (14 microns thick) and collected them on coated microscope slides.

Just ahead of the immunohistochemistry procedure, the slides were thawed at room temperature and rinsed with 0.8 M PBS (3 x 5 minutes). Normal donkey serum, diluted 1:20, was applied to prevent non-specific binding (30 minutes, room temperature). The affinity purified anti-melanopsin pas350 antibody was diluted 1:500 in 0.1M PBTA (0.1M PBS/NaN_3, 0.1% Triton, 5% bovine serum albumin) and applied to the positive control slides. Two negative controls were included in this experiment, a primary-negative control (PBTA without pas350) and a primary antibody that has been blocked by an antigen peptide, the epitope that the primary was designed to bind (pas350 diluted 1:500 and antigen peptide diluted 1:300, incubated together in PBTA for 20 minutes). These three solutions were applied to their respective slides for 2.5 hours at room temperature, followed by a series of PBTA rinses (5 x 3 minutes). The secondary antibody, donkey anti-rabbit Alexa 555 (DARabA555) (Invitrogen), was diluted 1:500 and applied overnight at 4°C. The slides were rinsed of the secondary antibody with distilled water (5 x 3 minutes), mounted using Prolong Gold Antifade Mountant with 4’,6-diamidino-2-phenylindole (DAPI; Thermo Fisher), and sealed with a coverslip.

Sablefish tissues were prepared very similarly to the zebrafish trial above. Again, while brains were the focus of this effort, eyes were included to compare to the expression patterns found by Davies et al. (and confirmed here). Whole eyes and brains (N=3) were dissected from juveniles euthanized at Golden Eagle Sablefish; brains were dissected with the same approach outlined for qPCR, and eyes were severed from the optic nerve and removed from the head. These fish were close to the same age as the juveniles used in qPCR, approximately 167 days old or 127 days post-ponding and these were also collected between 3-5 hours after lights have been turned on. This age group was chosen because of tissue size; they are well developed in terms of structure, but a full brain or eye section could still easily fit onto microscope slides. The isolated tissues were fixed in 4% paraformaldehyde in 0.08M phosphate buffered saline (PBS), immediately after dissection. The following day, the fixative was removed and the tissues were rinsed with 0.08 M PBS 3 x 2 hours. Subsequent cryoprotecting and embedding, and immunostaining steps exactly followed the protocol above for zebrafish. A series of dilutions of the primary and secondary antibodies were done (ranging from 1:500 to 1:2000), to achieve the best signal to noise ratio. Due to high levels of autofluorescence, 1:500 for both the pas350

primary and the DARabA555 was deemed best. Primary-negative and antigen-blocked primary antibodies were also used on sablefish tissues to confirm signals were from positive antibody staining, rather than from background noise or autofluorescence. Signals were compared between biological replicates (brains and eyes from 3 different fish).

To determine whether the pas350-stained cells were neuronal or glial cells, an anti-GFAP (glial fibrillary acidic protein) primary antibody was also applied to the tissues. GFAP plays many important roles in the central nervous system, including cell structure and communication (Yang & Wang, 2015), and is a common marker primarily used to identify astroglia (synonymous with zebrafish radial glia) (Jurisch-Yaksi et al., 2020) and astrocytes (Than-Trong & Bally-Cuif, 2015). Mammalian GFAP antibodies have successfully been used for teleost immunohistochemistry (Lam et al., 2009); here, I used a monoclonal rat anti-GFAP primary on sablefish because it was readily available on campus and kindly donated to me from the Swayne Lab. As the pas350 and GFAP antibodies were made in different hosts, they were applied together in the same protocol as outlined above. The GFAP primary antibody was diluted 1:100 and the secondary antibody, donkey anti-rat Alexa 488 (DARatA488), was diluted 1:500. In this case, as I was not supplied with an antigen peptide blocking control, I performed primary-negative controls only.

RESULTS

Opn4 expression levels and trends

I tested all primer sets initially using RT-PCR on cDNA and genomic DNA, running the product through a gel to confirm the predicted amplicon length, as well as cDNA specificity (see Chapter 2). All amplicons were the expected length however, despite rigorous primer design considerations, many exon-spanning primers worked on genomic DNA. Thus, as mentioned, I included a DNase step after RNA isolation, to reduce genomic DNA contamination. In preparation for qPCR, I calculated the PCR efficiency for each primer-probe multiplex set across a range of cDNA concentrations; each fell within the 90-110% efficient range. I quantified the housekeeping/reference genes (*Rps18* and *Eef1a*) across the different developmental stages and tissues, both of which were relatively stable and appropriate to be used for expression normalization, based on my tests. These points are all extensively discussed in Chapter 2.

I found low expression of melanopsins in extraretinal sablefish tissue compared to the *Eef1a* calibrator. Across the board (all life stages and all paralogs), Opn4 expression was no greater than 10% (RQ of 0.01) of that of *Eef1a* in the brain. Despite expression of Opn4 paralogs being low compared to the calibrator gene, the negative controls (which indicate any level of genomic DNA contamination contributing to the signal) failed to amplify with Opn4 targets, suggesting that the signals from the test samples reflect real gene expression.

To contextualize and orient the low melanopsin expression levels in the brain, cross-tissue comparisons were performed for one life stage; expression levels of all five Opn4 paralogs, as well as two visual opsins (*Rh1* and *Sws1*), were quantified and compared between juvenile brain and eye tissues. This showed that expression levels of the Opn4 paralogs do not differ much between the brain and eye, at least in the juvenile life stage. Specifically, only *Opn4m3* (t-test, P-value=0.000121) and *Opn4x2* (t-test, P-value=0.0001) differ noticeably between the two tissue types both of which are expressed at higher levels in the eye (Figure 21).

The traditional visual opsins, presumed to play roles in vision in the eye, were expressed at low levels in the juvenile brain. *Rh1* in particular did not amplify in 40 cycles for many of the brain samples (see Appendix H for complete list of non-detects). Comparatively, *Rh1* expression levels were very high in the eye, as expected. This is the only case, across all of the tested tissues, life stages and genes, in which expression was higher than the *Eef1a* calibrator. *Sws1* expression in the juvenile eye was higher than in the brain (t-test, P-value=<0.00001), but it was not as different as expected. Additionally, it was expressed at similar levels to, if not lower than, the Opn4 genes in the eye (Figure 21).

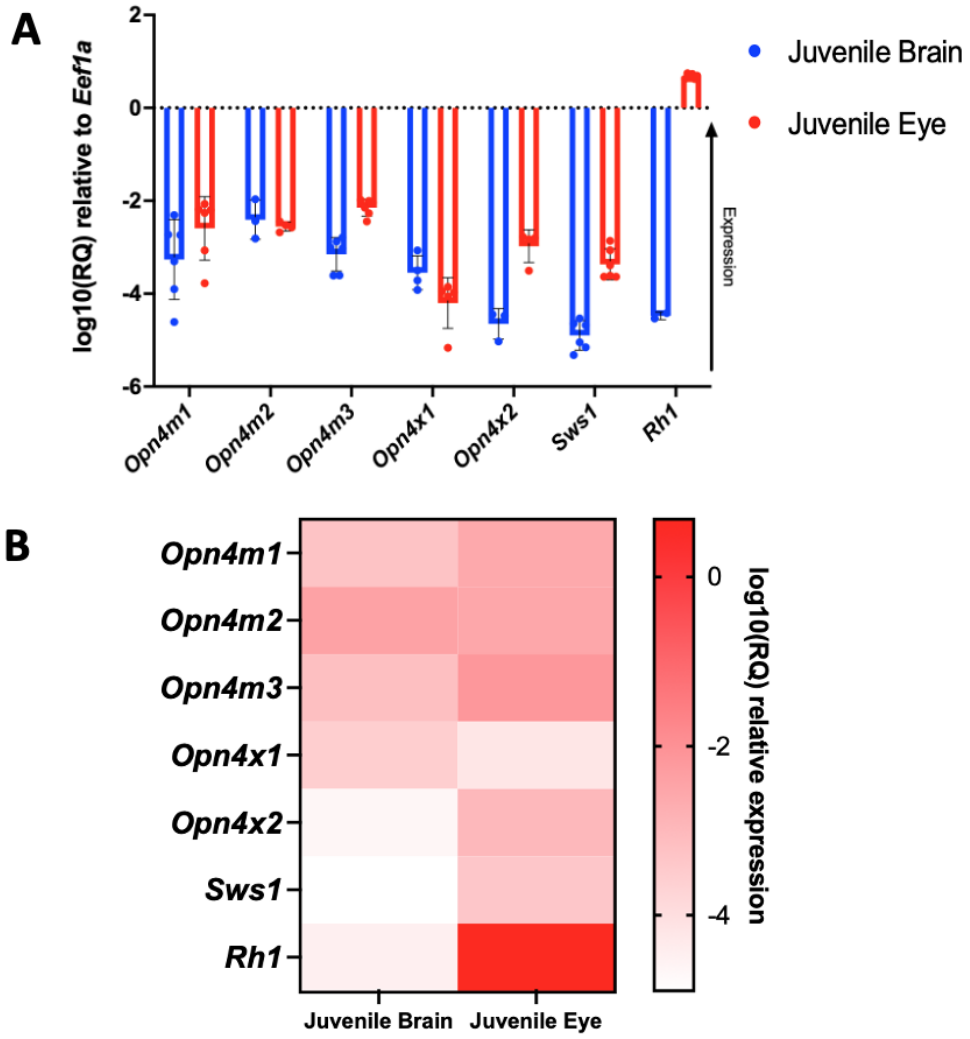


Figure 21.Comparison of opsin expression between juvenile sablefish brains and eyes (N=6). Expression data has been log10 transformed and is reported relative to *Eef1a*. **A)** Graph of mean log10(RQ) +/-SD. Individual points are plotted to show that there are non-detects in the data (particularly for *Rh1* in the brain). The less negative the number, the higher the expression. **B)** A heat map of the same data from A). High expression is shown by a deeper color.

Returning to the main experiment, Opn4 quantification in brain tissue among life stages, this perspective of relative expression levels is important to keep in mind. This work reveals that all Opn4 paralogs are expressed in the brain to varying extents (Figure 22). It is also clear that many of these melanopsin genes are transcribed during life stages that occur in the dark: the yolk-sac larval and wild adult stages. It should be noted that only the enucleated yolk-sac larvae were used in these comparisons; yolk-sac larvae with eyes unsurprisingly have different opsin expression levels compared to their enucleated tank-mates (evaluated visually, Figure 23).

Because of this confounding factor, yolk-sac larvae with eyes were removed from subsequent cross-life stage brain comparisons.

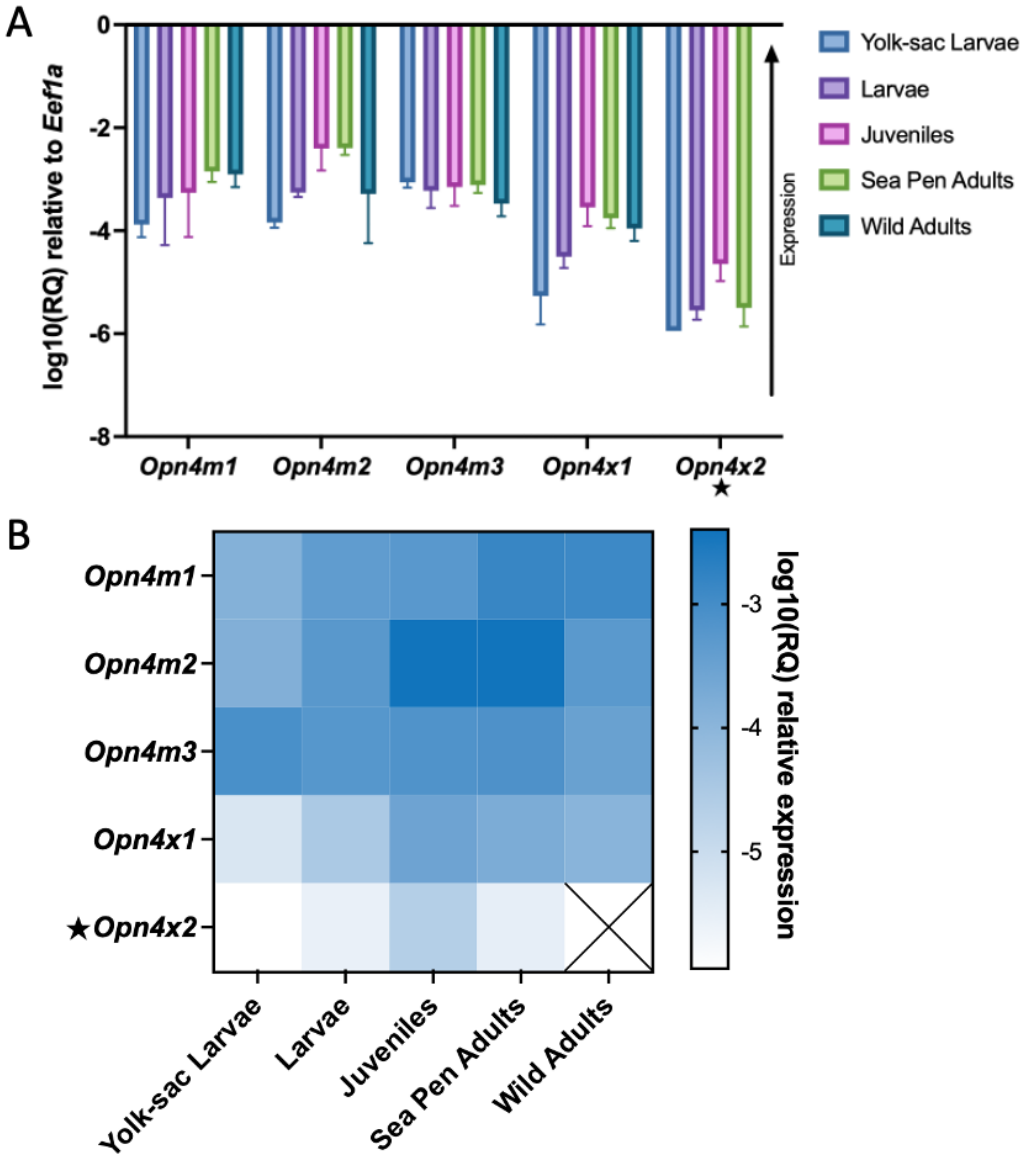


Figure 22. log10 transformed expression data of Opn4 paralogs in sablefish brain across life stages, relative to the expression of *Eef1a*. An exception is the yolk-sac larval stage, which contains brain tissue, as well as other extraocular tissue. N=6 biological replicates (each run in triplicate) for each opsin at every developmental stage. **A)** Plot of mean log10(RQ) +/- SD, **where the more negative the value, the lower the expression** (0=expression level the same as the calibrator). **B)** Heatmap of the same brain expression data for easier comparison across stages and paralogs. Asterisk by *Opn4x2* because there were many non-detects for this paralog in all life stages – sample size is dramatically reduced, and means do not account for non-detects.

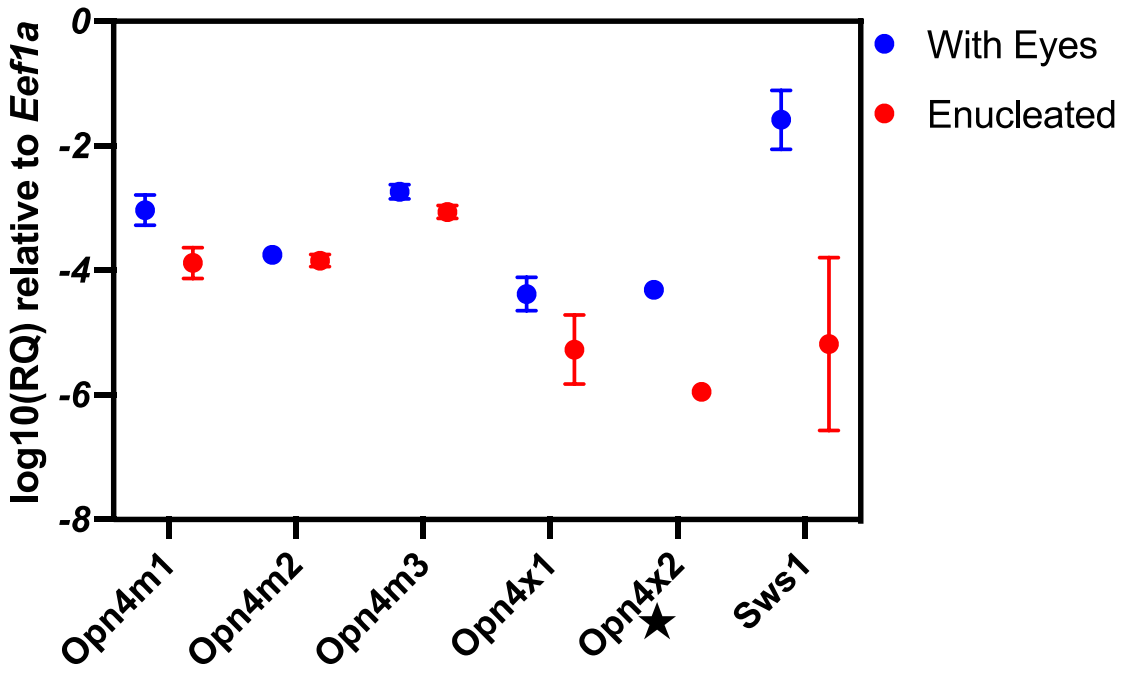


Figure 23. log10 transformed expression data relative to *Eef1a* (+/-SD) of Opn4 paralogs and a visual opsin (Sws1) in whole and enucleated yolk-sac larvae. Five biological replicates (each run in triplicate) for each opsin and tissue. Asterisk by *Opn4x2* because there were many non-detects for this paralog in enucleated yolk-sac larvae; sample size is dramatically reduced, and means do not account for non-detects.

Across this analysis, one trend is obvious: the Opn4m paralogs are expressed at higher levels in the brain than the Opn4x paralogs (Figure 22). It is worth noting that observable trends in the data do not always hold statistical significance. As biological replicates at a given life stage can be variable in expression, standard deviation and error bars are generally fairly large. Consequently, even if mean expression levels appear to be very different, the statistical tests for significance tend to fail to reject the null hypothesis (that the means are all equal).

Partly because of the aforementioned Opn4m trend, there are multiple statistically significant differences between Opn4 paralogs at any given life stage (Appendix I). Generally, the Opn4m's statistically differ from the Opn4x's, though this is not always true. For example, at the juvenile stage, the expression of any of the Opn4m paralogs does not differ significantly from *Opn4x1*. Expression among the Opn4m class of opsins (*Opn4m1*, *2* and *3*) at any particular life

stage tends not to differ significantly (with some exceptions, such as *Opn4m3* being expressed at lower levels than *Opn4m1* and *Opn4m2* in yolk-sac larvae).

While differences in the expression of melanopsin paralogs at a given life stage is informative in its own way, the main hypothesis going into this set of experiments was that Opn4 expression would be higher in the stages that have more light exposure. To test the hypothesis, comparisons were made for *each* Opn4 paralog individually across the stages of development (one-way ANOVA and Tukey Multiple comparisons). For some of the paralogs, expression levels were not detectably different across life stages. For example, expression of *Opn4m1* and *Opn4m3* was not statistically different among the five brain stages (one-way ANOVA, P-value>0.05) (Table 13). Qualitatively, however, *Opn4m3* shows a unique pattern of expression that appears to be somewhat inversely correlated with age (higher expression earlier in life). *Opn4m1* shows a qualitative pattern fairly consistent with patterns observed for *Opn4m2* and *Opn4x1*, mentioned below.

Opn4m2 and *Opn4x1* both show lower expression in the earlier life stages compared to the old fish, specifically between the yolk-sac larvae and the others. When plotted side by side, *Opn4m2* and *Opn4x1* appear to follow the same patterns of expression changes throughout development (Figure 24). These data show a rough correlation of *Opn4m2* and *Opn4x1* expression with light exposure and age qualitatively, though the trend is imperfect quantitatively (for example, *Opn4m2* expression in dark-reared yolk-sac larvae do not differ significantly from that of larvae under 24-hour light*;* Tukey Multiple Comparison, P-value>0.05).

Table 13. Summary of statistically significant differences in whole brain Opn4 expression between life stages: yolk-sac larvae (YEP), larvae (LB), juveniles (JB), sea pen adults (AB) and wild adults (WAB). Several One-Way ANOVAs determined whether the mean log10(RQ) of a particular Opn4 gene differed between life stages (H_o: $\mu_{yolk\text{-}sac}=\mu_{larvae}=\mu_{juvenile}=\mu_{sea\ pen\ adult}=\mu_{wild\ adult}$); these results are reported in the second column. For genes that did show statistically significant difference across life stages, subsequent Tukey Multiple Comparisons revealed *which* life stages differed from the others in a pairwise manner (H_o: $\mu_{yolk\text{-}sac}=\mu_{larvae,}$ $\mu_{yolk\text{-}sac}=\mu_{juvenile}$, etc.); these results are reported in the third column. Only life stages that showed significantly different levels of expression are shown in column three; all other combinations were not significant. Statistical significance was set to 0.05. P-value: >0.05 (ns), <0.05(*), <0.01(**), <0.001 (***), <0.0001(****).

Target	One-Way ANOVA	Tukey Multiple Comparison
Opn4m1	ns	ns
Opn4m2	**	YEP-JB * YEP-AB **
Opn4m3	ns	ns
Opn4x1	****	YEP-LB * YEP-JB **** YEP-AB **** YEP-WAB **** LB-JB ** LB-AB **
Opn4x2	*	YEP-JB * LB-JB *

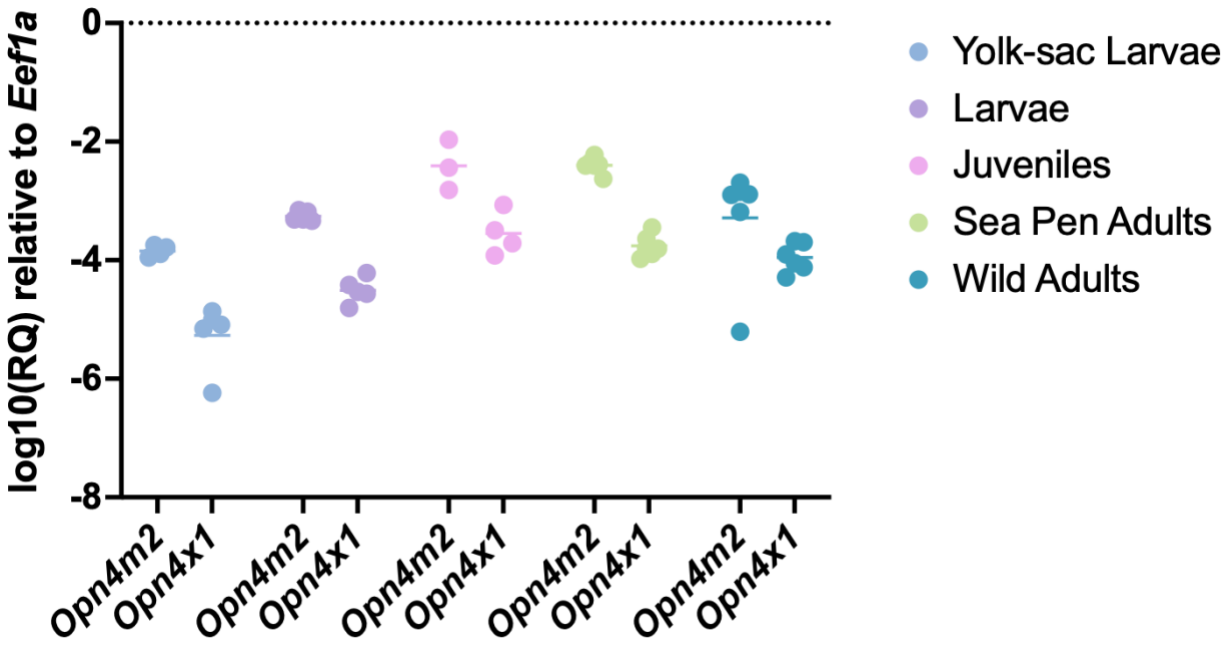


Figure 24. Plot of *Opn4m2* and *Opn4x1* expression data (log10(RQ) relative to *Eef1a*) showing the same qualitative pattern through development both genes, despite having different expression levels. Age increases from left to right.

Opn4x2 expression was lacking among biological replicates at any given stage, with no detection in most samples (i.e., no amplification in 40 cycles). Those that did amplify for the *Opn4x2* target had very low expression relative to the other genes, but amplified around cycles 31 through 36. Thus, while some biological replicates went 40 cycles without amplifying, others amplified as much as 9 cycles sooner, which signifies a 512-fold difference in expression. There are several approaches one can take to address the issue of non-detects in qPCR. For reasons outlined in Chapter 2, non-detects were excluded from the dataset, lowering the sample size and consequently having more conservative statistical outcomes.

Some opsin paralogs across different life stages appear to be more variable than others. Excluding *Opn4x2* (since the majority of the replicates were non-detects), *Opn4m1* had the highest overall standard deviation (mean SD of biological replicates for each life stage, averaged) and *Opn4m3* had the lowest. It is worth noting that variability was not consistent across the life stages; each Opn4 gene had more variable expression at certain point in development. For example, *Opn4m1* had a SD of 0.92 in larval brains, but a SD of 0.20 in adult brains (Table 14). Also worth noting is that this is referring to variation after normalization, thus

I am optimistic that these deviations are real and not just the product of technical variation (e.g., sample loading, RNA integrity, concentrations; these have already been accounted for).

Table 14. Descriptive statistics of log10 transformed expression, relative to *Eef1a*. *There were many *Opn4x2* non-detects, thus standard deviation is based on very few detectable values.

		Yolk-sac larvae	Larvae	Juveniles	Sea Pen Adults	Wild Adults	Average SD
Opn4m1	Mean	-3.88	-3.36	-3.26	-2.85	-2.91	0.495
	SD	0.247	0.918	0.860	0.206	0.246	
Opn4m2	Mean	-3.84	-3.26	-2.41	-2.40	-3.29	0.338
	SD	0.0979	0.0839	0.423	0.129	0.955	
Opn4m3	Mean	-3.06	-3.23	-3.15	-3.11	-3.48	0.239
	SD	0.103	0.332	0.364	0.152	0.243	
Opn4x1	Mean	-5.27	-4.51	-3.55	-3.76	-3.96	0.313
	SD	0.552	0.217	0.364	0.190	0.242	
Opn4x2*	Mean	-5.95	-5.55	-4.65	-5.50	n/a	0.218
	SD	0.00	0.183	0.327	0.360	n/a	

***Opn4* localization in the brain and eye**

Using the pas350 antibody, proteins encoded by Opn4m paralogs were localized in sablefish brain and retinal regions. First, the antibody was tested in its original target species, zebrafish. There was clear immunolabelling in the retina through all inner nuclear layer subtypes: horizontal cells, bipolar cells and amacrine cells. A subpopulation of retinal ganglion cells was also labelled, as well regions within the photoreceptor layer (Figure 25B). These findings mirror those of Davies et al. (2011), thus the antibody can be considered functional and the protocol successful (Figure 25A).

Sablefish retinal sections treated with the pas350 antibody show some of the same patterns as zebrafish. Cell type was not explicitly confirmed here (for example, using other antibodies), but rather based on the appearance and location among the retinal layers. The clearest pas350 positive signal in sablefish appears to be in bipolar and horizontal cells. The photoreceptor layer is bright with labelling, but this region remains fairly bright in the primary-negative and primary-blocked controls (Figure 25D). The planes of sectioning for Figure 25C and Figure 25D differ slightly, thus it is difficult to discern whether the signal in the negative control is occurring in the exact same layers as the test sample, but in general it appears that much (if not all) of the fluorescence in the photoreceptor layer is likely due to auto-fluorescence.

Compared to zebrafish, there is a notable absence of immunolabelling in the inner plexiform layer, where projections of bipolar cells terminate. These results show that Opn4m proteins expressed in many of the different cell types within the sablefish retina.

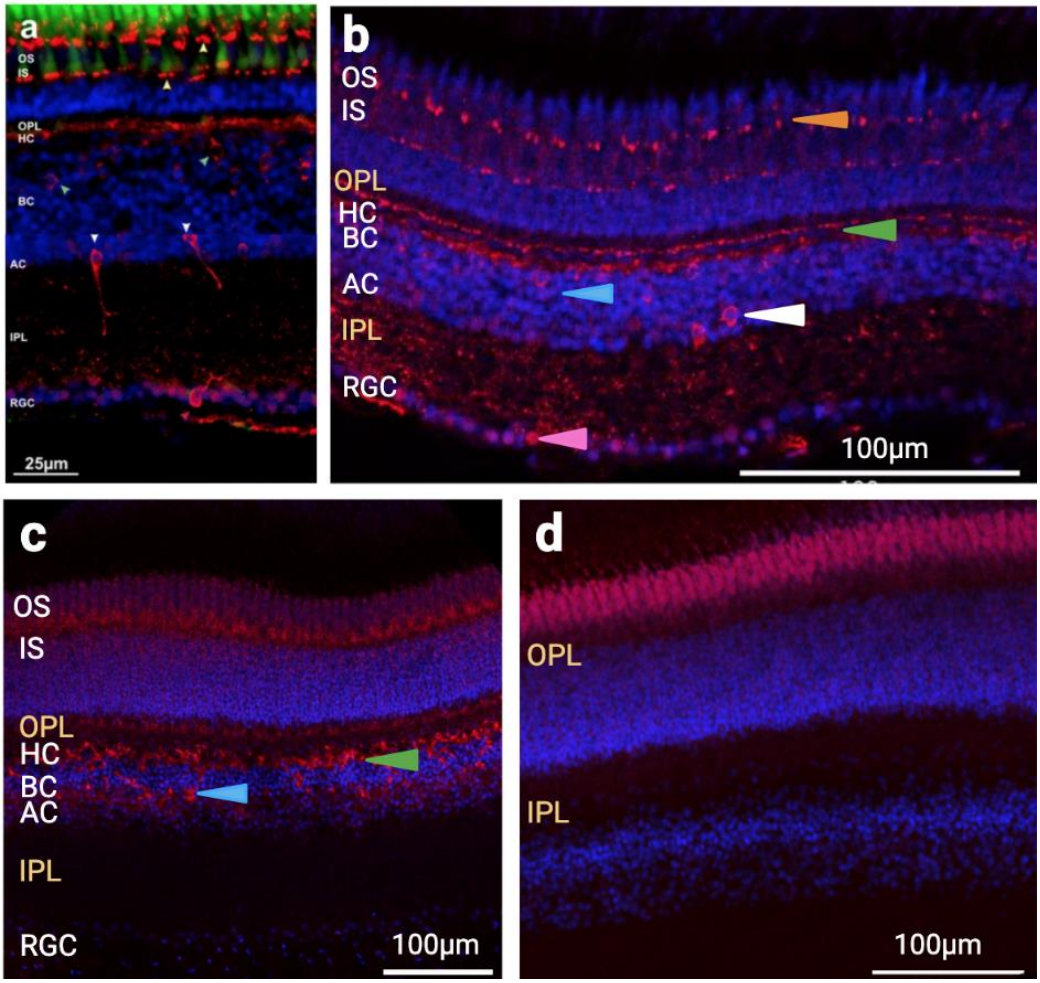


Figure 25. Immunofluorescent labeling of sablefish and zebrafish retinas with the anti-Opn4m antibody, pas350 (red) and counterstaining with DAPI (blue). **A)** Image from Davies et al. (2011) of a zebrafish radial retina section showing labelling of rings at near the inner segments, horizontal cells (HC), bipolar cells (BC), amacrine cells (AC) and retinal ganglion cells (RGC). In this example, the green cells are simply auto-fluorescent photoreceptors under a green filter. Adapted and included with publisher's permission. **B)** New image of a zebrafish radial retina section to test the antibody (14 microns thick), showing labelling horizontal cells (HC), bipolar cells (BC), amacrine cells (AC) and retinal ganglion cells (RGC). **C)** Image of a sablefish radial retina section (14 microns thick) showing Opn4m labeling in horizontal cells (HC) and bipolar cells (BC). **D)** Sablefish retina primary negative control showing that the signal near the photoreceptors here is not entirely eliminated, though it does appear different, potentially because of the plane of sectioning. OPL: outer plexiform layer, IPL: inner plexiform layer. Orange arrowheads point to inner segment labeling, green arrowheads to horizontal cells, blue to bipolar cells, white to amacrine cells and pink to retinal ganglion cells.

In the sablefish brain, there are also clear pas350-positive cells showing Opn4m protein expression (Figure 26). While sequential coronal sections were collected across the whole brain, the relatively large size of the tissue promoted folding and lifting during the washing steps of the immunohistochemistry protocol, making it impossible to accurately visualize antibody labelling in most regions of the brain. One region that was consistently intact was the optic tectum. Limited by time, this was the only region thoroughly investigated, though it is very likely that there are melanopsin proteins expressed elsewhere in the brain. The Opn4m immunolabeling by the pas350 antibody spans nearly the entire optic tectum, from the innermost region of the periventricular grey zone to the stratum opticum/marginale, with projections connecting the two. In the basal layer, some of the pas350 labeling overlaps with the DAPI nuclear counterstain, while other labeling does not appear to be associated with nuclei. Negative controls confirm that these signals are attributed to binding of the primary antibody (Figure 27).

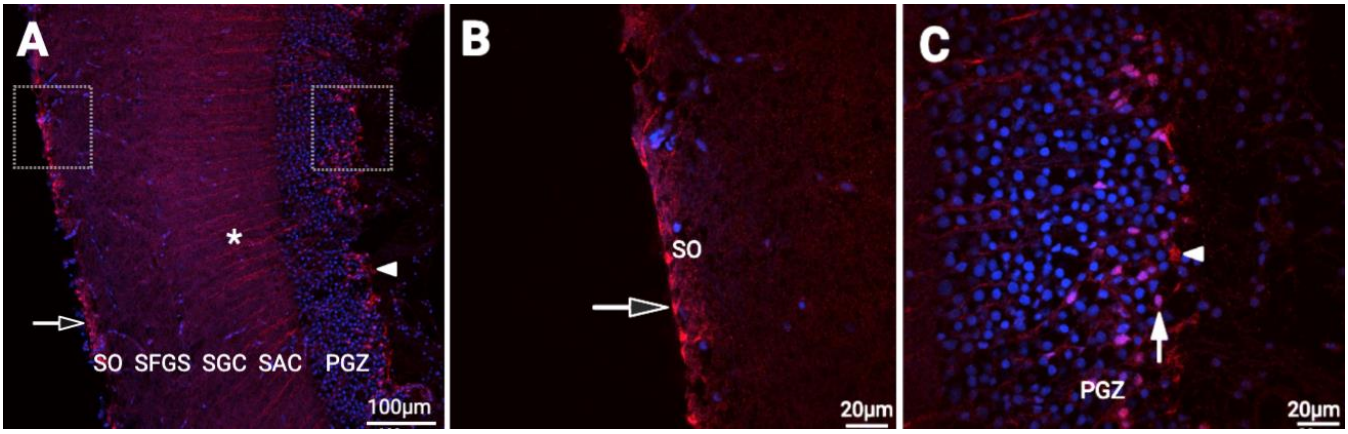


Figure 26. Z-stack images of anti-pas350 (Opn4m melanopsins) immunoreactivity in the sablefish optic tectum (20 microns thick). The anti-Opn4m labelling (pas350) is red, with DAPI as a blue nuclear counterstain. **A)** Image showing the apico-basal polarity of the pas350 labelling **B)** End-foot –like labelling in the superficial marginal zone, forming a distinct tile-like pattern independent of nuclei. **C)** Labelling of cell bodies in the deepest layer of the optic tectum, the periventricular grey zone (PGZ). White arrowheads point to immunoreactivity that is not associated with a DAPI-stained nucleus in the PGZ and white arrows point to immunoreactivity that overlaps with the DAPI staining. The asterisk shows the process extending from the basal layer to the apical region. The unfilled arrows show end-foot like processes. Layers of the optic tectum are labelled as follows: stratum opticum (SO), stratum fibrosum griseum superficial (SFGS), stratum grisum centrale (SGC), stratum album centrale (SAC) and periventricular grey zone (PGZ).

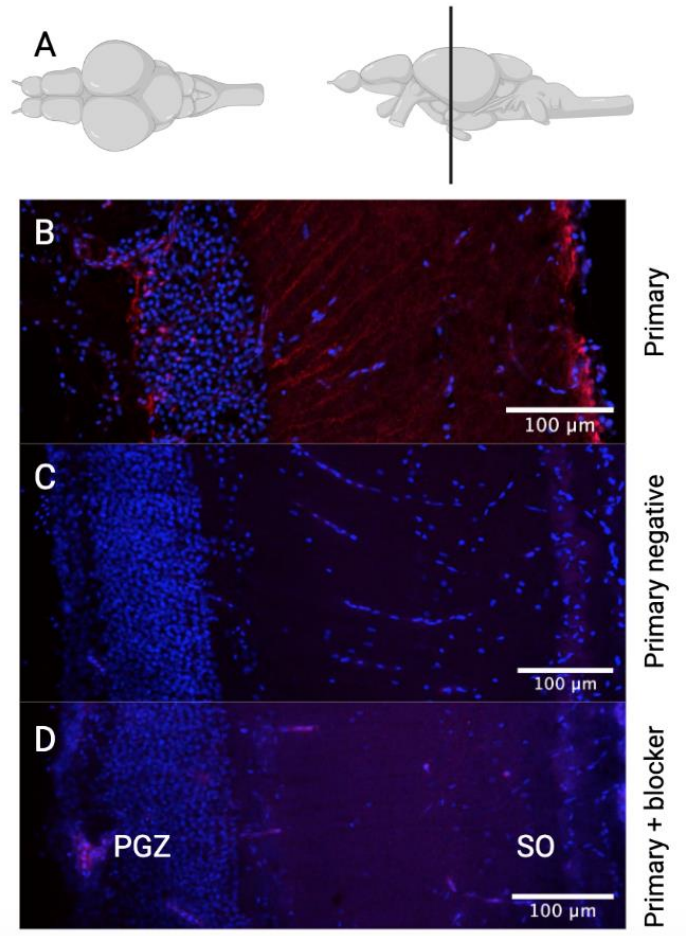


Figure 27. Images of the sablefish optic tectum (20 microns thick) under different immunohistochemical conditions to confirm that the signal is from binding of the primary antibody (pas350). Labelling from pas350 is red and the DAPI counter stain is blu blue. These images are shown in the reverse orientation compared to Figure 26 (from the other side of the brain). **A)** The approximate region of the brain that was sectioned for this experiment. **B)** A section with both pas350 primary antibody and donkey anti-rabbit Alexa555 secondary antibody applied, showing clear labeling in the basal periventricular grey zone, with projections leading to the apical region. **C)** A section with only the donkey anti-rabbit secondary antibody applied, showing only a very faint red signal, likely to be auto-fluorescence. **D)** A section with an antigen-blocked primary pas350 antibody and donkey anti-rabbit secondary antibody applied, showing a faint signal, also likely to be auto-fluorescence. Layers of the optic tectum are labelled as follows: stratum opticum (SO), stratum fibrosum griseum superficial (SFGS), stratum grisum centrale (SGC), stratum album centrale (SAC) and periventricular grey zone (PGZ).

The pattern displayed by IHC in the optic tectum is typical of teleost radial glia cells, where there is a basal layer that projects and terminates in ‘end-feet’ in the apical layer. While I attempted to confirm the cell type, the rat GFAP primary antibody was unsuccessful at staining glia (or anything) in sablefish so melanopsin localization to glial cells has yet to be experimentally confirmed.

DISCUSSION

Zebrafish opsins were first characterized in the early 90's, marking the species' debut as a model in the field. The initial studies (Robinson et al., 1993; Vighteichmann, Ma & Vigh, 1992), and those that followed, focused on visual opsins. The 10-gene opsin repertoire (all visual opsins initially) (Chinen et al., 2003) grew slowly as homologs of vertebrate non-visual opsins, such as melanopsin (Opn4; Davies et al., 2011), teleost multi tissue opsin (TMT; Fischer et al., 2013) and vertebrate ancient opsins (VA; Soni & Foster, 1997) were uncovered in zebrafish. This was further expanded to a great extent when Davies et al. (2015) took advantage of the availability of a zebrafish genome (Howe et al., 2013) to identify homologs from several other opsin families, as well as ten new opsins which they called novopsins (Opn6, Opn7, Opn8 and Opn9 genes). These non-visual opsins were studied using a quantitative barcode assay, which measures mRNA molecules directly without amplification, and in-situ hybridization to measure and localize mRNA transcripts. While the eye showed high levels of expression for most opsins, the brain showed a striking amount of opsin expression as well. Indeed, nearly every opsin characterized was shown to be expressed to some degree in the brain; by far the most out of the extra-retinal tissues surveyed. Their interpretation of these observations was summarized by the statement, "the zebrafish brain is likely to be a particularly important photoreceptive organ, more than previously appreciated and perhaps only second to the eye concerning the degree of light sensitivity".

Here, I describe the opsin expression domains of another teleost, a distant relative of zebrafish and a species that inhabits a very different photic environment, the sablefish. As reported in Chapter 1, sablefish have 36 opsins, 29 of which are non-visual. Following up on the speculation by Davies et al. (2015) that the brain is a centre of light sensitivity, I set out to study opsin expression in sablefish brains. I focused especially on melanopsins (Opn4 genes), as they are known deep brain photoreceptors in other vertebrates and they are an interesting family of five divergent gene duplicates.

qPCR – temporal expression of Opn4's

i) *Opn4 paralogs*

The five sablefish Opn4/melanopsin genes (*Opn4m1*, *Opn4m2*, *Opn4m3*, *Opn4x1* and *Opn4x2*) were detected in the brain using qPCR, though expression levels differed among paralogs and among life stages. Opn4's in the sablefish brain are expressed at low levels relative to the calibrator housekeeping gene, *Eef1a.* This is not surprising, as this gene encodes a Eukaryotic translation elongation factor and is required by all cells. It is known to be expressed at high levels; in zebrafish, it is considered one of the most highly expressed housekeeping genes (McCurley & Callard, 2008; Idigo et al., 2020). Among qPCR studies, it is rare that a target gene is expressed at higher levels than the normalizing housekeeping genes due to many genes being expressed only in a subset of cells rather than every cell type. Thus, while the expression of melanopsin genes here is low, it is likely still representative of significant transcription levels.

While all melanopsin expression levels in the brain are fairly low, some paralogs are expressed at higher levels than others. One trend that is consistent among the different life stages studied is the higher expression of the three Opn4m class of opsin genes compared to the Opn4x class. *Opn4x* and *Opn4m* genes are paralogs, generated by an ancient duplication event that occurred in the ancestor of vertebrates (see Chapter 1). Even the duplicates within these two lineages are more than 300 million years old, produced by a retrotransposition event and subsequent teleost whole genome duplication (Beaudry et al., 2017).

A survey of transcription factor binding sites upstream of the Opn4 paralogs in zebrafish showed that there are many more binding sites associated with circadian regulation in the promoter region of Opn4m-class opsins (Davies et al., 2011). At the sequence level, sablefish Opn4m's and Opn4x's differ from one another at many sites, but within either class, the genes share a high degree of amino acid similarity. Thus, there is likely a functional division between the two Opn4 classes, as a result of the duplication events that led to the paralogs, that could explain Opn4m's being more highly expressed in the brain.

The finding that Opn4m genes are expressed at higher levels than the Opn4x genes is not as defined in the zebrafish brain (quantified only at the adult stage), where *Opn4x1* is as highly expressed as the Opn4m genes (Davies et al., 2015). In mudskipper, which lack an *Opn4m2* gene, *Opn4m1* is expressed at much higher levels than the Opn4x's across different brain regions, but *Opn4m3* has low expression (Ma et al., 2022), again showing different expression patterns among species.

Of the Opn4 proteins, those encoded by *Opn4m1* and *Opn4m3* are bistable in fish (Davies et al., 2011), meaning that they can regenerate (re-isomerize) their own chromophore after activation by light. Thus, the data generally indicate that bistable Opn4 paralogs have higher levels of expression in the brain. Bistable melanopsins have been found to have slower kinetics than other monostable opsins (i.e., more sustained responses to light) (Dacey et al., 2005), which may be useful in the brain; opsins expressed there would not have roles in image forming vision, which requires immediate and fast signaling. That said, the monostable *Opn4m2* is also expressed at levels similar to *Opn4m1* and *Opn4m3*, so the distinction appears to relate more to the Opn4m and Opn4x evolutionary lineages than it does to chromophore valency. Nonetheless, bistability may still be preferred in the brain, for the reasons mentioned above and because they have no requirement for separate photoisomerase opsins to be expressed nearby to regenerate the chromophore. As of right now, there is no defined brain centre that expresses opsins with roles in photoisomerization as there is in the retinal pigment epithelium, though some studies have found high expression levels of photoisomerase opsins (such as *Rgr1*) in the brain (Davies et al., 2015).

Opn4m2, the highly expressed gene in the Opn4m class encoding a monostable opsin, is interesting in its own right; it is intronless and a likely product of retroduplication of the Opn4m progenitor. Retroduplicates are often non-functional due to spliced out upstream regulatory regions (Casola & Betran, 2017), but they can occasionally persist with necessary genetic elements. A well-known retroduplicate in teleosts is *Rh1* (encoding the protein rhodopsin, a dim-light visual opsin in rod cells in the retina); it maintained the ability to express in the retina over its intron-containing progenitor, possibly because transcription was more efficient without the need for splicing (Bellingham et al., 2003). By the same logic, Opn4m2 in sablefish may have an advantage over the other Opn4x monostable melanopsins, leading to its higher expression.

ii) A perspective on Opn4 expression levels in the brain

To gain perspective on the general low expression levels the Opn4 genes in the sablefish brain (relative to *Eef1a*), Opn4 transcripts were also quantified in juvenile sablefish eyes, along with two visual opsins, for comparison. Eyes are highly specialized light-sensitive organs and Opn4's are known to play important roles in the intrinsically-photosensitive retinal ganglion cells (ipRGCs), including circadian rhythm photoentrainment (via clock genes such as *Per2*; Dkhissi-Benyahya et al., 2013) and pupillary constriction (Hankins et al., 2008). It is also widely

accepted that visual opsins are expressed at high levels in the eye, as they are required by vertebrates to see; as such, comparing the expression of visual and non-visual opsin levels between the brain and eye helps to orient the Opn4 brain data.

Between whole juvenile sablefish eyes and brains, the expression of *Rh1* (rhodopsin) was strikingly different, amplifying, on average, 17 cycles sooner using eye cDNA as a template compared to brain cDNA. There is little evidence of expression of the visual opsins in the brain for any vertebrates (Davies et al., 2015), so this difference is not surprising. Compared to the other genes quantified in the eye, including Opn4's, *Rh1* expression is also exponentially higher; it is even more highly expressed than the *Eef1a* calibrator in the eye. This shows just how highly these genes can be expressed in regions that depend on their protein's function. Interestingly, the expression levels of the *Sws1* gene (encoding another visual opsin, localized to cone cells) in the juvenile eye were comparable to, and sometimes lower than, Opn4 gene expression. This may hint at rod-dominated vision in this species or coincide with recent findings suggesting that *Sws1* plays more of a developmental role in fish eyes than it does in vision (Flamarique et al., 2021).

The expression of the Opn4 paralogs does not differ much between juvenile sablefish brain and eye tissue (only *Opn4x2* and *Opn4m3* are expressed at higher levels in the eye). In adult zebrafish, expression of all melanopsin paralogs is much higher in the eye than in the brain (Davies, 2015). It is difficult to make precise comparisons between the sablefish findings here and those already published for different species, as conditions and life stages vary depending on the question being asked in the work; however, it is very clear that, while opsin repertoires may be fairly well-conserved across fish species, expression profiles of each gene can differ immensely. Assuming melanopsins are functional in the sablefish eye, as they are in many other fish species, then finding similar expression levels among brain and eye tissue supports my assertion that the range of Opn4 expression observed here in the brain (though low relative to *Eef1a*) is still biologically relevant.

iii) Expression levels among life stages

To begin to unfold the function of opsins in the brain of sablefish, I explored the hypothesis that the Opn4 genes may be expressed at higher levels during the life stages that are more readily exposed to the light. While melanopsin gene regulation in the brain is poorly understood, I worked under the assumption that gene expression would be up-regulated in the

light, when opsin proteins would have more stimuli to respond to, and therefore require more receptors. The studied life stages occupy diverse photic environments (both in the wild and in aquaculture); for example, yolk-sac dependent larvae and mature adults live up to thousands of meters deep in the ocean, whereas free-feeding larvae and juveniles are near the surface and in-shore. Thus, sablefish provide the opportunity to survey non-visual opsin expression plasticity in a single species, rather than comparing two species living in different environments.

When comparing expression of the melanopsin paralogs among life stages, two of the opsin genes roughly followed the anticipated trend of increased expression during the light-exposed stages. That is, sablefish *Opn4m2* and *Opn4x1* both show qualitative trends of increasing expression through development and with exposure to light. Intriguingly, though *Opn4m*2 is expressed at higher levels than *Opn4x1*, plots of their mean expression levels among age groups are roughly parallel, suggesting their expression may be influenced by the same factors. Though quantitatively (i.e., supported by statistical significance) specific trends are not as defined, there is an association of expression levels with light exposure for *Opn4m2* and *Opn4x1* in that both genes are expressed at significantly lower levels in yolk-sac larvae living in absolute darkness compared to the later stages that were either sampled directly from well-lit environments or had once occupied a well-lit environment.

This finding may also indicate that yolk-sac larvae and wild adult sablefish have different light exposures, contrary to what was proposed going into this set of experiments. Though the zone that the wild adult fish inhabit is certainly dark (hundreds to thousands of meters deep), the fish occasionally migrate vertically in the water column. Thus, there seems to be a distinction between a fish that has never been exposed to light and one that lives in the dark with periodic exposure to low levels of light. It is important to consider the caveats of the wild adult and yolk-sac larval samples used in this experiment when making these comparisons. Wild adults (by necessity) were fished from the deep, traversing the into light-abundant ocean zones before being flash frozen. Yolk-sac larvae were the only stage from which brains should not be dissected, so they were simply enucleated. These factors likely contribute to the observations here, but it is not clear to what extent.

Though this series of experiments does not show a perfect correlation with light exposure, it shows with certainty that *Opn4m2* and *Opn4x1* expression levels differ across life stages (which are associated with different light environments) in similar patterns. With this

experimental set up, age and light exposure cannot be separated as two independent variables; as this work is largely explorative, future experiments should be modeled to more carefully observe and quantify the correlations of Opn4 expression with these variables (i.e., light manipulation experiments and observing the effects of age under steady light environments). There is plenty of evidence that opsin expression levels can be regulated by hormones during ontogeny (Flamarique et al., 2019; Veldhoen et al., 2006), thus it is currently unclear whether the main driver of the observed differences in *Opn4m2 and Opn4x1* expression is the external environment, internal developmental cues, or both.

Most opsin developmental time series follow only very early stages, especially in fish (Takechi & Kawamura, 2005; Kojima et al., 2007). Sablefish, having a long lifespan and a prolonged period of embryonic development compared to zebrafish, are a good model to study over longer periods of time and they may provide much better resolution for the timing of opsin expression during development. As shown here, *Opn4x1* and *Opn4m2* may serve as a good ontogenetic candidate in future studies, even though, for the time being, it is very difficult to know whether the aforementioned trends among life stages (even if statistically significant) are biologically relevant.

While expression levels of *Opn4m2* and *Opn4x1* vary across life stages, the levels of *Opn4m1* and *Opn4m3*, encoding the bistable melanopsins, do not show detectable differences statistically. These do not seem to be up- or down- regulated in response to the light environment. Again, these life stages correspond to highly varied light environments, both in the wild and in the aquaculture, ranging from total darkness to 24-hour light. Thus, it is surprising that expression levels for these opsins are not highly variable across them.

Despite the statistical outcome, it is worth emphasizing that *Opn4m1* and *Opn4m3* expression levels do not qualitatively appear entirely stable when plotted. Mean *Opn4m1* expression across life stages follows the same general trend of *Opn4m2* and *Opn4x1*, in that it shows a moderate increase with age and light exposure. *Opn4m3* shows an expression pattern unlike those of the other paralogs, as it is qualitatively inversely correlated with age. With a fairly small sample size, it is difficult to speculate whether the statistics or the trend more accurately describe the temporal gene expression patterns of these genes, though typically, more confidence is placed in the former. Future work should increase the number of biological replicates to see how this impact the results.

There may be several plausible explanations for the relative stability of *Opn4m1* and *Opn4m3* genes observed here. First, melanopsin regulation (at the gene expression level) is fairly complex and poorly understood. In mammalian models, mRNA levels can be regulated by both light and darkness; rat Opn4 mRNA levels were demonstrated to be higher in the dark than during light exposure (Hannibal, 2005). In zebrafish, Opn4 gene expression in the eye maintains a diurnal rhythm, but transcription of the different paralogs peaks and dips in varying patterns and at different times relative to the onset of light exposure (Matos-Cruz et al., 2011). Such experiments survey short-term (e.g., 24 hour circadian), light-induced opsin gene expression changes rather than long-term changes, as I have done with sablefish at varying life stages. While some of my hatchery-reared life stages are in constant light environments (e.g., yolk-sac dependent larvae in complete darkness and yolk-sac depleted larvae in 24 hour light), it is possible that Opn4 expression does oscillate in a circadian rhythm, particularly for the juveniles and sea pen adults under light-dark cycles. Despite my best efforts, I may have sampled at times when expression of a particular gene was peaking for one life stage and dipping low in another, which could create a false pattern of stability. The exact magnitude of diurnal opsin expression changes is not well documented, so future work should explore this potential limitation.

Another explanation for the lack of detectable changes in expression across life stages for two of the opsin paralogs is that their expression levels may not be regulated by light. While it is intuitive to assume this for genes encoding light-sensitive proteins, this prediction may not be accurate for all opsins. Light-sensors are inherently also dark-sensors (a lack of cellular response), so some genes may need to be expressed stably regardless of the precise photic environment to monitor light information and send output to other regions of the brain. The proteins encoded by the comparatively stable *Opn4m1* and *Opn4m3* genes, for example, may act as constant light level monitors, isomerizing and re-isomerizing the chromophore in response to different wavelengths of light, though this is purely speculative.

This ties into an interesting theme in opsin research; though the proteins are well-studied as light sensors, gene expression levels do not always reflect this function well. In fact, through my experiments, I uncovered similar interesting findings beyond my initial hypothesis. Namely, Opn4 genes are expressed in yolk-sac-dependent larvae that have never been exposed to light. Presumably, without ever having a light cycle to adjust to, this is not regulated by circadian rhythms, but possibly basal transcription of the genes in ‘anticipation’ of light. That said, the

expression levels of some Opn4 genes in yolk-sac larvae are comparable to (or statistically the same as) those of other light-exposed stages.

In recent years, light-independent roles of opsins have also been uncovered, such as in chemo- and thermo-sensation, or as phospholipid flippases (transporting lipids across the membrane (Menon et al., 2011). It is therefore conceivable that some melanopsin genes with these seemingly unconventional roles would not require light to mediate their upregulation. In the case of sablefish, it seems possible that opsins would either not be used, or co-opted for other roles during the decades that adults spend in the deep. While my main hypothesis was related to photic zones in the ocean, at different ranges, other stimuli such as temperature and chemicals (e.g., oxygen levels), are also immensely variable and could influence gene expression and protein function as well. Nonetheless, it is likely that sablefish Opn4's do still maintain the ability to be stimulated by light, as the proteins encoded by orthologous genes have been shown experimentally to be light-sensitive (Fischer et al., 2013; Torii et al., 2011; Matos-Cruz et al., 2011), and they are similar in sequence at many of the important chromophore and G-protein binding sites (see Chapter 1).

The qPCR expression results across life stages and light environments are intriguing, and the implication of these findings it worth exploring further. As this study used whole brain samples as a starting point for sablefish opsin research, subsequent experiments should aim to dissect regions of the brain prior to RNA isolation. Limited by the maximum tissue weight of the extraction kit used, whole sea pen and wild adult brains were ground up and mixed thoroughly, using only a portion of the homogenate in subsequent steps, in order to get a sample representative of the entire brain (younger life stage brains were small enough to be lysed whole). Despite this effort, it is possible that the RNA sampled was not perfectly representative of the whole brain and may have represented certain regions more than others. Dissecting brains from all life stages into fore-, mid- and hindbrain prior to RNA extraction may reduce the size of the tissue enough so that no grinding and mixing is required in the older stages. Additionally, monitoring expression changes in dissected brain regions may also uncover region-specific changes from stage to stage that may have been masked in the whole brain survey. In the case of the stable *Opn4m1* and *Opn4m3* genes, they may be up- and down-regulated throughout life stages in sub-regions of the brain, but not in the brain as a whole (Figure 28). Even with these potential limitations, the whole brain results reported here are interesting and suggest that some

Opn4 genes are temporally expressed, in relation to light exposure, while others may not be. The strong trends, such as much higher expression of the Opn4m's in the brain, provide support for differential expression of the duplicated opsin gene family and can be used to begin to divulge functional information.

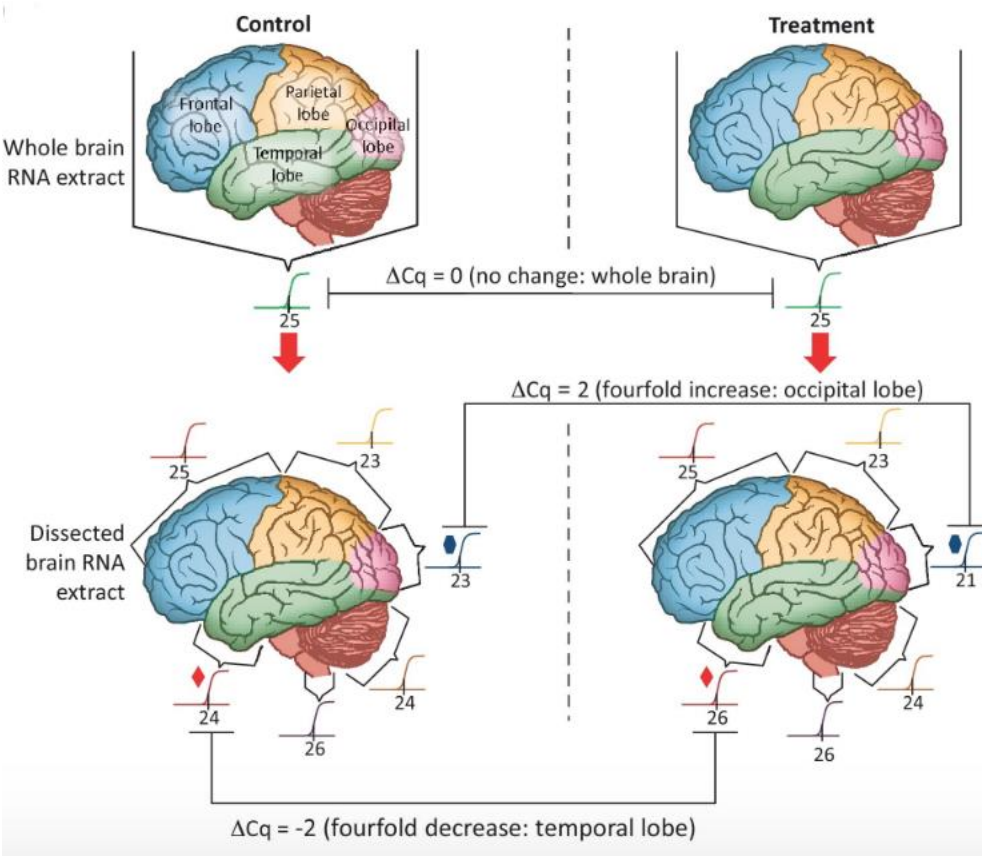


Figure 28. Diagram from Taylor et al. (2019) showing the influence of tissue selection on qPCR results. It highlights the fact that while expression may look stable when comparing whole brains across different treatments, there may be significant changes in sub-regions that are being masked. Included with publisher's permission.

Immunohistochemistry – a leap toward uncovering function

In addition to the temporal Opn4 gene expression survey, proteins encoded by a sub-set of Opn4 genes (Opn4m's) were localized in the brain using immunohistochemistry. The motivation for this addition was to carry this project from the novel characterization of opsins in sablefish through to the final step of protein synthesis. This serves as an important supplement to the qPCR data, as gene expression levels are not always a perfect proxy for protein levels, and because the Opn4 genes were expressed at low levels in the brain, leading to the question of

whether those levels correspond to functional products and if the low levels relative to the housekeeping gene calibrator indicate localized expression to sub-regions of the brain (as seen in some fish ISH experiments; Fernandes, 2012; Sandbakken et al., 2012).

Beyond the motivation mentioned above, the antibody-based IHC method used here was chosen over ISH due to its relative ease, with a suitable antibody already having been made. While the immuno-probe is not specific to one Opn4 paralog, the pas350 antibody is likely to react with Opn4m1, Opn4m2 and Opn4m3 as it does in zebrafish (Davies et al., 2011), based on the sequence alignment of the epitope to the sablefish melanopsin sequences.

The first sablefish IHC experiment was done on juvenile retinal tissue, mostly to confirm that the antibody works similarly to how it does in zebrafish, though the results are interesting in their own right. The pattern of Opn4m labeling in the retina is very similar to zebrafish (Davies et al., 2011); different cell subtypes in the inner nuclear layer showed some immunoreactivity to the pas350 antibody, namely, bipolar and horizontal cells. This confirms that the relatively low expression levels of Opn4's in the juvenile eye from qPCR connect to real protein expression and function. It also shows that while ciliary opsins are important in rod and cone cells, rhabdomeric opsins are also important and widespread in the eye. In fact, zebrafish photoreceptors show peculiar ring-like labeling around the cone cells in my study and in the original Davies et al. (2011) study. Of note, Davies et al. wrote little about this and were happy to hear that their unexpected observation of Opn4 in rings around the base of photoreceptors was confirmed ten years later. This pattern was not observed in sablefish, however, perhaps as a result of it being masked by high levels of autofluorescence in the photoreceptors layers in all samples, including negative controls. This stretches the fine line between the traditional classification of 'visual' and 'non-visual' opsins even thinner.

In the sablefish brain, Opn4m immuno-localization experiments were limited to one region due to time constraints, the optic tectum. The reactivity in this region forms a very distinct pattern; the morphology is bipolar, with perinuclear labeling (as well as some not associated with nuclei) in the periventricular grey zone (PGZ), and basal processes that lead to the surface, forming club or tile-like structures in the marginal stratum opticum (SO) layer, which are not localized to nuclei. IHC protein labeling in this region is new for melanopsins, though recently, mRNA of all 5 zebrafish Opn4 paralogs was localized to the PGZ and SO layers of the optic tectum with ISH, among other regions (Dekens et al., 2022). While it is not possible to

distinguish patterns of the particular Opn4m paralogs in this sablefish experiment, the aforementioned new zebrafish study shows the co-expression of each of them in this region, as well as the Opn4x's. Other opsins have also been associated with this pattern in the optic tectum; mRNA of all ten of the 'novopsins', the more recently discovered group of opsins that fall into the OPN5 family, was also localized to the PGZ in zebrafish (Davies et al., 2015), as well as TMT family opsins in medaka (Fisher et al., 2013). The IHC pattern here in sablefish supports my assertion that while the expression of the Opn4's is low in the whole brain, it represents localized expression and corresponds to the translation of functional proteins.

The IHC experiment conducted on the sablefish brain has an unexpected added benefit that previous ISH opsin experiments do not. Opsin transcripts, including Opn4 transcripts in zebrafish, have consistently been localized to either just the PGZ, or both the PGZ and the SO for a particular gene. That is, the labeling tends to look more like two different cell types. Using an antibody to that reacts with opsin proteins, I have also labeled processes connecting the two regions, suggesting that this is a single cell type. This may be partially due to the fact that opsin proteins are transmembrane and can thus occupy membranous space anywhere along the cell body, or because mRNA can be transported between apical and basal regions through the long cell processes in a way that can potentially hide it from an in-situ probe (Ehses et al., 2016).

Fish optic tectum cell types are not well characterized; labeling to discern neuronal cell types has only been done in four species: goldfish, trout, zebrafish and electric fish (Meek, 1978; Kinoshita et al., 2006; Nevin et al, 2010; Heligenberg & Rose, 1987). Among reported cell types in these species, some neurons were shown to have somas in the PGZ with dendritic processes extending to the marginal layers (including the SO), known as periventricular efferent neurons. However, these cells show complex dendritic branching, which is notably absent in the sablefish IHC labeling. In the more recent zebrafish paper, the authors characterize radial glia, which have a structure that more closely resembles the sablefish Opn4m-positive cells; a soma in the PGZ with an unbranched inner process leading to end-feet in the SO (Nevin et al., 2010, Figure 29A). The immunoreactivity in the sablefish brain also closely resembles the pattern observed in zebrafish with fluorescent GFAP (glial fibrillary acidic protein, a common glial cell type marker) (Than-Trong & Bally-Cuif (2015), Figure 29B). Radial glia cells have been identified as a particularly important cell type, acting as progenitors of neurons on top of their long understood structural and supportive roles (Jurisch-Yaksi et al., 2020). Non-visual opsins (Opn3 and Opn5)

have been identified in Mueller cells, a type of radial glia, in the chicken retina (Rios et al., 2019). My attempt to co-localize GFAP and Opn4m proteins in the sablefish optic tectum through IHC was unsuccessful, but as I was using an antibody created for rat GFAP, this is not entirely surprising. It is worth further pursuing GFAP IHC (using a more suitable antibody, such as one made for zebrafish or a custom sablefish antibody), or to use a neuronal counterstain to confirm the cell type. There are also additional markers that have been successfully used in zebrafish that may work in the place of GFAP (summarized in Figure 30). Than-Trong & Bally-Cuif (2015) note that tectal retinal glia cells also express brain lipid binding protein (BLBP), protein S100β, and Aromatase B, making them all good candidates for sablefish optic tectum glial cell labeling in the future.

As the discovery of opsins in radial glia would be novel, it is difficult to hypothesize what roles they might play there. One possibility is light-driven neurogenesis. Based on my observations, sablefish brains develop rapidly early in development, but plateau in growth and remain comparatively small when the fish reach the adult stage. As neuronal progenitors, radial glia may contribute to brain growth in a light-driven way. Alternatively, just as certain visual opsins (such as Sws1) can act as structural/developmental landmarks in the eye (for example, to guide the patterning in retinal mosaics; Flamarique et al., 2021), the brain may also be structured by opsin receptors.

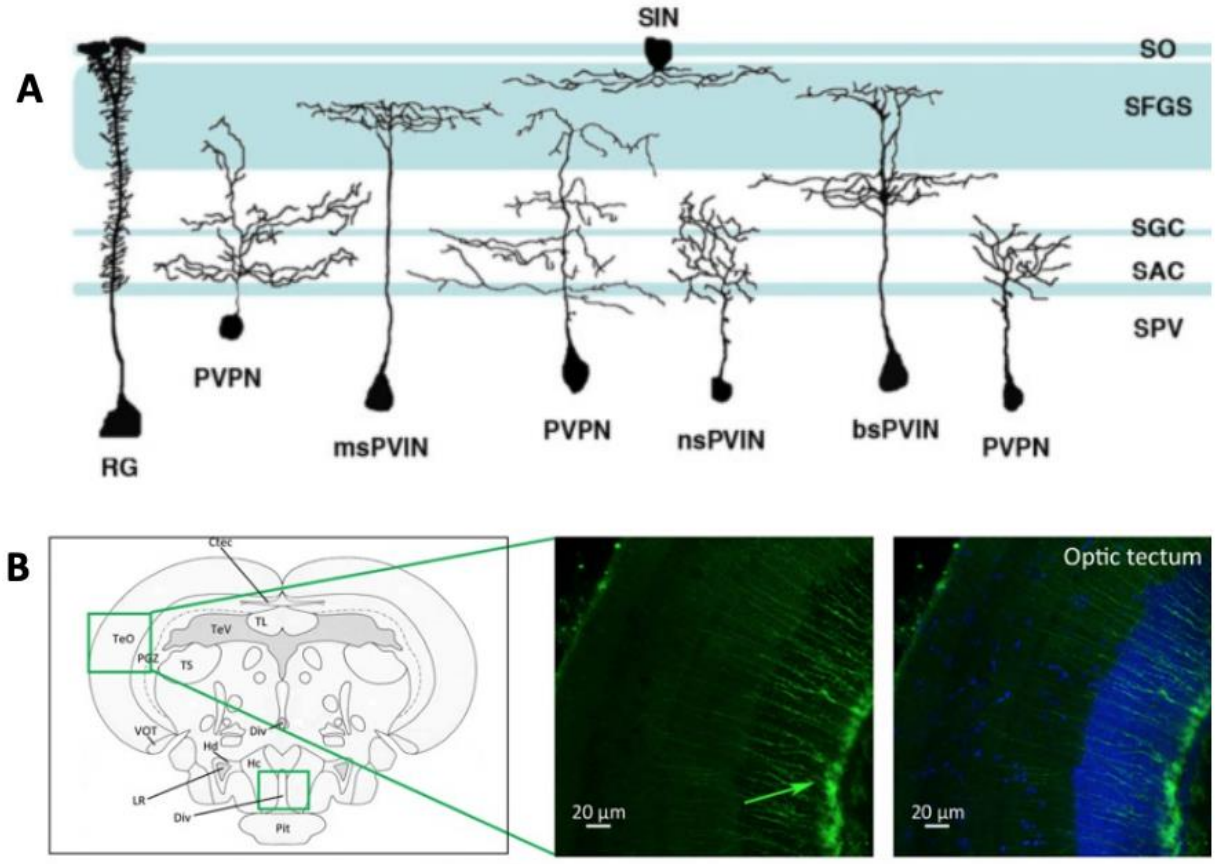


Figure 29. Radial glia cell morphology and localization in zebrafish. **A)** Image from Nevin et al. (2010) showing the optic tectum layers in larval zebrafish and different types of interneurons – radial glia (RG), periventricular projection neurons (PVPN), periventricular interneurons (PVIN) and superficial interneurons (SIN). SPV is another acronym for the periventricular grey zone. **B)** Image from Than-Trong & Bally-Cuif (2015) showing radial glia in a transgenic GFAP:egfp (green) zebrafish brain and DAPI counterstain (blue). These images appear to be consistent with anti-Opn4m immunolabeling. Both images included with publishers' permissions.

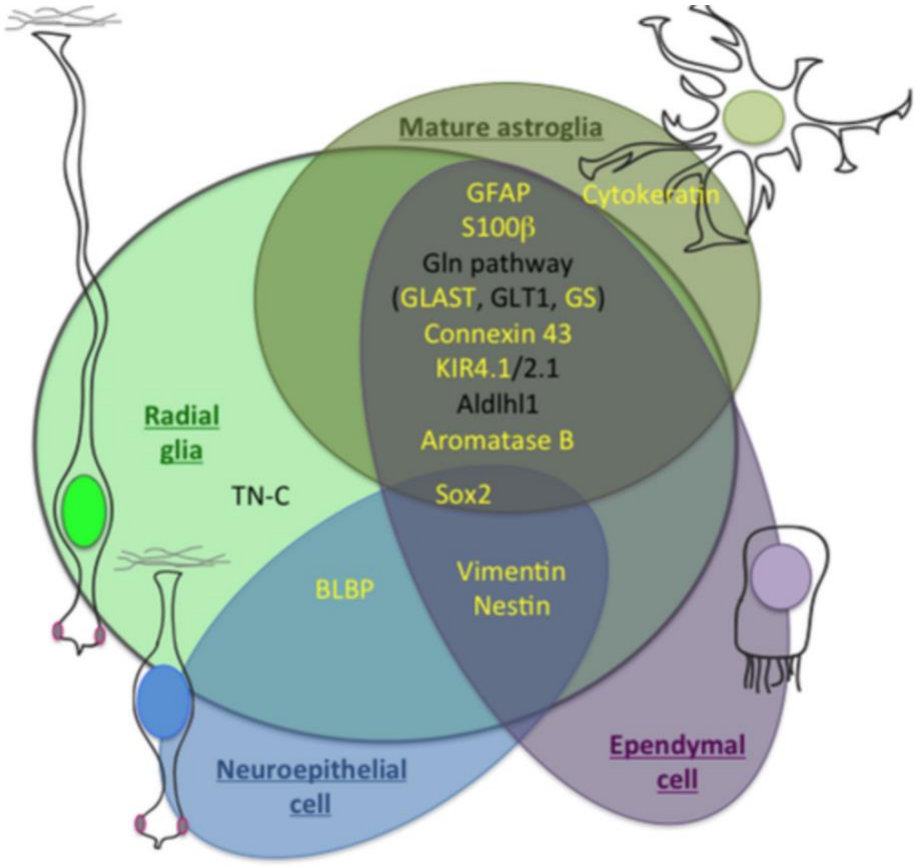


Figure 30. Diagram from Than-Trong & Bally-Cuif (2015) showing zebrafish glial markers. These may be good candidates for glial immunohistochemistry experiments in sablefish. Overlap shows that multiple glial cell types can be labeled by a single antibody. The paper argues that sometimes these genes are only expressed in radial glia at certain stages of development and in glia of particular brain regions. Included with publisher's permission.

Regardless of the cell type labeled here, the PGZ and SO layers of the optic tectum seem to be important opsin-expressing zones, though the functional implication is not yet known. Teleost brains are sensitive to light; multiple light-responsive clock genes and *c-fos*, a common marker for neuronal activity, oscillate in expression under various light regimes (Moore and Whitmore, 2014). mRNA of these genes has been localized to several regions in the brain including the PGZ of the optic tectum. While sablefish brains are hidden under layers of skin, fluid and skull, they are still fairly permissible to light. In early larval stages, the fish are transparent. As they age, they develop more tissue around the brain and become pigmented, but the brains are not far below the surface. Additionally, they have what is a referred to as a 'pineal window'; a non-pigmented region of skin atop of the pineal extension of the brain as do other

fish species (Strivastava et al., 2003; Nordtug et al., 1990). Thus, the localization of Opn4m proteins to projections in the superficial layer of the optic tectum (which resemble glial end-feet) might be expected from a cell responsible for measuring light levels. This spatial optimization of opsins may be particularly important in environments where light levels are weak, for example only from bioluminescence in the deep. It is worth noting that the IHC experiments were conducted on one of the light-dwelling stages; it will be intriguing to see if the same pattern exists in adult fish from the deep.

The SO is one of the tectal layers to which retinal ganglion cells (making up the optic nerve) innervate in fish. One can therefore speculate that opsin-expressing cells in the SO may receive, integrate and regulate signals from retinal ganglion cells, while independently responding to light stimuli themselves. Retinal ganglion cells themselves accumulate light information from other retinal cell types, as well as from their own intrinsic photosensors, to send to the brain. Opsins in located in retinorecipient regions of the brain (if light-sensitive) may provide additional information to this cumulative signal. Glial cells (if that ends up being the confirmed Opn4m-positive cell type in the future) are not able to fire action potentials like neurons can, but they can be excited and receive signals from neurons and influence light-evoked neuronal activity (Newman, 2004).

With this chapter laying the groundwork, there are many avenues that can be taken to uncover more information about melanopsins in the sablefish brain. To test whether these proteins are acting as light-sensors or if they have light-independent roles, one could use an expression system (such as *Xenopus* oocytes) to look for light-stimulated cell currents, search for the presence of chromophores in brain cells, or use techniques like calcium imagining now that there are specific regions of the sablefish brain to target. As we move toward uncovering the ultimate function of melanopsin in these regions, techniques such as single-cell RNA-seq and proteomics may be worth exploring, as well as knock-outs, and subsequent expression profiling and behavioral tests, as has recently been done for zebrafish melanopsins (Dekens et al., 2022).

CONCLUSION

The findings in this chapter emphasize the potential of melanopsins as important sensory receptors in the sablefish brain and support the diversification of function of the teleost melanopsin duplicates; all are expressed to varying degrees and in varying patterns. The Opn4m

class of opsins genes are expressed at much higher levels in the brain and may prove to be of particular importance in this tissue. Among life stages, some Opn4 genes (*Opn4x1* and *Opn4m2*) were expressed in patterns roughly correlated to age and light exposure. Others (*Opn4m1* and *Opn4m3*, which encode bistable proteins) were stable. Sablefish, which spend most of their lives in the deep, have large opsin repertoires, express opsin genes in the dark and express some relatively stably across highly diverse photic environments; thus, we are forced to rethink our basic understanding of opsins as light-sensors. At least one melanopsin protein (more likely three – the Opn4m's) is synthesized and playing a functional role in the sablefish brain. The question remains whether these function as light-sensors because of their unexpected temporal expression patterns. That said, the localization to a region of the brain that receives light information from the eye and expresses many other opsins (evidenced in other fish species), and within a cell type that projects to the most superficial part of the brain may support this role; this can be tested in the future. Some of the limitations in this chapter should also be improved upon in the future as well; namely, isolating just the optic tectum for qPCR gene expression assays, extending the immunohistochemistry experiment to cover more brain regions (fore- and hindbrain) and life stages, and performing controlled experiments to test the effect of light and age on opsin expression throughout developmental stages.

OVERARCHING SUMMARY AND PERSONAL REFLECTION

As is the case with most research, the more I uncovered about opsins in sablefish, the more questions arose. From the gene sequences, through transcription and translation, many of my findings challenged me to rethink what I knew about opsins.

Opsin repertoires and sequences

Large opsin repertoires are common amongst teleost fish, even in species living in the dark for much of their lives, such as sablefish. The visual opsin repertoires are generally more variable than their non-visual counter parts (due to lineage-specific duplications and losses), despite both encoding light-sensing proteins. The stability of non-visual opsin repertoires therefore challenges the notion that the collection of opsin genes in a species' genome is suited to their occupied environment. Whether they are being lost at a very slow rate or functionally maintained across various photic environments, there is much to be discovered on the topic.

One point is clear: without a comparative genomics approach, uncovering the roles of non-visual opsins is very difficult. My work shows the value in comparing entire opsin repertoires and sequences amongst distantly related fish species. This expanded scope allowed me to identify well-conserved subfamily signatures and motifs that have evolved in different lineages of the opsin tree. Many of the residues I identified as specific to (or characteristic of) the different subfamilies were not yet discussed in the literature, apart from those belonging to the well-studied OPN1/visual opsin subfamily.

Opsin expression

Following the trend of zebrafish (Davies et al., 2015), opsin expression in sablefish is widespread. Using pooled brain and eye samples as RT-PCR templates, each opsin primer set designed thus far successfully amplified transcripts. Though lower expressed genes were not well represented in the available sablefish transcriptomes, there was some evidence of opsin expression among additional non-visual tissues, including ovaries, testes, kidney and liver.

The brain continues to prove its importance as a major opsin-expressing zone, second only to the eye. The family of five genes that I studied, the melanopsins, are expressed at moderate to low levels in the sablefish brain. I hypothesized that, for a deep-sea fish to make use

of 36 opsins, they are likely being expressed during the stages in their life cycle that occur in the light. I tested this hypothesis on the melanopsin family in the brain by quantifying their gene expression across five life stages that correspond to different photic environments.

Expression levels of two of the melanopsin genes (*Opn4m1* and *Opn4m3*) did not differ significantly across all life stages, another two (*Opn4x1* and *Opn4m2*) varied in similar patterns that were roughly correlated with age and light exposure, and one gene *(Opn4x2*) was expressed at very low and inconsistent levels in the brain at all stages. Though more controlled tests are required in the future to separate age and light exposure, these findings suggest that the genes making up this subfamily play different functional roles, and that melanopsin gene regulation may not be entirely guided by light exposure.

Studies that focus on the mechanisms of non-visual opsin regulation are very rare. Many show that expression levels oscillate (and are maintained with circadian rhythm) with light-dark cycling, but different opsins peak and dip with different patterns and timing (Matos-Cruz et al., 2011). The idea that guided my hypothesis (that expression would be highest in the light) thus needs to be re-evaluated. Could the disconnect between light exposure and the expression stability of some melanopsin genes suggest light-independent roles in the brain (for example, chemo- or thermo-sensory)? Are expression levels surprisingly high in wild caught, deep-sea sablefish to compensate for and to be able to detect very low levels of light? Figuring out what factors influence opsin gene expression is a big step toward uncovering the roles of this diverse gene family.

The vision-centric approach to opsin research has shaped our understanding of non-visual opsins, but not always for the better. Though much of the foundation of opsin sequence, structure, function and diversity was uncovered using visual opsins, we must remember that vertebrate visual opsins represent only one branch in the very ancient opsin gene tree and to expand our perspective to think of the big picture.

The expression of G-protein coupled receptors (GPCRs) outside of their ‘typical’ sensory systems (the system in which they were first discovered) is not uncommon. For example, taste receptors, odorant receptors and opsins were all found in the bladder of mice (Smith etf al., 2021). Just as there are non-visual opsins, there are also non-olfactory odorant receptors and non-gustatory taste receptors. As is the case with opsins, we are left wondering if these ‘ectopic’ GPCRs use the same ligands and stimuli as in their sensory organs and what their roles may be.

While it is clear that these receptor families have diversified (for example, my Chapter 1 sequence analysis shows sequence evolution among opsin subfamilies), there also remains a lot of conservation and similarity (e.g., conserved residues among all opsins or all Class A GPCRs). Thus, while we question the ways in which these genes have evolved to be differentially expressed, we should also consider that they may play conserved or basal roles as receptors. This is especially pertinent as the divisions between different GPCRs (chemo- and photo-sensing, visual and non-visual, etc.) are converging and becoming increasingly blurred.

Localization

There are very few studies that localize opsin proteins in fish brains; most studies use in situ hybridization as a method to identify regions of opsin gene expression (mRNA). As non-visual opsins are enigmatic due to their unknown function, it makes most sense to focus on and localize the functional units (proteins) instead.

While I first felt limited by my narrowed focus on one brain region, I believe I benefitted from this in the end. Most opsin papers never connect expression patterns to functional output; there are plenty of studies that show widespread expression of opsins in fish brains, but we rarely see it functionally explained. Part of the reason may be due to the overwhelming number of regions in which non-visual opsins are differentially expressed. For example, Dekens et al. (2022) recently identified melanopsin transcripts in many more regions of the zebrafish brain than previously known, which is difficult to connect to physiological function when taken as a whole. Though I also only got as far as speculating function in this thesis, my focus on the optic tectum has allowed me to make steps toward identifying the specific cell type containing melanopsin proteins (radial glia). This sets up my successors in the Taylor Lab with a system to investigate the light-sensitive glia hypothesis and study non-visual opsin function (e.g. through calcium imaging and knock-outs). I am confident that the optic tectum of fish, with so many opsins co-expressed in the same layers (Davies et al., 2015; Dekkens et al. 2022), will prove to be a very interesting centre of light-sensitivity or some other form of stimulus reception.

Sablefish as a model

Sablefish will continue to be a strong model in the field of opsin research. Much of what we know about how light influences opsin expression comes from light manipulation

experiments in one life stage of one species, or through the comparison of multiple species living in different photic zones. Sablefish represent a single species that naturally occupies different spectral environments throughout its lifetime in a way that is correlated with development. As this life history may be amenable to developmental regulation using light as a stimulus, the study of sablefish opsins is certainly worth pursuing further. The availability of sablefish in aquaculture at Golden Eagle Sablefish is invaluable to this work.

Beyond that, many of the opsin genes studied in this thesis (except for *Opn4x2*) are expressed in wild caught adults from the deep, showing a promising avenue for elucidating non-visual opsin regulation and function where light is severely restricted.

Sablefish have very long lifespans (often over 50 years) and prolonged development; while zebrafish hatch in as little as two days, sablefish hatch around 15 days post-fertilization. Other members of the Taylor Lab have begun quantifying melanopsin expression in these very early stages, which were not covered in my work. This drawn-out development provides better resolution of when opsins are expressed throughout fish development – this is made particularly interesting as eggs and yolk-sac larvae are reared in the dark. I hope that, as the knowledge around sablefish opsins expands in coming years, some findings can be applied to increase sablefish production efficiency in the hatchery, particularly when it comes to how light may mediate development.

The Koop Lab, in collaboration with the Taylor Lab, is currently working on an improved version of the sablefish genome, with chromosome-level assembly. This should improve some remaining uncertainties in the opsin gene sequences that I characterized (e.g., intron/exon structure, as single exons are sometimes split amongst contigs in the current assembly) and also offer clarity into the relationships (e.g., synteny, linkage and origins) of the sablefish opsin genes. Additionally, this level of assembly could be used in the future to survey the regulatory regions of sablefish opsins; one could gain insight into the gene's function by looking at transcription factor binding sites, for example that could bind transcription factors involved in neural development. Though working with a non-model organism can certainly be challenging due to a general lack of information and transgenic unavailability, it has been very rewarding to discover so much new information over the past few years. Model organisms, such as zebrafish, are incredibly useful, but my work here highlights key differences between the two species worth exploring.

BIBLIOGRAPHY

Adamian, L., and J. Liang. 2002. Interhelical hydrogen bonds and spatial motifs in membrane proteins: Polar clamps and serine zippers. *Proteins-Structure Function and Bioinformatics* **47**:209-218.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *Journal of Molecular Biology* **215**:403-410.

Arendt, D. 2003. Evolution of eyes and photoreceptor cell types. *International Journal of Developmental Biology* **47**:563-571.

Audet, M., and M. Bouvier. 2012. Restructuring G-Protein-Coupled Receptor Activation. *Cell* **151**:14-23.

Backfisch, B., V. B. V. Rajan, R. M. Fischer, C. Lohs, E. Arboleda, K. Tessmar-Raible, and F. Raible. 2013. Stable transgenesis in the marine annelid Platynereis dumerilii sheds new light on photoreceptor evolution. *Proceedings of the National Academy of Sciences of the United States of America* **110**:193-198.

Basili, D., G. Gioacchini, V. Todisco, M. Candelma, L. Marisaldi, L. Pappalardo, and O. Carnevali. 2021. Opsins and gonadal circadian rhythm in the swordfish (Xiphias gladius) ovary: Their potential roles in puberty and reproductive seasonality. *General and Comparative Endocrinology* **303**.

Bateman, A., M. J. Martin, S. Orchard, M. Magrane, R. Agivetova, S. Ahmad, E. Alpi, E. H. Bowler-Barnett, R. Britto, B. Bursteinas, H. Bye-A-Jee, R. Coetzee, A. Cukura, A. Da Silva, P. Denny, T. Dogan, T. Ebenezer, J. Fan, L. G. Castro, P. Garmiri, G. Georghiou, L. Gonzales, E. Hatton-Ellis, A. Hussein, A. Ignatchenko, G. Insana, R. Ishtiaq, P. Jokinen, V. Joshi, D. Jyothi, A. Lock, R. Lopez, A. Luciani, J. Luo, Y. Lussi, A. Mac-Dougall, F. Madeira, M. Mahmoudy, M. Menchi, A. Mishra, K. Moulang, A. Nightingale, C. S. Oliveira, S. Pundir, G. Y. Qi, S. Raj, D. Rice, M. R. Lopez, R. Saidi, J. Sampson, T. Sawford, E. Speretta, E. Turner, N. Tyagi, P. Vasudev, V. Volynkin, K. Warner, X. Watkins, R. Zaru, H. Zellner, A. Bridge, S. Poux, N. Redaschi, L. Aimo, G. Argoud-Puy, A. Auchincloss, K. Axelsen, P. Bansal, D. Baratin, M. C. Blatter, J. Bolleman, E. Boutet, L. Breuza, C. Casals-Casas, E. de Castro, K. C. Echioukh, E. Coudert, B. Cuche, M. Doche, D. Dornevil, A. Estreicher, M. L. Famiglietti, M. Feuermann, E. Gasteiger, S. Gehant, V. Gerritsen, A. Gos, N. Gruaz-Gumowski, U. Hinz, C. Hulo, N. Hyka-Nouspikel, F. Jungo, G. Keller, A. Kerhornou, V. Lara, P. Le Mercier, D. Lieberherr, T. Lombardot, X. Martin, P. Masson, A. Morgat, T. B. Neto, S. Paesano, I. Pedruzzi, S. Pilbout, L. Pourcel, M. Pozzato, M. Pruess, C. Rivoire, C. Sigrist, K. Sonesson, A. Stutz, S. Sundaram, M. Tognolli, L. Verbregue, C. H. Wu, C. N. Arighi, L. Arminski, C. M. Chen, Y. X. Chen, J. S. Garavelli, H. Z. Huang, K. Laiho, P. McGarvey, D. A. Natale, K. Ross, C. R. Vinayaka, Q. H. Wang, Y. Q. Wang, L. S. Yeh, J. Zhang, and C. UniProt. 2021. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Research* **49**:D480-D489.

Baum, L. E., and T. Petrie. 1966. STATISTICAL INFERENCE FOR PROBABILISTIC FUNCTIONS OF FINITE STATE MARKOV CHAINS. *Annals of Mathematical Statistics* **37**:1554-&.

Beaudry, F. E. G., T. W. Iwanicki, B. R. Z. Mariluz, S. Darnet, H. Brinkmann, P. Schneider, and J. S. Taylor. 2017. The non-visual opsins: eighteen in the ancestor of

vertebrates, astonishing increase in ray-finned fish, and loss in amniotes. *Journal of Experimental Zoology Part B-Molecular and Developmental Evolution* **328**:685-696.
Beitz, E. 2000. TeXshade: shading and labeling of multiple sequence alignments using LaTeX2e. *Bioinformatics* **16**:135-139.
Bellingham, J., S. S. Chaurasia, Z. Melyan, C. M. Liu, M. A. Cameron, E. E. Tarttelin, P. M. Iuvone, M. W. Hankins, G. Tosini, and R. J. Lucas. 2006. Evolution of melanopsin photoreceptors: Discovery and characterization of a new melanopsin in nonmammalian vertebrates. *Plos Biology* **4**:1334-1343.
Berson, D. M., F. A. Dunn, and M. Takao. 2002. Phototransduction by retinal ganglion cells that set the circadian clock. *Science* **295**:1070-1073.
Bertolesi, G. E., and S. McFarlane. 2018. Seeing the light to change colour: An evolutionary perspective on the role of melanopsin in neuroendocrine circuits regulating light-mediated skin pigmentation. *Pigment Cell & Melanoma Research* **31**:354-373.
Blackshaw, S., and S. H. Snyder. 1997. Parapinopsin, a novel catfish opsin localized to the parapineal organ, defines a new gene family. *Journal of Neuroscience* **17**:8083-8092.
Borges, R., W. E. Johnson, S. J. O'Brien, V. Vasconcelos, and A. Antunes. 2012. The Role of Gene Duplication and Unconstrained Selective Pressures in the Melanopsin Gene Family Evolution and Vertebrate Circadian Rhythm Regulation. *Plos One* **7**.
Braasch, I., A. R. Gehrke, J. J. Smith, K. Kawasaki, T. Manousaki, J. Pasquier, A. Amores, T. Desvignes, P. Batzel, J. Catchen, A. M. Berlin, M. S. Campbell, D. Barrell, K. J. Martin, J. F. Mulley, V. Ravi, A. P. Lee, T. Nakamura, D. Chalopin, S. H. Fan, D. Wcisel, C. Canestro, J. Sydes, F. E. G. Beaudry, Y. Sun, J. Hertel, M. J. Beam, M. Fasold, M. Ishiyama, J. Johnson, S. Kehr, M. Lara, J. H. Letaw, G. W. Litman, R. T. Litman, M. Mikami, T. Ota, N. R. Saha, L. Williams, P. F. Stadler, H. Wang, J. S. Taylor, Q. Fontenot, A. Ferrara, S. M. J. Searle, B. Aken, M. Yandell, I. Schneider, J. A. Yoder, J. N. Volff, A. Meyer, C. T. Amemiya, B. Venkatesh, P. W. H. Holland, Y. Guiguen, J. Bobe, N. H. Shubin, F. Di Palma, J. Alfoldi, K. Lindblad-Toh, and J. H. Postlethwait. 2016. The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons (vol 48, pg 427, 2016). *Nature Genetics* **48**:700-700.
Burge, C., and S. Karlin. 1997. Prediction of complete gene structures in human genomic DNA. *Journal of Molecular Biology* **268**:78-94.
Bustin, S. A. 2010. Why the need for qPCR publication guidelines?-The case for MIQE. *Methods*, *50*(4), 217-226. https://doi.org/10.1016/j.ymeth.2009.12.006
Cavallari, N., E. Frigato, D. Vallone, N. Frohlich, J. F. Lopez-Olmeda, A. Foa, R. Berti, F. J. Sanchez-Vazquez, C. Bertolucci, and N. S. Foulkes. 2011. A Blind Circadian Clock in Cavefish Reveals that Opsins Mediate Peripheral Clock Photoreception. *Plos Biology* **9**.
Chen, S. K., T. C. Badea, and S. Hattar. 2011. Photoentrainment and pupillary light reflex are mediated by distinct populations of ipRGCs. *Nature* **476**:92-+.
Chinen, A. 2005. Reconstitution of Ancestral Green Visual Pigments of Zebrafish and Molecular Mechanism of Their Spectral Differentiation. *Molecular Biology and Evolution* **22**:1001-1010.
Clarke, N. 2020. Personal communication (manuscript unpublished).
Cortesi, F., Z. Musilova, S. M. Stieb, N. S. Hart, U. E. Siebeck, M. Malmstrom, O. K. Torresen, S. Jentoft, K. L. Cheney, N. J. Marshall, K. L. Carleton, and W. Salzburger. 2015. Ancestral duplications and highly dynamic opsin gene evolution in percomorph

fishes. *Proceedings of the National Academy of Sciences of the United States of America* **112**:1493-1498.
Crespo, C., D. Soroldoni, and E. Knust. 2018. A novel transgenic zebrafish line for red opsin expression in outer segments of photoreceptor cells. *Developmental Dynamics* **247**:951-959.
Dann, S. G., Allison, W. T., Levin, D. B., Taylor, J. S., and Hawryshyn, C. W. 2004. Salmonid opsin sequences undergo positive selection and indicate an alternate evolutionary relationship in Oncorhynchus. *Journal of Molecular Evolution*, *58*(4), 400-412. https://doi.org/10.1007/s00239-003-2562-y
Dartnall, H. J. A. 1952. VISUAL PIGMENT-467, A PHOTOSENSITIVE PIGMENT PRESENT IN TENCH RETINAE. *Journal of Physiology-London* **116**:257-289.
Davies, W. I. L., T. K. Tamai, L. Zhen, J. K. Fu, J. Rihel, R. G. Foster, D. Whitmore, and M. W. Hankins. 2015a. An extended family of novel vertebrate photopigments is widely expressed and displays a diversity of function. *Genome Research* **25**:1666-1679.
Davies, W. I. L., L. Zheng, S. Hughes, T. K. Tamai, M. Turton, S. Halford, R. G. Foster, D. Whitmore, and M. W. Hankins. 2011. Functional diversity of melanopsins and their global expression in the teleost retina. *Cellular and Molecular Life Sciences* **68**:4115-4132.
Dekens, M. P. S., B. M. Fontinha, M. Gallach, S. Pflugler, and K. Tessmar-Raible. 2022. Melanopsin elevates locomotor activity during the wake state of the diurnal zebrafish. *Embo Reports* **23**.
Denton, E. J., and F. J. Warren. 1956. Visual pigments of deep-sea fish. *Nature* **178**:1059-1059.
Deupi, X., J. Standfuss, and G. Schertler. 2012. Conserved activation pathways in G-protein-coupled receptors. *Biochemical Society Transactions* **40**:383-388.
Dkhissi-Benyahya, O., Coutanson, C., Knoblauch, K., Lahouaoui, H., Leviel, V., Rey, C., . . . Cooper, H. M. 2013. The absence of melanopsin alters retinal clock function and dopamine regulation by light. *Cellular and Molecular Life Sciences*, *70*(18), 3435-3447. https://doi.org/10.1007/s00018-013-1338-9
Dulai, K. S., M. von Dornum, J. D. Mollon, and D. M. Hunt. 1999. The evolution of trichromatic color vision by opsin gene duplication in New World and Old World primates. *Genome Research* **9**:629-638.
Ernst, O. P., D. T. Lodowski, M. Elstner, P. Hegemann, L. S. Brown, and H. Kandori. 2014. Microbial and Animal Rhodopsins: Structures, Functions, and Molecular Mechanisms. *Chemical Reviews* **114**:126-163.
Fairgrieve, M. R., Y. Shibata, E. K. Smith, E. S. Hayman, and J. A. Luckenbach. 2016. Molecular characterization of the gonadal kisspeptin system: Cloning, tissue distribution, gene expression analysis and localization in sablefish (Anoplopoma fimbria). *General and Comparative Endocrinology* **225**:212-223.
Fasick, J. I., Algrain, H., Serba, K. M., and Robinson, P. R. 2020. The retinal pigments of the whale shark (Rhincodon typus) and their role in visual foraging ecology (vol 36, E011, 2019). *Visual Neuroscience*, *37*, Article E011. https://doi.org/10.1017/s0952523820000103
Fernandes, A. M., K. Fero, A. B. Arrenberg, S. A. Bergeron, W. Driever, and H. A. Burgess. 2012. Deep Brain Photoreceptors Control Light-Seeking Behavior in Zebrafish Larvae. *Current Biology* **22**:2042-2047.
Feuda, R., Menon, A. K., and Gopfert, M. C. 2022. Rethinking Opsins. *Molecular Biology and Evolution*, *39*(3), Article msac033. https://doi.org/10.1093/molbev/msac033

Fischer, R. M., B. M. Fontinha, S. Kirchmaier, J. Steger, S. Bloch, D. Inoue, S. Panda, S. Rumpel, and K. Tessmar-Raible. 2013. Co-Expression of VAL- and TMT-Opsins Uncovers Ancient Photosensory Interneurons and Motorneurons in the Vertebrate Brain. *Plos Biology* **11**.
Flamarique, I. N., Ahmed, A. S., Cheng, C. L., Molday, R. S., and Devlin, R. H. 2019. Growth hormone regulates opsin expression in the retina of a salmonid fish. *Journal of Neuroendocrinology*, *31*(11), Article e12804. https://doi.org/10.1111/jne.12804
Flamarique, I. N., R. Fujihara, R. Yazawa, K. Bolstad, B. Gowen, and G. Yoshizaki. 2021. Disrupted eye and head development in rainbow trout with reduced ultraviolet (sws1) opsin expression. *Journal of Comparative Neurology* **529**:3013-3031.
Foster, R. G., M. S. Grace, I. Provencio, W. J. Degrip, and J. M. Garciafernandez. 1994. Identification of vertebrate deep brain photoreceptors. *Neuroscience and Biobehavioral Reviews* **18**:541-546.
Foster, R. G., I. Provencio, D. Hudson, S. Fiske, W. Degrip, and M. Menaker. 1991. Circadian photoreception in the retinally degenerate mouse (rd/rd). *Journal of Comparative Physiology a-Sensory Neural and Behavioral Physiology* **169**:39-50.
Fredriksson, R., and H. B. Schioth. 2005. The repertoire of G-protein-coupled receptors in fully sequenced genomes. *Molecular Pharmacology* **67**:1414-1425.
Frisch, K. V. 1911. Beitrge zur Physiologie der Pigmentzellen in der Fischhaut. *Pflger's, Archiv fr die Gesammte Physiologie des Menschen und der Tiere* **138**:319-387.
Giraldo-Calderon, G. I., M. J. Zanis, and C. A. Hill. 2017. Retention of duplicated long-wavelength opsins in mosquito lineages by positive selection and differential expression. *Bmc Evolutionary Biology* **17**.
Gooley, J. J., J. Lu, T. C. Chou, T. E. Scammell, and C. B. Saper. 2001. Melanopsin in cells of origin of the retinohypothalamic tract. *Nature Neuroscience* **4**:1165-1165.
Greenbaum, D., C. Colangelo, K. Williams, and M. Gerstein. 2003. Comparing protein abundance and mRNA expression levels on a genomic scale. *Genome Biology* **4**.
Gross, A. K., S. Madabushi, T. G. Wensel, and O. Lichtarge. 2004. Evolution guided mutagenesis of rhodopsin. *Investigative Ophthalmology & Visual Science* **45**:U183-U183.
Hao, W. S., and H. K. W. Fong. 1999. The endogenous chromophore of retinal G protein-coupled receptor opsin from the pigment epithelium. *Journal of Biological Chemistry* **274**:6085-6090.
Hardie, R. C. 2011. A brief history of trp: commentary and personal perspective. *Pflugers Archiv-European Journal of Physiology* **461**:493-498.
Hattar, S., H. W. Liao, M. Takao, D. M. Berson, and K. W. Yau. 2002. Melanopsin-containing retinal. ganglion cells: Architecture, projections, and intrinsic photosensitivity. *Science* **295**:1065-1070.
Howe, K., M. D. Clark, C. F. Torroja, J. Torrance, C. Berthelot, M. Muffato, J. E. Collins, S. Humphray, K. McLaren, L. Matthews, S. McLaren, I. Sealy, M. Caccamo, C. Churcher, C. Scott, J. C. Barrett, R. Koch, G. J. Rauch, S. White, W. Chow, B. Kilian, L. T. Quintais, J. A. Guerra-Assuncao, Y. Zhou, Y. Gu, J. Yen, J. H. Vogel, T. Eyre, S. Redmond, R. Banerjee, J. X. Chi, B. Y. Fu, E. Langley, S. F. Maguire, G. K. Laird, D. Lloyd, E. Kenyon, S. Donaldson, H. Sehra, J. Almeida-King, J. Loveland, S. Trevanion, M. Jones, M. Quail, D. Willey, A. Hunt, J. Burton, S. Sims, K. McLay, B. Plumb, J. Davis, C. Clee, K. Oliver, R. Clark, C. Riddle, D. Eliott, G. Threadgold, G. Harden, D. Ware, B. Mortimer, G. Kerry, P. Heath, B. Phillimore, A. Tracey, N. Corby, M. Dunn, C. Johnson, J. Wood, S. Clark, S. Pelan, G. Griffiths, M. Smith, R. Glithero, P. Howden, N. Barker, C. Stevens, J. Harley, K.

Holt, G. Panagiotidis, J. Lovell, H. Beasley, C. Henderson, D. Gordon, K. Auger, D. Wright, J. Collins, C. Raisen, L. Dyer, K. Leung, L. Robertson, K. Ambridge, D. Leongamornlert, S. McGuire, R. Gilderthorp, C. Griffiths, D. Manthravadi, S. Nichol, G. Barker, S. Whitehead, M. Kay, J. Brown, C. Murnane, E. Gray, M. Humphries, N. Sycamore, D. Barker, D. Saunders, J. Wallis, A. Babbage, S. Hammond, M. Mashreghi-Mohammadi, L. Barr, S. Martin, P. Wray, A. Ellington, N. Matthews, M. Ellwood, R. Woodmansey, G. Clark, J. Cooper, A. Tromans, D. Grafham, C. Skuce, R. Pandian, R. Andrews, E. Harrison, A. Kimberley, J. Garnett, N. Fosker, R. Hall, P. Garner, D. Kelly, C. Bird, S. Palmer, I. Gehring, A. Berger, C. M. Dooley, Z. Ersan-Urun, C. Eser, H. Geiger, M. Geisler, L. Karotki, A. Kirn, J. Konantz, M. Konantz, M. Oberlander, S. Rudolph-Geiger, M. Teucke, K. Osoegawa, B. L. Zhu, A. Rapp, S. Widaa, C. Langford, F. T. Yang, N. P. Carter, J. Harrow, Z. M. Ning, J. Herrero, S. M. J. Searle, A. Enright, R. Geisler, R. H. A. Plasterk, C. Lee, M. Westerfield, P. J. de Jong, L. I. Zon, J. H. Postlethwait, C. Nusslein-Volhard, T. J. P. Hubbard, H. Roest Crollius, J. Rogers, and D. L. Stemple. 2013. The zebrafish reference genome sequence and its relationship to the human genome. *Nature* **496**:498-503.

Huang, Z. Q., M. J. Fasco, and L. S. Kaminsky. 1996. Optimization of DNase I removal of contaminating DNA from RNA for use in quantitative RNA-PCR. *Biotechniques* **20**:1012-&.

Jeffery, C. J. 2014. An introduction to protein moonlighting. *Biochemical Society Transactions*, *42*, 1679-1683. https://doi.org/10.1042/bst20140226

Kang, S. W., B. Leclerc, S. Kosonsiriluk, L. J. Mauro, A. Iwasawa, and M. E. El Halawani. 2010. Melanopsin expression in dopamine-melatonin neurons of the premammillary nucleus of the hypothalamus and seasonal reproduction in birds. *Neuroscience* **170**:200-213.

Kelley, J. L., and W. I. L. Davies. 2016. The Biological Mechanisms and Behavioral Functions of Opsin-Based Light Detection by the Skin. *Frontiers in Ecology and Evolution* **4**.

Kimata, N., A. Pope, O. B. Sanchez-Reyes, M. Eilers, C. A. Opefi, M. Ziliox, P. J. Reeves, and S. O. Smith. 2016. Free backbone carbonyls mediate rhodopsin activation. *Nature Structural & Molecular Biology* **23**:738-+

Kojima, D., M. Torii, Y. Fukada, and J. E. Dowling. 2008. Differential expression of duplicated VAL-opsin genes in the developing zebrafish. *Journal of Neurochemistry* **104**:1364-1371.

Koenig, K. M., and Gross, J. M. 2020. Evolution and development of complex eyes: a celebration of diversity. *Development*, *147*(19), Article dev182923. https://doi.org/10.1242/dev.182923

Kottgen , E., and Abelsdorft, G. 1896. Absorption und Zersetzung des Sehpurpurs bei den Wirbeltieren. *Z. Psychol.; Physiol. Sinnesorg*.

Koyanagi, M., and A. Terakita. 2008. Gq-coupled rhodopsin subfamily composed of invertebrate visual pigment and melanopsin. *Photochemistry and Photobiology* **84**:1024-1030.

Koyanagi, M., A. Terakita, K. Kubokawa, and Y. Shichida. 2002. Amphioxus homologs of Go-coupled rhodopsin and peropsin having 11-cis- and all-trans-retinals as their chromophores. *Febs Letters* **531**:525-528.

Kozmik, Z., J. Ruzickova, K. Jonasova, Y. Matsumoto, P. Vopalensky, I. Kozmikova, H. Strnad, S. Kawamura, J. Piatigorsky, V. Paces, and C. Vlcek. 2008. Assembly of the cnidarian camera-type eye from vertebrate-like components. *Proceedings of the National Academy of Sciences of the United States of America* **105**:8989-8993.

Kuznetsova, T., K. Antos, E. Malinina, S. Papaioannou, and P. Medini. 2021. Visual stimulation with blue wavelength light drives V1 effectively eliminating stray light contamination during two-photon calcium imaging. *Journal of Neuroscience Methods* **362**.
Lam, C. S., M. Marz, and U. Strahle. 2009. gfap and nestin Reporter Lines Reveal Characteristics of Neural Progenitors in the Adult Zebrafish Brain. *Developmental Dynamics* **238**:475-486.
Lamb, T. D. 2013. Evolution of phototransduction, vertebrate photoreceptors and retina. *Progress in Retinal and Eye Research* **36**:52-119.
Leung, N. Y., D. P. Thakur, A. S. Gurav, S. H. Kim, A. Di Pizio, M. Y. Niv, and C. Montell. 2020. Functions of Opsins in Drosophila Taste. *Current Biology* **30**:1367-1379.e1366.
Li, P., S. Temple, Y. Gao, T. J. Haimberger, C. W. Hawryshyn, and L. Li. 2005. Circadian rhythms of behavioral cone sensitivity and long wavelength opsin mRNA expression: a correlation study in zebratish. *Journal of Experimental Biology* **208**:497-504.
Lucas, R. J., R. H. Douglas, and R. G. Foster. 2001. Characterization of an ocular photopigment capable of driving pupillary constriction in mice. *Nature Neuroscience* **4**:621-626.
Lupse, N., F. Cortesi, M. Freese, L. Marohn, J. D. Pohlmann, K. Wysujack, R. Hanel, and Z. Musilova. 2021. Visual Gene Expression Reveals a cone-to-rod Developmental Progression in Deep-Sea Fishes. *Molecular Biology and Evolution* **38**:5664-5677.
Matos-Cruz, V., J. Blasic, B. Nickle, P. R. Robinson, S. Hattar, and M. E. Halpern. 2011. Unexpected Diversity and Photoperiod Dependence of the Zebrafish Melanopsin System. *Plos One* **6**.
McCall, M. N., H. R. McMurray, H. Land, and A. Almudevar. 2014. On non-detects in qPCR data. *Bioinformatics* **30**:2310-2316.
McCurley, A. T., and G. V. Callard. 2008. Characterization of housekeeping genes in zebrafish: male-female differences and effects of tissue type, developmental stage and chemical treatment. *Bmc Molecular Biology* **9**.
Menon, I., T. Huber, S. Sanyal, S. Banerjee, P. Barre, S. Canis, J. D. Warren, J. Hwa, T. P. Sakmar, and A. K. Menon. 2011. Opsin Is a Phospholipid Flippase. *Current Biology* **21**:149-153.
Meredith, R. W., Gatesy, J., Emerling, C. A., York, V. M., and Springer, M. S. 2013. Rod Monochromacy and the Coevolution of Cetacean Retinal Opsins. *Plos Genetics*, *9*(4), Article e1003432. https://doi.org/10.1371/journal.pgen.1003432
Morra, G., A. M. Razavi, K. Pandey, H. Weinstein, A. K. Menon, and G. Khelashvili. 2018. Mechanisms of Lipid Scrambling by the G Protein-Coupled Receptor Opsin. *Structure* **26**:356-+.
Musilova, Z., F. Cortesi, M. Matschiner, W. I. L. Davies, J. S. Patel, S. M. Stieb, F. de Busserolles, M. Malmstrom, O. K. Torresen, C. J. Brown, J. K. Mountford, R. Hanel, D. L. Stenkamp, K. S. Jakobsen, K. L. Carleton, S. Jentoft, J. Marshall, and W. Salzburger. 2019. Vision using multiple distinct rod opsins in deep-sea fishes. *Science* **364**:588-+.
Musilova, Z., W. Salzburger, and F. Cortesi. 2021. The Visual Opsin Gene Repertoires of Teleost Fishes: Evolution, Ecology, and Function. *Annual Review of Cell and Developmental Biology, Vol 37* **37**:441-468.
Nakane, Y., K. Ikegami, H. Ono, N. Yamamoto, S. Yoshida, K. Hirunagi, S. Ebihara, Y. Kubo, and T. Yoshimura. 2010. A mammalian neural tissue opsin (Opsin 5) is a deep brain photoreceptor in birds. *Proceedings of the National Academy of Sciences of the United States of America* **107**:15264-15268.

Nguyen, M. T. T., S. Vemaraju, G. Nayak, Y. Odaka, E. D. Buhr, N. Alonzo, U. Tran, M. Batie, B. A. Upton, M. Darvas, Z. Kozmik, S. Rao, R. S. Hegde, P. M. Iuvone, R. N. Van Gelder, and R. A. Lang. 2019. An opsin 5-dopamine pathway mediates light-dependent vascular development in the eye. *Nature Cell Biology* **21**:420-+.

Oesterhelt, D., and W. Stoeckenius. 1973. Functions of a new photoreceptor membrane. *Proceedings of the National Academy of Sciences of the United States of America* **70**:2853-2857.

Owens, G. L., D. J. Windsor, J. Mui, and J. S. Taylor. 2009. A Fish Eye Out of Water: Ten Visual Opsins in the Four-Eyed Fish, Anableps anableps. *Plos One* **4**.

Panda, S., Nayak, S. K., Campo, B., Walker, J. R., Hogenesch, J. B., and Jegla, T. 2005. Illumination of the melanopsin signaling pathway. *Science*, *307*(5709), 600-604. https://doi.org/10.1126/science.1105121

Perez-Cerezales, S., S. Boryshpolets, O. Afanzar, A. Brandis, R. Nevo, V. Kiss, and M. Eisenbach. 2015. Involvement of opsins in mammalian sperm thermotaxis. *Scientific Reports* **5**:18.

Pfaffl, M. W. 2001. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Research* **29**.

Philp, A. R., J. M. Garcia-Fernandez, B. G. Soni, R. J. Lucas, J. Bellingham, and R. G. Foster. 2000. Vertebrate ancient (VA) opsin and extraretinal photoreception in the Atlantic salmon (Salmo salar). *Journal of Experimental Biology* **203**:1925-1936.

Ponchel, F., C. Toomes, K. Bransfield, F. T. Leong, S. H. Douglas, S. L. Field, S. M. Bell, V. Combaret, A. Puisieux, A. J. Mighell, P. A. Robinson, C. F. Inglehearn, J. D. Isaacs, and A. F. Markham. 2003. Real-time PCR based on SYBR-Green I fluorescence: An alternative to the TaqMan assay for a relative quantification of gene rearrangements, gene amplifications and micro gene deletions. *Bmc Biotechnology* **3**.

Prince, V. E., and F. B. Pickett. 2002. Splitting pairs: The diverging fates of duplicated genes. *Nature Reviews Genetics* **3**:827-837.

Perez, J. H., E. Tolla, I. C. Dunn, S. L. Meddle, and T. J. Stevenson. 2019. A Comparative Perspective on Extra-retinal Photoreception. *Trends in Endocrinology & Metabolism* **30**:39-53.

Rennison, D. J., G. L. Owens, and J. S. Taylor. 2012. Opsin gene duplication and divergence in ray-finned fish. *Molecular Phylogenetics and Evolution* **62**:986-1008.

Rice, P., I. Longden, and A. Bleasby. 2000. EMBOSS: The European molecular biology open software suite. *Trends in Genetics* **16**:276-277.

Robinson, J., E. A. Schmitt, R. J. Reece, and J. E. Dowling. 1993. An ultraviolet-sensitive opsin gene in zebrafish (brachydanio-rerio) - sequence and localization. *Investigative Ophthalmology & Visual Science* **34**:808-808.

Rondeau, E. B., A. M. Messmer, D. S. Sanderson, S. G. Jantzen, K. R. von Schalburg, D. R. Minkley, J. S. Leong, G. M. Macdonald, A. E. Davidsen, W. A. Parker, R. S. A. Mazzola, B. Campbell, and B. F. Koop. 2013. Genomics of sablefish (Anoplopoma fimbria): expressed genes, mitochondrial phylogeny, linkage map and identification of a putative sex gene. *Bmc Genomics* **14**.

Sato, K., T. Yamashita, Y. Haruki, H. Ohuchi, M. Kinoshita, and Y. Shichida. 2016. Two UV-Sensitive Photoreceptor Proteins, Opn5m and Opn5m2 in Ray-Finned Fish with Distinct Molecular Properties and Broad Distribution in the Retina and Brain. *Plos One* **11**.

Schmitt, E. A., G. A. Hyatt, and J. E. Dowling. 1999. Temporal and spatial patterns of opsin gene expression in the zebrafish (Danio rerio) (vol 12, pg 895, 1995). *Visual Neuroscience* **16**:601-605.
Sharma, A. K., D. A. Walsh, E. Bapteste, F. Rodriguez-Valera, W. F. Doolittle, and R. T. Papke. 2007. Evolution of rhodopsin ion pumps in haloarchaea. *Bmc Evolutionary Biology* **7**.
Sherina, V., H. R. McMurray, W. Powers, H. Land, T. M. T. Love, and M. N. McCall. 2020. Multiple imputation and direct estimation for qPCR data with non-detects. *Bmc Bioinformatics* **21**.
Shichida, Y., and T. Matsuyama. 2009. Evolution of opsins and phototransduction. *Philosophical Transactions of the Royal Society B: Biological Sciences* **364**:2881-2895.
Sigler, M. F., and K. B. Echave. 2019. Diel vertical migration of sablefish (Anoplopoma fimbria). *Fisheries Oceanography* **28**:517-531.
Simon, N., S. Fujita, M. Porter, and M. Yoshizawa. 2019. Expression of extraocular opsin genes and light-dependent basal activity of blind cavefish. *Peerj* **7**.
Smith, A. R., M. J. Van Staaden, and K. L. Carleton. 2012. An Evaluation of the Role of Sensory Drive in the Evolution of Lake Malawi Cichlid Fishes. *International Journal of Evolutionary Biology* **2012**:1-12.
Smith, T. A., B. N. Moore, A. Matoso, D. E. Berkowitz, J. J. DeBerry, and J. L. Pluznick. 2021. Identification of novel bladder sensory GPCRs. *Physiological Reports* **9**.
Soni, B. G., and R. G. Foster. 1997. A novel and ancient vertebrate opsin. *Febs Letters* **406**:279-283.
Spitschan, M., A. S. Bock, J. Ryan, G. Frazzetta, D. H. Brainard, and G. K. Aguirre. 2017. The human visual cortex response to melanopsin-directed stimulation is accompanied by a distinct perceptual experience. *Proceedings of the National Academy of Sciences of the United States of America* **114**:12291-12296.
Stahl, B. A., and J. B. Gross. 2017. A Comparative Transcriptomic Analysis of Development in Two Astyanax Cavefish Populations. *Journal of Experimental Zoology Part B-Molecular and Developmental Evolution* **328**:515-532.
Stanke, M., and B. Morgenstern. 2005. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Research* **33**:W465-W467.
Takechi, M., and S. Kawamura. 2005. Temporal and spatial changes in the expression pattern of multiple red and green subtype opsin genes during zebrafish development. *Journal of Experimental Biology* **208**:1337-1345.
Taylor, J. S., I. Braasch, T. Frickey, A. Meyer, and Y. Van de Peer. 2003. Genome duplication, a trait shared by 22,000 species of ray-finned fish. *Genome Research* **13**:382-390.
Taylor, S. C., K. Nadeau, M. Abbasi, C. Lachance, M. Nguyen, and J. Fenrich. 2019. The Ultimate qPCR Experiment: Producing Publication Quality, Reproducible Data the First Time. *Trends in Biotechnology* **37**:761-774.
Temple, S., N. S. Hart, N. J. Marshall, and S. P. Collin. 2010. A spitting image: specializations in archerfish eyes for vision at the interface between air and water. *Proceedings of the Royal Society B-Biological Sciences* **277**:2607-2615.
Terakita, A. 2005. The opsins. *Genome Biology* **6**.
Than-Trong, E., and L. Bally-Cuif. 2015. Radial glia and neural progenitors in the adult zebrafish central nervous system. *Glia* **63**:1406-1428.

Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. Clustal-w - improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**:4673-4680.

van Hazel, I., S. Z. Dungan, F. E. Hauser, J. M. Morrow, J. A. Endler, and B. S. W. Chang. 2016. A comparative study of rhodopsin function in the great bowerbird (Ptilonorhynchus nuchalis): Spectral tuning and light-activated kinetics. *Protein Science* **25**:1308-1318.

Veldhoen, K., Allison, W. T., Veldhoen, N., Anholt, B. R., Helbing, C. C., and Hawryshyn, C. W. 2006. Spatio-temporal characterization of retinal opsin gene expression during thyroid hormone-induced and natural development of rainbow trout. *Visual Neuroscience*, *23*(2), 169-179. https://doi.org/10.1017/s0952523806232139

Vighteichmann, I., M. A. Ali, and B. Vigh. 1992. Comparative ultrastructure and opsin immunocytochemistry of the retina and pineal organ in fish. *Progress in Brain Research* **91**:307-313.

Wald, G. 1939. The porphyropsin visual system. *Journal of General Physiology* **22**:775-794.

Wang, Y. N., Zhou, L., Wu, L. L., Song, C. B., Ma, X. N., Xu, S. H., . . . Li, J. 2021. Evolutionary ecology of the visual opsin gene sequence and its expression in turbot (Scophthalmus maximus). *Bmc Ecology and Evolution*, *21*(1), Article 114. https://doi.org/10.1186/s12862-021-01837-2

Woelders, T., T. Leenheers, M. C. M. Gordijn, R. A. Hut, D. G. M. Beersma, and E. J. Wams. 2018. Melanopsin- and L-cone-induced pupil constriction is inhibited by S- and M-cones in humans. *Proceedings of the National Academy of Sciences of the United States of America* **115**:792-797.

Wright, D. S., N. Demandt, J. T. Alkema, O. Seehausen, T. G. G. Groothuis, and M. E. Maan. 2017. Developmental effects of visual environment on species-assortative mating preferences in Lake Victoria cichlid fish. *Journal of Evolutionary Biology* **30**:289-299.

Yang, Z. H., and K. K. W. Wang. 2015. Glial fibrillary acidic protein: from intermediate filament assembly and gliosis to neurobiomarker. *Trends in Neurosciences* **38**:364-374.

Yau, K. W., and R. C. Hardie. 2009. Phototransduction Motifs and Variations. *Cell* **139**:246-264.

Ye, X., L. L. Zhang, H. Y. Dong, Y. Y. Tian, H. H. Lao, J. J. Bai, and L. Y. Yu. 2010. Validation of reference genes of grass carp Ctenopharyngodon idellus for the normalization of quantitative real-time PCR. *Biotechnology Letters* **32**:1031-1038.

Zhong, M., R. Kawaguchi, M. Kassai, and H. Sun. 2012. Retina, Retinol, Retinal and the Natural History of Vitamin A as a Light Sensor. *Nutrients* **4**:2069-2096.

APPENDIX

Appendix A. Translated peptide sequences of the complete sablefish opsin repertoire.

>Rh1
MNGTEGPYFYVPMVNTTGIVRNPYEYPQYYLVNPAAYAALGAYMFLLILVGFPVNALTLYVTLEHKKLRT
PLNYILLNLAVANLFMVLGGFTTTMYTSMHGYFVLGRLGCNLEGYFATLGGEIALWSLVVLAVERWMVV
CKPISNFRFGEDHAIMGLGLTWLMANACAPPLVGWSRYIPEGMQCSCGVDYYTRAEGFNNESFVIYMFTC
HFLIPLTIIFFCYGRLLCAVKEAAAAQQESETTQRAEREVSRMVVLMVIAFLVCWLPYAGVAWFIFLNQGSE
FGPVFMTIPAFFAKSSAIYNPMIYICMNKQFRHCMITTLCCGKNPFEEEEGASSTKTEASSVSTSSVSPA
>Exorh
MNGTEGPNFYVPMSNKTGVVRSPFEHTQYYLAEPWKYSLLAAYMLFLIITAFPINFLTLYVTVKNKKLRNP
LNFVLLNLAVADLFMVIGGFTVTLYTALHGYFILGVTGCNIEGFFATLGGEIALWSLVVLAIERYVVVCKP
MSNFRFGEKHAIAGLVFTWIMAMTCATPPLLGWSRYIPEGMQCSCGIDYYTPKPEIHNTSFVIYMFVLHFSI
PLFIIFFCYGRLLCTIRAAAALQQESETTQRAEKEVTRMVIVMVMSFLVCWVPYATVAWYIFDNQGTEFGP
VFMTIPAFFAKSSALYNPVIYILLNRQ
>Rh2
MVWEGGIEPNGTEGKNFYIPMSNRTGVVRSPYESAQYYLIDPIVFKFLAFYMFFLICTGTPINGLTLLVTAQ
NKKLRQPLNYILVNLAVAGLIMCAFGFTITITSAVNGYFILGPTACAVEGFMATLGGEVALWSLVVLAVER
YIVVCKPMGSFKFTGTHAGAGVLLTWVMAMACAAPPLFGWSRYIPEGMQCSCGPDYYTLAPGFNNESYV
MYMFTVHFFTPVFIIFFTYGSLVLTVKAAAAQQQESESTQKAEKEVTRMCILMVLGFLVAWVPYASFAGWI
FLNKGAPFSALTAAIPAFFAKSSALYNPVIYVLLNKQFRNCMLSTIGMGSAVDDEASVSASKTEVSSVS
>Sws1
MGKHFHLYENISKVSPFEGPQYYLAPHWAFHLQTIFMGLVLFAGAPLNFIVLFVTLKYKKLRVPLNYILVNI
SFAGFIFVTFSVSQVFVATMRGYYFLGYTLCAMEAAMGSIAGLVTAWSLAVLSFERYLIICKPFGAFRFGSN
HAMAAVAFTWFMGVGCAIPPFFGWSRYIPEGLGCSCGPDWYTHSEEFNCTSYTHFLMVTCFICPLSIIIFSYS
QLLGALRAVAAQQTESESTQKAEKEVSKMIIVMVGSFVSCYGPYALAALYFAYSTDENKDYRLVTIPAFFS
KSSCVYNPLISVFMNKQFNACIMETFFGKVMDESSEVSSKTEVSSVSTAS
>Sws2a1
MKHGRVTELPEDFWIPIPLDTDNITSLSPFLIPQDHLASSVTYYAMAMYMLFIFIVGTFINAL
TVACTIQNKKLRSHLNYILVNLAMSNLLVSGVGSFTAFCSFAARYFILGPLACKIEGFIVTLGGMVSLWSLA
VVAFERWLVICKPLGNFIFKPDHAIACCLFTWVFALIASVPPLVGWSRYIPEGLQCSCGPDWYTTNNKYNN
ESYVMFLFGFCFAVPLATIVFCYSQLLFTLKMAAKAQAESASTQKAEREVTRMVVIMVLGFLVCWLPYAS
FALWVVNNRGQSFDLRFASVPSVFSKSSAIYNPVIYVLLNKQFRTCMMKMLGMGGADDEESSTSSVTEVS
KVGPA
>Sws2a2
MELPEDFWIPVPLETNNITALSPFLVPQDHLGNLATFYSMAGFMFFTFVVGTGINVLTIACTVQYKKLRSHL
NYILVNLAVANLLVSCVGSFTAFCSFSTRYFIFGALACKIEGFLATLGGMVSLWSLAVVAFERWLVICKPLG
NFIFKPDHAIACCLFTWVFALIASVPPLVGWSRYIPEGLQCSCGPDWYTTNNKYNNESYVMFLGFCFAVPFS
TIVFCYSQLLFTLKMAAKAQAESASTQKAEREVTRMVVLMVMGFLVCWLPYASFALWVVNNRGQTFDLR
LATIPSCLSKGSAVYNPIIYVVFNKQFRLCMMKMLGMGGGDDEDSSTSASVTEVSKVGPA
>Sws2b
MRGIRPTELPEDFWIPIPLDTDNITSLSPFLVPQDHLGSLGIFYSLSAFMFFLFTTGTTINTLTIACTVKYKKLR
SHLNYILVNLAVANLLVASVGSFTCFYCFAFRYMILGPLGCKIEGFTAVGGMVSLWSLAVVAFERWLVICK
PLGNFAFKPSHAICCAMTWVFALMAAVPPLVGWSRYIPEGMQCSCGPDWYTTNNKYNNESYVMFLFCFC
FVPFCTIVFCYSQLLLMLKSAAKAQAESASTQKAEREVTRMVVVMVLGFLVCWMPYASFALWVINNRGN
PFDLRLATIPSCLSKASTVYNPVIYILLNKQFRSSLRKMLGMSEEDDEDSSTSQSVTEVSKVGPS
>Lws_
MAEEWGKQAFAARRYNEDTTRGSAFAYTNSNHTKDPFEGPNYHIAPRWVYNVATLWMFVVVALSVFTN
GLVLVATAKFKKLQHPLNWILVNLAIADLGETVFASTISVCNQYFGYFILGHPMCIFEGYVVSVCGITALWS
LTIISWERWIVVCKPFGNVKFDAKWATSGIVFSWVWSAVWCAPPVFGWSRYWPHGLKTSCGPDVFSGSED
PGVQSYMIVLMITCCLIPLAIIILCYLAVWLAIRAVAMQQKESESTQKAEREVSRMVVVMIVAYCVCWGPY
TFFACFAAANPGYAFHPLAAAMPAYFAKSATIYNPVIYVFMNRQFRTCIMQLFGKEADDGSEVSTSKTEVS
SVSPA

>VA1
MDTLSLSVNAVSYTVAAELPVTNDPFNGPINDIAPWNFTILAVLMFVVTSLSLSENFVVMFVTFKYKQLRQ
PLNYIIVNLAIADFLVSLTGGVISFLTNARGYFFLGKWACVLEGFAVTFFGIVALWSLAVLSFERFFVICRPL
GNIRLQARHAILGLLFVWTFSFIWTFPPVLGWNRYTISKIGTTCEPDWYSTNMIDHSYIITFFATCFILPLGVIF
FCYGKLLRKLRQVSHGRLASAKKPERQVTRMVVVMIVAFMVGWTPYAAFSILVTAYPTIKLDPRLSSPAFF
SKTAAVYNPIIYVFMNKQFRKCLIQLFSGKGTIPDSHLNQTSERPGITAESQTGEMSAIATRIPMGGTVSPRS
DDSPTECGSFAQLPIPENKVCPI
>Parapinopsin2
MQQSVFFSNASSYEGPNGEPPLSRTGFIILSIIMAFFTGPAIVLNATVIIVSLMHKQLRQPLNYALVNMAVAD
LGTAMTGGVLSVVNNAQGYFSLGRTGCVMEGFAVSLFGITSLCTVALIAVERMFVVCKPLGQITFQKKHA
LGGIALSWLWSLSWNLPPLFGWGRYELEGVGTSCAPDWHNRDPTNVSYILAYFAVCFAVPFALILASYTKL
MWTLHQVSKMACLEGGAVAKGEMKVASMVVLMVLTFLISWLPYASLALLVIYKPDVEIHPLVGTVPVYL
AKSSTMYNPIIYIYLNKQFRKYAVPFLLCGKEPLVDEEASEATTVETSPNKVSPA
>Parapinopsin
MERLPLHDGNSSADAELLSRSGYTLLAVIMGAFSVAGILLNVLVIVVTVRHRQLRQPLSYALVNLAVCDLG
CAMFGGLPTTVTSAMGYFSLGRVGCVLEGFAVAFFGISSLCTIAVISIERYIVVCPMGAVLFQTRHAVAGVV
LSWVWSFVWNTPPLFGWGSYELEGVEISCAPNWYSRDVGNMSYIIIYFSLCFAVPFSIIMVSYSRLLWTLRQ
VTKLQVSDAGSTNRVEVQVARMIVVMVLAFLVTWLPYAVMALAVIMDSSLYIDPVIATIPVYLAKTSTVY
NPIIYIFMNRQFRGYAVPTVLCGWNPWASDPQTSEEETTVASINRSQRVSSKGSLNE
>Parietopsin
MDSNSTPLSSGSPTPSIHAEMVTITPTIFPRVGYSILSFLMFINTVLTVFNNVLVITVLLRNPSLLQPMNVFILS
LAVSDLMIGLCGSLIVTITNYQGSFFIGHAACVFQGFAVNYFGLVSLCTLTLLAYERYNVVCRPRNSLKLSM
RRSIIGLLIVWIFCLFWAVAPLFGWSSYGPEGVQTSCSLAWEERSWSNYSYLILYTLLCFILPVAVIIYCYSKV
LTSMNKLNRSVELQGGRSSQKENDHAISMVLSMIIAFFVCWLPYTALSVVVVVDPALYIPPLVATMPMYFA
KTSPVYNPIIYFLSNKQFRDAALEMLSCGRYISHMPNTVSIGMRSLNRRSRLTSLSRNVNLHSKVLPL
>Opn3
MNPANETSPERSEDHSLFAVGTYKLLAFTIGTIGVFGFCNNVVVIILFCKFKRLRTPTNLLLVNISLSDLLVSV
IGINFTFVSCVNGGWSWSRATCTWDGFSNSLFGIVSIMTLAALAYERYIRVVHAQAVDFPWAWRAIGHIWL
YSLAWTGAPLLGWNRYTLEIHRLGCSLDWASKDPNDASFILLFLLACFFVPVGIMIYCYGNILYTVQMLRSI
QDLQTVQIIKILRYEKKVAVMFFLMISCFLLCWTPYAVVSMLEAFGRKNMVSPTVAIIPSIFAKSSTAYNPLI
YVFMS
>Tmt1a
MDSKMFSGQAVVNCSFNLSDDRELVDAAAGNAKLSPTGFVLSVILGFIMTFGFLNNLIVLLLFCKFKKLRTP
VNMLLLNISVSDMLVCLFGTTLSFASSIRGKWLLGRSGCSWYGFVNSCFGIVSLVSLAVLSYERYSTLMVC
NKRVSDYRKPLLAVGGSWLYSVVWTVPPLLGWSSYGLEGAGTSCSVTWTERSSGSHSYIVCLFLFCLGLP
VLVMVYCYGRLLYAVKQVLVGRIRRTSARRREFHILFMVLTSVVCYLLCWMPYGVVAMMATFGRPGVIS
PIAVVPSILAKSSTVINPVIYILMNKQFYRCFLILFRCEHQLTENAHSSMPSKTTVVQLHRRLGDSSPRHDAA
TGPGTAHVEQSSAAVSEKRLAAEPPD
>Tmt1b
MDSKMFSGQAVVNCSFNLSDDRELVDAAAGNAKLSPTGFVVLSVILGFIMTFGFLNNLIVLLLFCKFKKLR
TPVNMLLLNISVSDMLVCLFGTTLSFASSIRGKWLLGRSGCSWYGFVNSCFGIVSLISLVILSYDRYSTLAVY
NKRGADYRKPLLFIGCSWLYSLFWTVPPLLGWSSYGIEGAGTSCSVSWTVQTAQSHAYIICLFTFCLGLPIL
VMFYCYSRLLWAVKQVAKNRKSASRRREYHILFMVLTTAACYTLCWMPYGVVAMMATFGPPNIISPMAS
VVPSLLAKSSTVINPLIYILMNKQTFKFHILDVTVDGAEARKGDPSLPPEVILGACVDPLDQALLIQEGMVGT
QCTGSIVITLMVVAQFYRCFLILFHCDHWSAQNGNTSMPS
>Tmt2a
MVILLQHSNNLSTGHSSASDPAAASLGRAGHTAAAVVLGFILVLGFLGNFLVLTVFSRFPGLVTPVNVLLIN
ISASDMLICISGTPLSFAASVRGRWLTGSYGCRWYGFSNAFFGIVSLVSLSLLSLERYLALLVRTHSDSNPSQ
YRRARLAVAASWLYSLVWTLPPLLGWSSYGPEGPGTTCSVQWHQRSTAARSYISCLFVFCLLLPLLLMFFC
YGRILLAVRAVAGQVTRINRSSAERREGRVLLMVVSMVTGYLLCWMPYGVVAMLAAFGKPGVVPPAASL
IPSLLAKTSTVLNPVIYVLLNNQVPTIPCKHLTQRVHPKSLICRFLAPRSSLELEVVLRRHIDGRPKALISAPRS
EDRGARRDNRGAMSEDA
>Tmt2b
MVVHMRGCNFSTADSSPLSGSSLSTVATLRDSPASHGPGGLSRTGHTVVAVCLGFILVAGILSNFLALLIFA
KFRSLWTPINVILLNISLSDIVCVFGTPFSFAASLHGRWLIGEHGCKWYGFANSLFGIVSLVSLSVLSYERYT
AVLRSSQIDISDFRKAWLCVGGSWLYSLLWTLPPLLGWSSYGPEGPGTTCSVQWHLRSPTSVSYVLCLFIFC

LLLPFLLMVYSYGRILVAIRRVGKINLMAAQRREQHILTMVLSMVSCYMLCWMPYGIMALTATFGRSGLV
TPMASVVPSVLAKFSTVINPVIYMFFNNQFYRCFVAFMKCSSESQSVQGEEHPTPRTHFSGFSFVHRQASLG
SSQRQILGDSKSTALCSRNNDRHTLVVHYAP
>Tmt3a
MVVSNVSLSGCAGVNGALCAAAGEGHLRGGSHRSTLTPTGNLVVAVCLGFIGTLGLTNNLLVLVLFCRFK
MLRSPINLLLINISISDLLVCVLGTPFSFAASTQGRWLIGEGGCVWYGFANSLFGIVSLISLAVLSYERYSTMV
APTEADSSNYHKISLGITLSWVYSLIWTVPPLFGWSHYGPEGPGTTCSVDWTAKTANSISYIICLFVFCLIGPF
MVILFCYGKLLCAIRQVSGINPSMSRKREQRVLFMVVIMVICYLLCWLPYGIMALMATFGPPGMVTPVASII
PSVLAKTSTVINPVIYVFMNKQFYRCFKALLHCEVPRRGSSFKSSSKVATKAIRGVAVTGPRRTNEFLFMVA
SLGQPVATIPQLPSVEPTINITRGPSSDINKPVVVSLVAHFDG
>Tmt3b
MIVCNVSLSCAHCPGGGAGGTAATGTDAYEEVSGSLPPPTLSPRGHLVVAVCLGFIGTFGFLSNFLVMALF
CRYRALRTPMNLLLSISASDLLVSMLGTPFSFAASTQGRWLIGRAGCVWYGFVNACLGIVSLISLAVLSYER
YCTMMKPTIADGRDFRPALVGICFSWLYSMAWTVPPLLGWSKYGPEGPGTTCSVDWKTQTPNNISYIVCL
FTFCLVLPFGVILYSYGKLLQAIRQVRGVSSVVTRHREQRVLLMVVMMVVCYLVCWLPYGVAALLSTFGP
RDLLTPEASITPSLLAKFSTVINPFIYIFMNKQFYRCFRALLSCSVPERGSTLKTFSRPTKTLRTVRHDKEYHV
SAPAPSSALPTPDSIHRSSQGANHVSPSPSPSPSNRRSAASQTAAAAAAAANTPKPKLILVAHYRE
>Opn6a1
MSWFGNSIVLFVLCRQRASLQPTDYLTFNLAVSDASISVFGYSRGIIEIFNVFKDSDYLISSIWTCQVDGFFTL
VFGLSSINTLTVISITRYIKGCHPNRARHISRTSVSVSLLLIWITAGFWSGAPLLGWGSYKDRGYGTCEIDWA
KASYSSAYRSYIISIFIFCFFVPVLIMLFCYVSIINTVKRGNALSADGDLSDRQRRIERDVTIVSIVICTAFILAW
SPYAVVSMWSACGFHVPNLTSIFTRLFAKSASFYNPLIYFGLSSKFRKDVAVLLPCTRDARDSVRLKRFKPK
ADAHGRLKVPLNRSEKKYSPGYLDHPAAGTDPGYLDHP
>Opn6a2
MEITLKDFPVKVVNIPWRNNNLSNLNTDPPLSEQGETFIGVYLLVLGWLSWFGNSLVMFVLYRQRASLQST
DFLTLNLAISDASISIFGYSRGILEIFNIFKDDGYLITWIWTCQVDGFFTLLFGLASINTLTVISVTRYIKGCHPN
KAYCISLNTIAVSLICIWTGAMFWSVAPLLGWGSFTDRGYGTCEVDWSKANYSTIHKSYIISILIFCFFIPVMI
MLFSYVSIINTVKSTNAMSADGFLSTRQRKVERDVTRISIVICTAFIMAWSPYAVVSMWSAWGFHVPSTTSII
TRLFAKSASFYNPLIYFGMSSKFRKDVSVLLPCTRERKEVVHLQHFKNIKPMAEVPPPPAPLPVQKLEAKYA
PSKPDSDSGVNSTPQTPPTDPQGVLYSDLPSHIETSEYWCDRL
>Opn4m1
MVTEAPHPFPTVDVPDHAHYTIGSVILAIGITGMIGNFLVIYAFSKSRSLRTPANMFIINLAITDLLMCVTQAP
IFFTTSMHKRWIFGEKGCELYAFCGALFGICSMITLTVIAVDRYFVITRPLTSIGVMSRRRAFFILMGAWAYS
LGWSLPPFFGWSAYVPEGLLTSCTWDYMTFTPSVRAYTMLLFIFVFFLPLFIIIYCYVFIFRAIRSTNQSVGKI
NGSIHSHGSTRDSVRNFQRLQNEWKMAKIALIVILLYVVSWSPYSTVALTAFAGYADILTPYMNSVPAVIA
KASAIHNPIIYAITHPKYRIALAKYIPCLGILLCVHPRDLRSASSSFMSTRRSTVTSQTSDISSQLRRQSNRKSR
LGMTDTEADLSSMGSRPASRQVSCDISRDTTELPD
>Opn4m2
MSPPPPTRRTTLCSPGGNCTSFLTGALLHLPGRAPTPHAHLQPHLFPTVDVPDHAHYIIGSVVLLVGVTGML
GNALVVYVFCRCRMLRSPSNLLVVNLAVADFLMSLTQTPVFFVASLHRRWVFGELACELYAFCGALFGIT
SMMTLTAIAVDRCLAITRPLALPGWASRGRLLAVLGGVWLYSLGWSLPPFFGWSAYVPEGLQTSCSWDY
MSFTAAVRTYTLLLFTFVFFIPLGLIAGCYFAIFHAVRRAGREVGRLSGGETDKVYERLRSEWRLAKVALG
VILLFVVSWSPYSAVALIATAGYAHLLTPYMNSVPAVIAKASAIHNPIIYAITHPKYRAAIGRYVPLLRPLLR
LQDKDLRSSFSSGASSRRATLSSQSSSGVWLGAGVRGNSRFGKSRLSSASDSESCQTESEADGCSAGSTAFN
RRVSTDVTANATANE
>Opn4m3
MDSRLLIGPGVPTPDPTARAPWNISSISPHRLMEMSPVSTAGSVVSINPFPTVDVPDHAHYTIGVVILVVGIT
GMLGNFLVIYAFCRSRSLRTPSNIFIINLAVTDFLMCFTQTPIFFITSMHKRWIFGKKGCEMYAFCGALFGICS
MMTLMVIAVDRYVVITRPLASLGAMSRRKALSILAAAWVYSAGWSLPPFFGWSAYVPEGLMTSCSWDYM
TFTPSVRSYTMLLFTFVFFIPLFIIIFSYCCIFRAIRHTTRAITKINCHGTRDSAKRFHKMKTEWKMAKIALIVIL
LFVISWAPYSCAALTAFAGYADLLTPYMNSVPAVIAKASAIHNPIIYAITHPKYRSALGRYIPYLGVLLCVPG
RGRFSSSSFQSTRRSTLTSQSESKGKPRTSSLSESESARFSDTEEDSLSRTPASQLSGVRHANQRTKVRGHNT
GEFERTAAHNPTDPAICHLSQITTLTEHEDISMSDISLKPPGHLPATLDILKEESRPRTGRKTTASVIVTSISSPS
ILHSPPGFHGNLKLTKTYSPVVKEPLDITRLETTALP
>Opn4x1

MDVDHGFYRQVDVQAHAHYIVAFFVLVIGTVGVTGNALVMYAFFSNKKLRTAPNFFIMNLAVSDFLMAIT
QSPIFFVNSLYKGWIFGETGCKIYAFCGALFGITSMINLLAISLDRYIVITKPLQAIWTSKKRTCFIIALVWLYS
LAWSLAPLLGWSSYIPEGLMTSCTWDYVTSTPANKSYTLMLCCFVFFIPLGIISYCYLCMFLAIRHASKDIEK
LGYQVRKSTLIQQQSIKTEWKLAKIAFVVIIVFVLSWSPYACVTLIAWAGYGSYLSPYSKAVPAVIAKASAI
YNPFIYAIIHSKYRDTMAEKVPCLHFLAQAPRRKCISVSHSESSLRDLMLSRQSSDSRAKFHRVSSMSTTDTY
YNYKADVWSDVELDPMDQSHGVRSSYSLGALKDRDQKSLAKKTNEKGSQSQEQMHIPERGSLNRSDHTL
SPECVNTVMPPALGTRREDNPKCWNKETKQEVEKEGEHEQRETQEAEKEQQVLIQPDSLDNVVCVHSEAII
RPDSRNGSLDHSYCEEKQSTQSRGRELLLSLKPLDCSAQELESVQRFLS
>Opn4x2
MWSTERMEPEKANTQSSFFTKVDVPDHAHYIVALFVFVIGTLGITGNALVMFAVYSNKKLRNLPNYFIMNL
AVSDFLMAFTQSPIFFINCLYKEWAFGEMGCKMYAFCGALFGIASMINLLAISIDRYLVITKPLQTIHWGSKR
RTTLAILMVWLYSLAWSLAPLVGWSSYIPEGLMTSCTWDYVTYTLANRSYTMMLCCFVFFIPLGIIFYCYFF
MFLAVRKTSREVERLGTQVRKSTLIKQKSMRSEWKLAKIAFVVIVVYVLSWSPYACVTLISWAGHADILSP
YSKAVPAIIAKASAIYNPFIYAIIHNKYRMTLAEKFTCLRFLSPTPRKECISSSISESSFRDSIISRQSTASRTHFIT
ACPDMVLKDVEMEPLGRKSGDSFRSIPSYRQSSRGRSYKKQLQPKTQALEKVRNTKYKNCDFSLHVFWVI
EKTFSFSHCQHPSTMREPSNTCDHELVSSSLTVASLPLLVLTRRRSQSLTSEISDAWDRKGSIGDYHKSDSFD
SLNFGKIPSTDPRSPQGVPRIIVISPTSESSLIKHDSVCLEDSVEAMENNIFVSLNFSSEVFAAVELLP
>Opn5
MKQRLIQNAAALLVFNLPKFSYTTPLLHKPELAPRGCSNPIKDTGILSATGNGYVIYMTIKRKTKLKPPEIMT
VNLAVFDFGISVTGKPFFVVSSFAHRWLFGWEGCRFYGWAGFFFGCGSLITMTVVSLDRYLKICHIRYGTW
LKRQHAFLCLAFVWVYAAFWATMPLVGWGNYAPEPFGTSCTLDWWLAQASVSGQSFVMAILFFCLVLPT
GIIVFSYVMIIFKVKSSAKEISHFDARIRNSHNLEMKLTKVAMLICAGFLIAWIPYAVVSVVSAFGEPDSPISV
SVIPTLLAKSSAMYNPIIYQVVDLKNSCAKSSCLKGLNKRRHFRKSRFYTISGSLKDNTPAKEAHIEIATKYIL
QLFSHGLLAERLYIEVWIQKTICMIEESASLVFILMFHLPKVLCLDGLVGFYLASVVMDLSPAVVPIFFTPLL
AIP
>Opn9
MGENGSQKGSWFVPLSIHPLFLSQLSPSTDLAVAVFLSFTGVASVLGNGMVLLIYCRRRKKLRPSELMTINL
ALCDFGFSVLGGPFFITSSLCHAWLFGEMGCLCYGIQGFMFGIGSLLTTCLISLERCLKICCLRYGKGQWIER
RHVCLSIVLVWVYTLFWALLPAFGFGSYGPEPYGTSCTINWWSMRSSLNDRIYIFLILTLCFGFPTLTIVASY
LAILLTVYRSNRTLASIPSSSVRNNGKDLRLTKVAAVVCSTFLLAWMPYAAVSLLSALMPRDDQEASLQTA
AEESTDISLLLNWTAAEYYRQIYYHHENKWSNVNDMNSASITGLSEATFRSMLDKEAEPMTGSHQPFSSLP
PLVTLIPAMFAKFHSTINPFIYHIMNREFRDDVYAMLFGQEK
>Opn7a
MNGFLHHKQPIANQTVTDKVKSVDRQDRQNSNSFQMGPSDEDIAFRSNIPVPLDITVAVVYSVFGVCSLFG
NTTLLYVSYKKKHILKPAEYFIINLAISDLGLTLSLYPMAITSSFYHWLYGKTMCSVYAFCGILFGLCSLTTL
TLLSIVCFVKVCYPFYGNRFNSVGWLLIACAWLYALLFACSPLAHWGKYGPEPYGTACCIDWRLSNQHST
ARSYTVVLFILCYILPCCAIVASYTSILVTVHASRKNLEQHVSRLTHMSSIQTIIVKLSVAVCISFFAAWSPYA
VVSMWAAFGHIENIPPLAFALPAMFAKSSTIYNPIIYLMLRPNVRRMMYRDLGTLCNACLKDCFCHHWPA
KCCSKPEITVRIPSIHRQANQFPSSNSSAQPPIAALKDHRCEKCKGAFECFRHYPQMCNIANFAANNASMDQ
DNQVSQTHKKKTQSACHKRSLLAIMCAKRISEKDNFQINLEMVPRQAKVAWP
>Opn7b(cd)
MGNASDTSAVFASTISKEHDILMGSLYSLFCVLSLMGNCILLLVAYHKRSTLKPAEFFIINLSISDLGMTLSLF
PLAIPSSFKHRWLFGEITCQLYAMCGVLFGLCSLTNLTALSLVCCLKVCLPSHGNKFSSSHARLLVVGVWC
YASVFAVGPLAQWGRYSPEPYGTACCIDWHAPNHEFSALSYIVCLFVFCYALPCTAIFLSYTFILLTVRGSR
QAVQQHVSLQTKTTNAHALIVKLSVAVCFGFLGAWTPYAVVAMWAAFGDATLVPPDAFALAVFAKSSTI
YNPMVYLLCKPNFRECLYRDTSMLRQRIYRGSPQSDPRANFGSTSQRNKDMSVSMRFSNGQQGSYGAFLH
CPENAAPCHVTTPQRTACILTESTYREVTVNQLSDKPQADFL
>Opn8a
MATLSVFGNMAVLVSASRRAALLKAPELLTVNLAVTDIGMALSMYPLSIASAFNHAWIGGDTSCLYYGLM
GMIFSVTSIMTLAVMGMVRYLVTGSPPKTGIKFQRRTISIVISAIWLYASLWAIFPLLGWGSYGPEPFGLACSI
DWNGYGDSLNRSTFIMTLSVLCTFIPCLVILFTYFGIAWKLHRAYQSIQSNDFQYGNVEKKITLVGVMISSGF
LFAWTPYVAVSFWSMFHSQEHITPFVTLLPCLFAKSSTAYNPFIYFIFKRTSWQELLHLQRRIFCCSHRANSP
AEGRKLKGDMVFENGMDETCVGPGRQSWSQMMPLG
>Opn8b
MEELLSGLSLAGELEEKCILESSCANGTSREEEEEEPGAQKAVFSIVGNLLVLVMAVKRSERMKPPELLSVN
LAVTDLGAAVTMYPLAVASAWRHRWVGGDATCVYYAVAGFFFGIASIMNLTGLAIVRFIVSLNLQSPNEK

IGWRKVKLLCLWTWLYALIWAMFPFMGWGRYGPEPFGLSCSLAWGQMKHEGFYFVVSMFSFNLVIPSVII
VCCYFGIAIKLYFTYRRSETNSNRIPNIVKLHRRLLVIAVLISLGFIICWSPYAIVSMWSILHESSSIPPQVSLLP
CLFAKSSTVYNPLIYYIFSQSFRKEVKQLFLYFSSPICKVSNSINDTSIYMVSTNVKSNVHVVSTTQEIPE
>Opn8c
MTSSDNRTISAFTSKLTPAADICVGLAIVSVVVLSVLGNGLVLVICYRRRKKMVGSELLCVNLAVVDFLCCI
CFYPLSIMSSFNHMWLGENITCVYYGLGCFIFGLCSMFTITAISITRYLKTCIPVYGEGWLEGTNIRKACCAI
WLVAAVWSSFPLFGWGEYVPEPYGLSCTIAWRGYHTSAKDAFYVICSFACFTMVPVLLIVVSQCQILYKVS
RFSNSLAARGIRNKLRHAEKRLSMMFFCISLGFVVAWAPYAVVSLLFIFHRENQYMAPGGFVFPALFAKSS
HIYNPFIYFYFNKTFRKELRCLILSLWPKLGGNQVGVHIDTVHQLPIHIQLQERACVQKRTFSLTQDRTCSKG
KGKGKGKGKVLNSRSNPVPGRPVYACWGSTSNNAPITLDNKVAKDSPSVSI
>Peropsin
MGIDSGVNASDDVSLYGGKSAFTQLEHNIVAGYLITAGVISLASNIVVLLMFVKFKELRTATNFIIINLALTDI
GVAGIGYPMSAASDIHGSWKFGYTGCQIYAALNIFFGMASIGLLTVVAIDRYLTICRPDLGQKMTMRSYNL
LILAAWLNAVFWSAMPVVGWAGYAPDPTGATCTINWRQNDASFISYTMAVIAVNFVLPLSAMFYCYYNV
SATVRRYKASNCLSGVNIDWSDQMDVTKMSIVMIVMFLVAWSPYSIVCLWASFGDPKTIPAPMAIIAPLFA
KSSTFYNPCIYVIANKKFRRAIIGMVRCQTRQRITINTQVPMTTSQQPLTQ
>Rgr1
MVASYPLPDGFTDFDVFSLGSCLLVEGLLGFFLNAVTIAAFLKVRELRTPSNFLVFSLAMADIGISMNATVA
AFSSFLRYWPYGSDGCQTHGFQGFVTALASIHFIAAIAWDRYHQYCTRTKLQWSSAITLAVFVWLFTAFWS
AMPLIGWGEYDYEPLRTCCTLDYTKGDRNYVSYLIPMSIFNMAIQVFVVMSSYQSIAQKFKKTGNPRFNPN
TPLKMLFCWGPYGLLAFYAAVENATLVSPKLRMLAPILAKTSPTFNVFLYALGNENYRGGIWQFLTGEKIE
VPQIENKSK
>Rgr2
MAAFTLPEGFEFDMFTFGSALLVGGMLGFFLNAISIISFLSVKEMRTPSNFFVFNLAVADICLNINGLTAAYS
SYLRYWPFGQDGCACHGFQGMIAVLASISFMAAISWDRYHQYCTRQKLFWSTTLVMSGIIWILIFWAAVPL
MGWGVYDFEPMRTCCLDYTRGDRDYVTYMLTLCLLYLMFPAFTMLSCYDAIHKQFKKIHHHRFNTNVPL
RVMLMCWGPYVLMCVYACFENVKVVSPKLRMVLPVLAKTNPIFNALLYSFGNEFYRGGVWNFLTGQKIV
DPVIKKSN

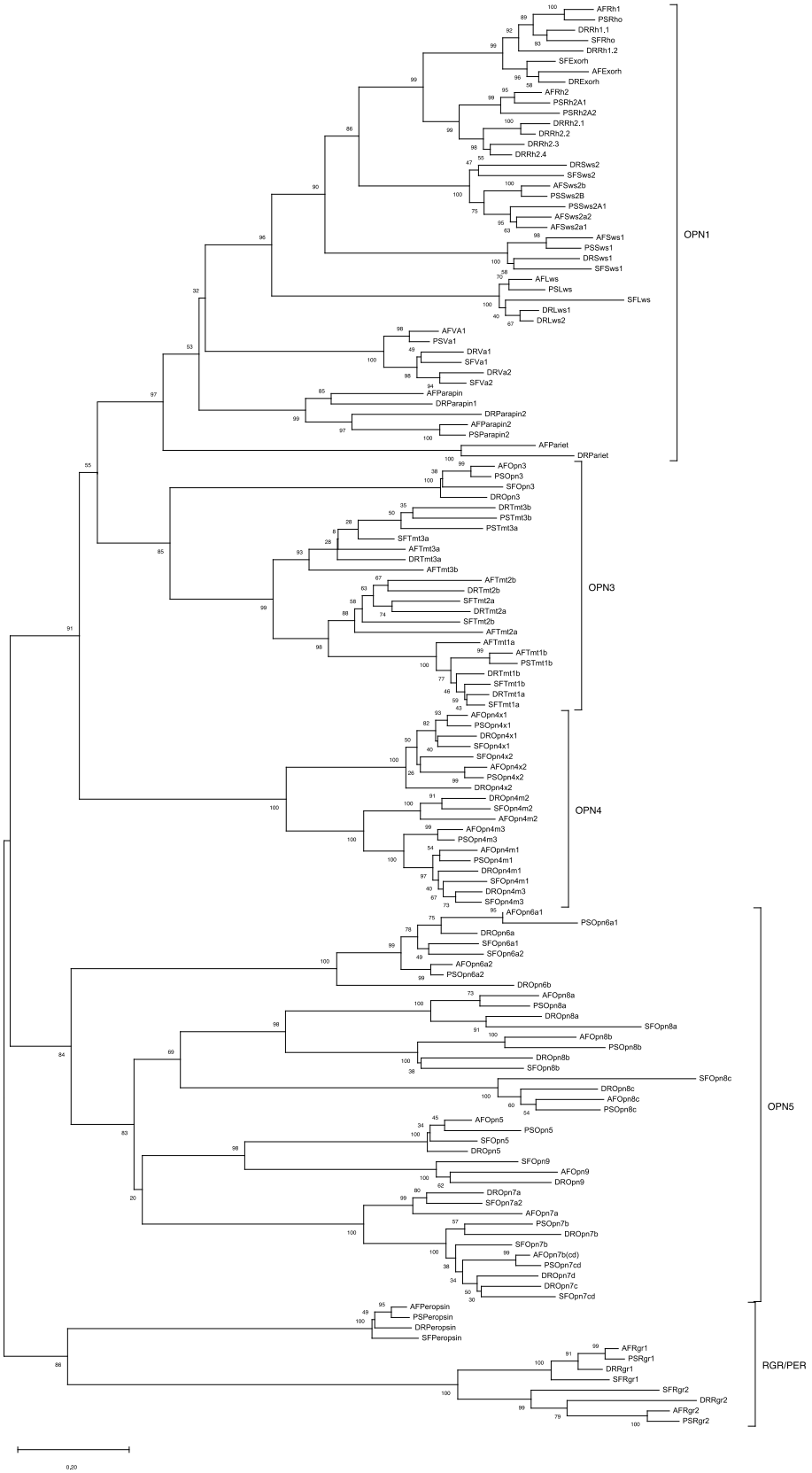
Appendix B. Sablefish opsins and accession numbers of contigs/scaffolds in the WGS database from which they came. Accession numbers are provided for both draft genomes from 2013 (GCA_000499045.1) and 2020 (GCA_000499045.1).

Opsin gene	Genome version	Accession numbers
Rh1	2013	AWGY01146807.1
	2020	AWGY02005742.1
Exorh	2013	AWGY01116468.1
	2020	AWGY02009385.1
Rh2	2013	AWGY01104235.1
	2020	AWGY02002618.1
Sws1	2013	AWGY01006425.1
	2020	AWGY02013549.1
Sws2B	2013	AWGY01100787.1
	2020	AWGY02001820.1
Sws2A1	2013	AWGY01147435.1, AWGY01191923.1
	2020	AWGY02003541.1
Sws2A2	2013	AWGY01147435.1
	2020	AWGY02001041.1
Lws	2013	AWGY01150092.1
	2020	AWGY02003541.1

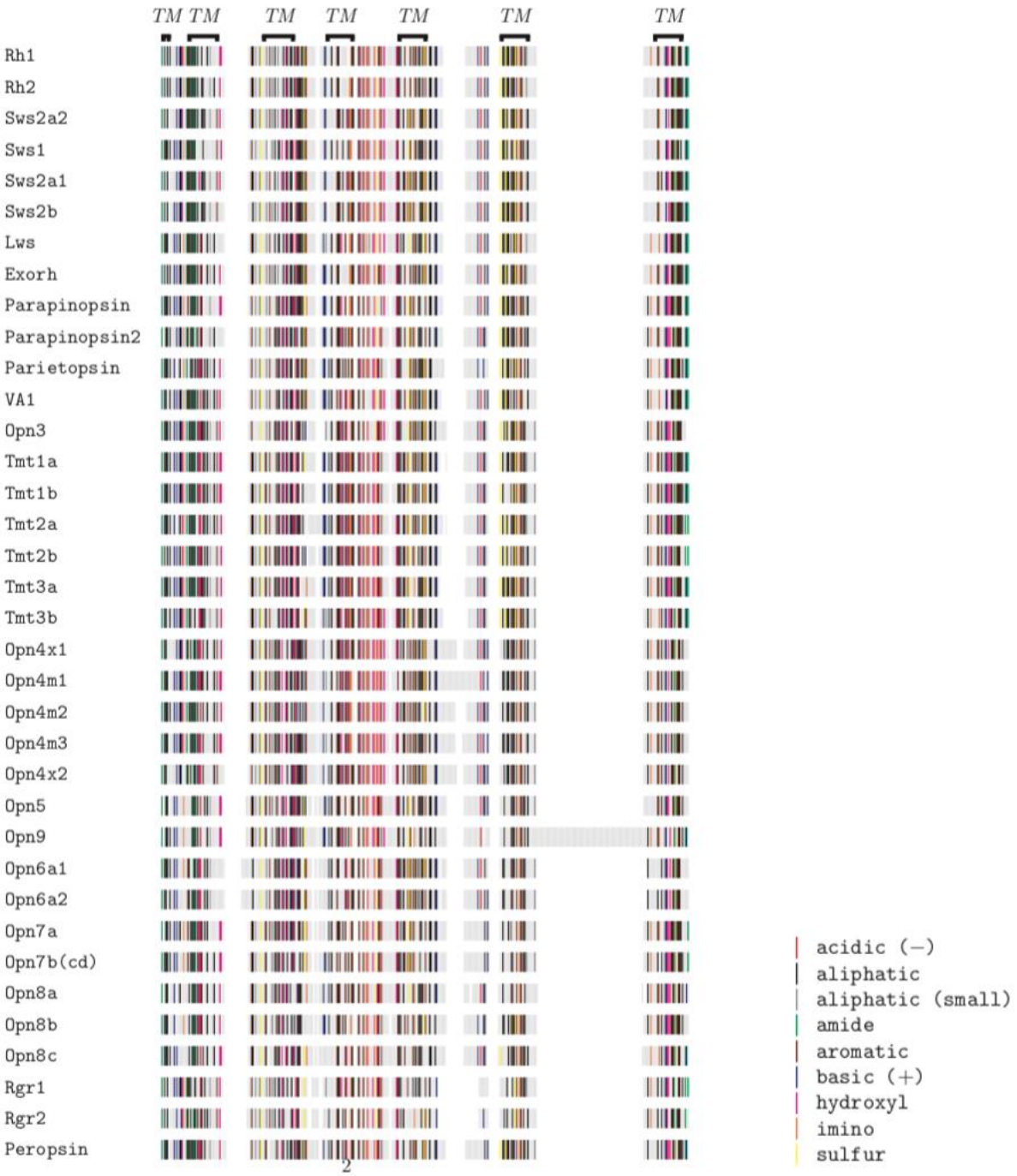
Va1	2013	AWGY01121641.1, AWGY01143958.1, AWGY01016223.1, AWGY01185742.1
	2020	AWGY02003497.1
Parapinopsin	2013	AWGY01178070.1
	2020	AWGY02010756.1
Parapinopsin2	2013	AWGY01194211.1
	2020	AWGY02000738.1
Parietopsin	2013	AWGY01129323.1
	2020	AWGY02006640.1
Opn3	2013	AWGY01098822.1, AWGY01136267.1
	2020	AWGY02006359.1
Tmt2b	2013	AWGY01193267.1, AWGY01193266.1
	2020	AWGY02002760.1
Tmt2a	2013	AWGY01143362.1
	2020	AWGY02007072.1
Tmt1a	2013	AWGY01118159.1, AWGY01168828.1
	2020	AWGY02001068.1, AWGY02008073.1
Tmt1b	2013	AWGY01062929.1, AWGY01042679.1
	2020	AWGY02005339.1
Tmt3a	2013	AWGY01080096.1, AWGY01141120.1
	2020	AWGY02003291.1
Tmt3b	2013	AWGY001050406.1, AWGY01028286.1
	2020	AWGY02007494.1
Rgr1	2013	AWGY01080391.1
	2020	AWGY02006468.1
Rgr2	2013	AWGY01030941.1
	2020	AWGY02006519.1
Peropsin	2013	AWGY01153505.1, AWGY01153506.1
	2020	AWGY02002813.1
Opn6a1	2013	AWGY01058453.1
	2020	AWGY02003527.1
Opn6a2	2013	AWGY01162125.1, AWGY01009804.1
	2020	AWGY02002783.1
Opn8a	2013	AWGY01030717.1
	2020	AWGY02016171.1
Opn8b	2013	AWGY01091392.1
	2020	AWGY02016171.1
Opn8c	2013	AWGY01036237.1, AWGY01036630.1
	2020	AWGY02004488.1
Opn5	2013	AWGY01176470.1
	2020	AWGY02000093.1
Opn9	2013	AWGY01054469.1
	2020	AWGY02000439.1
Opn7a	2013	AWGY01152197.1
	2020	AWGY02000038.1
Opn7b(cd)	2013	AWGY01203930.1

	2020	AWGY02003387.1
Opn4m3	2013	AWGY01103679.1
	2020	AWGY02000384.1
Opn4m1	2013	AWGY01160359.1
	2020	AWGY02006067.1
Opn4m2	2013	AWGY01080690.1
	2020	AWGY02011360.1
Opn4x1	2013	AWGY01011920.1
	2020	AWGY02000315.1
Opn4x2	2013	AWGY01076119.1
	2020	AWGY02003321.1, AWGY02012153.1

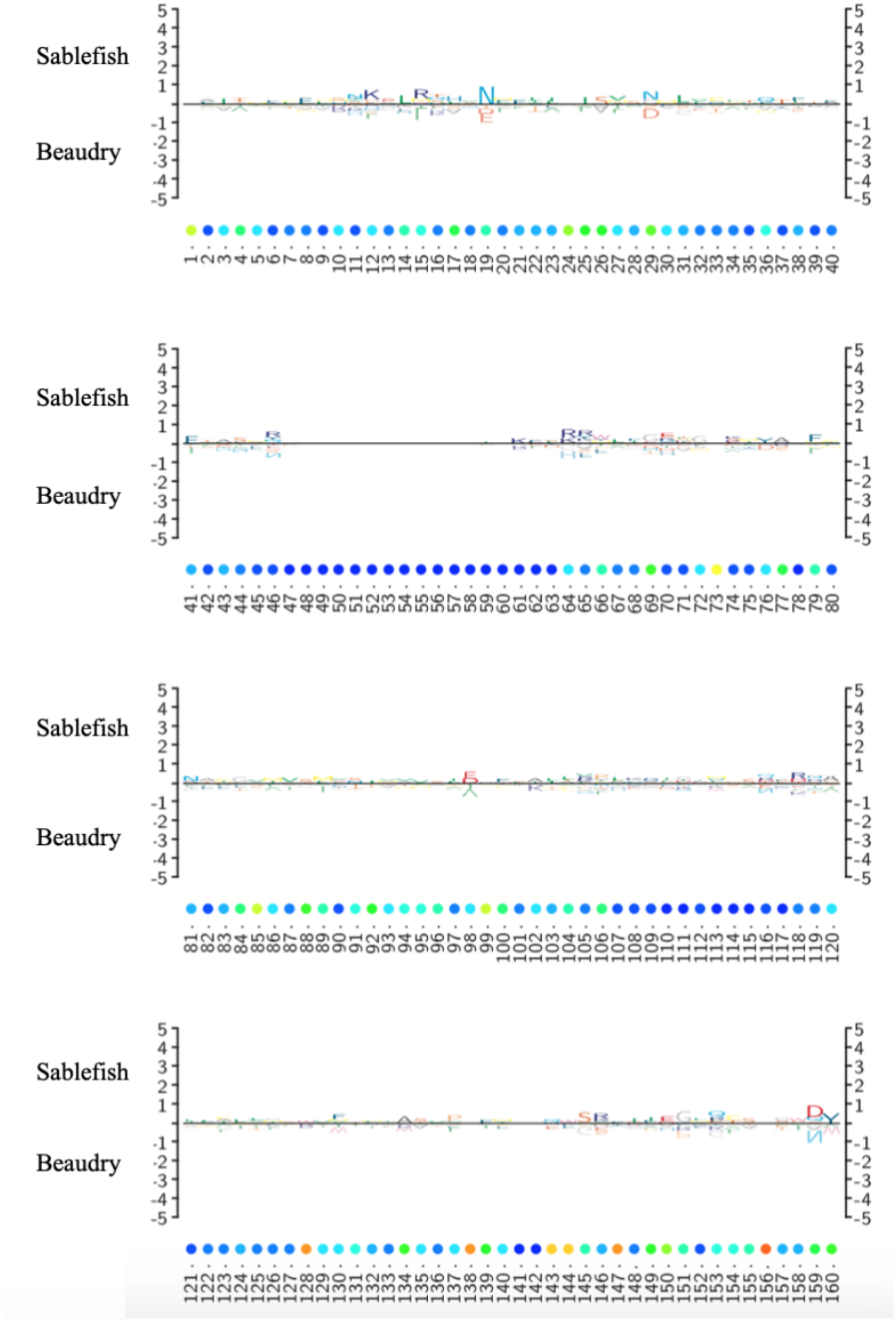
Appendix C. (Continued on next page) Evolutionary relationships among sablefish (*Anoplopoma fimbria,* AF), zebrafish (*Danio rerio*, DR), starry flounder (*Platichthys stellatus,* PS), and Asian arowana (*Scleropages formosus,* SF) opsins. Genetic distance was computed using the Jones-Taylor-Thornton (JTT) amino acid substitution model and was reconstructed from these distance data using the Neighbor-Joining method. The root was placed at the base of the Rgr/Peropsins. Evolutionary analyses were conducted in MEGA. The percentage of 800 replicate trees in which the associated taxa clustered together in the bootstrap test are shown next to the nodes. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The zebrafish repertoire was obtained from Davies et al. (2015); Asian arowana and starry flounder from Beaudry et al. (2017). The fish repertoires from Beaudry et al. were not comprehensive and some genes for these species are likely to be missing from the tree.


AFRh1
PSRho
DRRh1.1
SFRho
DRRh1.2
SFExorh
AFExorh
DRExorh
AFRh2
PSRh2A1
PSRh2A2
DRRh2.1
DRRh2.2
DRRh2.3
DRRh2.4
DRSws2
SFSws2
AFSws2b
PSSws2B
PSSws2A1
AFSws2a2
AFSws2a1
AFSws1
PSSws1
DRSws1
SFSws1
AFLws
PSLws
SFLws
DRLws1
DRLws2
AFVA1
PSVa1
DRVa1
SFVa1
DRVa2
SFVa2
AFParapin
DRParapin1
DRParapin2
AFParapin2
PSParapin2
AFPariet
DRPariet
OPN1
AFOpn3
PSOpn3
SFOpn3
DROpn3
DRTmt3b
PSTmt3b
PSTmt3a
SFTmt3a
AFTmt3a
DRTmt3a
AFTmt3b
AFTmt2b
DRTmt2b
SFTmt2a
DRTmt2a
SFTmt2b
AFTmt2a
AFTmt1a
AFTmt1b
PSTmt1b
DRTmt1b
SFTmt1b
DRTmt1a
SFTmt1a
OPN3
AFOpn4x1
PSOpn4x1
DROpn4x1
SFOpn4x1
SFOpn4x2
AFOpn4x2
PSOpn4x2
DROpn4x2
DROpn4m2
SFOpn4m2
AFOpn4m2
AFOpn4m3
PSOpn4m3
AFOpn4m1
PSOpn4m1
DROpn4m1
SFOpn4m1
DROpn4m3
SFOpn4m3
OPN4
AFOpn6a1
PSOpn6a1
DROpn6a
SFOpn6a1
SFOpn6a2
AFOpn6a2
PSOpn6a2
DROpn6b
AFOpn8a
PSOpn8a
DROpn8a
SFOpn8a
AFOpn8b
PSOpn8b
DROpn8b
SFOpn8b
SFOpn8c
DROpn8c
AFOpn8c
PSOpn8c
AFOpn5
PSOpn5
SFOpn5
DROpn5
SFOpn9
AFOpn9
DROpn9
DROpn7a
SFOpn7a2
AFOpn7a
PSOpn7b
DROpn7b
SFOpn7b
AFOpn7b(cd)
PSOpn7cd
DROpn7d
DROpn7c
SFOpn7cd
OPN5
AFPeropsin
PSPeropsin
DRPeropsin
SFPeropsin
AFRgr1
PSRgr1
DRRgr1
SFRgr1
SFRgr2
DRRgr2
AFRgr2
PSRgr2
RGR/PER
0,20

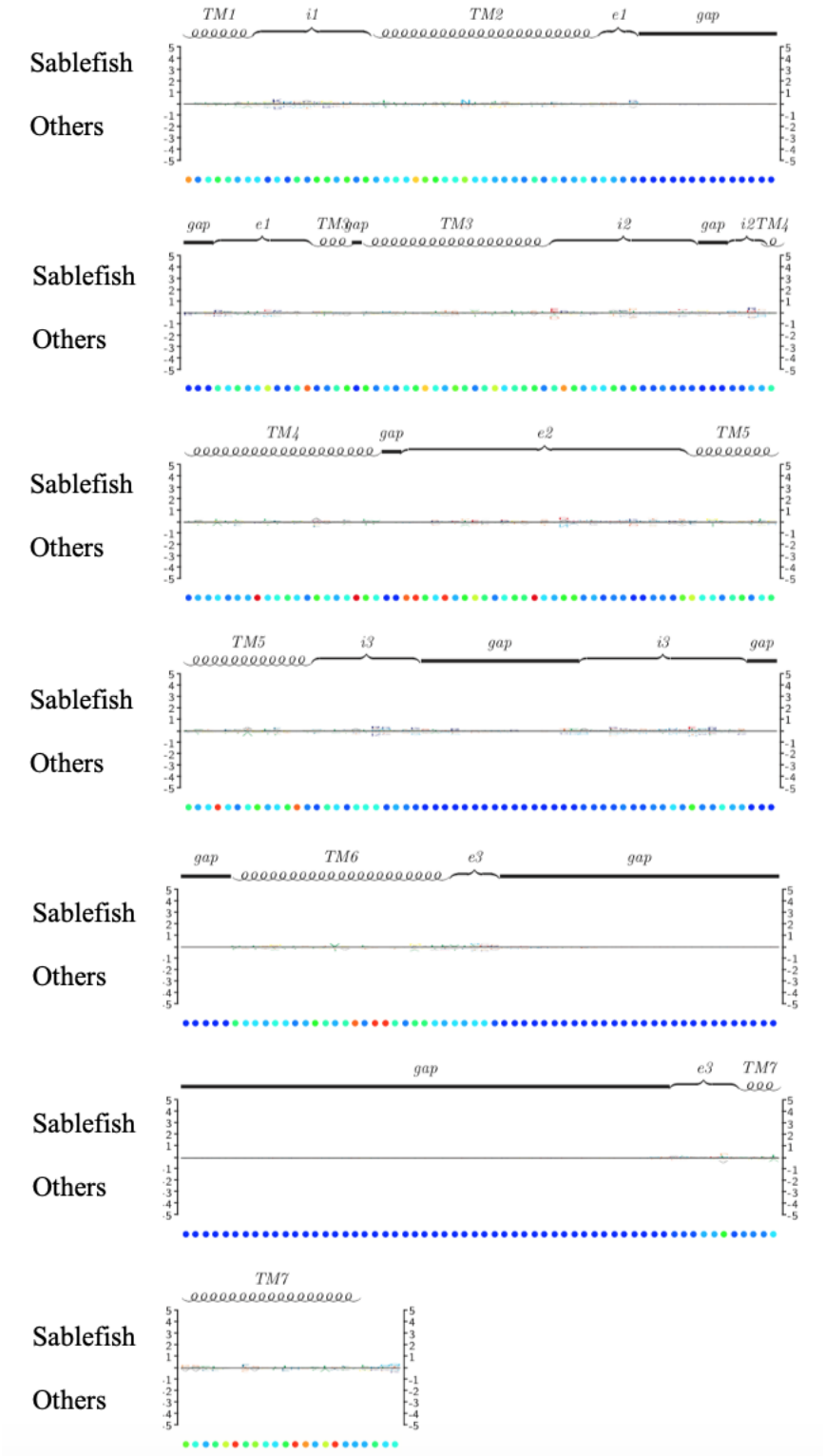
Appendix D. TEXshade fingerprint showing all sablefish opsins. Each row depicts an opsin sequence and each vertical line along it represents one amino acid. The shading shows the different chemical properties of amino acids among all of the 36 sablefish opsins. The transmembrane regions depicted represent those of *Bos taurus* rhodopsin.



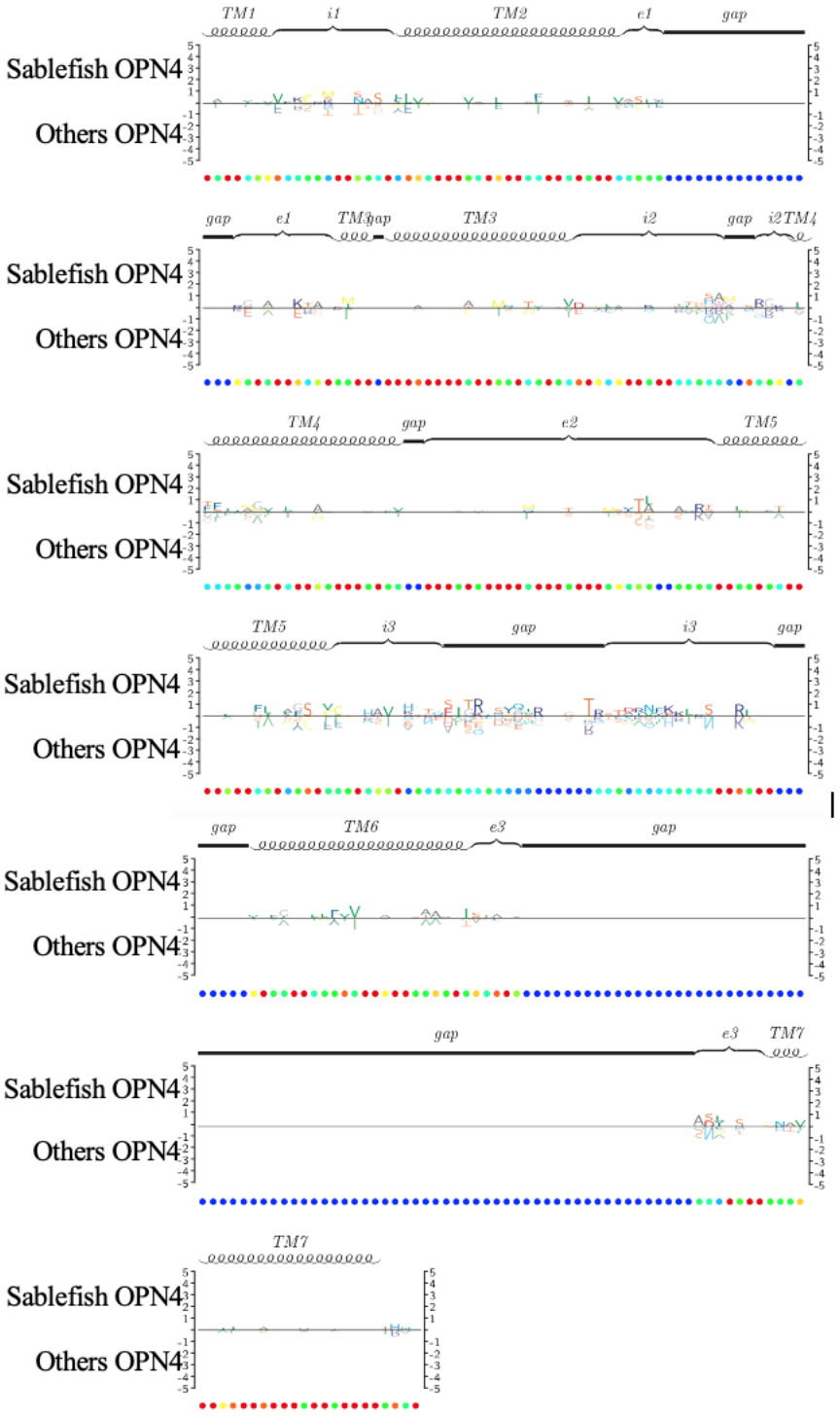
Appendix E. All sablefish opsins compared to the Beaudry et al. (2017) alignment, filtered to remove sequences with over 75% sequence redundancy, using TEXshade subfamily logo in an atypical way (see Methods). This is cropped from the full alignment to show only the first 160 positions in the alignment. Proximity to the mid-line (as is shown at every position here) indicates that there are no meaningful deviations in sablefish opsins and those of the other fish species. The Y-axis is measured in bits. The ruler below shows the position in the alignment, not the position according to bovine Rho. Conservation is shown as bullets on a hot-cold scale, with red meaning high conservation in the entire alignment at that position.



Appendix F. A full-length version of Figure 4, showing all sablefish opsins compared to those of Asian arowana, starry flounder and zebrafish using TEXshade subfamily logo.



Appendix G. Full length TEXshade subfamily logo showing sablefish Opn4 opsins compared to Asian arowana, starry flounder and zebrafish Opn4's.



Appendix H. Complete list of non-detects from Chapter 3's qPCR experiments.

Experiment	Tissue	Life stage	Opsin (#non-detects out of N=6)
Juvenile eye vs. brain (Figure 21)	Eye	Juvenile	*Opn4m2*(2) *Opn4x1*(1) *Opn4x2*(2)
	Brain	Juvenile	*Rh1*(4)
Opn4 in brains - all life stages	Brain	Yolk-sac larva	*Opn4x2*(4) *Opn4m2*(1)
		Larva	*Opn4m2*(1) *Opn4x1*(1) *Opn4x2*(2)
		Juvenile	*Opn4m2*(3) *Opn4x2*(3)
		Wild adult	*Opn4x2*(6)

Appendix I. Table summarizing statistically significant differences in whole brain Opn4 expression among paralogs at a given life stage. A series of one-way ANOVA determined whether the mean log10(RQ) if expression of the different Opn4 paralogs differed within each life stage taken alone (H_o: $\mu_{Opn4m1}=\mu_{Opn4m2}=\mu_{Opn4m3}=\mu_{Opn4x1}=\mu_{Opn4x2}$). Tukey Multiple Comparisons revealed *which* melanopsins differed from the others at that life stage in a pairwise manner . Statistical significance was set to 0.05. P-value: >0.05 (ns), <0.05(*), <0.01(**), <0.001 (***), <0.0001(****).

Lifestage	One-Way ANOVA	Tukey Multiple Comparison
Yolk-sac larvae	****	*Opn4m1-Opn4m3* ** *Opn4m1-Opn4x1***** *Opn4m1-Opn4x2**** *Opn4m2-Opn4m3** *Opn4m2-Opn4x1***** *Opn4m2-Opn4x2**** *Opn4m3-Opn4x1***** *Opn4m3-Opn4x2*****
Larvae	****	*Opn4m1-Opn4x1*** *Opn4m1-Opn4x2***** *Opn4m2-Opn4x1*** *Opn4m2-Opn4x2***** *Opn4m3-Opn4x1*** *Opn4m3-Opn4x2***** *Opn4x1-Opn4x2**
Juveniles	**	*Opn4m1-Opn4x2** *Opn4m2-Opn4x2*** *Opn4m3-Opn4x2**
Sea pen adults	****	*Opn4m1-Opn4m2*** *Opn4m1-Opn4x1***** *Opn4m1-Opn4x2***** *Opn4m2-Opn4m3*****

		*Opn4m2-Opn4x1****** *Opn4m2-Opn4x2****** *Opn4m3-Opn4x1****** *Opn4m3-Opn4x2****** *Opn4x1-Opn4x2******
Wild adults	*	*Opn4m1-Opn4x1**