Duplicated genes in a divided retina: Opsin gene expression in the four-eyed fish, *Anableps anableps*

by

Gregory Lawrence Owens
B.Sc., University of Victoria, 2008

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

MASTER OF SCIENCE

in the Department of Biology

Supervisory Committee

Duplicated genes in a divided retina: Opsin gene expression in the four-eyed fish, *Anableps anableps*

by

Gregory Lawrence Owens
B.Sc., University of Victoria, 2008

Supervisory Committee

Dr. John S. Taylor (Department of Biology)
Supervisor

Dr. Robert L. Chow (Department of Biology)
Departmental Member

Dr. Steven J. Perlman (Department of Biology)
Departmental Member

Abstract

Supervisory Committee

Dr. John S. Taylor (Department of Biology)
Supervisor

Dr. Robert L. Chow (Department of Biology)
Departmental Member

Dr. Steven J. Perlman (Department of Biology)
Departmental Member

The filtering of light by water is contingent on depth, direction and clarity. Consequently, fish must contend with a much more variable spectral world than terrestrial species. The gene family responsible for light sensitivity, the opsins, has expanded in fish. The duplication events responsible for large fish opsin gene repertoires have been characterized as part of this thesis research. The four-eyed fish, *Anableps anableps*, swims at the surface with its eyes at the waterline. Among many unusual adaptations, these eyes have two pupils, one above and one below the surface, giving it simultaneous access to broad spectrum aerial light and filtered aquatic light. It also has a nine cone opsin genes including duplications in three of the four cone opsin subfamilies. *In situ* hybridization was used to localize opsin transcripts in the retina. My data show that *A. anableps* expresses SWS1, SWS2 and RH2 opsins and has broad spectral sensitivity across its entire retina. In addition, I discovered that the region of the retina exposed to aquatic light expresses LWS and is, therefore, additionally red sensitive to match the longer wavelength available in cloudy water. By comparing this pattern with its normal eyed sister species, *Jenynsia onca*, I found that this increased red sensitivity is accomplished through the reduction of green sensitive pigments, which in *A. anableps*

(but not *J. onca*) are expressed only in the ventral region of the retina that is exposed to aerial light.

Table of Contents

List of Tables

List of Figures

List of Appendices

List of Abbreviations

BAC:	bacterial artificial chromosome
bp:	base pairs
DIG:	digoxigenin
dN/dS	ratio of the rate nonsynonymous to synonymous substitutions
E3:	extracellular domain 3
ExoRh:	exo-rhodopsin
ISH:	*in situ* hybridization
JC:	Jukes and Cantor
LNA:	locked nucleic acid
LRT:	likelihood ratio test
LWS:	long wavelength sensitive cone opsin
ML:	maximum likelihood
MP:	maximum parsimony
MY:	million years
mya:	million years ago
n/s:	non-significant
NJ:	neighbour-joining
NNI:	nearest neighbour interchange
PBS:	phosphate buffered saline
qPCR:	quantitative polymerase chain reaction
RH1:	rhodopsin type 1 rod opsin
RH2:	rhodopsin type 2 middle wavelength sensitive opsin
SPR:	subtree pruning and regraft
SWS1:	short wavelength sensitive type 1 opsin
SWS2:	short wavelength sensitive type 2 opsin
TM6:	transmembrane domain 6
TN:	Tamura-Nei
UV:	ultraviolet
λ_{max}:	wavelength of maximal sensitivity

Acknowledgments

I wish to thank Dr. John Taylor for his unending support and optimism. His uncompromising vision helped me achieve success. I must also thank Diana Rennison for working with me daily over the course of the degree. Together we achieved more than the sum of our intellects. I also thank Chris Laver, Justin Mui and Riki Dayan for discussions and input. I thank Corey Watson, Ben Sandkam and Felix Breden for helpful discussion and debates on the mysteries of opsins. I thank my committee members, Dr. Robert Chow and Dr. Steve Perlman, for helpful advice and critiques of my work. I thank Dr. W. Ted Allison for supplying an *in situ* hybridization protocol, and the University of Victoria advanced imaging lab for microscopy assistance. This work was funded by a University of Victoria Graduate Fellowship, an NSERC CGS-M and Dr. Taylor's NSERC grant. I thank my parents for their support and encouragement. Lastly, I thank Horatio, the Anableps.

Chapter 1 – Vision from sea to sky, from cornea to cones

Evolutionary adaptations of the eye

Light reception originated as a non-directional assessment of ambient luminance and functioned in relatively simple tasks such as circadian entrainment and gauging depth. This required nothing more than a photoreceptive molecule and a downstream response pathway, and can be seen in single celled organisms (Björn, 2008). In metazoan evolution, several key advances in vision evolution were made: gene duplication in the opsin family of G protein coupled receptors was the first step. This was followed by the emergence and refinement of pigment cells, morphologically specialized photoreceptors and light focusing lens tissue (Nilsson, 2009). These advancements allowed light reception to be fast, directional, sensitive and focused, respectively, and allowed for the formation of, what we now know as, the eye.

In order to be effective, an eye must be sensitive to environmentally important wavelengths and must be able to focus incoming light onto the retina. The vertebrate eye has also evolved to cope with a diversity of light levels. Since the turn of the last century the remarkable physiological adaptations to these challenges have been studied in a variety of animals but it is only within the last quarter century that technology has allowed the examination of molecular adaptations (Walls, 1967; Nathans & Hogness, 1983). It is modern science's challenge to coalesce these findings to demonstrate how selection has shaped vision at all levels.

Morphological adaptations to low light vision

Nocturnal animals are presented with light up to ten million times dimmer than diurnal animals and have evolved different eye morphology to cope with this (Munz & McFarland, 1977). The eye of nocturnal animals, such as the opossum, typically have a larger pupil diameter to increase light catch (Walls, 1967). This also occurs in the mesopelagic zone of the deep sea (200-1000 m deep) where light is dim and of restricted wavelength composition. Here, the last remnants of sunlight are filtered into a largely monochromatic peak around 470 nm to 480 nm. For each 100 metres of depth, irradiance is reduced by >10 times (Kampa, 1970). In order to effectively capture this weak and

progressively monochromatic light, extreme morphological adaptations have evolved in deep-sea fish.

A number of fish have evolved asymmetrical eyes for preferential vision upwards in order to detect silhouettes of overhead animals (Land, 2000). Contrary to the typical round fish eye, these eyes are modified into what is known as a tubular eye (Figure 1.1). At the dorsal most tip of a tubular eye sits a large spherical lens. The eye extends downward in a tube shape with lateral wall formed from iris and the medial wall formed from an extension of the retina, called the accessory retina. The ventral most portion of the eye's inner lining is covered by the main retina. In contrast to the accessory retina, which directly abuts the lens, the main retina is placed far enough away from the lens to allow for image focusing (Locket, 1977). In effect, the eye has two distinct retinal halves. It is hypothesized that the advantage of tubular eyes lies in the fact that a tubular eye allows for a far greater lens diameter in a similar sized eye when compared to the normal 'rounded' eye and, consequently, greater light capture (Munk, 1966).

While tubular eyes allow for enhanced collection of downwelling light, the underwater environment is 3D and bioluminescence can also occur from below. Bioluminescence, the production of light from an organism itself, occurs in over 80% of deep-sea species (Herring, 1996). Deep-sea fish have evolved remarkable ways to sense this other light source. *Scopelarchus analis*, the short fin pearleye, has a lens pad which sits under its pupil and directs upwelling light onto the accessory retina, while *Dolichopteryx longpipes* has evolved an entirely separate retinal chamber that focuses upwelling light onto the retina using a mirror (Figure 1.1a-b) (Brauer, 1902; Locket, 1977; Pearcy et al., 1965; Wagner et al., 2009; Schwab et al., 2001). *Bathylychnops exilis* has evolved an additional eyeball attached to the main eyeball. This accessory eyeball points downward and has evolved its own lens from scleral tissue (Figure 1.1c) (Pearcy et al., 1965; Schwab et al., 2001). Through amazing morphological adaptations, these deep-sea fish have evolved double vision to cope with a dim and highly directional spectral environment.

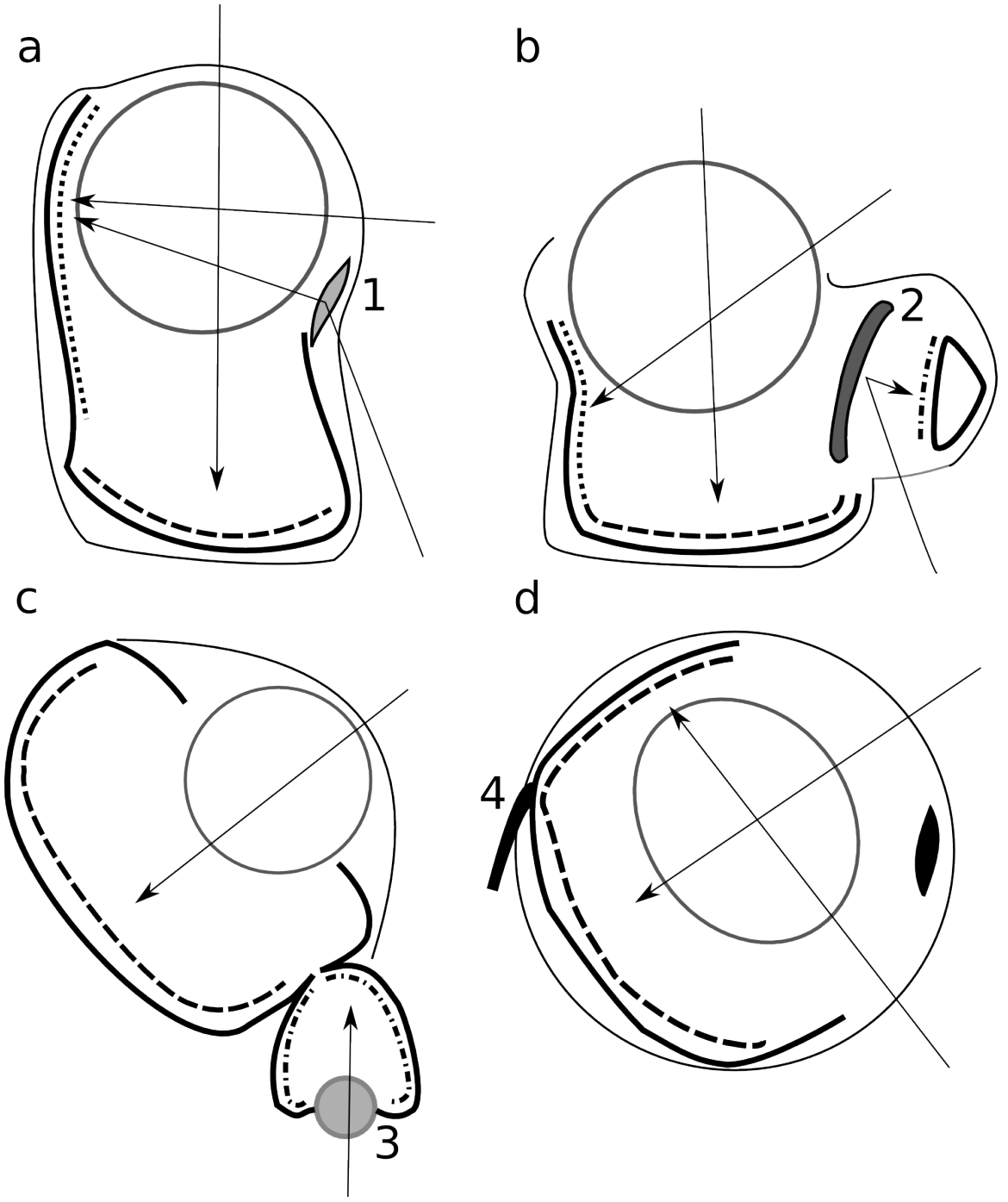


Figure 1.1: A schematic cross section of eyes adapted to double vision

Eye adaptations found in deep-sea fish and the surface dwelling *A. anableps.* In each, dashed lines represent retina, dotted lines represent accessory retina and dotted-dashed lines represent retinal diverticulum (specially adapted retinal tissue). Arrows represent the path of light from different directions. Dorsal direction is up in all images. A) *Scopelarchus analis* eye. 1 – the lens pad acts as a wave guide for incoming light. B) *Dolichopteryx longpipes* eye. 2 – the reflective retinal tapetum reflects upwelling light onto the retinal diverticulum. C) *Bathylychnops exilis* eye. 3 – the sclerally derived lens-like structure focuses upwelling light into the secondary globe. D) *Anableps anableps* eye. 4 – the optic nerve is positioned at the boundary between the dorsal and

ventral retina. Downward aerial light is projected onto the ventral retina while upwelling aquatic light is projected onto the dorsal retina (Sivak, 1976; Frederiksen, 1973; Munk, 1966).

Morphological adaptations to amphibious vision

Double vision is also found at the water's surface. In this environment, the light levels are much higher and absolute photon capture is not the main problem. Here the differences between air and water present challenges for focusing and for wavelength sensitivity due to water's refracting and wavelength filtering properties. One animal, the four-eyed fish, has evolved exquisite adaptations to effectively see in both mediums simultaneously.

Anableps anableps, the four-eyed fish, is a surface living member of the order Cyprinodontiformes. It is found in the mangrove forests along the northeast coast of South America (Schultz & Stern, 1948). *A. anableps* feeds primarily on intertidal red algae, insects and marsh crabs, all items found at or above the water line (Brenner & Krumme, 2007). The moniker "four-eyed" is derived from the fact that it keeps its bulbous eyes at the water surface, giving it four fields of view; the dorsal half of the eye is exposed to air, while the ventral half is underwater (Figure 1.1d). During bright light, medial extensions of the iris extend across the center of the retina at the water line dividing the dorsal and ventral pupils and preventing glare (Schwassmann & Kruger, 1965). It has been suggested that the dorsal retina is too close to the lens for clear vision and the optomotor response has been found to be lacking from the aquatic field of view (Swamynathan et al., 2003; Albensi & Powell, 1998; Saidel & Fabiane, 1998). Despite this, foraging is most frequent during daylight and periods of optimal water clarity ideal for aquatic vision (Brenner & Krumme, 2007).

The ability to use simultaneous aerial and aquatic vision presents challenges due to the differences in refractory index between the two media. For aerial vision, refraction occurs as light passes from air into the cornea due to the difference in refractory index (1.0 for air, 1.37-1.4 for cornea) (Patel et al., 1995; Leonard & Meek, 1997). For aquatic vision, this refraction does not occur because the refractory index in water (1.33) is very close to that of cornea. This difference in refraction can lead to myopia (nearsightedness) in aerial vision, and hyperopia (farsightedness) in aquatic vision for amphibious animals (Brett, 1957).

To compensate for this issue, there have been morphological changes to both the cornea and lens in the four-eyed fish. The dorsal cornea is flattened to decrease the refraction that occurs. Furthermore, the lens is egg-shaped and placed at an angle, which enables aerial and aquatic light to encounter different axis of the lens (Sivak, 1976). Aquatic light passes through the ventral iris and through the longer axis of the lens with a greater surface curvature. This provides the refraction necessary for light focusing, despite the lack of corneal refraction. On the other hand, aerial light passes through the dorsal iris and then through a relatively flat lens axis, which is very similar to what is found in terrestrial animals. Thus, through one lens, two different focusing tasks occur simultaneously (Sivak, 1976). The spectral composition of light also differs between the aerial and aquatic environments. Though little is known about the adaptations to this variation in *A. anableps*, opsins are the genes most-likely to respond to such selection.

Visual pigments and opsin genes

Visual pigments are encoded by opsin genes expressed in the photoreceptor cells of the retina in vertebrates. The proteins encoded by these genes bind a chromophore called retinal. When exposed to light, the chromophore undergoes isomerisation from 11-cis to all-trans, causing a conformational change in the bound opsin protein. This conformational change influences binding to the associated intracellular heterotrimeric G-protein, transducin. Transducin binding leads to hyperpolarization of the photoreceptor cell and a signal transmitted to the bipolar cells. These cells synapse onto the ganglion cell layer, which is connected to the brain through its axons. During the signal transmission through the retina, the signal is modulated via horizontal and amacrine cells. Together, the retina performs some of the first steps in visual processing for such tasks as edge detection and colour discrimination (Kuffler, 1953; Kolb et al., 2001; Wheeler, 1982).

The wavelengths of light to which each photoreceptor is sensitive are determined by the type of chromophore and the amino acid sequence of the opsin protein expressed. There are two types of chromophores, A1 (retinal) and A2 (3-dehydroretinal). Opsins bound to A2 have a red shifted wavelength of maximal sensitivity (λ_{max}) when compared to A1 bound opsins (Knowles & Dartnall, 1977). Amino acid sequence variation has been used to sort opsin proteins into five monophyletic clades with clustered but

overlapping λ_{max} values. The long wave sensitive (LWS) subfamily absorb yellow and green light (490-570 nm), the rhodopsin (RH1) subfamily absorb green light (460-530 nm), the rhodopsin-like (RH2) subfamily also absorb green light (480-530 nm), the short wave sensitive type 2 (SWS2) subfamily absorb blue and violet light (410-490 nm) and the short wave sensitive type 1 (SWS1) subfamily absorb violet and UV light (355-440 nm) (Yokoyama, 1994; Bowmaker, 2008). The phylogenetic relationship among subfamilies is thought to be ((((RH2, RH1), SWS2), SWS1), LWS). The gene duplications that lead to the five-gene repertoire occurred very early in vertebrate evolution as representatives of all subfamilies can be found in the genome of the agnathan, *Geotria australis* (Collin & Trezise, 2004). Subsequent gene duplications have led to expanded repertoires, most commonly in teleost fish.

While there are clear differences in λ_{max} between opsin subfamilies, there can also be differences between orthologous genes between species and between paralogous genes within a species. For example, while human SWS1 absorbs blue light (λ_{max} = 414 nm), in mouse the orthologous protein absorbs UV light (λ_{max} = 359 nm) (Oprian et al., 1991; Yokoyama et al., 1998). Another example is zebrafish (*Danio rerio*), which have a total of four RH2 genes with λ_{max} values from 467 nm to 505 nm (Chinen et al., 2003). Studies have attempted to pin point what amino acid changes are responsible for these differences by reconstituting the opsin-chromophore complex and measuring absorption *in vitro*. Once the original protein is measured, point mutations are introduced to see the effect of individual amino acids changes. Furthermore, phylogenetically reconstructed ancestral pigments can also be created and measured to determine the maximal absorbance of opsins in ancestral animals (Shi et al., 2001; Yokoyama & Radlwimmer, 2001; Chinen et al., 2005a; Chinen et al., 2005b). This work has shown that there are a minority of amino acid sites which play a disproportionately large role in determining the λ_{max}, so called key sites (Yokoyama & Yokoyama, 1990; Chang et al., 1995). These key sites vary by opsin subtype and typically are located in the retinal-binding pocket of the protein (Kochendoerfer et al., 1999; Yokoyama, 2008). Substitutions to shift λ_{max} can work individually or strictly synergistically and are not always additive (Takenaka & Yokoyama, 2007; Yokoyama, 2008).

Visual pigments and the spectral environment

Through these key sites, evolution can act to tune the λ_{max} of each opsin to the spectral needs of the animal. Indeed, many studies have examined the spectral absorption of the photoreceptors in the context of light environment and found correlates. This is most striking in the rods of deep-sea fish. Their rod cells express an RH1 opsin that is blue shifted to 470-480 nm (Lythgoe, 1972; Partridge et al., 1989). This is shorter than the ~500 nm sensitive RH1 opsins common in the rod cells of surface dwelling fish (Yokoyama, 2000). This blue shift functions to match most closely the wavelength of photoreceptor absorption with the wavelength of light that penetrates seawater the most effectively (Partridge et al., 1988). The tendency for the amino acid sequence of rod pigments to change with water depth is also seen in the freshwater fish of lake Baikal, one of the deepest lakes in the world. In this environment, the littoral species have rod pigments that absorb at 516 nm, while the closely related abyssal species absorb at 484 nm (Bowmaker et al., 1994; Hunt et al., 1996). Interestingly, three genera of deep-sea dragon fish (*Malacosteus*, *Aristostomias* and *Pachystomias*; order Stomiiformes, family Stomiidae) defy this trend. In these species, the rhodopsin genes are red shifted (515 nm). This appears to be an adaptation to the far-red bioluminescence that they produce themselves (Denton et al., 1970; O'Day & Fernandez, 1974; Denton et al., 1985; Bowmaker et al., 1988; Crescitelli, 1990).

There are several other examples of correlations between opsin tuning and environmental variation. Atlantic eels, *Anguilla anguilla*, begin their lives in the Sargasso sea and then make their way into freshwater rivers and lakes on the European, North African and North American coastlines. After several years in these fresh water environments, they return to the deep waters of the Atlantic as sexual mature silver eels (Tesch, 1977). During this migration, the eels switch their rod pigment absorption from 523 nm to 501 nm by switching from the chromophore A2 to its analogue A1 (Carlisle & Denton, 1959). Following this, the opsin protein also switches from RH1-FW to its paralog RH1-DS. This ultimately produces a rod cell with a maximal sensitivity for 482 nm (Beatty, 1975; Wood & Partridge, 1993; Hope et al., 1998). The transition from 523 nm to 501 nm to 482 nm coincides with maximal rod sensitivities for fish living in freshwater, coastal and deep-sea environments respectively (Lythgoe, 1988).

While tuning to the spectral environment can be seen when comparing species, it is also seen within the eye itself. This is seen in both the uneven distribution of chromophore types and in differential opsin gene expression across the eye. In bullfrogs, the dorsal retina, used in downward aquatic vision, uses primarily A2 chromophore, giving it longer wavelength sensitivity (Reuter et al., 1971). This is thought to correlate with the fact that turbid water attenuates shorter wavelengths of light to a greater degree, leaving longer wavelengths to be important in aquatic vision (Lythgoe & Partridge, 1989). A convergent pattern of intraretinal pigment differentiation has evolved independently in cichlids where shorter wavelength cones are more common in the ventral half of the eye (Levine et al., 1979).

Expression analysis and *in situ* hybridization

There are several ways to characterize opsin gene expression within a retina. Microspectrophotometry (MSP) measures light absorption in photoreceptor cells and is most common. Unfortunately, it has limited sampling and can overlook sparsely expressed genes. A more sensitive technique for understanding the fine scale patterning of retina is *in situ* hybridization which is a technique for visualizing cells that are expressing a gene of interest. It involves creating a labelled nucleotide probe that is complimentary to an mRNA transcript of interest. The probe is applied to whole mount or sectioned tissue where it binds to the mRNA. This probe is then visualized through the use of its antigen label, often digoxigenin, and a subsequent fluorescent or enzymatic colour reaction. This technique was first pioneered on *Xenopus* eggs but has subsequently been used in everything from virology to neuroscience (Pardue & Gall, 1969; Brigati et al., 1983; Bloch et al., 1986; Morris et al., 1986; Unger et al., 1986).

The utilization of *in situ* hybridization on visual opsin genes has been performed on a variety of animals including ice cod, zebrafish, tuna, flounder, mouse, cow, chicken, and monkey (Bumsted et al., 1997; Bruhn & Cepko, 1996; Brann & Young, 1986; Applebury et al., 2000; Hoke et al., 2006; Miyazaki et al., 2008; Takechi & Kawamura, 2005; Pointer et al., 2005). These results were particularly interesting in zebrafish due to the fact that it has undergone several gene duplication events to produce four RH2 and two LWS opsins. *In situ* hybridization found that these genes were differentially expressed during ontogeny and across the retina. They found that shorter wavelength paralogs were

expressed earlier and, in the adult, were expressed in the more central dorsal regions of the eye, when compared to their longer wavelength paralogs (Takechi & Kawamura, 2005). These authors suggested that this variation might be due to the heterogeneity of the light environment.

A. anableps is an excellent animal for examining opsin gene expression in the eye due to both its peculiar eye morphology and expanded opsin repertoire. The dorsal and ventral halves of the *A. anableps* retina are exposed to different media and thus different light compositions. Light hitting the dorsal retina passes through water, which preferentially filters certain wavelengths depending on the clarity of the water, while the ventral retina receives comparatively unfiltered aerial light. While previous work had found no differences in pigments between the dorsal and ventral retinas, more recent molecular analysis has shown a far greater number of opsin genes expressed in the eye than previously recognized (Avery & Bowmaker, 1982; Owens et al., 2009). MSP analysis only found four distinct photoreceptor cell types, while molecular analysis (my honour's thesis) showed a total of ten: one SWS1 and RH1 genes, two SWS2 and RH2 genes, as well as four LWS genes. At least eight of these genes were recovered from cDNA, indicating that they are expressed in the eye at some level. With this in mind, I hypothesized that *A. anableps* unique morphological adaptations to simultaneous aerial and aquatic vision have been accompanied by changes at the opsin gene expression level. I tested this hypothesis by mapping the expression of all opsin genes onto the retinal landscape using *in situ* hybridization. I discovered that opsin expression differs in the dorsal and ventral retina in *A. anableps* and then tested the hypothesis that this pattern coincided with the morphological evolution by examining the opsin expression in the close relative with normal eye morphology, *Jenynsia onca*.

Thesis overview

This thesis examined the molecular evolution of opsin gene duplicates at the sequence and expression level. Before focusing on the four-eyed fish in Chapter 3, I reviewed opsin gene duplication and divergence in all fish in Chapter 2. Specifically I asked the general question: Does opsin gene repertoire expansion lead to opsin protein functional diversification? To answer this question, I completed a thorough phylogeny of fish opsin genes to identify all opsin gene duplication events. I then inferred selection pressure on

these genes by examining dN/dS ratios across the phylogeny. In Chapter 3, I used *in situ* hybridization to visualize the expression patterns of cone opsin gene duplicates in *A. anableps.* I was curious to see if an expanded opsin gene repertoire played a role in the specialization of the eye for simultaneous above and below water vision. By also studying close relative *Jenynsia onca,* I was able to test the hypothesis that any partitioning of opsin expression domains in *A. anableps* is associated with the evolution of its distinct eye morphology. In Chapter 4, I closed by proposing a reason why opsin gene duplication is skewed to longer wavelength subfamilies and reviewing how opsin gene duplication affects gene expression.

Chapter 2 – Opsin gene duplication: Searching for selection

ABSTRACT

Of all vertebrates, only the ray-finned fish have notably expanded their visual opsin repertoire. To understand why this has occurred, we must first know when and where gene duplication has occurred in fish evolution. By reconstructing the phylogenetic relationships among fish opsins, I have shown that duplication events span the age of the taxon Teleostei. Furthermore, they are predominantly tandem and in the longer wavelength subfamilies (RH2 and LWS). Following gene duplication, relaxed selection is seen. Positive selection was detected in some codons after duplication but only rarely was positive selection observed at codons known to influence spectral sensitivity. One notable exception is the gene SWS1-1 in ayu (*Plecoglossus altivelis*), where one of the duplicates appears to have been positively selected for a red shift in spectral sensitivity.

INTRODUCTION

Gene duplication and divergence

Gene duplication plays an important role in the generation of evolutionary novelty. The rate of gene duplication has been estimated to be 0.01/gene/MY (Lynch & Conery, 2003; Lynch & Force, 2000). At this rate, from the origin of teleosts 300 million years ago, each gene present in the ancestor has been duplicated on average 3 times (Hurley et al., 2007). Although the majority of duplicated genes are lost within four million years, those that remain may undergo neofunctionalization and/or subfunctionalization (Lynch & Conery, 2003). In neofunctionalization, one, or both, duplicates gain a new function not present in the original gene. For example, hominoids and Old World monkeys have two eosinophil-derived neurotoxin genes, EDN and ECP. Both genes are in the RNase A gene superfamily. EDN in hominoids and Old World monkeys, as well as, the single-copy ancestral gene (pro-ortholog) in New World monkeys are RNases. ECP, on the other hand, has little RNase activity but has become a potent toxin against bacteria (Zhang et al., 1998).

A second possible consequence of gene duplication is subfunctionalization (Force et al., 1999). In this, the genes produced by duplication subdivide the multiple roles played by the single-copy ancestral gene. This has occurred in the zebrafish genes *engrailed-1* and *engrailed-1b*. The pro-ortholog in mouse, *Engrailed-1*, is expressed in the pectoral appendage bud, the hindbrain and the spinal cord. In zebrafish, *engrailed-1* is expressed in the pectoral appendage bud, while *engrailed-1b* is expressed in the hindbrain and spinal cord (Force et al., 1999). Subfunctionalization was originally meant to describe role-partitioning in duplicates at the level of gene expression but is now also applied to post-duplication divergence at coding sequences. For example, in humans, SYN2 has two isoforms, one which is characterized by truncation at the 3' end. In *Takifugu rubripes*, one duplicate, Syn2B, has undergone a similar truncation while Syn2A is always full-length. In this sense, what the human genome has accomplished through alternative splicing, *T. rubripes* has accomplished by gene duplication and coding sequence divergence (Yu et al., 2003).

Where gene duplication provides the raw material for increased genetic novelty, mutation creates the diversity. All mutations in coding sequence are either synonymous, if they don't change the amino acid that the codon codes for, or nonsynonymous, if they do. Both types of mutation occur at equal pace, but the rate at which nonsynonymous substitutions are fixed in a population depends on selection. Since synonymous substitutions do not result in protein-level changes, they are considered to be free from selection and accrue at a relatively constant rate (Miyata & Yasunaga, 1980). During purifying selection, nonsynonymous substitutions are fixed less often because a majority of amino acid changes are detrimental and selected against. In this case, synonymous substitutions are more common than nonsynonymous. On the other hand, during positive selection, fixation of nonsynonymous substitutions is favoured because the change of amino acid provides a selective advantage. Consequently, positive selection is indicated by a greater frequency of nonsynonymous substitutions than synonymous ones.

Analyzing the selection pressure a gene evolved under is done by comparing sequences and determining the number of synonymous and nonsynonymous substitutions that occurred during its evolution. This determines an approximate rate of nonsynonymous substitutions (dN) and synonymous substitutions (dS) which, when divided, produces the

ratio ω. A ω ratio of <1, =1 or >1, indicates purifying, neutral or positive selection respectively. A ω ratio of greater than 1 has been seen in proteins where the generation of novel structures is favoured including the human MHC locus, a protein involved in antigen recognition, and sperm lysin, which functions in species specific gamete recognition (Hughes & Nei, 1988; Klein & Horejsi, 1997; Lee & Vacquier, 1992; Lee et al., 1995; Yang et al., 200b).

While early methods of detecting positive selection estimated ω across an entire phylogeny, it soon became clear that positive selection does not necessarily occur equally in all lineages or in all codons (Hughes & Nei, 1988; Yang & Nielsen, 1998; Yang et al., 2000a). With this in mind, models have been developed for identifying positive selection in specific branches, at specific sites or both (branch-site) (Yang & Nielsen, 2002). These models use maximum likelihood estimates to reconstruct ancestral sequences and identify the ω ratio under which the sequence evolved.

The visual opsin gene family

One particularly interesting gene family for the study of adaptive evolution is the vertebrate visual opsins. Opsin gene duplication at the dawn of vertebrates has allowed for broad spectral sensitivity and colour (i.e., wavelength) discrimination in nearly all vertebrates. Lineage-specific duplications have led to much variation in sensitivity and wavelength discrimination among taxa, including trichromatic vision in humans (Nathans et al., 1986b) (contrasting dichromatic vision in most mammals) and even variation over ontogeny within a species (e.g., the adaptation to freshwater and deep sea spectral environments in eels (Wood & Partridge, 1993; Hope et al., 1998)).

Opsin genes are members of a monophyletic family of G protein coupled receptors that have been subdivided into five subfamilies in vertebrates; Long-Wavelength Sensitive (LWS), Short-Wavelength Sensitive (SWS1 and SWS2), Rhodopsin (RH1) and Rhodopsin-like (RH2) (Yokoyama, 1994). Opsin proteins are expressed in the photoreceptors of the eye and, together with a chromophore, they undergo a conformational change when exposed to light. The wavelength of light to which an opsin is most sensitive to (or λ_{max}) is determined by its amino acid sequence. Much work has been done towards understanding the molecular basis of differences in opsin absorbencies between homologous proteins; it has been found that a majority of the

differences can often be attributed to a few amino acid residues, dubbed "key sites." These key sites are often, but not exclusively, residues in close contact to the chromophore (Kochendoerfer et al., 1999; Yokoyama, 2008).

The close and partially transparent relationship between genotype (amino acid sequence) and phenotype (λ_{max}) has made opsins an excellent model for the study of adaptive evolution. Yokoyama et al. (2008) studied the RH1 genes of a series of fish to reveal how λ_{max} related to the fish's spectral environment and what amino acid substitutions were responsible for changes in λ_{max}. They found that a majority of λ_{max} changes occurred from substitutions in only 12 codons. However, Nozawa et al. (2009) studied all opsin subfamilies in vertebrates using a variety of methods designed to detect codons under positive selection and found that sites under positive selection were unlikely to be involved in spectral tuning. Lastly, Briscoe et al. (2010) found that a recently duplicated UV opsin in butterfly was under positive selection and that this was correlated with increased UV wing colouring.

Differences in λ_{max} are found between orthologs and, perhaps more interestingly, also between paralogs. Through gene duplication, some animals have expanded their opsin repertoire to include genes with a wide variety of spectral sensitivities. In hominoids, a tandem duplication and subsequent sequence divergence within the LWS subfamily produced two opsins, one sensitive to red and the other to green light (Nathans et al., 1986b; Vollrath et al., 1988). It is through the signals generated by cone photoreceptor cells expressing these opsins, and the shorter wavelength sensitive blue SWS1 opsin, along with subsequent neuronal integration, that colour vision, as we know it, is achieved. Humans that lack one of the LWS genes are colour blind and have greater difficulty in wavelength discrimination.

Within tetrapods, opsin gene duplications are extremely rare. On the other hand, fish have undergone many opsin gene duplications and consequently have much more diverse opsin repertoires (Hofmann & Carleton, 2009). Thus, teleosts provide an ideal group to study the evolution of opsin genes.

In this chapter, I used multiple methods to construct a phylogenetic tree for each opsin subfamily. From this tree, I inferred when gene duplication events occurred relative to speciation events. Innan and Gojobori (2009) recently proposed that middle-wave opsins,

the RH2 and SWS2 subfamilies, have more and older duplication events and accelerated amino acid evolution when compared to boundary-wave opsins, the LWS and SWS1 subfamilies. To test this hypothesis, I analysed duplication event timing and used the program PAML to examine dN/dS ratios across each subfamily to identify sites under positive selection. I also compared dN/dS ratios between gene duplication and speciation branches. This allowed me to test whether gene duplication or speciation is a greater driver of protein divergence. I also asked whether post-duplication branches have codons under positive selection. In this way, I revealed if spectral differentiation following gene duplication is positively selected.

METHODS

Phylogenetic analysis

The majority of fish opsin gene sequences used were obtained with BLASTn (Altschul et al., 1997). The databases queried were the NCBI nucleotide database and Ensembl genome sequence databases. Opsin gene sequences from *Danio rerio* and *Oryzias latipes* were employed as queries. Sequences were chosen for three reasons: species with multiple opsin subfamilies surveyed, species with within subfamily duplications, and species chosen to clarify duplication event timing. Bream opsins included in this survey (*Acanthopagrus berda*, *Acanthopagrus schlegeli* and *Pagrus major*) were obtained from Dr F. Y. Wang (Personal correspondence). Coding sequences were aligned by hand using BioEdit (Hall, 1999). All accession numbers are listed in Appendix 1.

A single “all-opsins” multiple sequence alignment was used in the first phylogenetic analysis. This analysis included non-visual opsins (pinopsins and VA opsins) as well as invertebrate visual opsins as outgroups. PAUP was used to generate a neighbour-joining phylogenetic tree based upon distances estimated using the Jukes-Cantor model of DNA sequence evolution (Jukes & Cantor, 1969; Swofford, 2002). While relationships among the major classes of opsins are well established, this analysis allowed us to confirm that some of the especially divergent genes had been assigned to subfamilies correctly.

Opsin subfamilies

Sequences for each of the five subfamilies of vertebrate visual opsins were utilized in separate analyses because this allowed us to use slightly longer alignments and different substitution models for each subfamily.

Optimal model parameters for maximum likelihood analysis were generated from Modeltest in PAUP* 4.8B10 (Posada & Crandall, 1998; Swofford, 2002). Maximum likelihood (ML) trees using these parameters were constructed using PhyML (Guindon & Gascuel, 2003). Tree improvement was done using the best of NNI and SPR (Hordijk & Gascuel, 2005). Support for nodes on this tree was estimated using an approximate likelihood ratio test (Anisimova & Gascuel, 2006). Neighbour joining trees using Jukes and Cantor and Tamura-Nei distances with bootstrap (1000 replicates) were also reconstructed using PAUP (Jukes & Cantor, 1969; Saitou & Nei, 1987; Felsenstein, 1985; Tamura & Nei, 1993; Swofford, 2002). Lastly, a strict consensus maximum parsimony tree was created also using PAUP. Pair-wise deletion was used for instances of missing nucleotides in all analyses. For each ML tree, nodes representing gene duplication events and their position relative to speciation events were noted.

Measuring and mapping all duplication events

Gene duplication events were identified from phylogenetic trees. Duplication events were numbered based on distance between paralogs from oldest to youngest using roman numerals. In several cases, tree reconciliation suggested, possibly erroneously, duplication events in which one paralog was lost in nearly all members of a large group of species (RH1-II, RH1-III, RH2-III, RH2-IV and LWS-I). In these cases, only species which contained both paralogs were used for distance measures. Mean group distance was calculated using the Tamura-Nei method in MEGA4 for all sites, first and second positions and third positions (Tamura & Nei, 1993; Tamura et al., 2007). In cases where gene conversion among paralogs was suspected (LWS-V, RH2-I), suspected genes were not included in analysis. Number of duplicates for each age category, as determined by amount of divergence in third positions, and subfamily was charted. A species tree was used to summarize the duplication data for all opsin subfamilies. The topology used reflects fish taxonomy (Nelson, 2006). The tree was generated from a maximum likelihood analysis of RH1 sequences with duplication nodes removed and terminal

branches added for *Jordanella floridae*, *Scophthalmus maximus*, *Zacco pachycephalus*, *Candidia barbatus*, *Clupea harengus* and *Nannostomus beckfordi*, species that lacked RH1 gene sequences, in a manner that was consistent with fish taxonomy.

To explore the substitution pattern of the highly divergent 'green' LWS clade, a sliding window analysis was completed. In this analysis, the average amino acid identity from each 'green' LWS to *Danio rerio* LWS-1 was charted. The average pair-wise amino acid distance was also charted for all LWS opsins not in the 'green' clade. The program Swaap 1.0.3 was used with a window size of 30 (Pride, 2000).

Duplication events and gene organization

In order to discriminate among several possible modes of gene duplication, I recorded the location, either on chromosomes or on long-insert clones, of opsin genes for six fish species. Data for stickleback (*Gasterosteus aculeatus)*, tiger pufferfish (*Takifugu rubripes)*, green spotted pufferfish (*Tetraodon nigroviridis)*, Japanese medaka (*Oryzias latipes)* and zebrafish (*Danio rerio)* were obtained from the Ensembl genome browser. Data for tilapia (*Oreochromis niloticus)* was from Hofmann and Carleton (2009) and Karen Carleton (personal communication). For *Xiphophorus helleri*, data was from BAC clone sequence (Watson et al., 2010). Gene orientation was determined by neighbouring reference genes shared among species; in zebrafish synaptoporin (ENSDARG00000044278) and solute carrier family 6 member 13 (ENSDARG00000000730) were used to anchor RH2 orientation. These data were used to infer a series of gene rearrangement events to explain the RH2 orientation in representative species.

Positive selection and dN/dS ratios

The phylogenetic analysis package, PAML (Yang, 2007) was used to compare the ratio of nonsynonymous substitutions to synonymous substitutions (the dN/dS ratios or ω) among several subsets of data: i) the overall ω for all branches was compared between opsin subfamilies; ii) the ω for post duplication branches was compared to ω for post speciation branches within subfamilies; and iii) average ω between paralogous clades within two opsin subfamilies, SWS2 and RH2. These analyses used only full-length opsin

genes (58 species, Appendix 1). Phylogenetic trees used for these analyses are shown in Appendix 3-7 and post duplication branches are indicated.

To identify individual codons under positive selection across an entire subfamily, I used two separate tests; M1a vs. M2a and M8 vs. M8a (Nielsen & Yang, 1998; Wong et al., 2004; Yang et al., 2005). M1a divides the sites into two categories, one under neutral selection ($\omega=1$) and one under negative selection ($\omega<1$), while M2a adds a third category under positive selection ($\omega>1$). M8 assumes a beta distribution from 0 to 1 of ω for sites and an additional class of sites under positive selection ($\omega>1$), while M8a acts as a null model by fixing this last class of sites at $\omega=1$. Following these analyses, a likelihood ratio test was conducted on each model pair to determine if there were significant likelihood gains by allowing positive selection. Both models M2a and M8 can be affected by local optima (Yang et al., 2000a; Anisimova et al., 2001). To ameliorate this issue, starting ω values of less than and greater than one were used.

To ask the question, do opsin genes undergo positive selection for spectral diversification following gene duplication, I examined all 65 branches following gene duplication to look for specific codons under positive selection using the branch-site model B (Yang et al., 2005). This model divides branches into foreground (those specified) and background (all others). Branches tested in this way are indicated in Appendix 3-7. Codons are then divided into categories which allow a subset of foreground branch codons to evolve under positive selection ($\omega>1$) while the same codons in background branches are under purifying or neutral selection ($\omega<1$, $\omega=1$). This is tested against a null model which does not allow any codon to be under positive selection using a likelihood ratio test. To account for multiple testing on each subfamily, Bonferroni's correction was applied (Miller, 1981; Anisimova & Yang, 2007).

RESULTS

Phylogenetic analysis

Highly conserved residues allowed vertebrate opsin genes coding sequences to be easily aligned by hand. Phylogenetic analyses (NJ) generated a tree with five visual and two non-visual opsin clades (Figure 2.1). Visual opsins appear to be a paraphyletic group. For subfamily analysis, maximum likelihood trees are discussed in detail below.

Neighbour-joining or maximum parsimony trees are not discussed except when they disagreed with maximum likelihood analysis. Across all subfamilies, a total of forty-two opsin gene duplication nodes were identified. These duplication events were mapped to a species tree in Figure 2.2. Mean distance for each duplication was determined using the Tamura-Nei algorithm (Appendix 2) (Tamura & Nei; 1993). Third position distance was unable to be calculated for RH1-II because of the high level of divergence. Duplication events were sorted by subfamily and age (Figure 2.3 & Figure 2.4). In some cases, duplications appeared very old based on genetic distance between paralogs but a lacked orthologs reduced confidence in their exact placement (LWS-I, RH1-II and RH1-III). In these cases, duplication nodes were placed in Figure 2.2 to encompass only species with both paralogs to recognize the uncertainty of the duplication event timing. Analyses showed that opsin gene duplications occurred primarily in the RH2 subfamily, and, to a lesser extent, the LWS and RH1 subfamilies. Analyses also showed that a majority of opsin duplication events are recent and have diverged less than 0.1 distance at third positions.

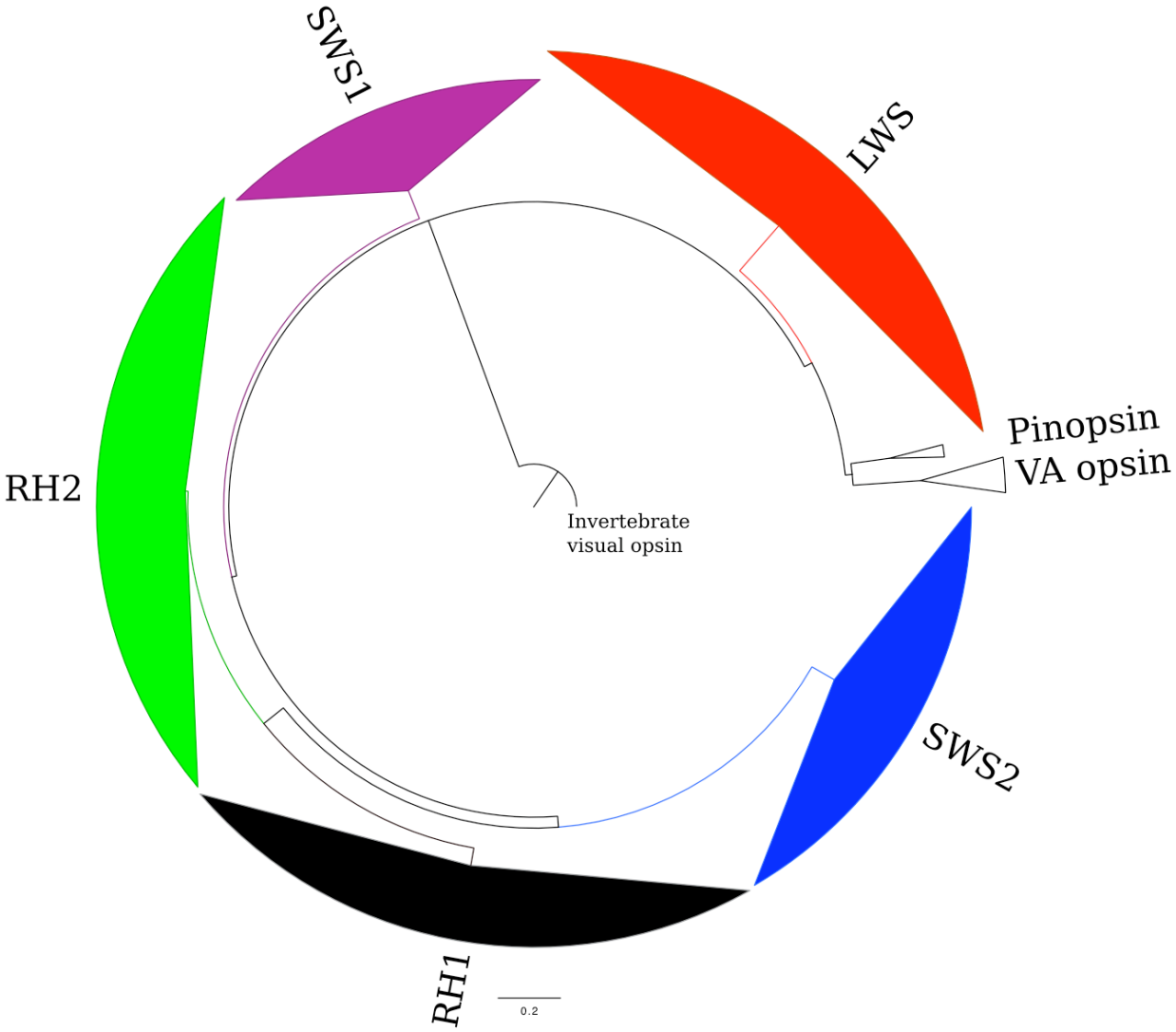


Figure 2.1: Phylogenetic tree of all opsins in fish.

The tree was created using the neighbour-joining method with the Jukes and Cantor model of evolution (Jukes & Cantor. 1969; Saitou & Nei, 1987). Accession numbers are listed in Appendix 1. *Drosophila melanogaster* Rh4 opsin was used as an outgroup. Visual opsin subfamilies were collapsed and colour coded based on spectral absorbance.

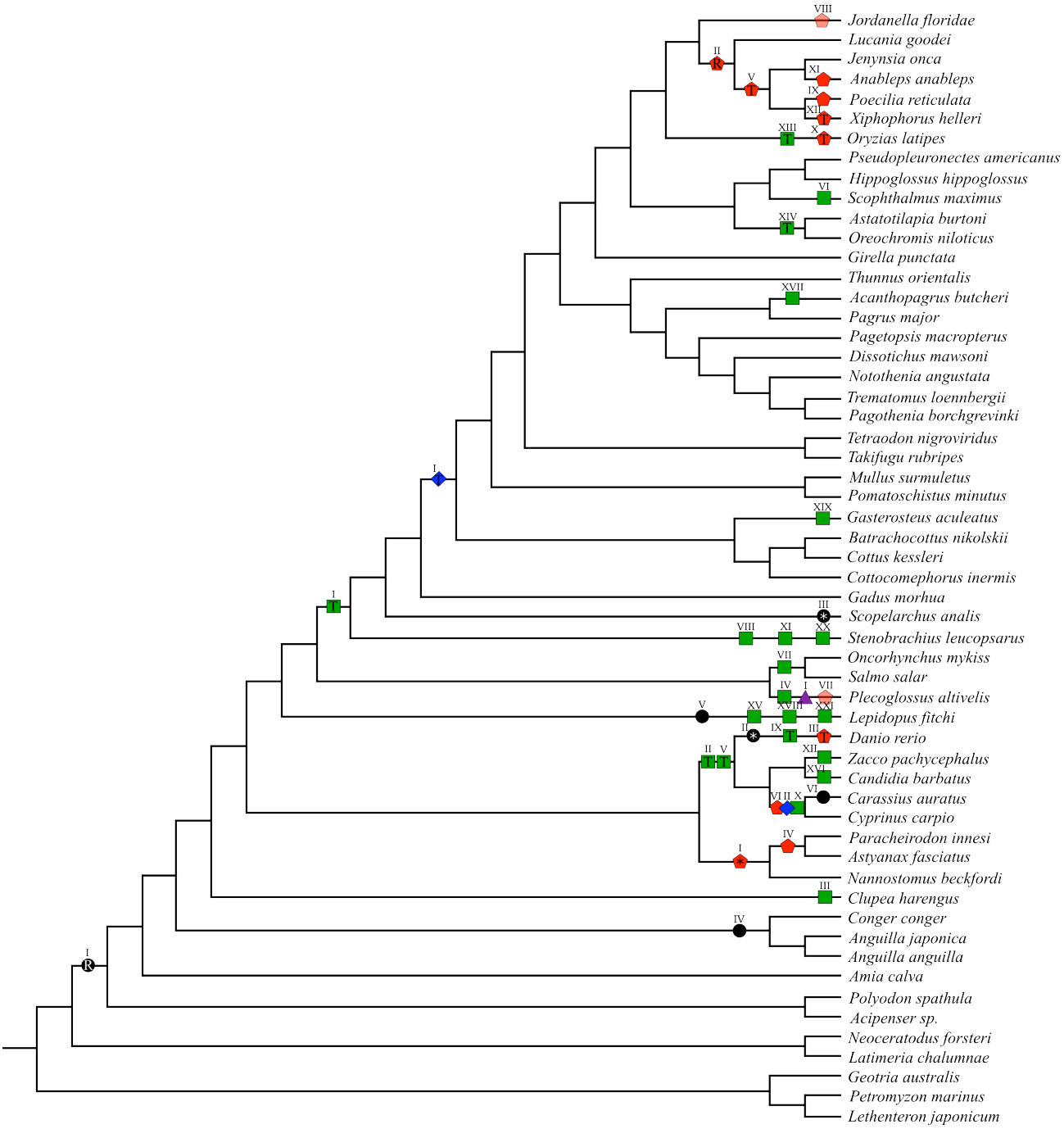


Figure 2.2: Fish opsin duplications and pseudogenization events.

The tree was constructed as a composite of a maximum likelihood RH1 gene tree and established species taxonomy. Gene duplication events are mapped onto the tree. Filled purple triangles, blue diamonds, black circles, green squares and red pentagons represent duplications in the SWS1, SWS2, RH1, RH2 and LWS subfamilies respectively. Roman numerals correspond to duplication numbers in Appendix 2. 'T' within a shape represents that the duplication is known to be a tandem duplication, 'R' means that it is a retrotransposition event while '*' means that the duplication

event is older than shown, but is unable to be confidently placed due to lack of orthologs in other species. Transparent shapes represent duplication events that may be allelic variation.

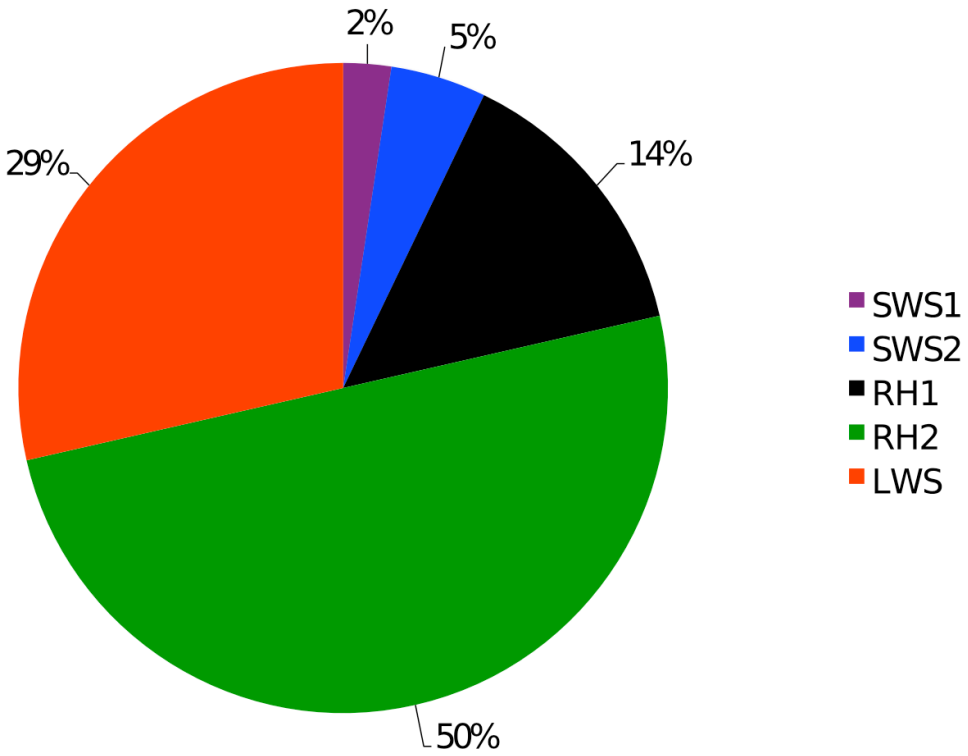


Figure 2.3: Opsin duplication events by subfamily.

Percentage of opsin duplication events that occurred in each subfamily.

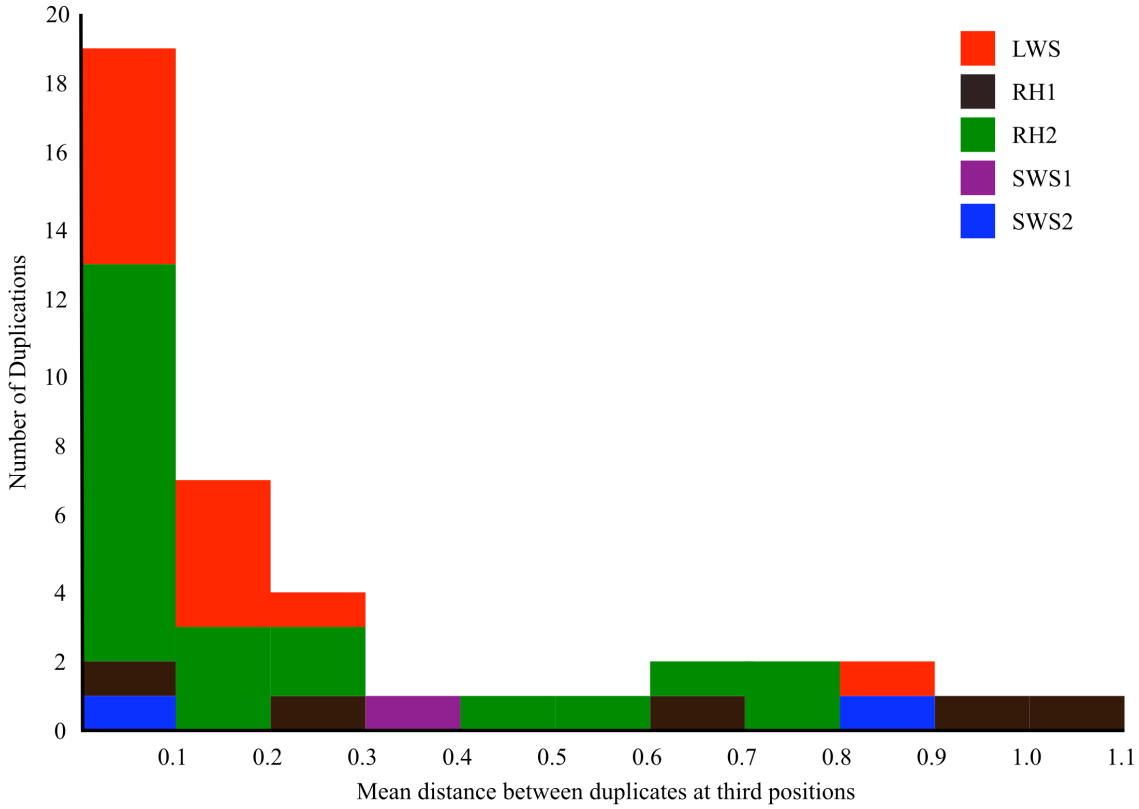


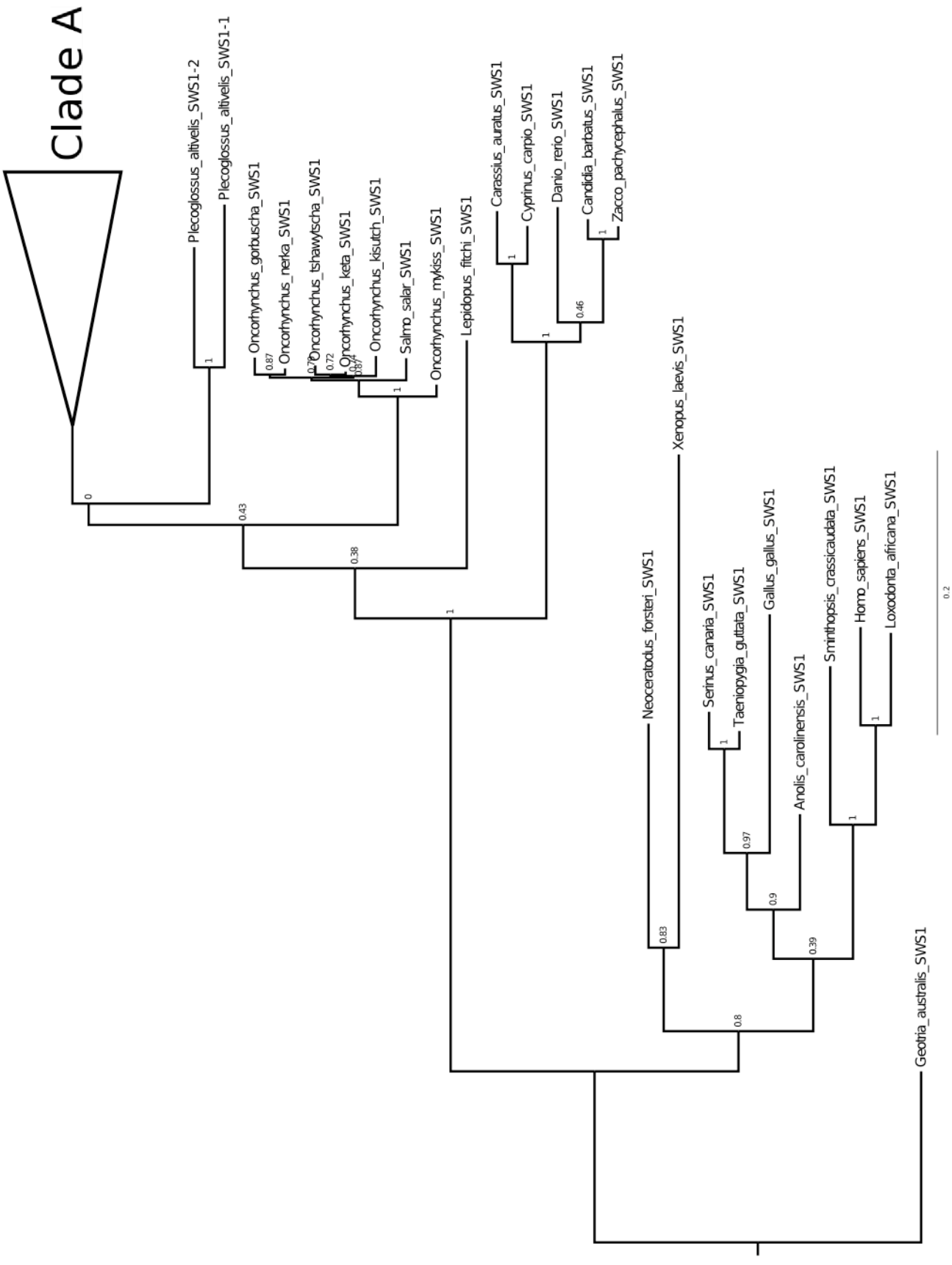
Figure 2.4: Histogram of fish opsin gene duplications by age.

Histogram of duplication prevalence as sorted by divergence at third positions. Duplication events are colour-coded based on opsin subfamily. Divergence is measured using the Tamura-Nei correction in MEGA4 (Tamura & Nei, 1993; Tamura et al., 2007). Duplication RH1-II is not included (See results).

SWS1

I analyzed SWS1 gene sequences from thirty-seven fish species. The topology of the SWS1 gene tree was largely consistent with fish taxonomy with the exception of scabbardfish, *Lepidopus fitchi* (See discussion) (Figure 2.5).

Only one of the forty-two opsin gene duplication nodes occurred on the SWS1 tree, SWS1-I. Ayu smelt (*Plecoglossus altivelis*) have two SWS1 genes (AYU-UV1 and AYU-UV2). These paralogs (85% identical at the nucleotide level) are almost as different from one another as they are from single-copy SWS1 genes from species in the family Salmonidae (79% and 75% identical to *O. kisutch* respectively). This observation suggests that the duplication event occurred very early during the evolution of the smelt family Osmeridae.


Clade A
Plecoglossus_altivelis_SWS1-2
Plecoglossus_altivelis_SWS1-1
Oncorhynchus_gorbuscha_SWS1
Oncorhynchus_nerka_SWS1
Oncorhynchus_tshawytscha_SWS1
Oncorhynchus_keta_SWS1
Oncorhynchus_kisutch_SWS1
Salmo_salar_SWS1
Oncorhynchus_mykiss_SWS1
Lepidopus_fitchi_SWS1
Carassius_auratus_SWS1
Cyprinus_carpio_SWS1
Danio_rerio_SWS1
Candidia_barbatus_SWS1
Zacco_pachycephalus_SWS1
Xenopus_laevis_SWS1
Neoceratodus_forsteri_SWS1
Serinus_canaria_SWS1
Taeniopygia_guttata_SWS1
Gallus_gallus_SWS1
Anolis_carolinensis_SWS1
Sminthopsis_crassicaudata_SWS1
Homo_sapiens_SWS1
Loxodonta_africana_SWS1
Geotria_australis_SWS1
0.2

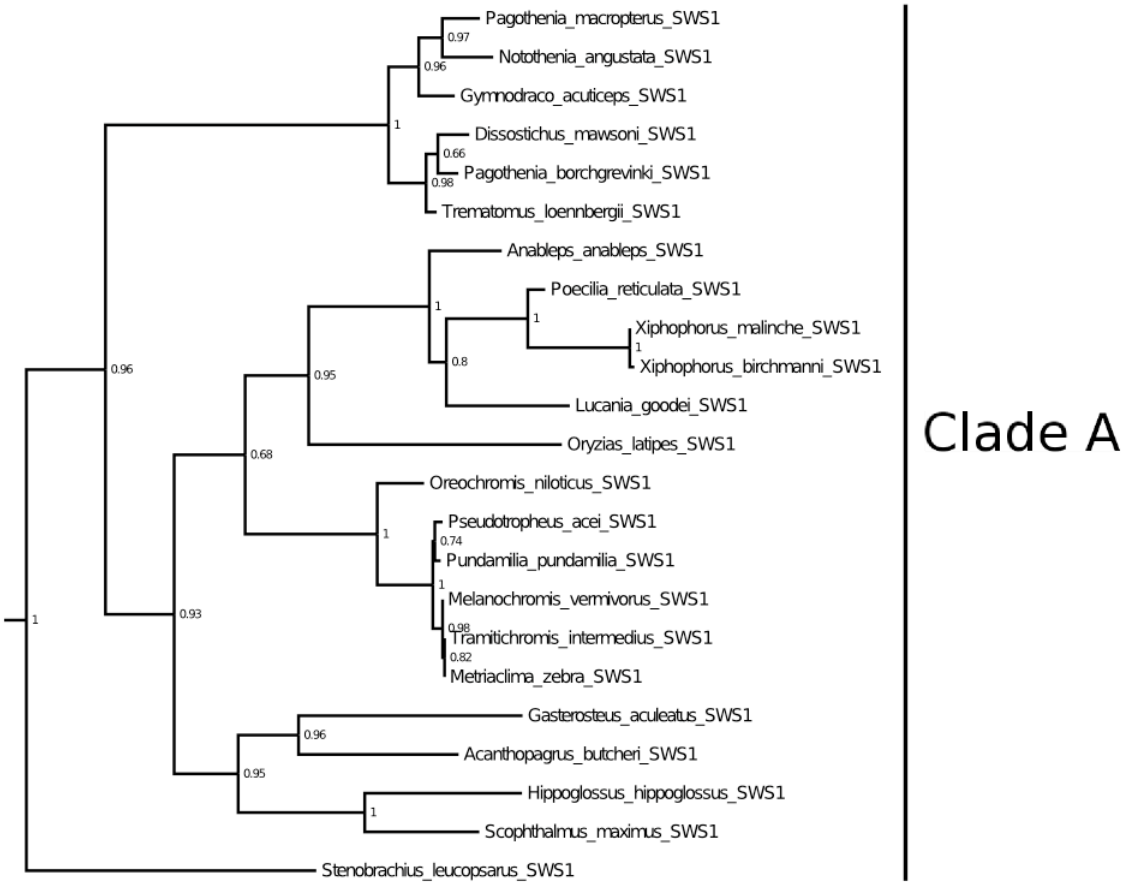


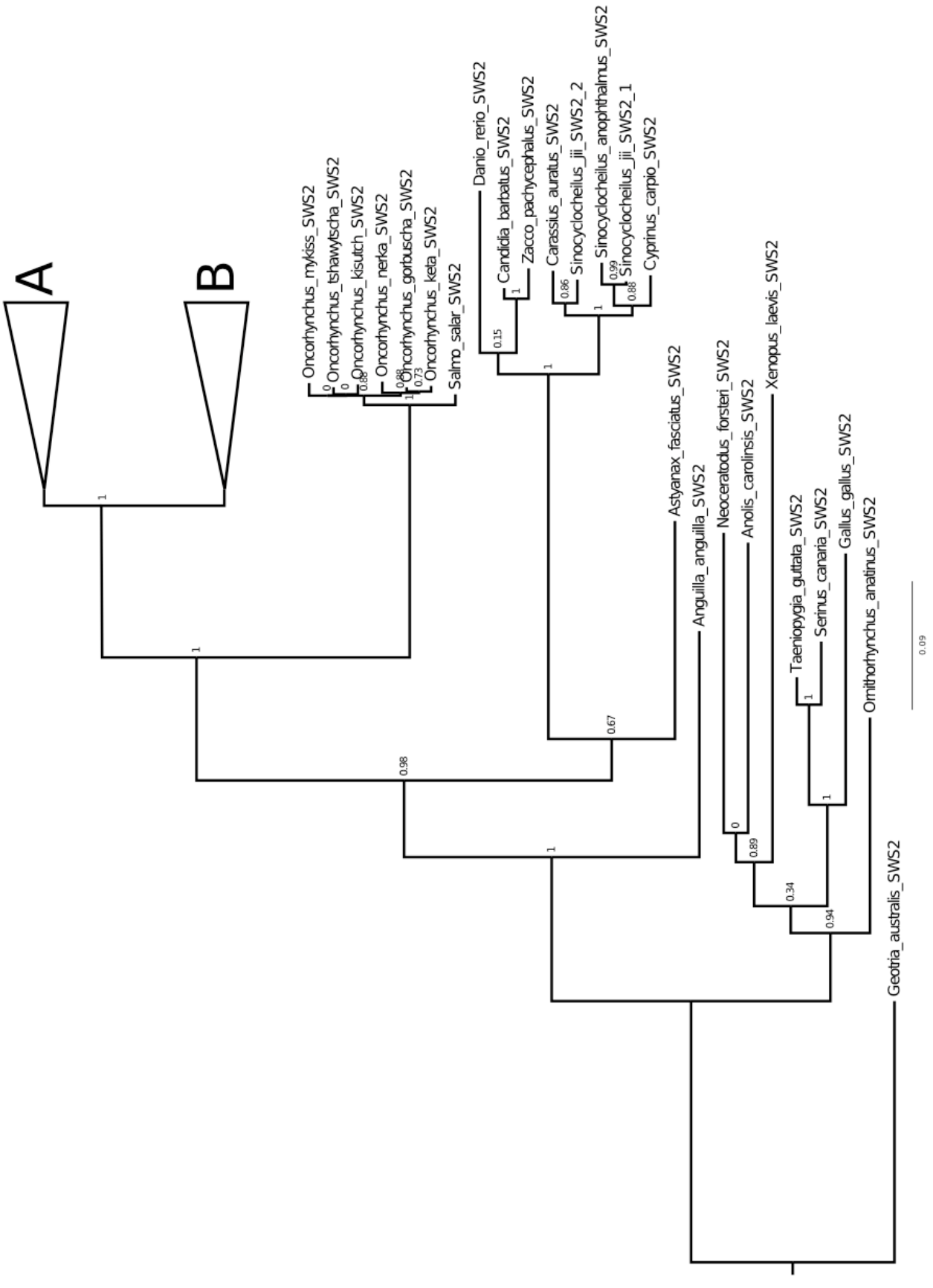
Figure 2.5: Phylogenetic tree of SWS1 opsins in fish.

The tree was created using the maximum-likelihood method. Accession numbers are listed in Appendix 1. SWS1 opsin from lamprey (*G. australis*) was used as a root. PhyML was used to estimate genetic distances, based on Modeltest's best-fit model of evolution, and complete phylogenetic analysis (Guindon & Gascuel, 2003; Posada & Crandall, 1998). Tree topology was tested using the best of NNI and SPR. Numbers at nodes represent aLRT values (Anisimova & Gascuel, 2006). The model of evolution was determined to be HKY85+I+G (I = 0.2299, G = 1.1920). Clade A encompasses Neoteleostei.

SWS2

SWS2 opsin genes were analyzed for thirty-nine species. Sequence relationships followed established species taxonomy (Figure 2.6). Two gene duplication events are represented on the tree, one in the group Cyprinini-*Sinocyclocheilus* (SWS2-II) and the other encompassing a large portion of Percomorpha (SWS2-I).

For the younger duplication event, this analysis indicates that it occurred in the clade Cyprinini-*Sinocyclocheilus* including the genera *Sinocyclocheilus*, *Cyprinus* and *Carassius* (Yang et al., 2010). This duplication node was supported by an aLRT value of .99, and 100% bootstrap support in NJ trees. The second duplication, producing the paralogous clades SWS2A and SWS2B, likely occurred in the ancestor of Holacanthopterygii, a taxonomic group that includes Paracanthoptergyii (represented by *Gadus morhua)* and Acanthoptergyii (cichlids, livebearers and others). The ML tree indicates that the *Gadus morhua* sequence is within the SWS2A clade, with poor support (0.105 aLRT) while NJ and MP methods placed it as the outgroup to the duplication node. Thus, the duplication event might have occurred after acanthopterygians diverged from *Gadus* and other paracanthoptergygians. The orders Beloniformes, Cyprinodontiformes, and Perciformes have representatives containing both duplicates, while Atheriniformes, Tetraodontiformes, Scorpaeniformes, Pleuronectiformes, and Gasterosteiformes have representatives of one of the two SWS2 duplicates. Interestingly, in stickleback (Order: Gasterosteiformes) an SWS2B pseudogene is present in the genome, indicating a relatively recent gene loss.


A
B
Oncorhynchus_mykiss_SWS2
Oncorhynchus_tshawytscha_SWS2
Oncorhynchus_kisutch_SWS2
Oncorhynchus_nerka_SWS2
Oncorhynchus_gorbuscha_SWS2
Oncorhynchus_keta_SWS2
Salmo_salar_SWS2
Danio_rerio_SWS2
Candidia_barbatus_SWS2
Zacco_pachycephalus_SWS2
Carassius_auratus_SWS2
Sinocyclocheilus_jii_SWS2_2
Sinocyclocheilus_anophthalmus_SWS2
Sinocyclocheilus_jii_SWS2_1
Cyprinus_carpio_SWS2
Astyanax_fasciatus_SWS2
Anguilla_anguilla_SWS2
Neoceratodus_forsteri_SWS2
Anolis_carolinsis_SWS2
Xenopus_laevis_SWS2
Taeniopygia_guttata_SWS2
Serinus_canaria_SWS2
Gallus_gallus_SWS2
Ornithorhynchus_anatinus_SWS2
Geotria_australis_SWS2
0.09

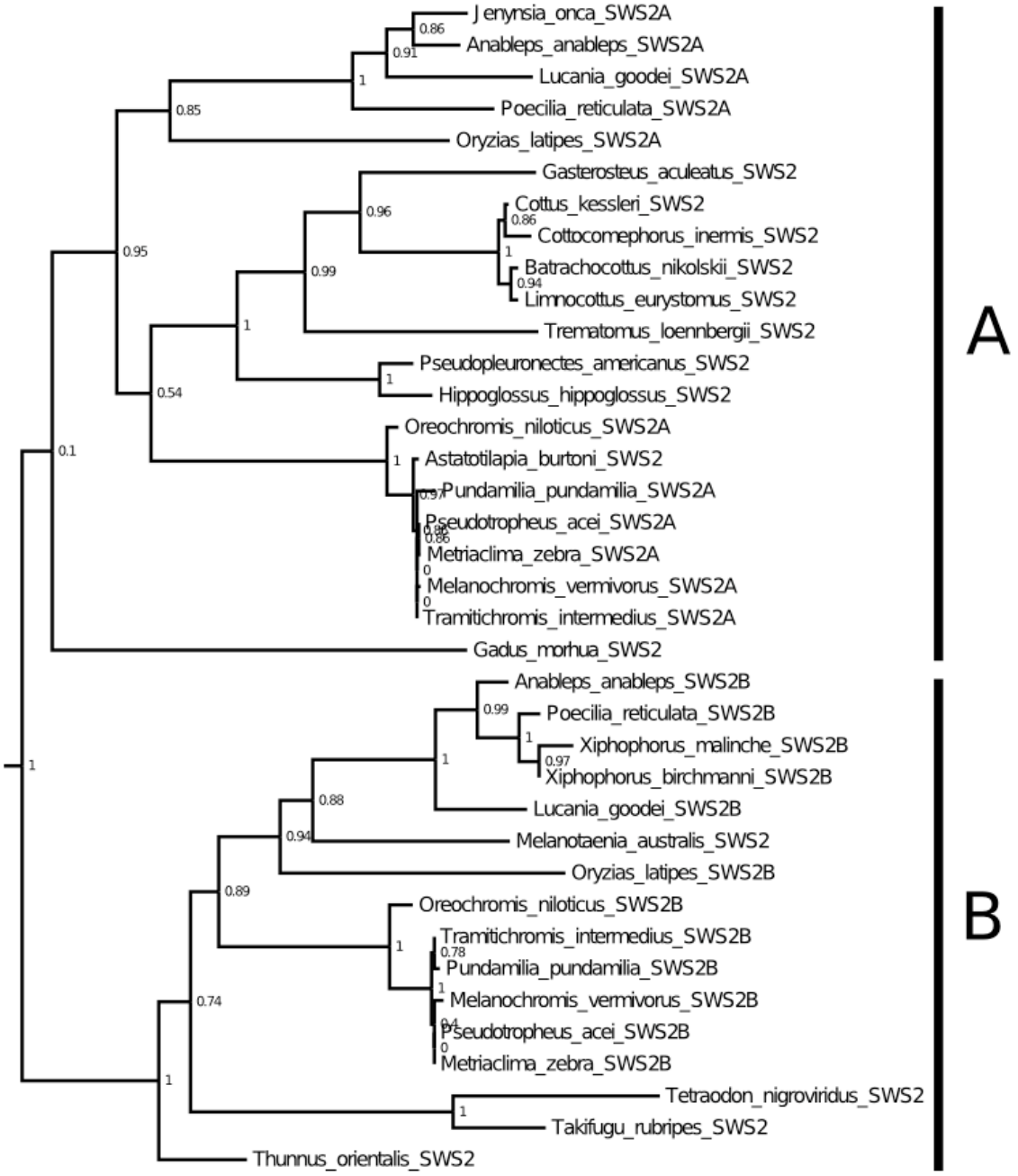


Figure 2.6: Phylogenetic tree of SWS2 opsins in fish.

The tree was created using the maximum-likelihood method. Accession numbers are listed in Appendix 1. SWS2 opsin from lamprey (*G. australis*) was used as a root. PhyML was used to estimate genetic distances, based on Modeltest's best-fit model of evolution, and complete phylogenetic analysis (Guindon & Gascuel, 2003; Posada & Crandall, 1998). Tree topology was tested using the best of NNI and SPR. Numbers at nodes represent aLRT values (Anisimova & Gascuel, 2006). The model of evolution was determined to be HKY85+I+G (I = 0.2238, G = 1.1396). The letters A and B indicate SWS2A and SWS2B clades respectively.

RH1

RH1 opsin gene sequences were used for seventy species of ray-finned fish, and represent six gene duplication events. Sequence relationships largely match species level taxonomy although sarcopterygian sequences do not form a monophyletic clade (Figure 2.7). This is likely because the model of evolution used in analysis is based on fish RH1 genes, which may differ from the evolutionary parameters present in sarcopterygians. The first duplication (RH1-I) separates actinopterygian RH1 and exo-rhodopsin genes. This node reflects a retroduplication of the ancestral RH1 gene. RH1 opsin retroduplication appears to have occurred in the common ancestor of all actinopterygians, as sturgeon (*Acipenser sp.*), paddlefish (*Polyodon spathula*), and bowfin (*Amia calva*) all possess the single-exon RH1. The monophyly of actinopygian RH1 and ExoRh is not supported in NJ or MP analysis. These two trees support the hypothesis that the ExoRh duplication occurred before actinopterygians and sarcopterygians diverged. However, the placement of the retroduplication node at the base of Actinopterygii has an aLRT value of 0.97 using ML with a model of evolution based on the sequences.

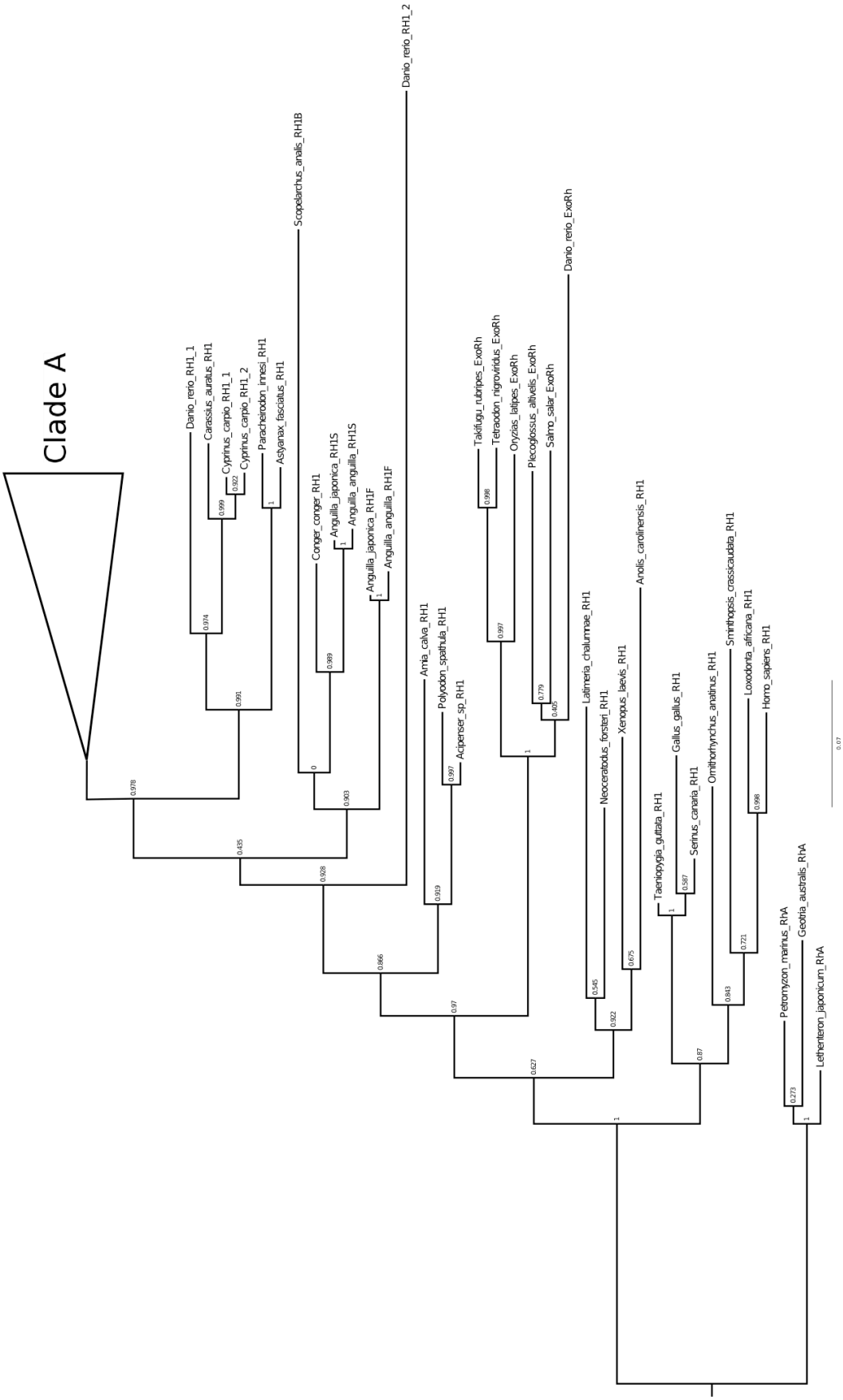
Zebrafish (*Danio rerio*) have two single-exon RH1 genes in addition to exo-rhodpsin. One of these RH1 paralogs (RH1-2) is the sister sequence to most other teleost RH1 genes. Tree reconciliation would place a duplication node at this point in the tree and suggest that orthologs of RH1-2 had been lost in all other species. However, I suspect that a tree reconstruction error explains the position of this gene at the base of the teleost RH1 clade (see discussion). This is also the case for the RH1 duplicate in the pearl eye (*Scopelarchus analis*), which is placed as sister group to the deep-sea eel RH1 genes, a position that is inconsistent with well-established taxonomy. Both duplicates have low aLRT values for their placement and are part of a large polytomy in the NJ and MP trees.

The third node from the base of the ray-finned fish RH1 tree separates five eel RH1 genes from all but one of the orthologs from non-elopomorph teleosts. Within the eels, RH1 was duplicated before the genera *Anguilla* and *Conger* diverged to produce the differentially expressed freshwater and deep-sea paralogs (Hope et al., 1998).

Two additional gene duplication events are present in this tree, one (RH1-VI), which occurred in the carp (genus *Cyprinus*), after it diverged from goldfish (genus *Carassius*)

and the other (RH1-V) in the scabbard fish (*Lepidopus fitchi*). As seen in the SWS1 tree, the scabbard fish sequences group with salmonid sequences contrary to established species relationships.

Strangely, RH1 from the catfish (*Ictalurus punctatus*) is expected to occur in a clade with goldfish and zebrafish (all occur in the taxon Ostariophysi, within Otocephala), but instead it forms a monophyletic clade with RH1 sequences from the breams (e.g., genera *Pagrus* and *Acanthopagrus*). Due to the high support (aLRT value of 1.0), I suspect the sample was mislabelled.


Clade A
Danio_rerio_RH1_1
Carassius_auratus_RH1
Cyprinus_carpio_RH1_1
Cyprinus_carpio_RH1_2
Paracheirodon_innesi_RH1
Astyanax_fasciatus_RH1
Scopelarchus_analis_RH1B
Conger_conger_RH1
Anguilla_japonica_RH1S
Anguilla_anguilla_RH1S
Anguilla_japonica_RH1F
Anguilla_anguilla_RH1F
Danio_rerio_RH1_2
Amia_calva_RH1
Polyodon_spathula_RH1
Acipenser_sp_RH1
Takifugu_rubripes_ExoRh
Tetraodon_nigroviridus_ExoRh
Oryzias_latipes_ExoRh
Plecoglossus_altivelis_ExoRh
Salmo_salar_ExoRh
Danio_rerio_ExoRh
Latimeria_chalumnae_RH1
Neoceratodus_forsteri_RH1
Xenopus_laevis_RH1
Anolis_carolinensis_RH1
Taeniopygia_guttata_RH1
Gallus_gallus_RH1
Serinus_canaria_RH1
Ornithorhynchus_anatinus_RH1
Sminthopsis_crassicaudata_RH1
Loxodonta_africana_RH1
Homo_sapiens_RH1
Petromyzon_marinus_RhA
Geodia_australis_RhA
Lethenteron_japonicum_RhA
0.978
0.374
0.999
0.922
0.991
1
0
0.989
1
0.903
1
0.435
0.928
0.997
0.919
0.866
0.998
0.997
1
0.779
0.405
0.97
0.545
0.922
0.675
1
0.587
0.87
0.843
0.721
0.998
0.627
1
0.273
1
0.07

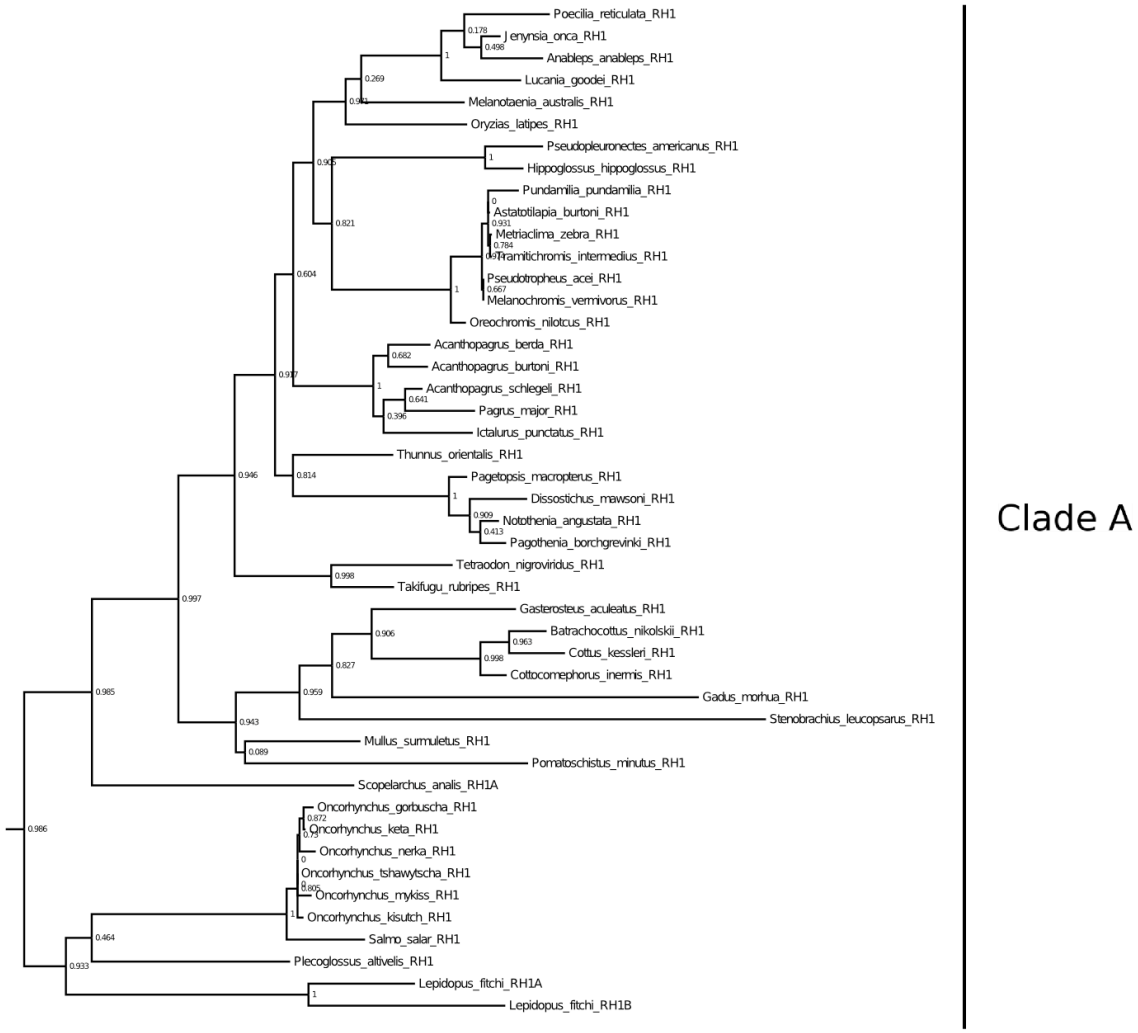


Figure 2.7: Phylogenetic tree of RH1 opsins in fish.

The tree was created using the maximum-likelihood method. Accession numbers are listed in Appendix 1. RhA opsins from lamprey (*G. australis, P. marinus* and *L. japonicum*) were used as a root. PhyML was used to estimate genetic distances, based on Modeltest's best-fit model of evolution, and complete phylogenetic analysis (Guindon & Gascuel, 2003; Posada & Crandall, 1998). Tree topology was tested using the best of NNI and SPR. Numbers at nodes represent aLRT values (Anisimova & Gascuel, 2006). The model of evolution was determined to be GTR+I+G (I = 0.2746, G = 1.1396). Clade A encompasses Euteleostei.

RH2

Forty-seven species are represented in the RH2 tree. This gene-tree matches well with the established species-tree with a few noted exceptions (Figure 2.8). Contrasting the

SWS1, SWS2, and RH1 opsin subfamilies, where the duplication appears to be rare, the RH2 opsin subfamily has twenty-one of the forty-two duplication nodes.

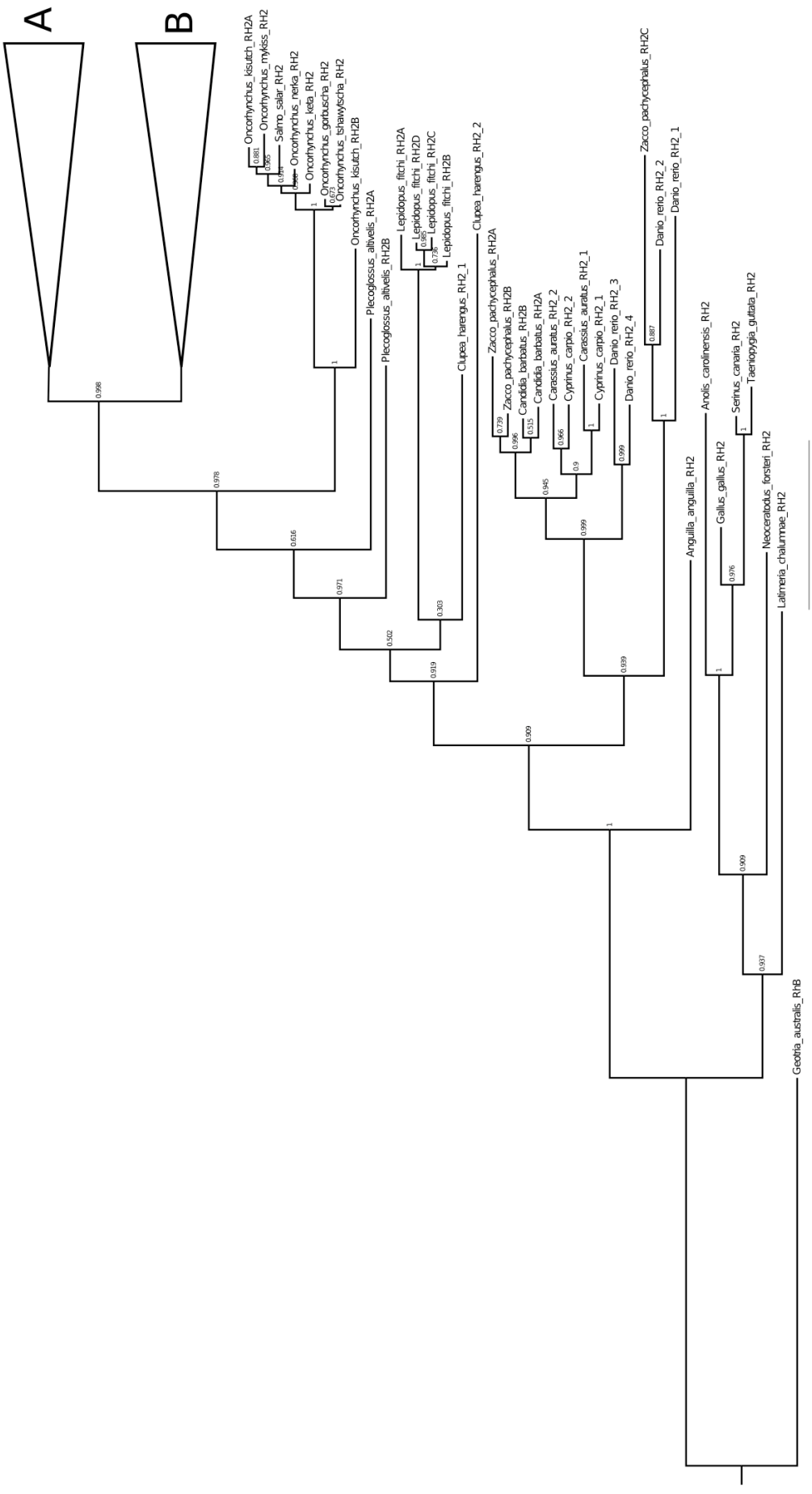
Within the order Cypriniformes, there are a six duplication events; two shared by all species (RH2-II and RH2-V), one in the carp and goldfish ancestor (genera: *Cyprinus* and *Carassius*) (RH2-X) and one for each of the following three genera *Danio* (RH2-IX), *Zacco* (RH2-XII) and *Candidia* (RH2-XVI). The Atlantic herring (*Clupea harengus*) also has two RH2 genes. These occur in phylogenetically confusing locations in this tree. Herring are within the order Clupeiformes and group in Otocephala together with the order Cypriniformes. This relationship is not presented in the gene tree. The tree shows one herring gene (RH2_2) as sister to all other fish sequences and the other (RH2_1) as sister to a clade of four RH2 genes from a euteleost, the scabbardfish, *Lepidopus fitchi*. The scabbardfish has three independent RH2 duplication events (RH2-XV, RH2-XVIII and RH2-XXI) and all are, again, in a position in the tree inconsistent with its species taxonomy (See discussion). Coho salmon (*Oncorhynchus kisutch*) has two RH2 paralogs stemming from a gene duplication event (RH2-VII) before the divergence of the salmonids, although so far not found in any other salmonid species.

Ayu (*Plecoglossus altivelis*) has two RH2 genes that are, similar to herring, not sister sequences. Tree reconciliation would attribute both ayu and herring paralogs to ancestral duplication events and subsequent gene loss of one paralog in all other species. I suspect the timing of these duplication events are not accurately reflected in this phylogenetic analysis. In the MP trees, each pair of paralogs forms a monophyletic group, a pattern that does not infer an enormous number of independent gene loss events in other taxa.

Another RH2 duplication node (RH2-I) near the base of the tree, marks the generation of the paralogous RH2A and RH2B gene trees. Though the paralogs produced by this event were originally called RH2-1 and RH2-2 in pufferfish, I have adopted the more commonly used RH2A-RH2B notation. This tandem duplication (see below) occurred in the ancestor of fish in the taxon Neoteoleosti (i.e., in Euteleostei after procanthopterygians such as smelt, pike and salmonids diverged). Some surveyed species lack RH2B although they are not members of a single monophyletic group, thus RH2B gene loss may have occurred independently several times. The lanternfish (*Stenobrachius leucopsarus*) has undergone three independent RH2B duplication events (RH2-VIII,

RH2-XI and RH2-XX) to produce four RH2 genes (one gene is pseudogenized and not included in this analysis).

In the RH2A clade, the first two nodes separate the tuna (*Thunnus orientalis*) and red seabream (*Pagrus major*) RH2B from all others. As the gene name implies, these sequences are likely orthologous with RH2B genes from other species and their inclusion in the RH2A clade may be a case of long-branch attraction or gene conversion (see discussion). Within the RH2A clade there have been further gene duplications in stickleback (*Gasterosteus aculeatus*) (RH2-XIX), seabreams (genus: *Acanthopagrus*) (RH2-XVII), medaka (*Oryzias latipes*) (RH2-XIII), turbot (*Scophthalmus maximus*) (RH2-VI) and cichlids (family: Cichlidae) (RH2-XIV). Although this tree predicts independent duplications in Nile tilapia (*Oreochromis niloticus*) and the African cichlids (i.e. *Pseudotropheus acei*), this is not supported by other analyses and may be an artefact due to the model used (Spady et al., 2006; Shand et al., 2008).


A
B
Oncorhynchus_kisutch_RH2A
Oncorhynchus_mykiss_RH2
Salmo_salar_RH2
Oncorhynchus_nerka_RH2
Oncorhynchus_keta_RH2
Oncorhynchus_gorbuscha_RH2
Oncorhynchus_tshawytscha_RH2
Oncorhynchus_kisutch_RH2B
Plecoglossus_altivelis_RH2A
Plecoglossus_altivelis_RH2B
Lepidopus_fitchi_RH2A
Lepidopus_fitchi_RH2D
Lepidopus_fitchi_RH2C
Lepidopus_fitchi_RH2B
Clupea_harengus_RH2_1
Clupea_harengus_RH2_2
Zacco_pachycephalus_RH2A
Zacco_pachycephalus_RH2B
Candidia_barbatus_RH2B
Candidia_barbatus_RH2A
Carassius_auratus_RH2_2
Cyprinus_carpio_RH2_2
Carassius_auratus_RH2_1
Cyprinus_carpio_RH2_1
Danio_rerio_RH2_3
Danio_rerio_RH2_4
Zacco_pachycephalus_RH2C
Danio_rerio_RH2_2
Danio_rerio_RH2_1
Anguilla_anguilla_RH2
Anolis_carolinensis_RH2
Gallus_gallus_RH2
Serinus_canaria_RH2
Taeniopygia_guttata_RH2
Neoceratodus_forsteri_RH2
Latimeria_chalumnae_RH2
Geobia_australis_RH2
0.998
0.978
0.616
0.971
0.502
0.303
0.619
0.909
0.999
0.939
0.945
0.999
0.999
0.887
0.976
0.909
0.937
0.09

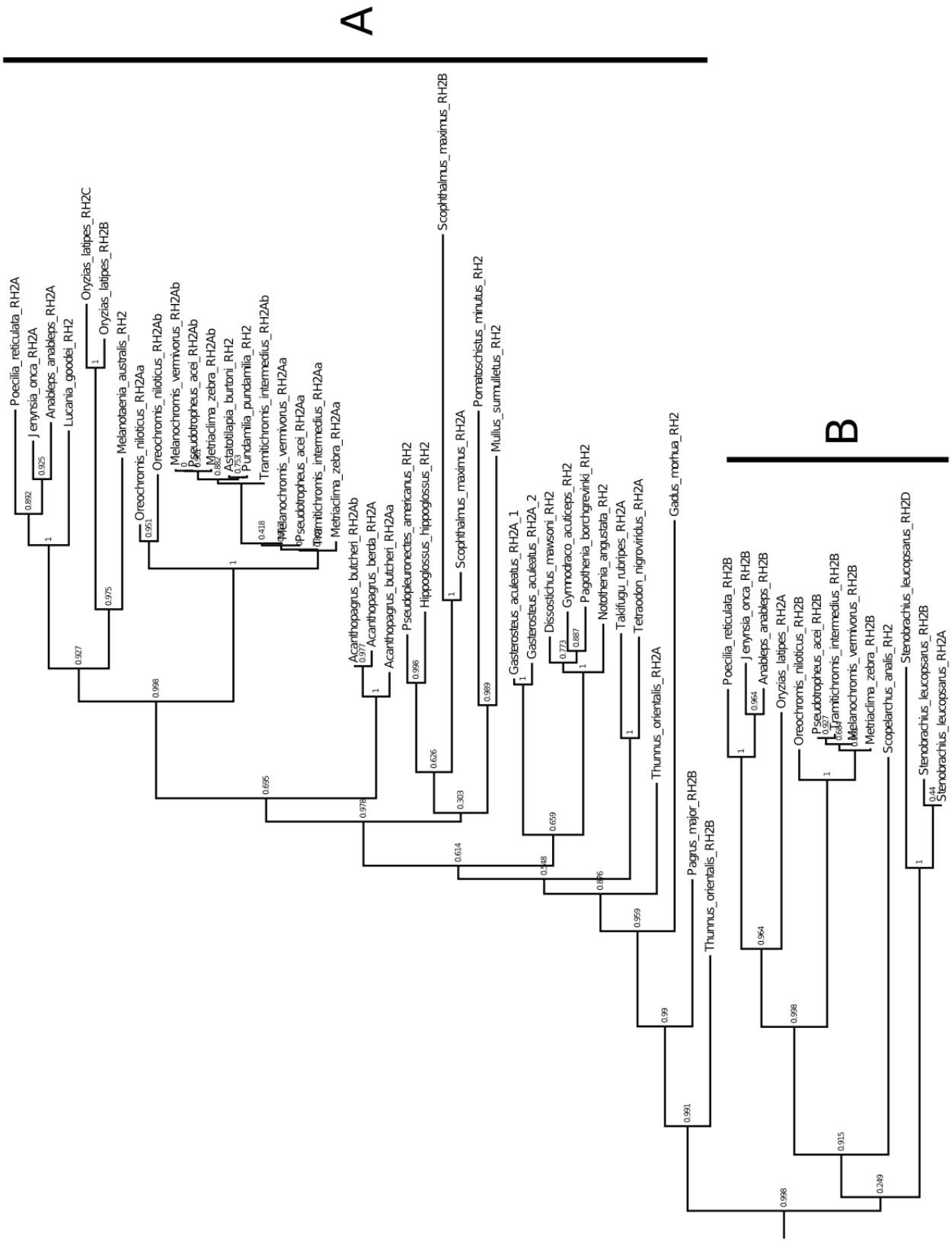

A
Poecilia_reticulata_RH2A
Jenynsia_onca_RH2A
Anableps_anableps_RH2A
Lucania_goodei_RH2
Oryzias_latipes_RH2C
Oryzias_latipes_RH2B
Melanotaenia_australis_RH2
Oreochromis_niloticus_RH2Aa
Oreochromis_niloticus_RH2Ab
Melanochromis_vermivorus_RH2Ab
Pseudotropheus_acei_RH2Ab
Metriaclima_zebra_RH2Ab
Astatotilapia_burtoni_RH2
Pundamilia_pundamilia_RH2
Tramitichromis_intermedius_RH2Ab
Melanochromis_vermivorus_RH2Aa
Pseudotropheus_acei_RH2Aa
Tramitichromis_intermedius_RH2Aa
Metriaclima_zebra_RH2Aa
Acanthopagrus_butcheri_RH2Ab
Acanthopagrus_berda_RH2A
Acanthopagrus_butcheri_RH2Aa
Pseudopleuronectes_americanus_RH2
Hippoglossus_hippoglossus_RH2
Scophthalmus_maximus_RH2B
Scophthalmus_maximus_RH2A
Pomatoschistus_minutus_RH2
Mullus_surmuletus_RH2
Gasterosteus_aculeatus_RH2A_1
Gasterosteus_aculeatus_RH2A_2
Dissostichus_mawsoni_RH2
Gymnodraco_acuticeps_RH2
Pagothenia_borchgreviniki_RH2
Notothenia_angustata_RH2
Takifugu_rubripes_RH2A
Tetraodon_nigroviridis_RH2A
Thunnus_orientalis_RH2A
Gadus_morhua_RH2
Pagrus_major_RH2B
Thunnus_orientalis_RH2B
B
Poecilia_reticulata_RH2B
Jenynsia_onca_RH2B
Anableps_anableps_RH2B
Oryzias_latipes_RH2A
Oreochromis_niloticus_RH2B
Pseudotropheus_acei_RH2B
Tramitichromis_intermedius_RH2B
Melanochromis_vermivorus_RH2B
Metriaclima_zebra_RH2B
Scopelarchus_analis_RH2
Stenobrachius_leucopsarus_RH2D
Stenobrachius_leucopsarus_RH2B
Stenobrachius_leucopsarus_RH2A

Figure 2.8: Phylogenetic tree of RH2 opsins in fish.

The tree was created using the maximum-likelihood method. Accession numbers are listed in Appendix 1. RhB opsin from lamprey (*G. australis*) was used as a root. PhyML was used to estimate genetic distances, based on Modeltest's best-fit model of evolution, and complete phylogenetic analysis (Guindon & Gascuel, 2003; Posada & Crandall, 1998). Tree topology was tested using the best of NNI and SPR. Numbers at nodes represent aLRT values (Anisimova & Gascuel, 2006). The model of evolution was determined to be GTR+I+G (I = 0.2485, G = 0.8147). The letters A and B indicate RH2A and RH2B clades respectively.

LWS

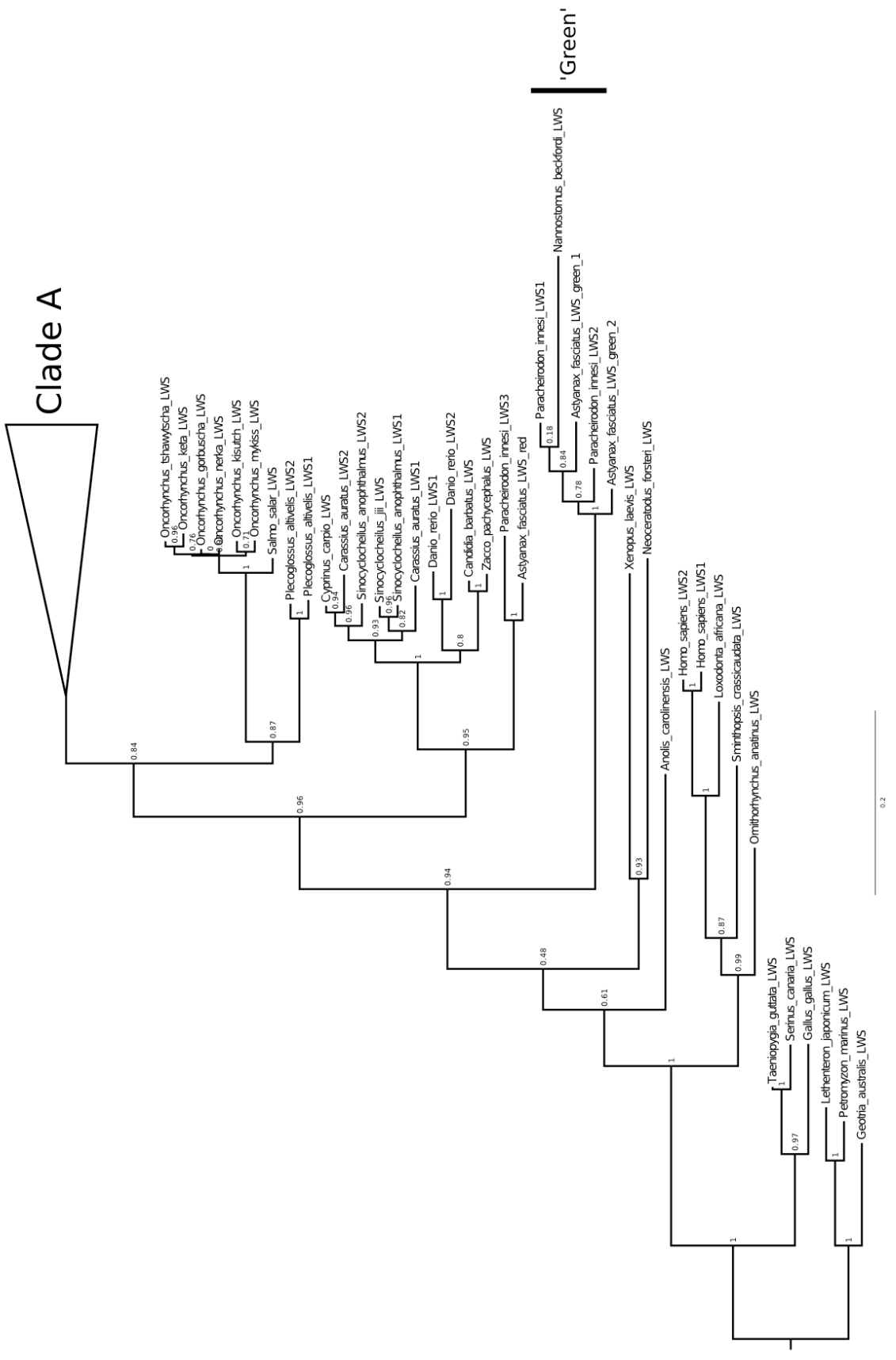
In the LWS opsin gene analysis, sarcopterygians do not form a monophyletic group. This is most likely due to the use of a model of evolution based only on fish opsins. Phylogenetic analysis of LWS opsins (sequences from fifty-eight species) produced a tree with a topology that was largely consistent with species-level relationships among teleosts (Figure 2.9). There was one especially interesting deviation: Blind cave (*Astyanax fasciatus*) and neon tetra (*Paracheirodon innesi*) are members of the taxon Ostariophysi, together with zebrafish. Both species possess a gene that is similar to LWS opsins from other ostariophysians in the survey, i.e., cavefish and neon tetra LWS opsins form the sister group to LWS opsins from Cypriniformes. However, both species also have a pair of genes that differ from the LWS opsins found in all other fish. My lab sequenced an ortholog of these unusual opsins in the golden pencilfish, *Nannostomus beckfordi*. The sister group relationship between this five-gene clade and the LWS opsin genes in other fish places a duplication node (LWS-I) at the base of the fish LWS tree with strong support. I did not find orthologs of these genes in a BLAST search of any of the whole genome sequences available for ray-finned fish. By including the new data from pencilfish, *Nannostomus beckfordi*, in my analysis, I show that the cavefish and neon tetra green LWS opsin duplicates were produced by an event preceding the radiation of the family Characiformes, at minimum. When compared to other LWS opsin sequences, these five genes differ in specific domains and are almost identical in others (Figure 2.10). Specifically, the transmembrane 6 (TM6) and extracellular 3 (E3) domains differ where nearly all other sequences are high conserved. Although ML analysis does little to order internal relationships in the five gene clade, NJ and MP analysis suggest that neon tetra LWS1 and LWS2 and cavefish LWS green-1 and LWS green-2 are the

product of a shared duplication event (LWS-IV) that occurred after both species diverged from golden pencilfish.

Within the family Cypriniformes, there are two duplication events. One is only present in zebrafish (*Danio rerio*) (LWS-III), while the other appears to have occurred in the common ancestor of the clade Cyprinini-*Sinocyclocheilus* (including the genera *Cyprinus*, *Carassius*, and *Sinocyclocheilus*) (LWS-VI) (Yang et al., 2010). The smelt, *Plecoglossus altivelis*, has two LWS genes (98% identical) that were sequenced in separate studies and may represent alleles or a recent gene duplication (LWS-VII). The medaka (*Oryzias latipes*) also has a recent gene duplication event (LWS-X). These duplicons are 99% identical duplicates, not shared by any other species.

A large number of LWS opsin duplication events occur in the taxon Cyprinidontoidei, which includes the livebearers (e.g., guppies, swordtails, four-eyed fish and one-sided livebearer), splitfins, flagfish, and killifish. The first event in this taxon appears to have been a retroduplication producing LWS S180r (LWS-II) (Ward et al., 2008; Watson et al., 2010). It occurred after the American flagfish, *Jordanella floridae* lineage (Family: Cyprinodontidae) diverged from the other species surveyed from Cyprinidontoidei. *Jordanella floridae* also contains two LWS genes (98% identical), which either represent allelic differences or a recent gene duplication (LWS-VIII).

The LWS opsin gene tree also suggests that there have been three independent duplication events in anablepids and three in poeciliids. However gene conversion has obfuscated true orthology (see discussion). *J. onca* P180, *A. anableps* S180γ and Poeciliid P180 all stem from the same gene duplication event (LWS-V). Subsequently, the progenitor S180 gene duplicated independently in *Anableps anableps* (LWS-XI), *Poecilia reticulata* (LWS-IX) and *Xiphophorus helleri* (LWS-XII) to produce S180α/S180β, S180/A180 and S180-1/S180-2 respectively.


Clade A
'Green'
Oncorhynchus_tshawytscha_LWS
Oncorhynchus_keta_LWS
Oncorhynchus_gorbuscha_LWS
Oncorhynchus_nerka_LWS
Oncorhynchus_kisutch_LWS
Oncorhynchus_mykiss_LWS
Salmo_salar_LWS
Plecoglossus_altivelis_LWS2
Plecoglossus_altivelis_LWS1
Cyprinus_carpio_LWS
Carassius_auratus_LWS2
Sinocyclocheilus_anophthalmus_LWS2
Sinocyclocheilus_jii_LWS
Sinocyclocheilus_anophthalmus_LWS1
Carassius_auratus_LWS1
Danio_rerio_LWS1
Danio_rerio_LWS2
Candidia_barbatus_LWS
Zacco_pachycephalus_LWS
Paracheirodon_innesi_LWS3
Astyanax_fasciatus_LWS_red
Paracheirodon_innesi_LWS1
Nannostomus_beckfordi_LWS
Astyanax_fasciatus_LWS_green_1
Paracheirodon_innesi_LWS2
Astyanax_fasciatus_LWS_green_2
Xenopus_laevis_LWS
Neoceratodus_forsteri_LWS
Anolis_carolinensis_LWS
Homo_sapiens_LWS2
Homo_sapiens_LWS1
Loxodonta_africana_LWS
Sminthopsis_crassicaudata_LWS
Ornithorhynchus_anatinus_LWS
Taeniopygia_guttata_LWS
Serinus_canaria_LWS
Gallus_gallus_LWS
Lethenteron_japonicum_LWS
Petromyzon_marinus_LWS
Geotria_australis_LWS
0.2

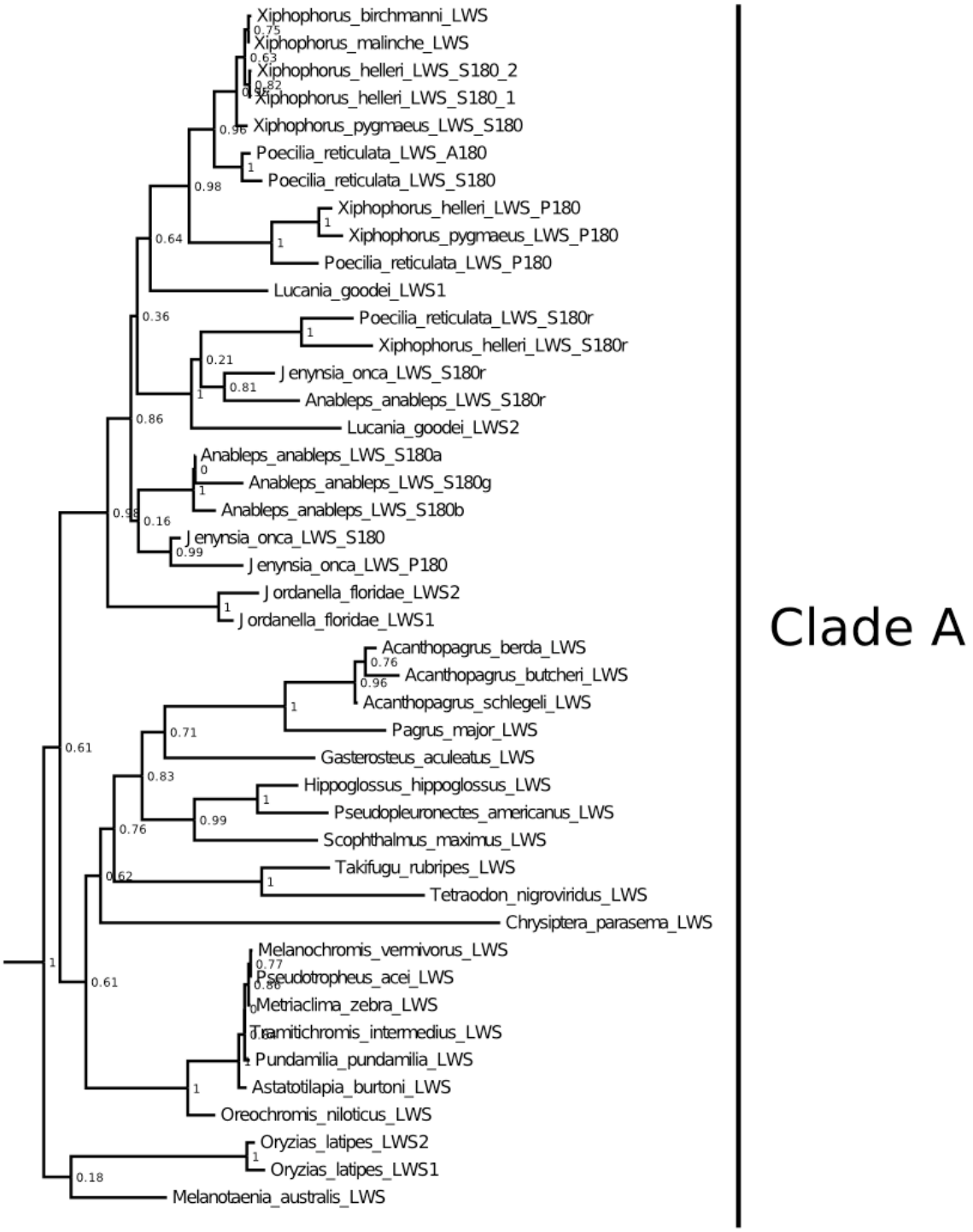


Figure 2.9: Phylogenetic tree of LWS opsins in fish.

The tree was created using the maximum-likelihood method. Accession numbers are listed in Appendix 1. LWS opsins from lamprey (*G. australis, P. marinus* and *L. japonicum*) were used as a root. PhyML was used to estimate genetic distances, based on modeltest's best-fit model of evolution, and complete phylogenetic analysis (Guindon & Gascuel, 2003; Posada & Crandall,

1998). Tree topology was tested using the best of NNI and SPR. Numbers at nodes represent aLRT values (Anisimova & Gascuel, 2006). The model of evolution was determined to be GTR+I+G (I = 0.3378, G = 1.2425). The 'green' clade is indicated (see discussion). Clade A encompasses Acanthopterygii.

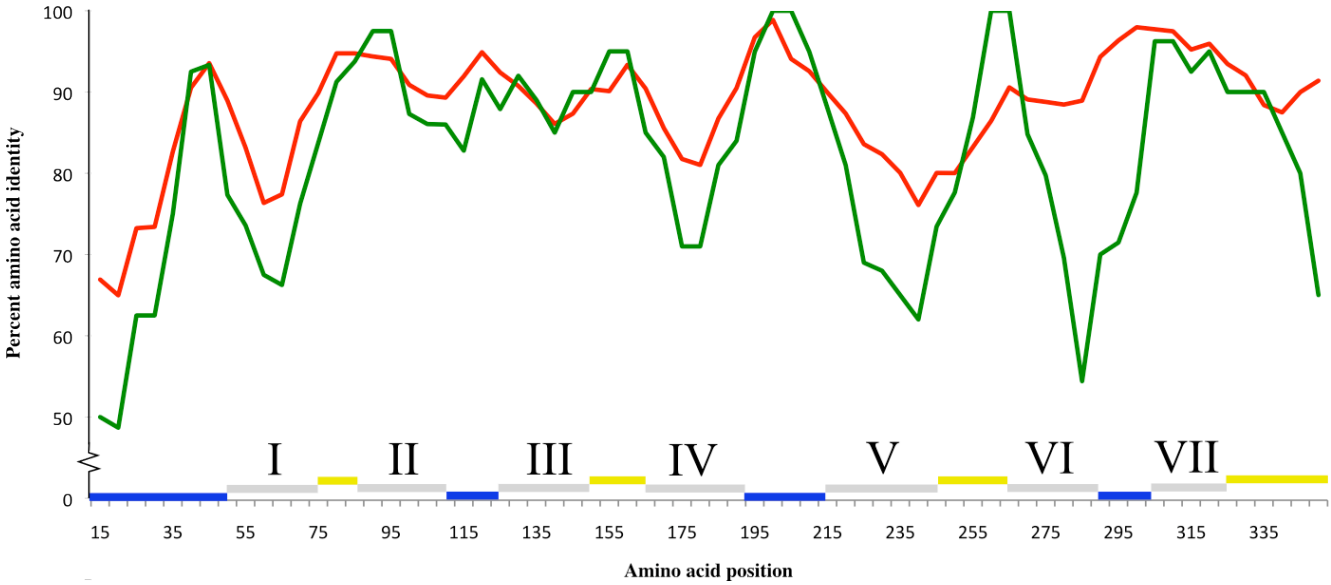


Figure 2.10: Sliding window analysis of 'green' LWS amino acid divergence.

The green line represents a sliding window analysis of average amino acid distance from *D. rerio* LWS-1 to each member of the 'green' LWS clade. Red line represents averaged pair-wise sliding window amino acid distance between all LWS sequences except the 'green' clade. Blue, grey and yellow bars represent external, transmembrane and internal domains respectively. Transmembrane domains are numbered. A window size of 30 was used for sliding window analysis.

Gene orientation and duplication

Tandem duplication appears to be the most common mode of opsin gene family expansion in fishes. Of the thirteen duplication events in species with genomic resources, twelve are tandem duplication events (Figure 2.11). The SWS2 paralogs, produced in the ancestor of Neoteleostei (SWS2-I) occur next to one another in a head to tail orientation in the Nile tilapia (*Oreochromis niloticus*), medaka (*Oryzias latipes*), green swordtail (*Xiphophorus helleri*) and the three-spine stickleback (*Gaterosteus aculeatus*). RH2A and RH2B occur in an inverted (tail to tail) orientation in the two pufferfishes and the cichlid, *Oreochromis nilioticus*. Several additional tandem duplication events have taken place at this locus. After the tandem duplication event producing RH2A and RH2B, RH2A was tandemly duplicated in *Oreochromis niloticus* (RH2-XIV), leaving this lineage with three

linked RH2 genes (RH2B, RH2Aa and RH2Ab). In stickleback, RH2B was lost (see above) and RH2A was tandemly duplicated. The RH2A paralogs in stickleback (RH2A-1 and RH2A-2) are oriented in a head to head pattern (Figure 2.11). They are 99% identical in coding region, suggesting that this second tandem duplication event was very recent. Stickleback RH2A-2 has a 39 bp deletion according to gene prediction and may be a pseudogene. In medaka, the RH2A and RH2B genes were originally miss-labelled. RH2A (called RH2-B) appears to have experienced inverted tandem duplication. This, followed by the loss of the progenitor gene, resulted in a head to tail arrangement of the RH2 gene pair with RH2A now upstream of RH2B. A tandem duplication of this repositioned RH2A leads to the three-gene repertoire and orientation present in medaka (Figure 2.12). The single copy RH2 gene in the ancestor of zebrafish experienced three tandem duplication events producing the four-gene locus present in this species (Figure 2.11).

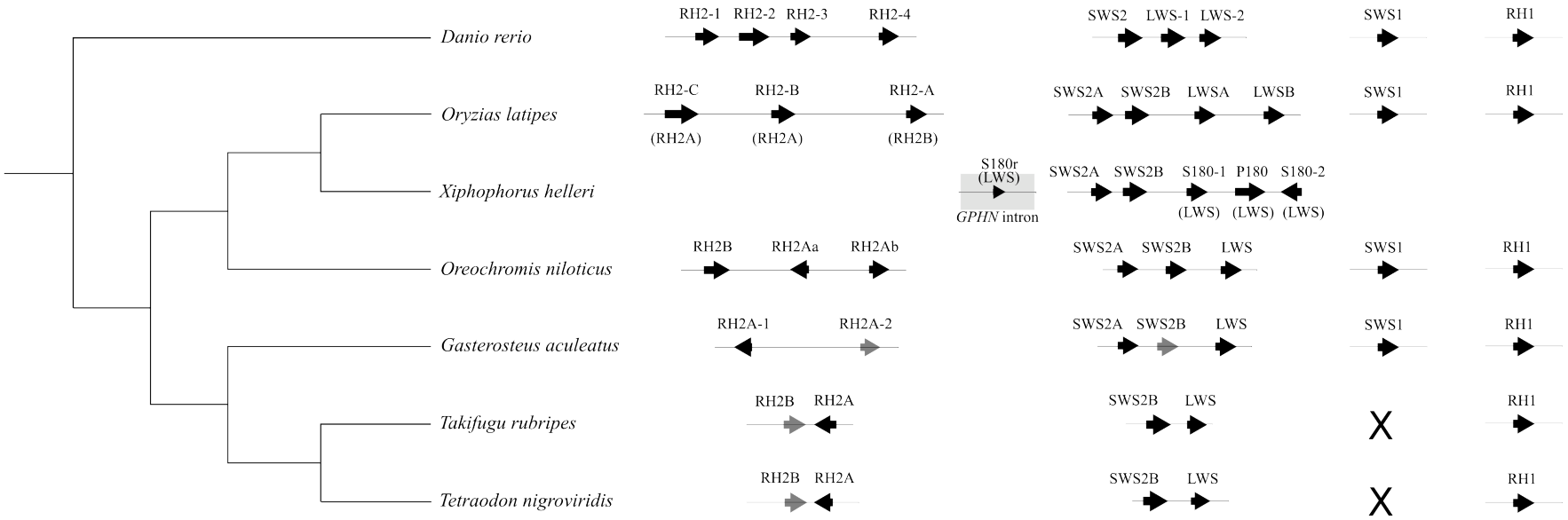

Danio rerio
Oryzias latipes
Xiphophorus helleri
Oreochromis niloticus
Gasterosteus aculeatus
Takifugu rubripes
Tetraodon nigroviridis
RH2-1 RH2-2 RH2-3 RH2-4
SWS2 LWS-1 LWS-2
SWS1
RH1
RH2-C RH2-B RH2-A
(RH2A) (RH2A) (RH2B)
SWS2A SWS2B LWSA LWSB
S180r (LWS)
GPHN intron
SWS2A SWS2B S180-1 P180 S180-2
(LWS) (LWS) (LWS)
RH2B RH2Aa RH2Ab
SWS2A SWS2B LWS
RH2A-1 RH2A-2
RH2B RH2A
SWS2B LWS
X

Figure 2.11: Opsin gene orientation and order.

Species tree based on established taxonomy with synteny of visual opsins in seven representative species. Gene size and intergenic regions are drawn to scale for all species except *O. niloticus*. Only LWS and SWS2 opsins are drawn for *X. helleri* due to a lack of information for other subtypes. Grey arrows are pseudogenes. X represents gene not present

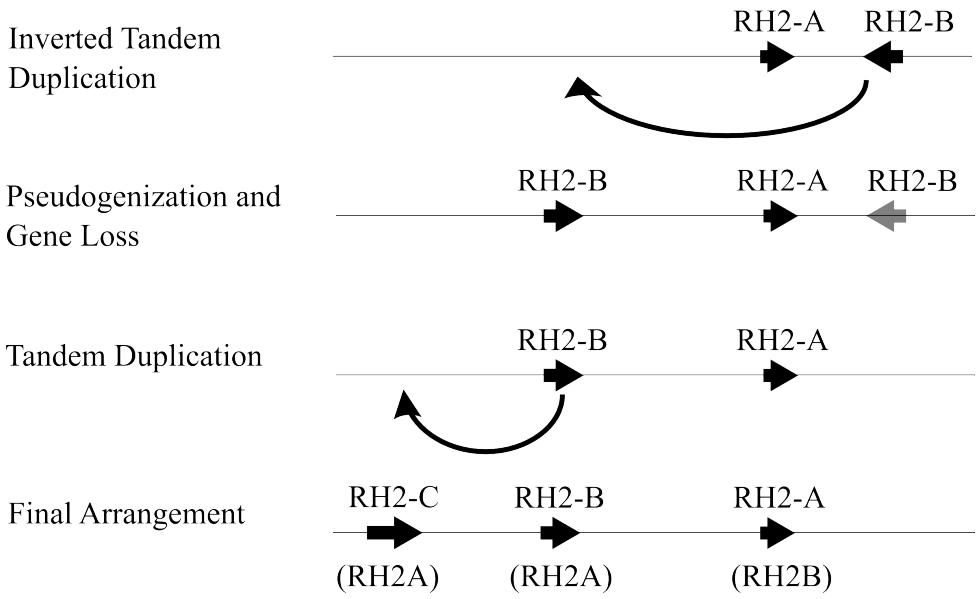


Figure 2.12: Hypothesized *O. latipes* RH2 rearrangement events.

Molecular events to account for the anomalous orientation and order of RH2 genes in *O. latipes.*

Tandem duplication is also the major contributor to LWS opsin gene subfamily amplification. Between-gene PCR and sequencing and screening large BAC clones indicate that LWS S180-2 and LWS P180 in livebearers are in an inverted position (tail to tail orientation) and that LWS S180-1 is upstream of LWS P180 (Ward et al., 2008; Watson et al., 2010). Both zebrafish and medaka also possess two linked LWS in a head to tail orientation (Figure 2.11).

Genome duplication can also cause gene family expansion. It has been suggested that the LWS-VI and SWS2-II duplications in Cyprininae are a consequence of tetraploidy (Li et al., 2009). An RH2 gene duplication event in salmonids (RH2-VII) may also be a result of tetraploidy; however there are currently no linkage data to confirm this (Temple et al., 2008).

The observation that RH1 and LWS S180r sequences derived from genomic DNA are missing all (RH1) or most (LWS S180r) introns indicates that both genes were produced by retroduplication. Opsin retroduplication appears to be rare; however, the prevalence of this mode of duplication cannot be estimated precisely because most opsin sequences in NCBI are derived from processed mRNA.

Synonymous and nonsynonymous substitutions

There are far fewer nonsynonymous substitutions per nonsynonymous site than there are synonymous substitutions per synonymous site indicating that opsins are under strong purifying selection. Average ω values for SWS1, SWS2, RH1, RH2 and LWS opsins were 0.113, 0.167, 0.074, 0.131 and 0.118 respectively (Figure 2.13) (Table 2.1, Table 2.2). However, average values can conceal interesting patterns of DNA sequence substitution along specific branches or in regions within a gene. To investigate these possibilities I compared ω values among subsets of branches within opsin gene subfamilies and among codons.

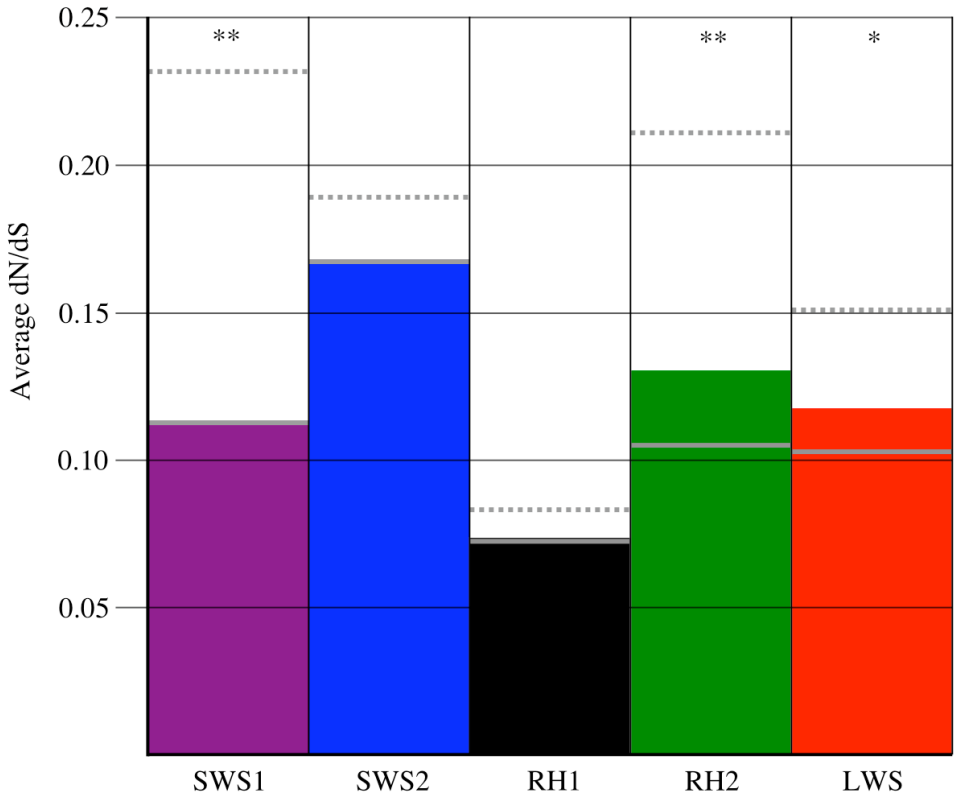


Figure 2.13: Average dN/dS ratio for each opsin subclass.

Solid colour bars represent the average for all sequences. Dotted grey lines denote the average for only gene duplication branches, while solid grey lines are the average for only speciation branches. In starred columns, gene duplication and speciation averages are significantly different (∗=p<0.05, ∗∗=p<0.01).

Table 2.1: Average ω for opsin subfamilies, subclades and branch types.

ω values were calculated using PAML and M0 or M2. In the branch type column, values in bold are significantly different from each other (p<0.05). For the subclades column, bolded font implies that the subclade has a significantly different ω from all other branches (Yang, 2007).

Gene	Overall	Branch type		Subclades	
SWS1	0.113	Gene duplication	**0.232**	-	-
		Speciation	**0.113**	-	-
SWS2	0.167	Gene duplication	0.164	SWS2A	0.187
		Speciation	0.167	SWS2B	0.173
RH1	0.074	Gene duplication	0.083	-	-
		Speciation	0.073	-	-
RH2	0.131	Gene duplication	**0.211**	RH2A	0.126
		Speciation	**0.114**	RH2B	**0.173**
LWS	0.118	Gene duplication	**0.153**	-	-
		Speciation	**0.111**	-	-

Table 2.2: Results for all PAML analysis using branch and site models.

Likelihood values, parameter estimates, LRT results for all branch and site PAML analyses. Positively selected sites are numbered using bovine rhodopsin as a reference (Nathans & Hogness, 1983). Since some opsin subfamilies have 5' extensions, numbering can go into negative to indicate codons before the bovine rhodopsin start codon. All listed sites have a posterior probability > 0.95, while >.99 is indicated with *. Bolded p values are significant ($p < 0.05$).

Gene	Model	lnL	Parameter estimates	Positively selected sites	Null model	*df*	2Δ lnL	*P* value
SWS1	M0 (one ratio)	-12105.663	$\omega = 0.113$	-	-	-	-	-
	Free Ratio	-11952.468		-	M0	69	306.390	**0**
	M2 (Gene duplication and speciation)	-12098.468	$\omega_{\text{Gene duplication}} = 0.232$ $\omega_{\text{Speciation}} = 0.113$	-	M0	1	14.391	**1.49E-04**
	M1a (nearly neutral)	-11933.744	$p_0 = 0.880$ $p_1 = 0.120$ $\omega_0 = 0.089$ $\omega_1 = 1.000$	Not allowed	-	-	-	-
	M2a (positive selection)	-11933.744	$p_0 = 0.880$ $p_1 = 0.079$ $p_2 = 0.042$ $\omega_0 = 0.089$ $\omega_1 = 1.000$ $\omega_2 = 1.000$	No selected sites	M1a	2	0.000	1
	M8a (beta&ωs =1)	-11710.461	$p_0 = 0.984$ $p = 0.408$ $q = 2.900$ $p_1 = 0.016$ $\omega = 1.000$	Not allowed	-	-	-	-
	M8 (beta&ω)	-11709.790	$p_0 = 0.990$ $p = 0.397$ $q = 2.676$ $p_1 = 0.010$ $\omega = 1.364$	34R	M8a	1	1.342	0.123
SWS2	M0 (one ratio)	-15203.596	$\omega = 0.167$	-	-	-	-	-
	Free Ratio	-15065.456		-	M0	85	276.280	**0**
	M2 (A, B and Background)	-15200.993	$\omega_A = 0.187$ $\omega_B = 0.173$ $\omega_{\text{Background}} = 0.151$	-	M0	2	5.206	0.074
	M2 (A and Background)	-15201.908	$\omega_A = 0.187$ $\omega_{\text{Background}} = 0.159$	-	M0	1	3.374	0.066
	M2 (B and Background)	-15203.462	$\omega_B = 0.173$ $\omega_{\text{Background}} = 0.165$	-	M0	1	0.266	0.606
	M2 (Gene duplication and speciation)	-15203.367	$\omega_{\text{Gene duplication}} = 0.189$ $\omega_{\text{Speciation}} = 0.166$	-	M0	1	0.456	0.5
	M1a (nearly neutral)	-14779.192	$p_0 = 0.792$ $p_1 = 0.208$ $\omega_0 = 0.094$ $\omega_1 = 1.000$	Not allowed	-	-	-	-
	M2a (positive selection)	-14777.371	$p_0 = 0.791$ $p_1 = 0.197$ $p_2 = 0.012$ $\omega_0 = 0.095$ $\omega_1 = 1.000$ $\omega_2 = 2.022$	No selected sites	M1a	2	3.643	0.162
	M8a (beta&ωs =1)	-14591.177	$p_0 = 0.940$ $p = 0.499$ $q = 2.996$ $p_1 = 0.060$ $\omega = 1.000$	Not allowed	-	-	-	-
	M8 (beta&ω)	-14586.497	$p_0 = 0.955$ $p = 0.476$ $q = 2.570$ $p_1 = 0.045$ $\omega = 1.381$	-3S, 1V*, 34S	M8a	1	9.362	**1.11E-03**
LWS	M0 (one ratio)	-12674.853	$\omega = 0.118$	-	-	-	-	-
	Free Ratio	-12533.947		-	M0	88	281.813	**0**
	M2 (Gene duplication and speciation)	-12671.545	$\omega_{\text{Gene duplication}} = 0.153$ $\omega_{\text{Speciation}} = 0.111$	-	M0	1	6.616	**1.01E-02**
	M1a (nearly neutral)	-12296.109	$p_0 = 0.830$ $p_1 = 0.170$ $\omega_0 = 0.052$ $\omega_1 = 1.000$	Not allowed	-	-	-	-
	M2a (positive selection)	-12296.109	$p_0 = 0.830$ $p_1 = 0.170$ $p_2 = 0.000$ $\omega_0 = 0.052$ $\omega_1 = 1.000$ $\omega_2 = 21.382$	No selected sites	M1a	2	0.000	1
	M8a (beta&ωs =1)	-12171.670	$p_0 = 0.953$ $p = 0.262$ $q = 2.109$ $p_1 = 0.047$ $\omega = 1.000$	Not allowed	-	-	-	-
	M8 (beta&ω)	-12171.670	$p_0 = 0.953$ $p = 0.262$ $q = 2.109$ $p_1 = 0.047$ $\omega = 1.000$	No selected sites	M8a	1	0.000	1
RH1	M0 (one ratio)	-18286.241	$\omega = 0.074$	-	-	-	-	-
	Free Ratio	-18113.481		-	M0	101	345.521	**0**
	M2 (Gene duplication and speciation)	-18285.586	$\omega_{\text{Gene duplication}} = 0.066$ $\omega_{\text{Speciation}} = 0.075$	-	M0	1	1.311	0.252
	M1a (nearly neutral)	-18079.609	$p_0 = 0.922$ $p_1 = 0.078$ $\omega_0 = 0.062$ $\omega_1 = 1.000$	Not allowed	-	-	-	-
	M2a (positive selection)	-18079.609	$p_0 = 0.922$ $p_1 = 0.078$ $p_2 = 0.000$ $\omega_0 = 0.062$ $\omega_1 = 1.000$ $\omega_2 = 60.184$	No selected sites	M1a	2	0.000	1
	M8a (beta&ωs =1)	-17610.262	$p_0 = 0.996$ $p = 0.349$ $q = 3.499$ $p_1 = 0.004$ $\omega = 1.000$	Not allowed	-	-	-	-
	M8 (beta&ω)	-17610.262	$p_0 = 0.996$ $p = 0.349$ $q = 3.499$ $p_1 = 0.004$ $\omega = 1.000$	No selected sites	M8a	1	0.000	1
RH2	M0 (one ratio)	-19543.705	$\omega = 0.131$	-	-	-	-	-
	Free Ratio	-19257.496		-	M0	137	572.417	**0**
	M2 (A, B and Background)	-19536.569	$\omega_A = 0.126$ $\omega_B = 0.173$ $\omega_{\text{Background}} = 0.121$	-	M0	2	14.272	**7.96E-04**
	M2 (A and Background)	-19546.090	$\omega_A = 0.187$ $\omega_{\text{Background}} = 0.133$	-	M0	1	0	1
	M2 (B and Background)	-19536.716	$\omega_B = 0.173$ $\omega_{\text{Background}} = 0.123$	-	M0	1	13.978	**1.85E-04**
	M2 (Gene duplication and speciation)	-19518.352	$\omega_{\text{Gene duplication}} = 0.211$ $\omega_{\text{Speciation}} = 0.114$	-	M0	1	50.705	**1.07E-12**
	M1a (nearly neutral)	-19053.243	$p_0 = 0.819$ $p_1 = 0.181$ $\omega_0 = 0.074$ $\omega_1 = 1.000$	Not allowed	-	-	-	-
	M2a (positive selection)	-19053.243	$p_0 = 0.819$ $p_1 = 0.181$ $p_2 = 0.000$ $\omega_0 = 0.074$ $\omega_1 = 1.000$ $\omega_2 = 29.492$	No selected sites	M1a	2	0.000	1
	M8a (beta&ωs =1)	-18688.089	$p_0 = 0.962$ $p = 0.328$ $q = 2.488$ $p_1 = 0.038$ $\omega = 1.000$	Not allowed	-	-	-	-
	M8 (beta&ω)	-18688.089	$p_0 = 0.962$ $p = 0.328$ $q = 2.488$ $p_1 = 0.038$ $\omega = 1.000$	19L	M8a	1	0.000	1

I tested the hypothesis that selection changes following duplication in the SWS2 and RH2 subfamilies. I found insignificant ($p = 0.066$) likelihood gains when branches in the SWS2A clade were allowed to have one average ω ($\omega = 0.187$) and the other branches a different ω ($\omega = 0.159$). For the RH2 subfamily, I found significant ($p = 1.07E-12$) likelihood gains when RH2B branches and all other branches were allowed to have independent ω's ($\omega = 0.173$ and $\omega = 0.123$ respectively) (Table 2.1, Table 2.2). Thus, I detected a slight increase in the ω following gene duplication, although in each case only for one duplicate. For both RH2A and SWS2B, an independent ω did not result in a better fit to the data ($p = n/s$)

I then examined whether genes diverged faster post duplication when compared to post speciation by allowing gene duplication events to have an independent ω. This produced significant likelihood gains in the SWS1 ($p = 1.49E-4$), LWS ($p = 9.61E-3$) and RH2 ($p = 1.07E-12$) subfamilies. In each case, ω was higher for the gene duplication branches (Figure 2.13)(Table 2.1, Table 2.2).

Despite recent criticism of the sites approach, I also attempted to identify codons under positive selection by using model M1a vs. M2a and M8 vs. M8a. For the SWS1, RH1, RH2 and LWS subfamilies there was no significant difference in likelihood when a separate class for sites with $\omega>1$ was included in the analysis, indicating no evidence for individual sites being under positive selection. For the SWS2 subfamily and the M1a vs. M2a test, allowing positive selection produced a non-significant gain in likelihood ($p=0.1617$) and indicated no sites under selection (Cut off posterior probability of 95%). For the M8 vs. M8a test, there was a significant gain in likelihood when positively selected codons were allowed ($p = 1.11E-3$) (Table 2.2). None of the so-called key sites, important for spectral tuning, were considered to be under positive selection.

While M1a vs. M2a and M8 vs. M8a test whether any site is under positive selection across an entire phylogeny, I also looked for positive selection on specific branches. I used this to test the hypothesis that following gene duplication, opsin genes are under positive selection for spectral divergence. A total of 65 post duplication branches were tested individually and eleven (17%) were found to have significant evidence for a proportion of codons evolving under $\omega>1$ (Table 2.3). Of those, two indicated that a key

spectral tuning site was under positive selection. Selected branches are indicated in Appendix 3-7.

Table 2.3: Significant results for all PAML analysis using branch-site models.

Likelihood values, parameter estimates and LRT results for all branch-site PAML analyses where the null hypothesis could be rejected. Branches numbers correlate with duplication event numbering while 'a' and 'b' arbitrarily indicate upper and lower branches identified in Appendix 3-7. Positively selected sites are numbered using bovine rhodopsin as a reference (Nathans & Hogness, 1983). All listed sites have a posterior probability > 0.95, while >.99 is indicated with *. Bolded sites are those identified as being key spectral tuning sites (Yokoyama, 2008). The p threshold was corrected for multiple testing using Bonferroni's correction (Miller, 1981; Anisimova & Yang, 2007).

Gene	Model	lnL	Parameter estimates	Positively selected sites	2Δ lnL	*p* value	Corrected *p* threshold
SWS1	Ia	-11925.32	p_0 = 0.852 p_1 = 0.113 ω_0 = 0.088 ω_2 = 9.116	**46F***, **86F*,** 92F	6.11	1.34E-02	2.50E-02
	Ia Null	-11928.37	p_0 = 0.792 p_1 = 0.106 ω_0 = 0.087 ω_2 = 1.000	N/A			
LWS	Vb	-12279.82	p_0 = 0.810 p_1 = 0.165 ω_0 = 0.050 ω_2 = 49.441	238K*, 240S*	13.45	2.45E-04	2.78E-03
	Vb Null	-12286.54	p_0 = 0.680 p_1 = 0.139 ω_0 = 0.050 ω_2 = 1.000	N/A			
	VIIa	-12290.71	p_0 = 0.830 p_1 = 0.166 ω_0 = 0.052 ω_2 = 999.000	None	10.80	1.02E-03	
	VIIa Null	-12296.11	p_0 = 0.830 p_1 = 0.170 ω_0 = 0.052 ω_2 = 1.000	N/A			
RH1	Ia	-18050.84	p_0 = 0.890 p_1 = 0.077 ω_0 = 0.061 ω_2 = 998.999	16T*, 36A*, 136W*, 248R, 308C*, 344S*	33.13	8.63E-09	4.17E-03
	Ia Null	-18067.40	p_0 = 0.828 p_1 = 0.070 ω_0 = 0.061 ω_2 = 1.000	N/A			
	IIa	-18070.56	p_0 = 0.908 p_1 = 0.079 ω_0 = 0.061 ω_2 = 999.000	**195A***, 197G	12.02	5.25E-04	
	IIa Null	-18076.58	p_0 = 0.830 p_1 = 0.070 ω_0 = 0.061 ω_2 = 1.000	N/A			
RH2	Ib	-19044.36	p_0 = 0.806 p_1 = 0.180 ω_0 = 0.074 ω_2 = 999.000	15S*, 124A	11.68	6.33E-04	1.72E-03
	Ib Null	-19050.20	p_0 = 0.575 p_1 = 0.128 ω_0 = 0.074 ω_2 = 1.000	N/A			
	VIIa	-19025.45	p_0 = 0.794 p_1 = 0.165 ω_0 = 0.074 ω_2 = 59.791	25E*, 33D, 105L*, 163M*, 220F*, 281A	34.75	3.76E-09	
	VIIa Null	-19042.83	p_0 = 0.729 p_1 = 0.156 ω_0 = 0.073 ω_2 = 1.000	N/A			
	VIIb	-19042.43	p_0 = 0.803 p_1 = 0.179 ω_0 = 0.073 ω_2 = 38.545	157V*, 300L, 326G*	11.77	6.02E-04	
	VIIb Null	-19048.32	p_0 = 0.746 p_1 = 0.165 ω_0 = 0.073 ω_2 = 1.000	N/A			
	VIIIa	-19023.45	p_0 = 0.780 p_1 = 0.160 ω_0 = 0.074 ω_2 = 88.309	25E*, 43Y*, 64Q*, 119L, 320T*	30.02	4.28E-08	
	VIIIa Null	-19038.46	p_0 = 0.655 p_1 = 0.135 ω_0 = 0.073 ω_2 = 1.000	N/A			
	XVa	-19030.95	p_0 = 0.798 p_1 = 0.171 ω_0 = 0.074 ω_2 = 242.968	34P*, 37Y, 38K, 40L*, 216V	31.15	2.39E-08	
	XVa Null	-19046.52	p_0 = 0.608 p_1 = 0.132 ω_0 = 0.074 ω_2 = 1.000	N/A			
	XXa	-19041.94	p_0 = 0.816 p_1 = 0.175 ω_0 = 0.075 ω_2 = 999.000	345A	15.88	6.75E-05	
	XXa Null	-19049.88	p_0 = 0.527 p_1 = 0.112 ω_0 = 0.075 ω_2 = 1.000	N/A			

DISCUSSION

The all opsin tree (Figure 2.1) divided all sequences into subfamilies. This tree presents the LWS subfamily as sister to the non-visual opsins, pinopsin and vertebrate ancient opsin. As a result, the visual opsins are a paraphyletic group. This has been seen before, but further study is needed to see exactly where the non-visual opsins group relative to the visual opsins (Terakita, 2005; Davies et al., 2010; Plachetzki et al., 2010).

In two cases, LWS and RH1, sarcopterygian sequences did not form a monophyletic group. This can be partially explained by the fact that the model of evolution used was based only on fish, although in the case of the LWS subfamily, avian sequences are particularly divergent. This may be the result of an ancient LWS duplication retained only in birds, although genomic analysis indicates that bird LWS opsin genes are linked with the SWS2 opsin, as is found in other vertebrates. This means that if there was an ancient duplication, it was tandem.

SWS1 and SWS2 duplication events

Only one duplication occurred in the SWS1 subfamily in fish. This duplication in ayu is the only known SWS1 duplication. Although only SWS1-2 is expressed in the eye of ayu smelt, both are under purifying selection indicating functional significance (Minamoto & Shimizu, 2005).

The SWS2 tree is dominated by one duplication, SWS2-I. This ancient duplication occurred early in Neoteleostei evolution, similarly to RH2-I, and paralogs produced from it have the potential to be found in more than 10,000 species of fish (Nelson, 2006). A second duplication is found in the carp (clade Cyprinini-*Sinocyclocheilus*). This clade also shares duplications in the LWS and RH2 subfamilies. It has been suggested that the LWS and SWS2 duplications are the result of tetraploidy that is shared by this group (Li et al., 2009; Yang et al., 2010). This is supported by the fact that both LWS and SWS2 duplicates have a similar 3rd position mean group distance, 0.105 and 0.092 respectively. The RH2 duplication, on the other hand, has a larger 3rd position distance, 0.173, indicating that it is likely an older duplication event. The LWS and SWS2 genes are closely linked and, therefore, simultaneous duplication requires only the duplication of a small genomic region (for example via segment duplication, transposons or a single

chromosome duplication (Bailey et al., 2002)) and may not necessarily be the product of tetraploidy.

RH1 duplication events

The oldest opsin duplicate found is that of ExoRh/RH1 (RH1-I). This event was a retroduplication of the ancestral RH1 gene still found in tetrapods. It produced an intronless duplicate that retained the RH1 name, while the ancestral copy gained the new name exo-rhodopsin (ExoRh). This name switch is due to the fact that intronless RH1 in fish is expressed in the rod cells of the retina as is the intron containing RH1 gene found in other vertebrate groups, while the intron containing ExoRh gene in fish is expressed in the pineal (Mano et al., 1999). Through the examination of the presence or absence of introns, this duplication is thought to have occurred just after bichir (genus *Polypterus*) diverged from the fish lineage near the base of Actinopterygii (Venkatesh et al., 1999). My phylogeny does not disagree with this hypothesis, but a lack of bichir RH1 sequence does not allow it to be confirmed.

Two RH1 duplicates present considerable difficulty in assigning duplication times: *Danio rerio* RH1-1/RH1-2 and *Scopelarchus analis* RH1A/RH1B. In both cases, one paralog groups with RH1 sequences from species it is closely related to, while the other paralog is located early in the tree in an unexpected position with a long branch. It is possible that both pairs are a product of the same duplication event, as the maximum parsimony method groups *D. rerio* RH1-2 and *S. analis* RH1B, but this relationship is likely a case of long branch attraction (Bergsten, 2005). Overall both paralogs pairs are highly divergent with Tamura-Nei distance values of ~0.3. Analysis of divergence at different positions suggests both pairs are under strong purifying selection as the divergence at 3rd positions is much higher than 1st and 2nd positions. Therefore, it is reasonable to conclude that both are from duplications early in teleost evolution although exact timing cannot be inferred until copies of these genes are uncovered in additional species.

The phylogenetic position of the deep-sea scabbardfish, *Lepidopus fitchi* presents problem in all trees in which it is included (RH1, RH2 and SWS1). The genus *Lepidopus* is in the order Perciformes and suborder Scombroidei but all *Lepidopus* opsin sequences fall between the salmonids and smelt (Salmoniformes and Osmeriformes) and the

cyprinids (Cypriniformes). Also within the same suborder as scabbardfish is the Pacific bluefin tuna, *Thunnus orientalis*, which groups predictably with other Perciformes sequences and not with *Lepidopus fitchi*. Although a taxonomically incorrect branch position may be the result of accelerated or divergent evolution, the consistent positioning of three independent loci makes this explanation improbable. Indeed, in Yokoyama et al.'s (2008) phylogenetic analysis, the scabbard fish RH1 sequences were sister to those of Stomiiformes, an order of deep-sea fish sister to smelt (Osmeriformes), to the exclusion of Perciformes sequences. When Stomiiformes RH1 sequences are added to the current phylogeny, they group with *Lepidopus fitchi* RH1 (Data not shown). This raises questions about either the phylogenetic position of *Lepidopus fitchi*, or the species identity of the fish from which DNA was isolated. Phylogenetic analysis using complete mitochondrial sequence have placed species in the genus *Lepidopus* together with *Thunnus orientalis* (Jondeung & Karinthanyakit, 2010). No other molecular studies have examined *Lepidopus fitchi.*

In a similar vein, the catfish (*Ictalurus punctatus*) RH1 gene is also in an inappropriate phylogenetic location. *I. punctatus* is within the order Siluriformes, sister order to Cypriniformes (zebrafish and carp). The fact that the catfish RH1 gene instead groups strongly with bream sequences (Genus: *Acanthopagrus*) suggests that the sequence is not actually from a catfish and is likely from a species of bream.

RH2 duplication events

The oldest duplication in the RH2 subfamily (RH2-I) produced RH2A/RH2B. This tandem duplication occurred early in Neoteleostei evolution. Although a majority of sequences fall within one of the two clades, both *Pagrus major* and *Thunnus orientalis* RH2B are found as outgroups to the RH2A clade. This would indicate either separate early duplication events within this clade lost by all others or partial gene conversion between paralogs confusing the phylogenetic signal.

Within the RH2A clade, five further gene duplication events have occurred in five different lineages, while in the RH2B clade, three further gene duplications have occurred in one lineage. This suggests that duplication of RH2A is more common than RH2B. This may be due to the higher retention of RH2A. In three lineages that duplicated RH2A, RH2B is not present.

Two deep sea fish, *Lepidopus fitchi* and *Stenobrachius leucopsarus* have four RH2 genes each. These genes are a product of three independent gene duplication events each shared by no other species. Also in each case, one duplicate has been pseudogenized (Yokoyama & Tada, 2010).

LWS duplication events

The first duplication on the LWS tree is LWS-I. This group of LWS genes, dubbed 'Green', falls out at the very base of teleost LWS genes tree, but is clearly more closely related to fish LWS opsins than they are to sarcopterygian LWS opsins or to opsin genes from any of the other subfamilies (Figure 2.9). This would indicate that LWS-I occurred at the base of teleostei and that orthologs were either lost in, or never sequenced in, all other species. It has been shown that members of this five-gene clade are expressed in skin tissue (Kasai & Oshima, 2006). It is possible that the position of these green LWS opsins is a consequence of selection for this unusual expression pattern that occurs only in the characid family (includes neon tetras, cavefish and the pencilfish). Although selection may bias substitution patterns in first and second nucleotide positions, third positions should evolve largely neutrally because a majority of changes in third positions do not alter the amino acid that the codon encodes. The large third position distance between duplicates of LWS-I indicate that this is a relatively old duplication event (3^{rd} position Tamura Nei distance = 0.83).

When examining amino acid sequence, the LWS 'green' clade displays unusual divergence at the TM6 and E3 domains (Figure 2.10). While in most LWS opsins these positions are highly conserved across broad evolutionary timescales, in the green opsins, though similar to each other, these have diverged from other opsins at up to 50% of their amino acid positions. TM6 plays a key role in G-protein binding as well as providing an opening for retinal to enter the binding pocket (Park et al., 2008; Scheerer et al., 2008). This hints that these green opsins may have evolved to interact with a different G-protein to coincide with their unusual expression pattern.

The order Cyprinidontiformes presents a complicated picture of LWS gene evolution. At face value, the duplication event that produced LWS P180 in the Poeciliids (*P. reticulata*, and genus *Xiphophorus*) is confined to the family Poeciliidae and the *J. onca* LWS P180 was produced by an independent event. However, Windsor and Owens (2009)

showed evidence that supports the hypothesis that the LWS P180 duplication event occurred before Poeciliidae and Anablepidae diverged and that homology has been obscured by gene conversion. The duplicon, proto-P180, was partially overwritten via gene conversion in *A. anableps* by its paralog S180a and in *J. onca* by its paralog S180. By examining the 3' region of the gene, which was not gene converted, I can recapitulate the original phylogenetic relationship placing *Xiphophorus* and *Poecilia* P180 together with *Jenynsia* P180 and *Anableps* S180γ.

Gene conversion has also been seen in the guppy, *Poecilia reticulata*. Ward et al. (2008) reported a hybrid opsin (a fifth sequence not shown here) from guppies collected in Cumana, Venezuela. This hybrid locus, combining the A180 gene with the 3' end of the P180 gene, does not appear to be expressed and may be population-specific as it has not been uncovered in surveys of other guppies (Hoffmann et al., 2007; Weadick & Chang, 2007). Gene conversion offers a mechanism for novelty generation in opsins through the recombination of distinct tandem loci to produce intermediate hybrids. Indeed, in *A. anableps* the gene conversion of LWS S180γ produced a gene with a unique keysite haplotype, suggesting a novel spectral absorption. In humans, gene conversion among opsin loci can produce hybrid genes. This conversion is often detrimental to colour vision as an intermediate pigment reduces the spectral differentiation between the red and green LWS genes but it is possible that an intermediate pigment may be advantageous under different circumstances (Nathans et al., 1986a).

Another interesting LWS duplication in Cyprinidontiformes produced LWS S180r (LWS-II). This was a retrotransposition event; it is missing four of its five introns and is located within intron XI of the unrelated gene *gephryn* (Watson et al., 2010). Typically successful retrogenes either travel with basic regulatory elements or use those of neighbouring genes to regulate their expression pattern. The expression pattern of the S180r genes is contentious in the literature. S180r was amplified from cDNA from *P. reticulata* and *L. goodei*, but no sequence spanned the first intron leaving open the possibility of genomic contamination. Indeed, the *L. goodei* LWS2 sequence reported in NCBI as being amplified from cDNA actually extends into the first intron indicating either an unspliced transcript or genomic contamination. Regardless of an apparent lack

of expression, S180r has retained its reading frame for over 72 million years indicating that the retention of its protein product provides a selective advantage (Spady, 2006).

In two cases in the LWS phylogeny, there is uncertainty whether two sequences represent independent loci or allelic variation. For *Jordanella floridae*, two sequences were obtained from one individual and for *Plecoglossus altivelis* two sequences were obtained in separate studies with separate populations. Each pair has a coding sequence pair wise distance of 0.02, which, while low, is above the divergence of confirmed independent loci in *Oryzias latipes* (pair-wise distance of 0.012), *Poecilia reticulata* (0.013) and *Xiphophorus helleri* (0.001). Gene duplication and gene conversion can create loci with identical sequences, so judging whether sequences represent independent loci from coding sequence alone is difficult.

General features of opsin duplication

By comparing the mean distance between duplicates, I can get an estimate of how old the gene duplication event is. Using this data, I found that the most common duplications are younger (Figure 2.4). The number of duplications decrease as age increases until it levels off past 0.3 third position Tamura-Nei distance. This indicates that young duplicates are prone to loss but, after a period, subsequent loss is unlikely**.** This may be a sampling issue as I am merely looked at gene duplication events, not retention. As a duplicate ages, its lineage will likely speciate and so the complete removal of evidence for the duplication event will require multiple gene loss or extinction events.

By looking at the positions of opsins on chromosomes or in large-insert clone sequences, I found that a majority of duplications are tandem. Frequent expansion of gene copy number is common in tandem arrays of odorant receptors through unequal crossing over and this is likely to also play a similar role in visual opsins (Levinson & Gutman, 1987; Young & Trask, 2002). Furthermore, the differences in gene orientation and relative location indicate that cryptic duplication or rearrangement events have occurred in the past.

Features of boundary and middle wave sensitive opsins

Gojobori and Innan (2009) hypothesized that duplicates of boundary wave opsins, that is those opsins that encode proteins sensitive to wavelengths at either end of the visible

spectrum (LWS and SWS1), are retained less often, are under greater purifying selection and are younger than middle wave opsins (RH2 and SWS2). Here I have tested each of these hypotheses and found some to be lacking support.

Measuring the average number of genes per subfamily is not informative as this number would be affected by sampling bias based on species, as well as, age of duplication event. Studies of opsin repertoires are not randomly distributed across fish phylogeny and tend to be focused on a few groups, including cyprinids, cichlids and livebearers. Furthermore, some duplications may be over represented because they occurred in the ancestor of a particularly specious clade. For example, although SWS2 has only two duplication events, its average number of genes is 1.3 while LWS, which has twelve duplication events has an average of 1.5 genes. It is more fruitful to look at the number of gene duplication events. Although it is likely that some of the recent duplication events will not be retained in the future, this bias should be equally represented in all subfamilies.

Half of all opsin duplications in fish are from the RH2 subfamily, a middle wave opsin, and of the remaining duplication events, a majority are from the LWS subfamily. The SWS2 subfamily, the other middle wave opsin, only has two duplication events. In fact, the trend, if taken at face value, suggests that longer wavelength opsins duplicates are retained more often.

To examine the age of opsin duplication events, I looked at divergence at third positions between duplicates to prevent selection from affecting these estimates. From this I asked the question, is the mean age of duplications different between subfamilies or between border and middle wave opsins? To test the first question, I completed a Kruskal-Wallis test, which found no significant difference between any pair of opsin subfamilies ($p = 0.1564$). To test the second question, duplication events for border and middle wave opsins were individually grouped and a Mann Whitney test was completed. This showed no significant difference between the two groups ($p = 0.4292$). These tests do not prove that no differences exist, but show that the current data does not support a difference. Indeed, with only one duplication event in the SWS1 subfamily and two in the SWS2 subfamily, it is impossible to draw any broad conclusions about duplications in those subfamilies.

When I examined the ω ratio for each opsin subfamily, I found that all subfamilies are under purifying selection (Table 2.1). Ranking subfamilies by average ω shows RH1 < SWS1 ~ LWS < RH2 < SWS2. This agrees with the hypothesis that border opsins are under greater purifying selection, but differences in ω are not large.

Duplication and selection

The average ω for each subfamily vary from 0.074 (RH1) to 0.167 (SWS2) (Table 2.1). These values are consistent with values found in other studies. Yang and Neilsen (1998) examined 48 nuclear proteins in rodents, primates and artiodactyls and found that average ω varied from 0.017 in ATP synthase β to as high as 0.836 in interleukin 6. A study of RH1 genes in mammals found that average ω is 0.040, indicating slightly higher purifying selection than was found in this study (Zhao et al., 2009).

I also examined two large duplication clades from the duplication events RH2-I and SWS2-I. In each case, I asked if after duplication one duplicate had undergone a significant change in selection pressure, as measured by ω, compared to the rest of the tree. I found that RH2B had significantly reduced purifying selection when compared to the rest of the branches (ω = 0.173 vs. 0.123). RH2B is the paralog which is more often lost and tends to be short wave shifted (λ_{max} = 452-488nm) compared to RH2A (λ_{max} = 492-555nm). The nearest out-group to this duplication is Salmonidae. Salmonid RH2 genes have λ_{max} values from 494nm to 534nm (Cheng et al., 2006). Assuming this is the ancestral phenotype, the reduction in purifying selection is associated with RH2B gaining a new role (sensitivity to shorter wavelengths).

Looking at all opsin duplications, I found that for the SWS1, RH2, and LWS subfamilies there is reduced purifying selection following gene duplication. This trend has been seen in a variety of other genes and organisms (Kondrashov et al., 2002). In opsins, this relaxed selection may result in changes to key spectral tuning sites, allowing duplicates to exhibit different spectral sensitivities. In all cases where an opsin gene duplication has diverged enough to be differentiated in expression analysis, paralogs were found to be expressed at different levels, in different cell types or at different times. This indicates that the relaxation of purifying amino acid selection is often accompanied by changes in expression regulation between duplicates.

Positive selection

Although reduced purifying selection can occur simply due to a loss of gene importance, positive selection indicates that there was an advantage in amino acid changes. In order to identify positive selection in opsin genes, I tested the data using site and branch-site models. The site model allowed positive selection ($\omega>1$) in certain codons for all branches. This was only found to cause significant likelihood gains and identified sites under positive selection in the SWS2 subfamily. Of the three sites selected, all are in the extracellular tail. None of the selected sites are "key sites", nor do they represent residues in close contact with the chromophore. Thus, it is unlikely that the codons found here are involved in selection for spectral sensitivity. The fact that more sites were not found to be under positive selection may be due to the large phylogenetic distances between sequences tested. In most cases, positive selection does not occur in all branches of a phylogeny (Bielawski & Yang, 2001; Raes & Van de Peer, 2003).

The alternative branch-site model searches for codons under positive selection on one specific branch (Zhang et al., 2005). This allows transient selection to be identified. It is important to note that recent criticism of the branch-site test for positive selection have arisen in the literature. Hughes & Friedman (2008) suggested that the positively selected sites identified by branch sites models may be instances of reduced synonymous substitutions and not, in fact, increased nonsynonymous substitutions. Nozuwa et al. (2009) used simulated data to test the false positive rate for branch-site models and found it to be 0.23%.

Nozawa et al. (2009) also examined all vertebrate visual opsin gene subfamilies to find sites under positive selection in all branches using the branch-site method. They then compared which sites were selected with key sites established through mutagenesis experiments. He found that very few sites were under positive selection in all lineages and that the average ω for key sites was <1, and therefore not indicating positive selection.

I disagree with Nozawa et al.'s conclusions for several reasons. First is the assumption that key sites are the only adaptive sites. Key sites are those that affect λ_{max}. Although λ_{max} is an easily studied functional aspect of opsins, it is not the only feature. Absolute sensitivity, interactions with G-proteins and other proteins may have selective

implications independent of λ_{max} and be determined by amino acids not indicated as key sites. Furthermore, keys sites are dependent on the background amino acid composition and identical substitutions produce variable effects in different sequences. This means that a key site substitution determined in one species may have a very different, or no, effect in a different organism. While the F46L mutation is expected to cause a -1 nm shift in the ancestral SWS2 pigment, the reverse mutation, L46F causes a -6 nm shift in bluefin killifish SWS2A, the complete reverse of what is expected (Yokoyama & Tada, 2003; Yokoyama et al., 2007). A key site substitution causing a λ_{max} shift in birds may not cause any shift in fish and is therefore unlikely to be under positive selection across the entire phylogeny. Many key sites only act in groups and the mutation of one would have no affect on λ_{max}. Lastly, although Nozawa et al. showed that key sites on average were not under positive selection, the ω values shown (from 0.31 to 0.67) are all substantially higher than average ω values I obtained for all codons (from 0.074 to 0.167). In fact, codon specific models found that 79-92% of codons had $\omega < 0.1$. This indicates that relative to most sites, key sites are under relaxed selection. Potentially, this relaxed selection may be instances of positive selection on certain branches followed by negative selection.

With these criticisms in mind, I specifically examined post-duplication branches and found eleven (17%) had codons evolving under positive selection. Five of the branches identified as being under positive selection had codons evolving under $\omega = 999$ which is the maximum that PAML can estimate. These are likely statistical errors stemming from a lack of synonymous substitutions. For the other six branches, ω estimates varied from 9.1 to 242.9 for a small subset of the codons (<5%). In five of these remaining branches, none of the sites identified as being under selection are known to affect spectral sensitivity. As mentioned above, this does not allow us to rule out adaptive advantages to these substitutions. In cases where these branches lead to a spectrally novel duplicate (particularly RH2-VII in salmonids and LWS-V in livebearers), these sites may provide targets for future mutagenesis experiments.

Of particular interest is the branch leading to ayu SWS1-1, identified as SWS1-Ia, which has both key sites identified as under positive selection and a feasible ω value. In this branch, two key sites are under positive selection which resulted in the mutations

F46T and F86A. In mammalian SWS1, F46T, in conjunction with several other mutations, causes a red shift from UV to violet absorption (Shi et al., 2001). While F86A has not been empirically tested, F86L has been found to also contribute to a red shift in mammalian SWS1 (Shi et al., 2001). Due to the close structural and chemical properties of alanine and leucine, the F86A substitution may also contribute to a red shift. Furthermore, the final codon under positive selection in the branch, codon 92, is adjacent to another key site in SWS1 opsins.

This evidence points to selection for a red shift in the SWS1-1 gene of *P. altivelis.* Although a Fisher's exact test for positive selection failed to identify positive selection on this branch, this is not unexpected from the estimated parameters. Fisher's exact test asks if all codons are under positive selection. While this branch has only 3.5% of its codons estimated to be under positive selection, 85% are evolving under a ω value of 0.09. This hypothesis needs to be confirmed by *in vitro* reconstitution experiments to measure the λ_{max} of both ayu SWS1 genes as well as intermediate proteins to see if the individual mutations purported to be the product of selection actually cause a phenotypic change in the protein. If confirmed, this would be the first instance of identified positive selection for a shift in λ_{max} seen in a vertebrate opsin.

CONCLUSIONS

Opsin genes in fish have been duplicated and retained relatively frequently. These duplications are disproportionately in longer wavelength subfamilies and occur via tandem duplication. Although tests for positive selection provided little evidence for adaptive evolution following gene duplication as a general feature in opsins, the SWS1-1 opsin in ayu does appear to have undergone positive selection for a change in spectral sensitivity. The general failure to identify adaptive evolution may be due to a lack of in-depth understanding of the molecular basis of spectral tuning in fish opsins. Alternatively, changes in opsin spectral sensitivity may not have a strong enough phenotypic change to induce detectable positive selection.

Although evidence for selection on spectral differentiation is elusive, opsin gene duplicates are often diverse in both spectral absorption and expression. In *A. anableps*, a large repertoire of spectrally diverse opsins are paired with a large retina with spectrally

diverse light sources. In Chapter 3, I examine the result of gene duplication in a morphologically divided retina.

Chapter 3 – Differential opsin gene expression in the four-eyed fish, *Anableps anableps* and its two-eyed sister, *Jenynsia onca*.

ABSTRACT

The four-eyed fish, *Anableps anableps*, has remarkable eyes that provide simultaneous vision above and below water. Among the many morphological adaptations for above and below water vision, is a divided retina. A close relative, the one sided livebearer (*Jenynsia onca*) has normal eyes for a fish. UV (SWS1), violet (SWS2B) and blue (RH2B) opsins are expressed throughout the retinas in both species. In *J. onca*, the green pigment (RH2A) is also expressed across the retina, whereas in *A. anableps* RH2A transcripts occur only in the ventral retina, which receives aerial light. In *J. onca*, the yellow opsin (LWS) is expressed in only medial or dorsal cone cells. In *A. anableps,* the dorsal retina, responsible for aquatic vision, similarly expressed LWS opsins but in this case the boundary between the region with LWS opsin transcripts and without was clearly delimited by the transition from dorsal to ventral retina. I hypothesize that *A. anableps* achieved this retinal specialization through the reduction of RH2A expression. Furthermore, this relative increase in long wavelength sensitivity for the aquatic field of view is correlated with the wavelength of maximal light transmittance in muddy water.

INTRODUCTION

The light environment presents a dizzying array of wavelengths, which combine to make the innumerable hues and colours we perceive. Although air allows for the relatively unobstructed passage of the entire visual spectrum, water does not (Lythgoe & Partridge, 1989). Consequently, animals living in water tune their visual sensitivity to wavelengths of importance. This is most strikingly found in the mesopelagic zone of the ocean (200 to 1000 m deep), which is nearly monochromatic due to the filtering properties of the water (Jerlov, 1968). Fish found here have tuned their rod photoreceptors to the only available wavelength of light through amino acid changes to the opsin proteins found in their photoreceptors (Crescitelli, 1990). This is also seen in the cottoid fish of Lake Baikal, where species living at depths greater than 100 m have blue shifted their overall visual sensitivity by both differential opsin usage and sequence

changes (Bowmaker et al., 1994). The variation observed among species in different environments is also sometimes observed within species. For example, eels have two rod pigments and switch which paralog is expressed during their migration from freshwater to the deep sea environment. The deep sea pigment is tuned to the wavelength of light most abundant in the deep sea, giving optimal photon capture (Wood & Partridge, 1993; Hope et al., 1998). Fish have also been found to alter their opsin expression when under altered light environments. In both bream and killifish, when short wavelength light is reduced, the proportion of long wavelength opsin expressed increases (Fuller et al., 2005; Shand et al., 2008).

The spectral environment can also vary within a single visual field, presenting an individual with two distinct spectral environments simultaneously. Indeed, intraretinal pigment variation has been shown in mammals, amphibians, birds and fish (Applebury et al., 2000; Reuter et al., 1971; Bruhn & Cepko, 1996; Takechi & Kawamura, 2005). In cichlids, blue sensitive cones are more abundant in the ventral retina, while in bullfrogs, visual pigments in the dorsal retina are red shifted when compared to the ventral retina (Reuter et al., 1971; Levine et al., 1979). This division of spectral environment is taken to the extreme in the four-eyed fish, *Anableps anableps*.

A. anableps has large frog-like eyes extending above its head allowing it to hold half of its eye above the waterline while keeping the rest of the body, and the other half of its eye, underwater. *A. anableps* always swims at the waterline, and is thus consistently exposing the ventral half of the retina to unfiltered aerial light. The dorsal retina receives only reflected light filtered through water and dissolved solutes that may drastically change the spectral content. Although all three species in the genus *Anableps* have the characteristic eye morphology, its sister genus *Jenynsia* has retained the ancestral normal eye morphology (Nelson, 2006). Recently both species had their visual opsin repertoire sequenced, revealing high conservation despite the morphological differences (Windsor & Owens, 2009; Owens et al., 2009).

Visual opsins are a monophyletic group of seven transmembrane domain G-protein coupled receptors that bind one of two isoforms of the light-sensitive, vitamin A-derived chromophore, retinal. Opsins are expressed in photoreceptor cells of the eye and are responsible for the first step in light perception. There are five subfamilies of visual

opsins; the two classes of short wave-sensitive opsins (SWS1 and SWS2), middle wavelength-sensitive rhodopsin and rhodopsin-like opsins (RH1 and RH2), and the long wave sensitive opsins (LWS) (Yokoyama, 1994). These subfamilies span the visual spectrum from UV to red, while the maximal absorption (λ_{max}) of each individual opsin is determined by its amino acid sequence (Yokoyama, 2008).

Most tetrapods (e.g., lobe-finned fish, amphibians and reptiles) have one gene from each subfamily. Mammals have lost their SWS2 and RH2 genes, a trait believed to be an adaptation to a nocturnal phase of mammalian evolution (Bowmaker, 2008). Many, ray-finned fish have gained additional opsins through gene duplication (see Chapter 2) and these gene duplicates have diverged in sequence to produce spectrally distinct proteins often with distinct expression domains. In neoteleostei, the SWS2 and RH2 opsins were tandemly duplicated approximately ~200 and ~230 mya respectively (Spady, 2006). These duplicates have subsequently spectrally diverged as much as 37 and 51 nm respectively in cichlids and are differentially expressed in many species (Parry et al., 2005). Quantitative PCR (or qPCR) has shown that in rift lake cichlids, the longer wavelength RH2A is almost universally expressed, while its shorter wavelength duplicate RH2B varies from zero to equalling its paralog's expression. Most interestingly, cichlids living in the turbid waters of Lake Victoria express only minimal RH2B, corresponding with a general shift towards longer wavelength pigments when compared to species in the clearer waters of Lake Malawi (Hofmann et al., 2009). In bluefin killifish (*Lucania goodei*), while SWS2B represents 20-25% of total opsin expression, SWS2A represents only 0.1% (Fuller et al., 2004; Fuller et al., 2005). In Anablepidae, the LWS subfamily has expanded; *A. anableps* has four LWS genes, while *J. onca* has three (Owens et al., 2009; Windsor & Owens, 2009). These genes are the product of recent duplications and gene conversions. Both Anablepidae and its sister family Poeciliidae (home to guppies, swordtails and others) share a retroduplicated LWS gene, S180r. In the guppy, *Poecilia reticulata*, the four LWS genes vary expression by over 1000-fold difference between the most least abundant transcript (Ward et al., 2008).

While qPCR gives a broad look at gene expression averaged across the whole eye, a more informative technique for examining opsin gene expression is *in situ* hybridization. This technique allows for cellular localization and reveals expression domains within the

retina. Although several studies have linked cell morphology with opsin subfamily expression, few have examined the expression domains of within subfamily opsin duplicates (Minamoto & Shimizu, 2005; Miyazaki et al., 2008). In zebrafish (*Danio rerio*) there are four RH2 genes, all of which are differentially expressed; the longer wavelength duplicates are expressed more ventrally and peripherally and also appear earlier in development (Takechi & Kawamura, 2005).

To better understand how *A. anableps* has adapted opsin gene expression to cope with the diverse spectral environments and in concert with the evolution of its distinct eye morphlogy, I have used *in situ* hybridization to localize gene expression of all cone opsin genes. Although early studies detected no difference in visual pigments between retinal halves, recent molecular work has identified a greater number of visual opsins than previously appreciated, allowing for the possibility of fine scale expression differences (Avery & Bowmaker, 1982; Owens et al., 2009). I compared expression patterns found in *A. anableps* with its sister species, *J. onca*.

METHODS

Probe preparation

Coding regions were PCR amplified from *A. anableps* eye cDNA for SWS1 (792 bp, FJ711153.1), SWS2A (532 bp, FJ711152.1), SWS2B (353 bp, FJ711151.1), RH2A (524 bp, FJ711149.1), RH2B (688 bp, FJ711150.1) and LWS (548bp, FJ711157.1). For SWS2 and RH2 subfamilies, probes were designed from the most divergent regions to prevent cross-reactivity. The LWS probe is designed in a region nearly identical between LWS S180α, LWS 180β and LWS S180γ and is expected to bind all equally well, and to a lesser extend bind LWS S180r. Amplicons were cloned in the pGEM-T easy vector and transcribed using a digoxigenin (DIG) RNA labeling kit (Roche, Basel, Switzerland) after linearization. A sense transcript of RH2A was also transcribed as a negative control. Since *J. onca* sequences are ~96% identical to orthologous *A. anableps* probes, the same probes were used for both species. The single LWS probe was used to detect all LWS genes in *J. onca*. Probe identity between targets and close paralogs is shown in Appendix 8.

Probe specificity was tested using an RNA-RNA dot blot in which the probe was allowed to bind to an RNA copy of the target gene (from *A. anableps* and *J. onca*) as well

as its most-similar paralogs under stringency conditions mimicking the *in situ* hybridization. For this experiment, targets were transcribed using SP6 or T7 polymerase for 2 hours at 37°C and were then diluted 10^{-2} to 10^{-4}. Hybridization and high stringency washes occurred at 65-68°C using 0.5X - 0.2X SSC with Tween-20. For the LWS probe, specificity was not tested against S180r for either species.

Slide preparation

Anableps anableps and *Jenynsia onca* were collected in South America by Spencer Jack from Afishionados and were kept in fresh water under natural day-night light cycle. A total of four *A. anableps* (two adult male and two juvenile) and two *J. onca* (one male and one female) were sacrificed at midday for these experiments. Fish were euthanized in buffered MS-222. Eyes were removed and immediately fixed in 4% paraformaldehyde in phosphate-buffered saline (PBS) at 4°C overnight with pupils pierced to allow retinal infiltration. Eyes were then cryoprotected in increasing concentrations of sucrose in PBS, ultimately incubating in 20% sucrose in PBS overnight at 4°C. Eyes were infiltrated and then frozen in a 2:1 mixture of 20% sucrose in PBS and Tissue-Tek OCT compound, orientated either for dorsal-ventral or nasal-temporal sections. A Leica CM1850 UV cryostat was used to cut 9-μm sections onto SuperFrost Plus slides (Fisher, Pittsburgh, PA). Slides were stored at -80°C until use.

In situ hybridization

Slides were air dried and then hydrated in a series of ethanol/water solutions. Tissue was then treated with ~13 μg/ml Proteinase K in PBSTween preheated to 37°C for 4 minutes, acetylated and then dehydrated. After air-drying, excess DIG-labelled RNA probe in Haupmann's buffer was applied to the slide, sealed with a chamber well slide cover and incubated in a moist chamber at 65°C overnight. Post hybridization washes were performed using 2X SSC at room temperature and 50% formamide in 2X SSC and 0.2X SS at 65°C. Blocking was performed using blocking reagent in maleate buffer for 2-3 hours. Detection was done using anti-DIG-AP (Roche, Basel, Switzerland) overnight at 4°C followed by visualization using NBT/BCIP.

RESULTS

Probe Specificity

The SWS1 and SWS2A probes showed strong binding to their target and minimal binding to non-specific targets (Appendix 9-10). The SWS2B probe bound only to its target in *A. anableps*, while it weakly bound both SWS2A and SWS2B in *J. onca* (Appendix 11). The RH2A probe preferentially bound its target in both species, but also showed weak binding to *J. onca* RH2B and the anti-sense negative control (Appendix 12). The RH2B probe preferentially bound its target in both species, and also bound *J. onca* RH2A to a lesser degree (Appendix 13). The LWS probe preferentially bound *A. anableps* LWS S180γ, *J. onca* LWS S180 and P180 but not the negative control (Appendix 14).

Of six probes tested, five showed positive binding to photoreceptor cells. Only the SWS2A and negative control probes did not show any binding (Figure 3.1).

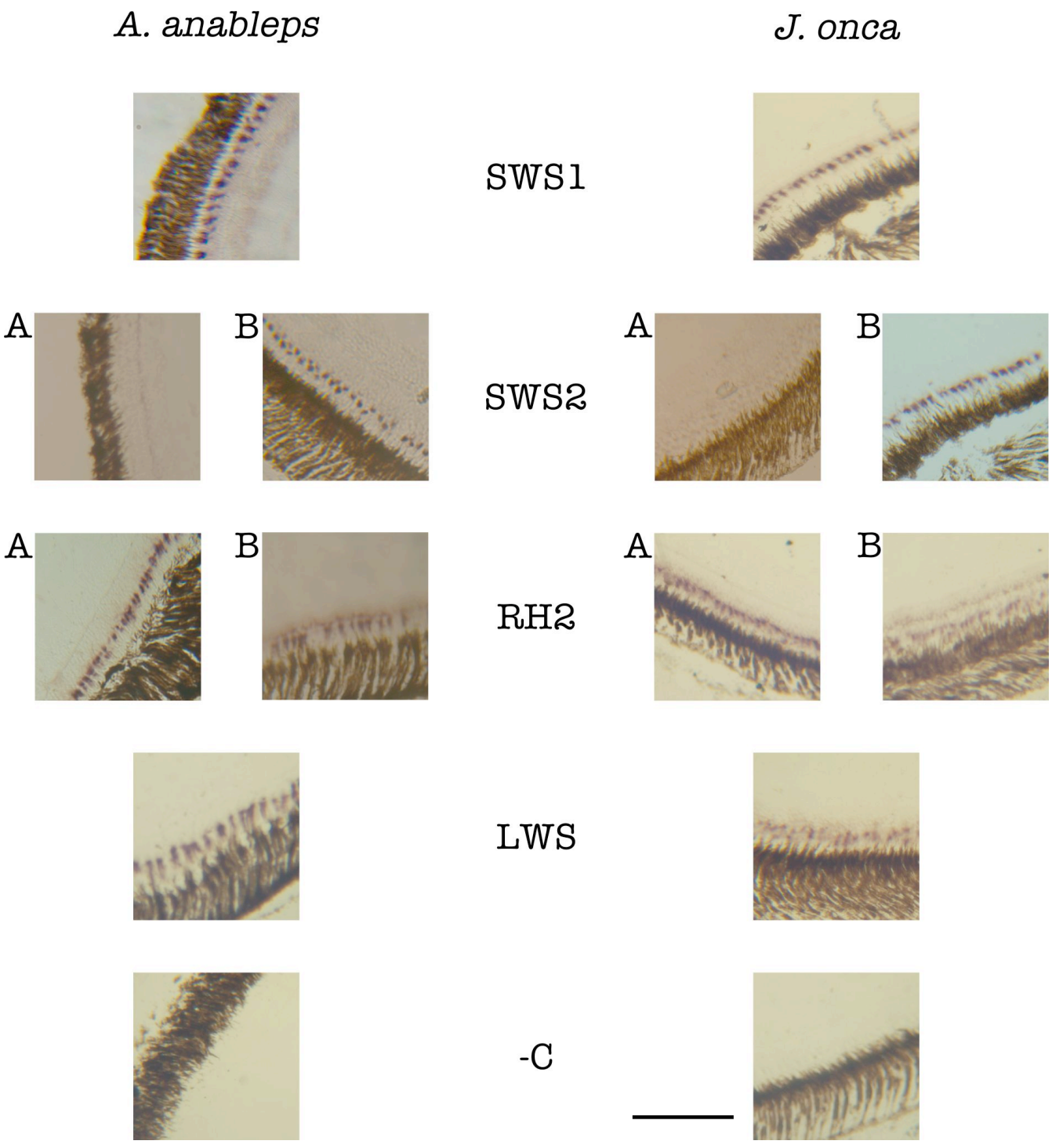


Figure 3.1: Cell level expression for each probe used.

In all positive cases, the probe only significantly bound the photoreceptor cell layer. Gene expression and probe binding produces the purple colouring. The brown layer is the pigment epithelium. –C is a sense RH2A probe. The scale bar represents 100 μm.

SWS1

In *J. onca*, SWS1 expression was uniform in the dorsal region of the eye, while in the ventral edge, cells expressing SWS1 were sparse and disperse (Figure 3.2). This pattern was seen in both individuals. In *A. anableps*, expression is strong and dense in the dorsal half of the eye. In the ventral half of the eye, the density of cells expressing SWS1 decreases towards the ventral tip (Figure 3.3). This pattern was seen in three of four individuals.

A)
D
V
B)
N
T

Figure 3.2: ISH of SWS1 probe to *J. onca* sections.

A) Dorsal-ventral cross section of male *J. onca*. B) Nasal-temporal cross section of female *J. onca*. All sections were hybridized with the SWS1 RNA probe. D is dorsal, V is ventral, N is nasal and T is temporal. The scale bar represents 500 μm. Black curves indicate areas showing expression, while the dotted line indicates reduced expression.

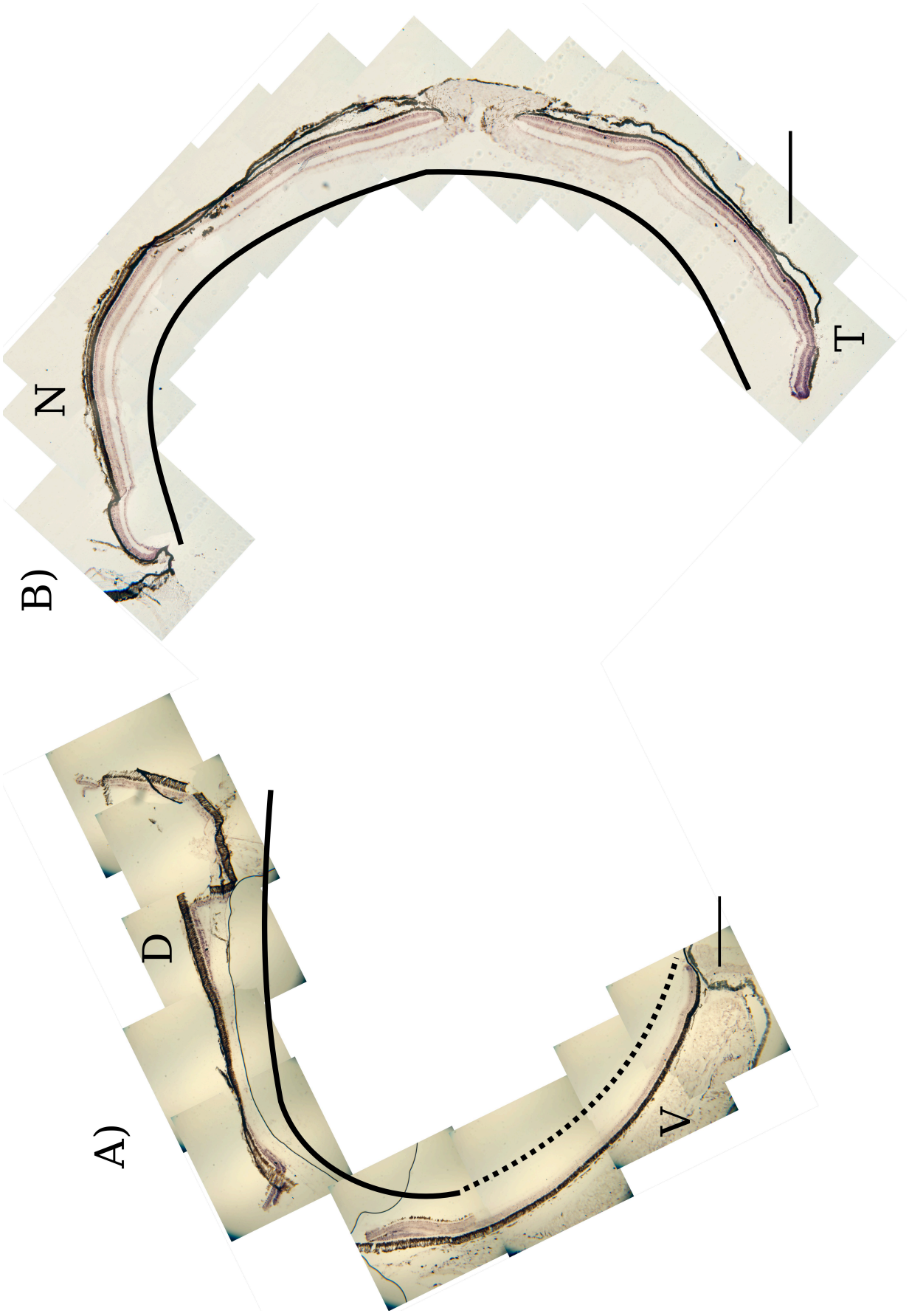

B)
N
T
A)
D
V

Figure 3.3: ISH of SWS1 probe to *A. anableps* sections.

A) Dorsal-ventral cross section of adult male A. *anableps.* B) Nasal-temporal cross section of juvenile *A. anableps.* All sections were hybridized with the SWS1 RNA probe. D is dorsal, V is ventral, N is nasal and T is temporal. The scale bar represents 500 μm. Black curves indicate areas showing expression, while the dotted line indicates reduced expression.

SWS2

The longer wavelength blue sensitive opsin, SWS2A, was not expressed in photoreceptors of either *J. onca* or *A. anableps* (Figure 3.4). In both species, cells expressing the violet sensitive paralog, SWS2B, are distributed across the eye (Figure 3.5 & Figure 3.6). Two *A. anableps* and all *J. onca* individuals showed reduced expression in the ventral portions of the retina.

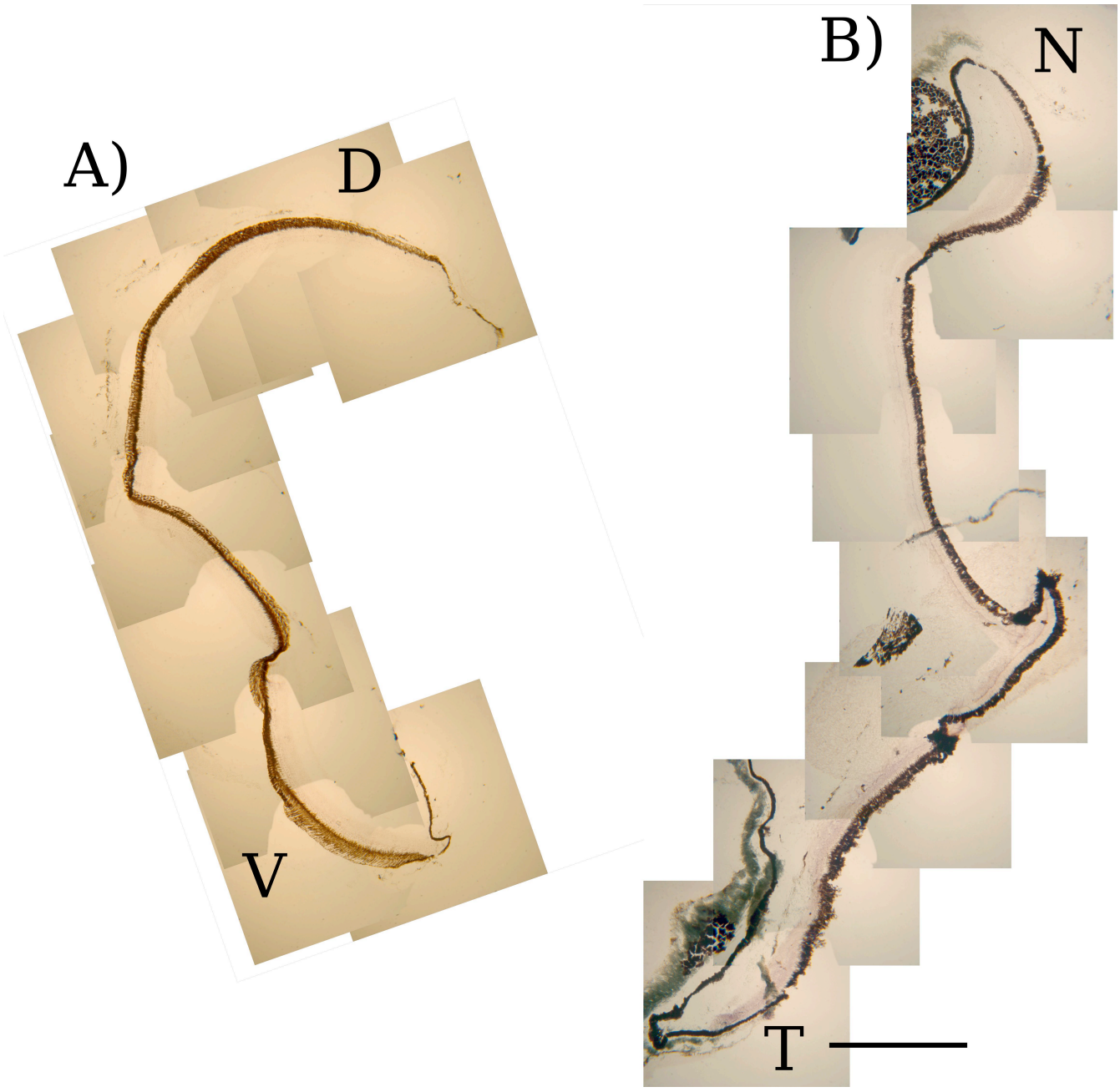

A)
D
V
B)
N
T

Figure 3.4: ISH of SWS2A probe to *J. onca* and *A. anableps* sections.

A) Dorsal-ventral cross section of male *J. onca*. B) Nasal-temporal cross section of adult male *A. anableps.* All sections were hybridized with the SWS2A RNA probe. D is dorsal, V is ventral, N is nasal and T is temporal. The scale bar represents 500 μm.

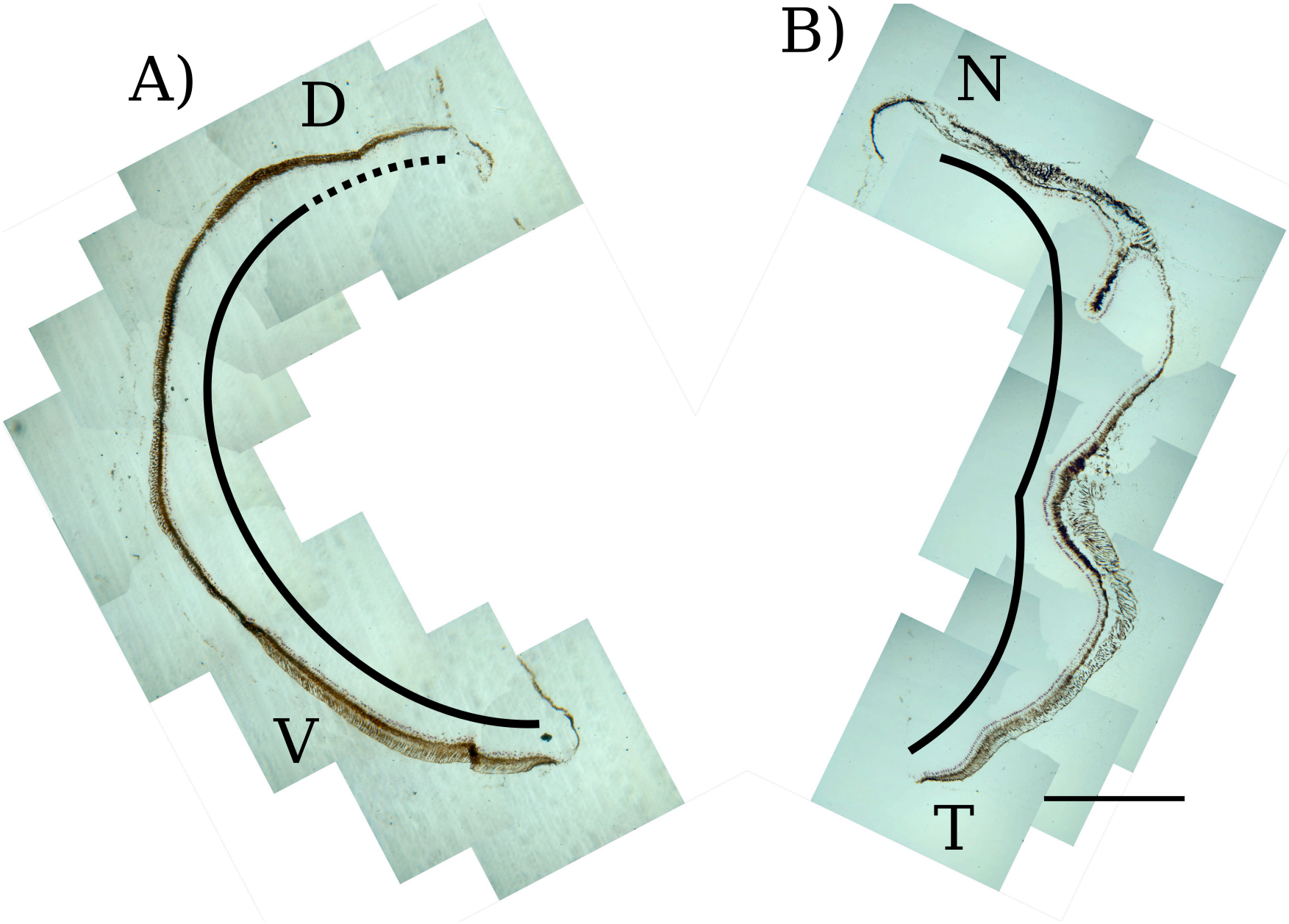

A)
D
V
B)
N
T

Figure 3.5: ISH of SWS2B probe to *J. onca* sections.

A) Dorsal-ventral cross section of male *J. onca*. B) Nasal-temporal cross section of female *J. onca.* All sections were hybridized with the SWS2B RNA probe. D is dorsal, V is ventral, N is nasal and T is temporal. The scale bar represents 500 μm. Black curves indicate areas showing expression, while the dotted line indicates reduced expression.

B)
N
T
A)
D
V

Figure 3.6: ISH of SWS2B probe to *A. anableps* sections.

A) Dorsal-ventral cross section of adult male A. *anableps.* B) Nasal-temporal cross section of adult male *A. anableps.* All sections were hybridized with the SWS2B RNA probe. D is dorsal, V is ventral, N is nasal and T is temporal. The scale bar represents 500 μm. Black curves indicate areas showing expression.

RH2

In *J. onca*, both RH2A and RH2B opsins had uniform expression across the retina in all slides (Figure 3.7 & Figure 3.8). In adult *A. anableps*, the longer wavelength paralog, RH2A, is densely expressed only in the ventral retina, except for a patch in the very dorsal tip which also shows expression. Localized ventral expression with minor dorsal tip expression was seen in both adult individuals (Figure 3.9). In juveniles, expression is sparse but confined to the ventral half of the eye, when present. The shorter wavelength paralog, RH2B, was found to have uniform expression across the eye in the adult *A. anableps* in all slides, and little to no expression in the juveniles (only one slide showed minor expression) (Figure 3.10).

A)
D
V
B)
N
T

Figure 3.7: ISH of RH2A probe to *J. onca* sections.

A) Dorsal-ventral cross section of female *J. onca*. B) Nasal-temporal cross section of female *J. onca.* All sections were hybridized with the RH2A RNA probe. D is dorsal, V is ventral, N is nasal and T is temporal. The scale bar represents 500 μm. Black curves indicate areas showing expression.

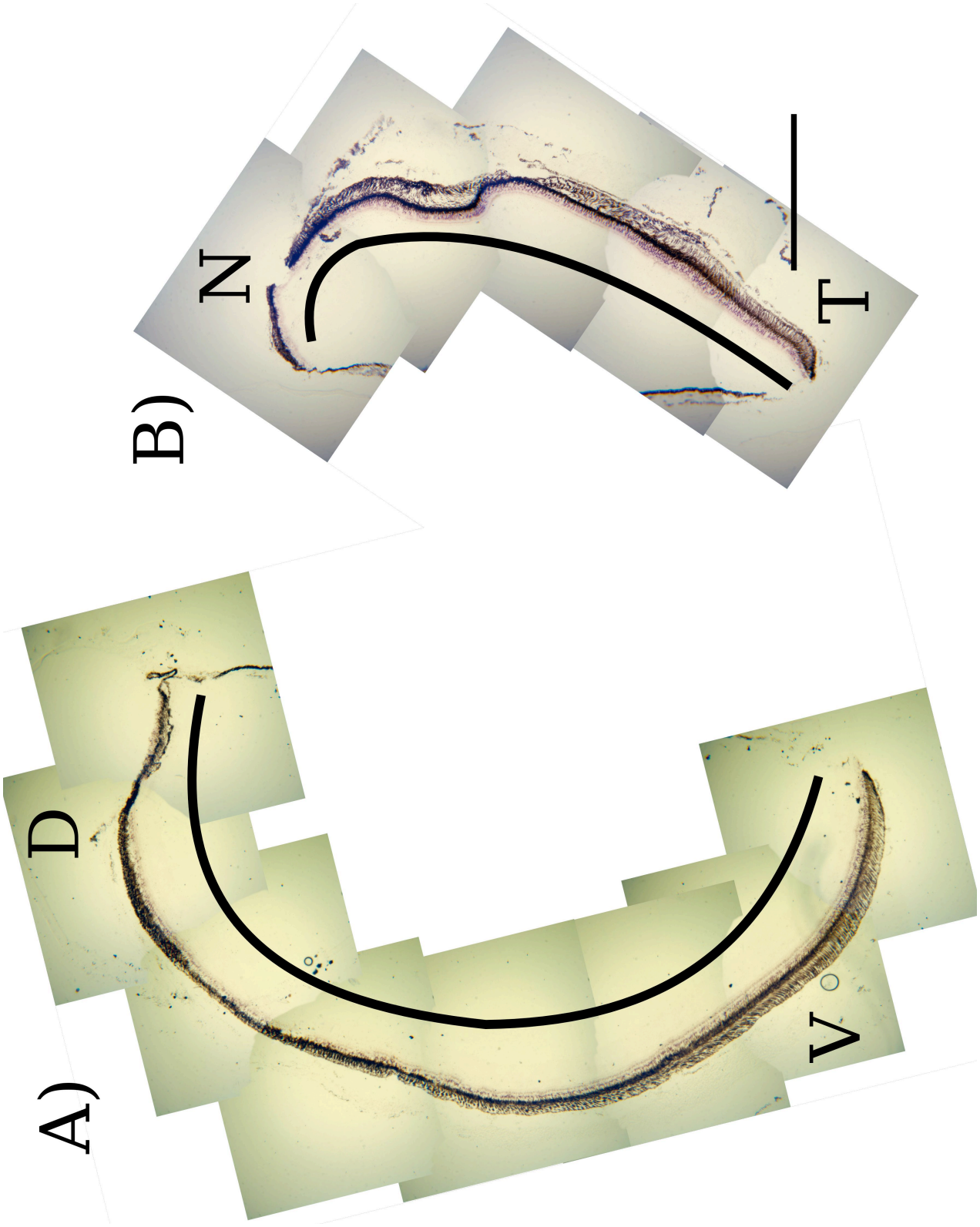

A)
D
V
B)
N
T

Figure 3.8: ISH of RH2B probe to *J. onca* sections.

A) Dorsal-ventral cross section of male *J. onca*. B) Nasal-temporal cross section of female *J. onca.* All sections were hybridized with the RH2B RNA probe. D is dorsal, V is ventral, N is nasal and T is temporal. The scale bar represents 500 μm. Black curves indicate areas showing expression.

A)
D
V
B)
N
T

Figure 3.9: ISH of RH2A probe to *A. anableps* sections.

A) Dorsal-ventral cross section of adult male A. *anableps.* B) Nasal-temporal cross section of adult male *A. anableps.* This section is cut at such an angle that the medial retina is more dorsal than the edges. All sections were hybridized with the RH2A RNA probe. D is dorsal, V is ventral, N is nasal and T is temporal. The scale bar represents 500 μm. Black curves indicate areas showing expression.

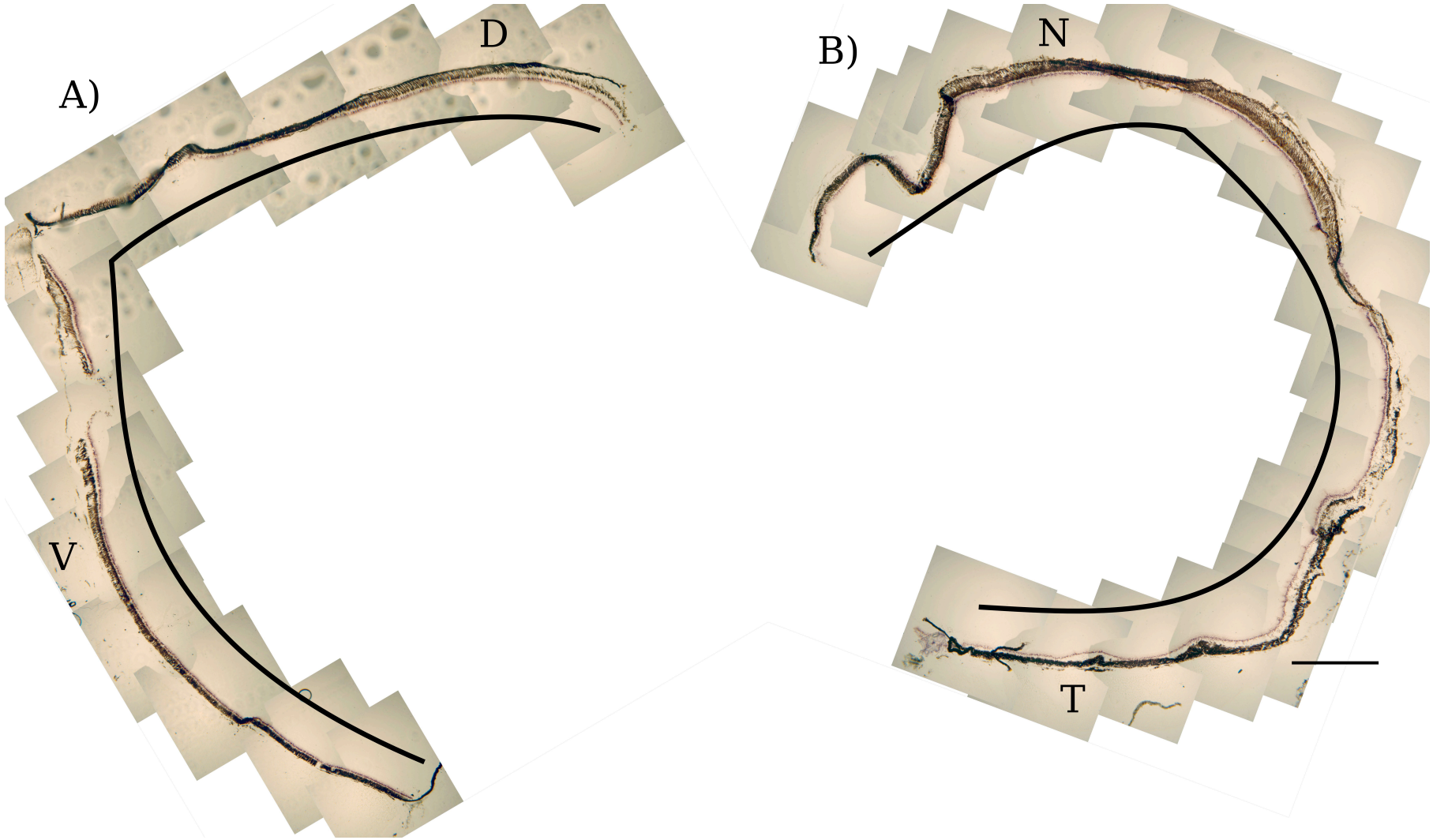

A)
D
V
B)
N
T

Figure 3.10: ISH of RH2B probe to *A. anableps* sections.

A) Dorsal-ventral cross section of adult male A. *anableps.* B) Nasal-temporal cross section of adult male *A. anableps.* All sections were hybridized with the RH2B RNA probe. D is dorsal, V is ventral, N is nasal and T is temporal. The scale bar represents 500 μm. Black curves indicate areas showing expression.

LWS

Using the LWS probe, different expression patterns were seen in male and female *J. onca*. Male *J. onca* showed a nasal-temporal stripe of expression mid eye (one individual) while female *J. onca* showed expression only in the dorsal corner of the eye (one individual) (Figure 3.11). In *A. anableps*, expression was strong and dense strictly in the dorsal half of the retina which projects slightly ventral past the bend dividing the dorsal and ventral halves (Figure 3.12). This pattern appears to be the inversion of the expression pattern of RH2A. Strictly dorsal expression was seen in three individuals.

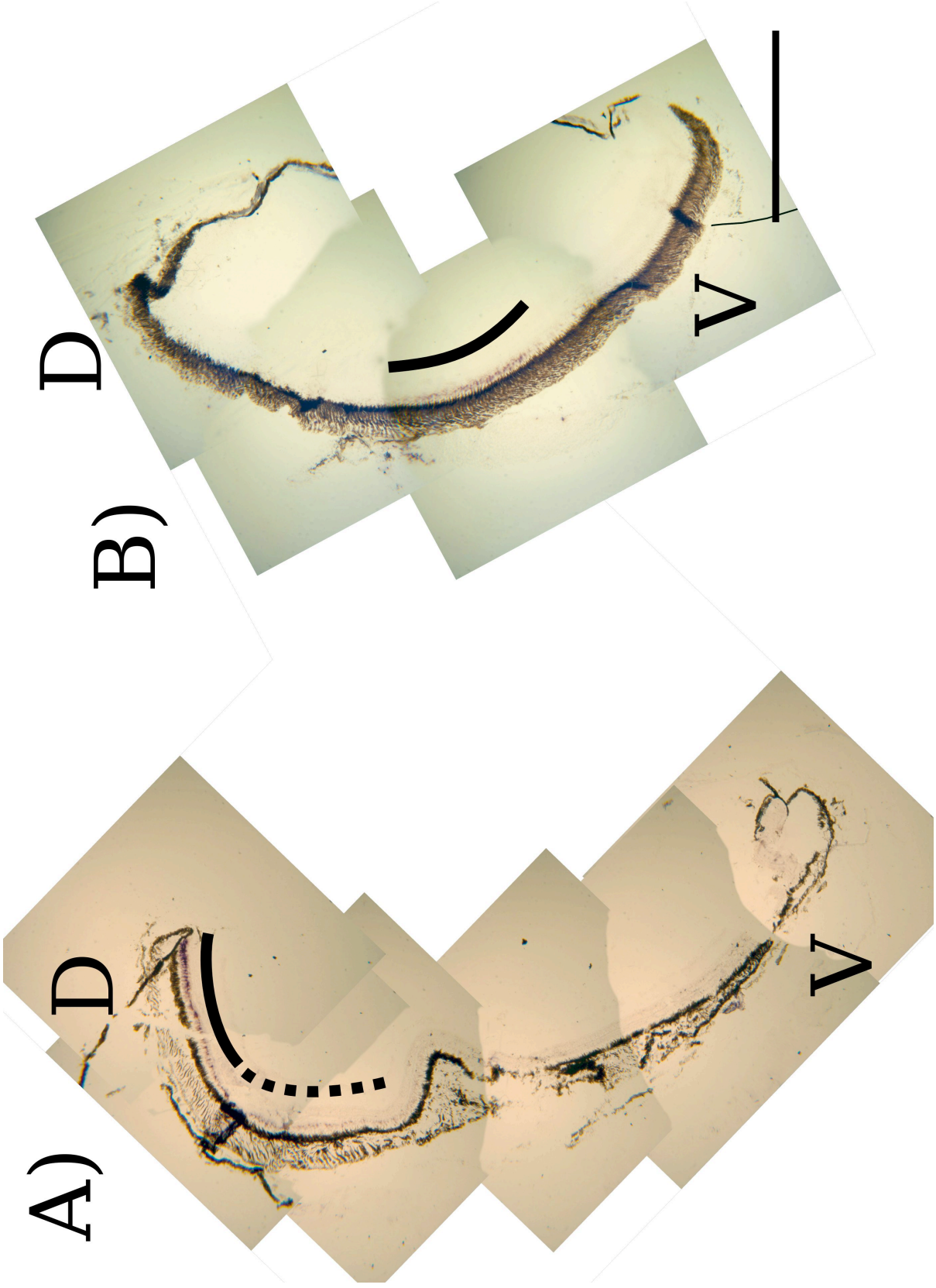

A)
D
V
B)
D
V

Figure 3.11: ISH of LWS probe to *J. onca* sections.

A) Dorsal-ventral cross section of female *J. onca*. B) Dorsal-ventral cross section of male *J. onca.* All sections were hybridized with the LWS RNA probe. D is dorsal, V is ventral, N is nasal and T is temporal. The scale bar represents 500 μm. Black curves indicate areas showing expression, while dashed lines indicate reduced expression.

A)
D
V
B)
N
T

Figure 3.12: ISH of LWS probe to *A. anableps* sections.

A) Dorsal-ventral cross section of adult male A. *anableps.* B) Nasal-temporal cross section of adult male *A. anableps.* This section is cut at such an angle that the medial retina is more dorsal than the edges. All sections were hybridized with the LWS RNA probe. D is dorsal, V is ventral, N is nasal and T is temporal. The scale bar represents 500 μm. Black curves indicate areas showing expression.

Negative controls

Sections probed with sense RH2A probe in both species showed no probe binding in all samples for both species (Figure 3.13).

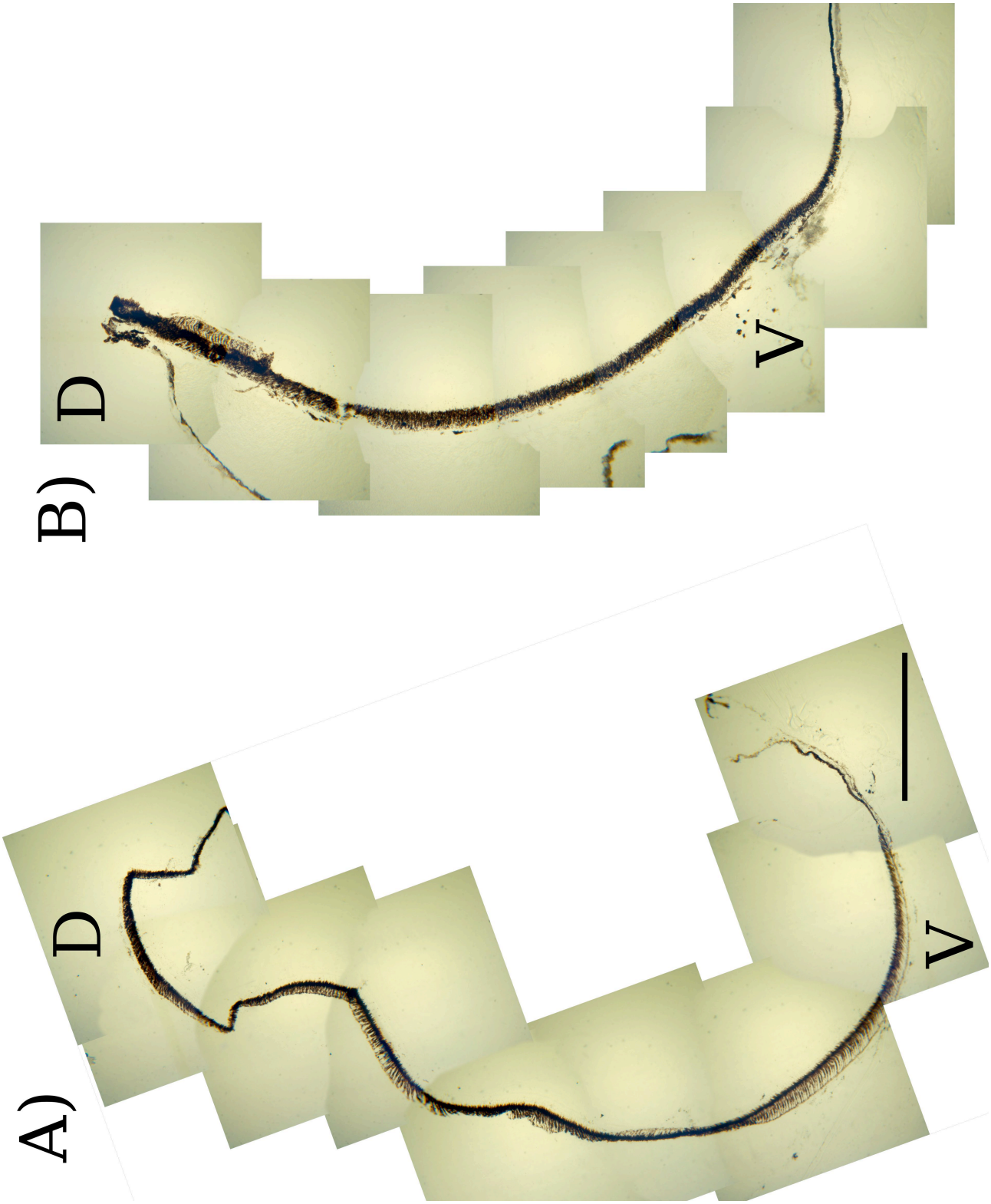

A)
D
V
B)
D
V

Figure 3.13: ISH of sense RH2A probe to *J. onca* and *A. anableps* sections.

A) Dorsal-ventral cross section of male *J. onca*. B) Dorsal-ventral cross section of adult male *A. anableps*. All sections were hybridized with sense RH2A probe. D is dorsal, V is ventral. The scale bar represents 500 μm.

DISCUSSION

The most important feature of a photoreceptor is the opsin gene it expresses because this determines the wavelength of light to which it is sensitive. *In situ* hybridization has often been used to determine the identity of opsins present in photoreceptor cells. These studies largely focused on cell type (Minamoto & Shimizu, 2005; Miyazaki et al., 2008) or retinal development (Stenkamp et al., 1996; Helvik et al., 2001), although in 2005 Takechi et al. discovered that the duplicated RH2 and LWS genes in zebrafish were expressed in different regions of the retina. They attributed this to the uneven spectral environment of the fish, i.e., light above the fish has different wavelength composition than light below it.

This feature is taken to an extreme in *A. anableps*, where there is a clear dichotomy between the dorsal and ventral retina. Each retinal half is exposed to a different light environment (aerial or aquatic light). Consequently, the retina has evolved differences in photoreceptor size as well as ganglion cell density between retinal halves but until now, no differences in photoreceptor sensitivities have been detected (MSP data showed a similar distribution of violet- blue- and yellow-sensitive cone cells in both regions of the retina) (Borwein & Hollenberg, 1973; Avery & Bowmaker, 1982; Oliveira et al., 2006).

Reconciling MSP and opsin expression

Early MSP work identified three cone types present in the *A. anableps* retina: 409, 463, and 576 nm. By comparing maximal absorbance values for orthologous genes in closely related fish, I have assigned those peaks to SWS2B, RH2B and LWS opsins respectively (Table 3.1). Although it is plausible that the SWS2A gene, and not RH2B, is responsible for the 463 nm cone in *A. anableps*, it is unlikely due to the lack of SWS2A expression found in this study. Furthermore, SWS2 opsins are typically found in single cones, while the 463 nm cone is a double cone. Although there is no MSP data for *J. onca*, key site

analysis indicates that its opsins have λ_{max} values very similar to *A. anableps* (Windsor & Owens, 2009).

Table 3.1: Cone opsin maximal absorbance values in fish related to *A. anableps*.

Maximal absorbencies for swordtail, bluefin killifish, medaka and tilapia for each cone opsin gene. MSP peaks from *A. anableps* are assigned to genes based on closest match. Bold indicates that the absorbance peak to gene assignment is uncertain (Watson et al., 2010; Fuller et al., 2004; Matsumoto et al., 2006; Spady et al., 2006; Avery & Bowmaker, 1982).

	Swordtail (*Xiphophorus helleri*)	Bluefin Killifish (*Lucania goodei*)	Medaka (*Oryzias latipes*)	Tilapia (*Oreochromis niloticus*)	Four-eyed fish *(A. anableps)*
LWS S180	568	573	561	561	576
LWS P180	**534**	-	-	-	
RH2A	**534**	539	492/516	518/528	
RH2B	**459**	-	452	472	463
SWS2A	**459**	455	439	456	
SWS2B	405	405	405	425	409
SWS1	365	359	356	360	

SWS opsins and RH2B

The UV sensitive opsin, SWS1, was expressed throughout the retina in *A. anableps* and *J. onca*. This gene is predicted to absorb ~360 nm light based upon data from related species. Although UV sensitive cones were not detected using MSP, this study did not scan into the UV region of the spectrum due to technical limitations (Avery & Bowmaker, 1982). My data suggests that both *A. anableps* and *J. onca* are sensitive to UV light. In *A. anableps*, UV sensitivity appears to be reduced in the aerial field of view (fewer SWS1 cones in the ventral retina). This may be correlated with the thickened dorsal cornea, which likely blocks UV wavelengths (Swamynathan et al., 2003).

SWS2B is expressed throughout the eyes of *A. anableps* and *J. onca*. Although SWS2A did not appear to be expressed, it does have a functional reading frame in both species, so it likely has, or recently had, some functional importance. Opsin expression is developmentally labile in fish and, therefore, it is possible that SWS2A is expressed earlier in development than was surveyed in this experiment (Spady et al., 2006; Shand et al., 2008).

The genes SWS2A and SWS2B are linked in a head to tail orientation in related species (including medaka and swordtail) (Matsumoto et al., 2006; Watson et al., 2010).

This gene pair is also linked to a locus control region (LCR) implicated in the control of opsin gene expression (Watson et al., 2010). Although in *A. anableps* and *J. onca* the gene predicted to be closest to the LCR (SWS2B) is more highly expressed, this is not always the case (Matsumoto et al., 2006).

The middle wavelength sensitive gene RH2B is expressed in the double cones across the retina for both species. Based on λ_{max} estimations, this gene is responsible for the 463 nm peak found using MSP. It is interesting to note that SWS2A, which is expected to absorb at a similar wavelength, was not expressed. In this way, the three uniformly expressed genes in both species are evenly spaced at ~360 nm, 409 nm and 463 nm.

LWS and RH2A

The most interesting patterns of opsin gene expression were found with the LWS and RH2A opsin probes in *A. anableps*. RH2A transcripts were detected only in the ventral portion of the retina and dorsal tip of the retina, while LWS opsin transcripts were detected only in the dorsal and medial retina (Figure 3.14). The expression domains of these two genes do not appear to overlap except in the dorsal distal tip. LWS expression extends slightly ventral to the optic nerve, which is placed at the demarcation between dorsal and ventral retinas. RH2A expression stops short of the optic nerve. When examining nasal-temporal sections, the non-overlapping pattern is apparent. Figure 3.9B and 3.12B are neighbouring sections of the same eye (<100 μm apart); these sections show RH2A expression in the nasal and temporal edges, while LWS is only medial. This is likely a section taken slightly off axis; the middle of the retina is more dorsal than the nasal and temporal edges. Although a small region of co-expression cannot be ruled out, it appears that these genes have distinct expression domains, with the exception of the dorsal distal region.

In *J. onca*, RH2A, like its paralog RH2B, is uniformly expressed across the retina. The LWS genes, though, show both spatial and individual level variation in *J. onca*. Females appear to express LWS genes only in the dorsal half of the retina, while males have a horizontal stripe which crosses the retina medially. Due to sample size, I cannot determine if these differences reflect true sex differences or if they are due to other factors such as age or individual level variation. In closely related guppies, the LWS genes expressed have considerable individual level variation (Archer & Lythgoe, 1990).

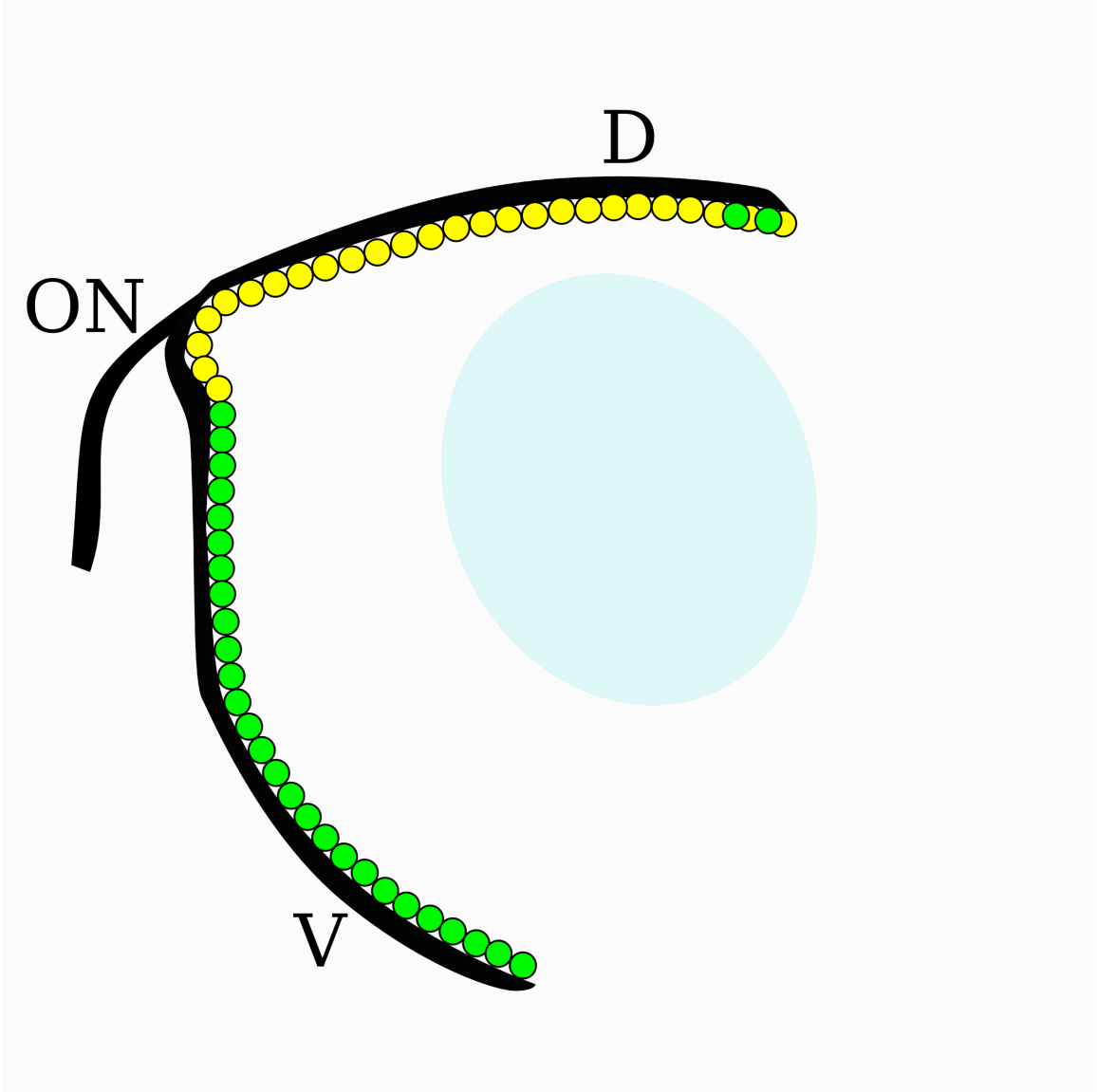


Figure 3.14: Summary of expression domains of RH2 and LWS in *A. anableps.*

A schematic of the *A. anableps* eye section dorsal-ventrally. Yellow circles represent areas where LWS is expressed, while green circles represent RH2A expression areas. The blue oval is the lens, ON is the optic nerve, D is dorsal and V is ventral.

The evolution of eye morphology and opsin expression

The expression patterns of *J. onca* allow us to make inferences on how opsin gene expression coevolved with the *A. anableps* eye morphology. First, it is important to note that the optic nerve in *A. anableps* is placed at the border between the dorsal and ventral retina. As the optic nerve is placed in the middle of the retina in a normal fish eye, this suggests that *A. anableps* gained its distinct retinal halves from dividing a normal retina, not from uneven extension of one retinal half.

The majority of opsins (SWS1, SWS2A, SWS2B, and RH2B) show similar patterns in both *J. onca* and *A. anableps* indicating that these patterns evolved prior to *A. anableps*' eye morphology (Figure 3.15). On the other hand, the expression pattern of RH2A in *J. onca* indicates that *A. anableps* evolved its ventrally limited RH2A expression pattern through the loss of expression in the dorsal region. As cone cells are added along the periphery of the eye, the dorsal distal region contains young photoreceptors (Stenkamp, 2007). I hypothesize that RH2A is expressed in this region because the youngest cells follow the ancestral expression pattern (RH2A in cones) before cell communication causes this expression to be lost only in the dorsal region.

The expression of RH2A in the dorsal region of the retina could be lost in several possible ways. Cells expressing RH2A may migrate to the ventral retina, they may undergo apoptosis or they may switch the opsin that they express, likely to LWS. The migration hypothesis is not supported by my data because no cells expressing RH2A are found in the middle of the dorsal retina (i.e., on their way south). If RH2A expressing cells were continually being generated in the dorsal retina and migrating across the dorsal region, I expect to see at least some cells expressing RH2A in the middle of the dorsal region. With current data I am unable to evaluate between the remaining two hypotheses. In salmon, the short corner cones express SWS1 in the alevins but following full yolk absorption these cones either die, or begin expressing SWS2 (Cheng et al., 2006; Cheng & Flamarique, 2007).

In *J. onca*, the expression pattern of LWS is variable but in all cases expression is confined to a small region of the retina that overlaps with RH2A expression. Therefore, the location of LWS expression is very similar to *A. anableps*, where LWS expression is medial and dorsal. Thus, it appears that subdivision of the retina between LWS and RH2A was accomplished through the reduction of RH2A expression in the dorsal retina.

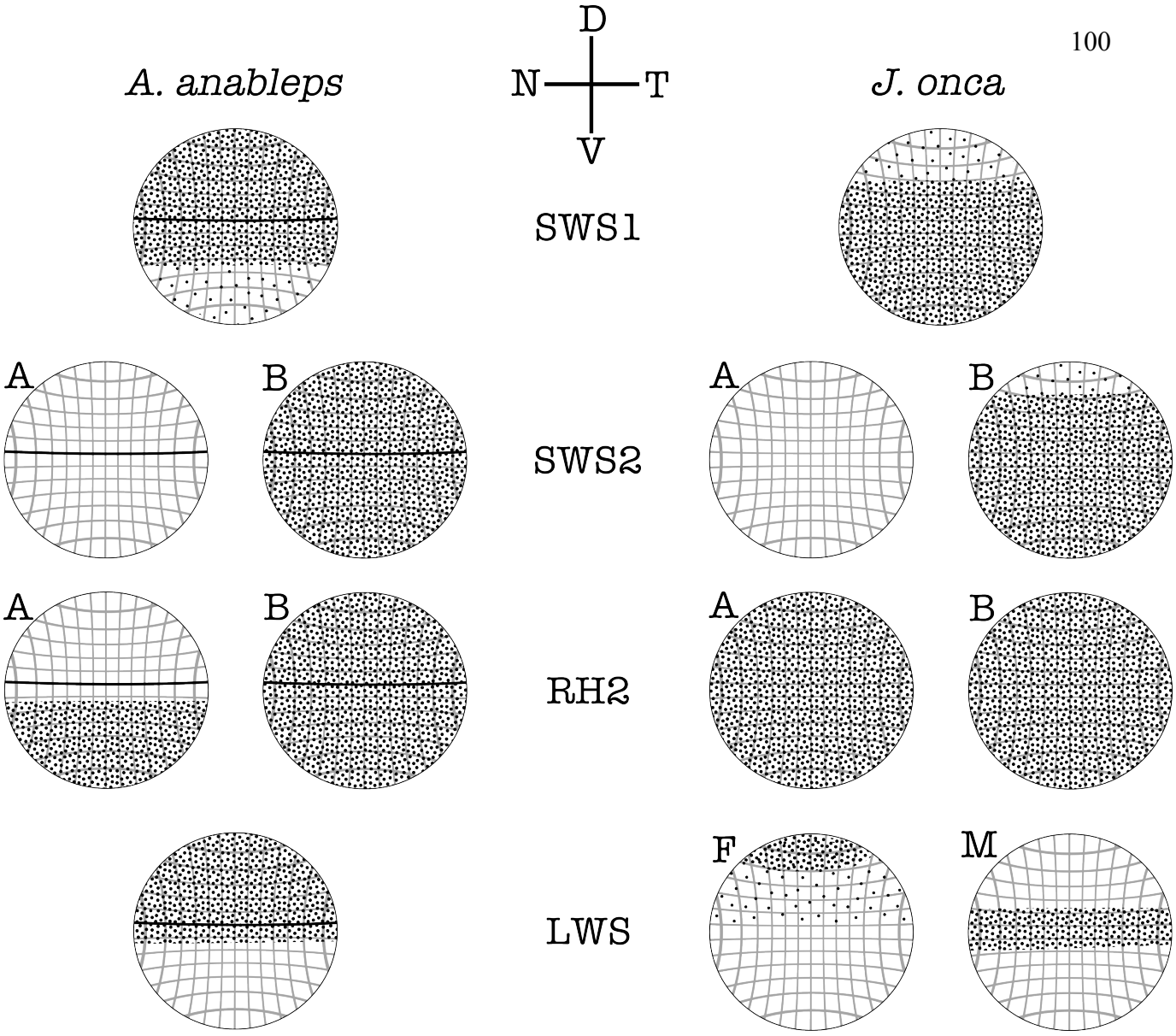


Figure 3.15: Summarized expression domains for all opsins

Dots represent areas where each gene is expressed uniformly. Widely spaced dots represent less dense expression. The line across the retina in *A. anableps* represents the bend in the retina dividing dorsal from ventral. M is male, F is female, D is dorsal, V is ventral, N is nasal and T is temporal.

LWS paralogs

This experiment used a single probe to examine all LWS paralogs. Percent identity between the probe and paralogs varied from 100% to 93%. Although binding to LWS S180r was not tested using dot blot analysis, a divergent guppy sequence based probe

with an 86% identity to *A. anableps* sequence successfully bound the photoreceptor layer of *A. anableps*. This indicates that with the current level of stringency, the LWS probe has the potential to bind all LWS paralogs in both *A. anableps* and *J. onca*.

The LWS repertoire is different between both tested species. *J. onca* has LWS S180, P180 and S180r, while *A. anableps* has LWS S180α, S180β, S180γ and S180r. The ancestral key site haplotype found in S180, S180α, S180β and S180r predicts a pigment absorbing at ~560 nm. The P180 and S180γ opsins have key site substitutions that are predicted to shift λ_{max} by -45 nm and -16 nm respectively (Yokoyama, 2008). Thus, it is possible that the differential expression in LWS paralogs may have large functional implications.

My work indicates that if these genes are differentially expressed, all expression is limited to the dorsal retina in *A. anableps*. To differentiate the expression patterns of highly similar LWS paralogs, smaller, more specific probes should be used. For example, the use of locked nucleic acid (LNA) modified DNA probes increases probe specificity and allows for much shorter probes (>12bp) (Braasch & Corey, 2001; Kloosterman et al., 2005). This would allow for the minor nucleotide differences between paralogs to be used for differentiation, where a longer probe would not succeed.

Vision implications

When taken all together, my opsin gene expression data indicate that *J. onca* has broad spectral sensitivity from 360 nm (UV) to 530 nm (green) in its whole retina. Furthermore, it has enhanced long wavelength sensitivity in a limited portion of its field of view, either below or beside the fish. Four or five visual pigments present together gives the potential for tetra- or pentachromatic vision assuming all cone cells function in wavelength discrimination and the requisite neural circuitry exists. This is not necessarily true; double cones have been hypothesized to function in movement, luminance or polarized light detection rather than wavelength discrimination (Osorio & Vorobyev, 2005; Wagner, 1990; Cameron & Pugh, 1991).

A. anableps has broad sensitivity from ~360 nm (UV) to ~460 nm (blue) in its whole retina. The ventral retina, responsible for aerial vision, has increased sensitivity to ~530 nm (green) light, while the dorsal retina, responsible for aquatic vision, has increased

sensitivity to ~560 nm (yellow) light. Thus, with the above caveats, both halves of *A. anableps'* eye have the potential for tetrachromatic vision.

Analysis of ganglion cell density shows that there is a strong visual streak just ventral of the demarcation between the ventral and dorsal retina (Oliveira et al., 2006). This retinal area corresponds to vision just above the waterline, and the increased cell density indicates high visual acuity in this area. My work shows that this area contains cone cells that express LWS opsins, instead of the RH2A opsin, although a region of overlap cannot be ruled out.

The differences in opsin expression may be correlated with the differences in light environment between the retinal halves. The brackish waters of the mangrove forests *A. anableps* inhabit often contain dissolved organic matter that affects light transmission, especially during the rainy season (Farmer et al., 1993). In pure water, light transmits most readily at 470 nm, but with the addition of dissolved organic matter this value shifts to a longer wavelength up to 570 nm (Lythgoe & Partridge, 1989; Davies-Colley et al., 1988; Nicol & Somiya, 1996). As upwelling light presented to the dorsal retina would necessarily be filtered through water, it is likely that this light will have a greater proportion of long wavelength when compared to unfiltered aerial light. Indeed, during the rainy season of the Orinoco river, the wavelength of maximal penetration was 580 nm (Farmer et al., 1993). Light measurements completed for the mangrove dwelling archerfish found that downwelling light had maximal irradiance at ~500 nm while upwelling had the highest irradiance at ~570 nm (Temple et al., 2010). Thus, it seems likely that the restricted expression of LWS in the dorsal retina functions to better match the longer wavelength light it receives.

CONCLUSIONS

This study provides evidence that *A. anableps* has adapted its opsin gene expression to coincide with its unique eye morphology. The distinct expression patterns between opsins and the relatively clear functional implications make *A. anableps* an excellent model for studying the development of intraretinal pigment variation and the evolutionary pressures that facilitated it. Furthermore, behavioural tests on wavelength discrimination in the aerial and aquatic fields of view may shed light on how differences in pigment absorption affect colour vision.

Chapter 4 – Seeing clearly: Molecular evolution of visual opsin genes in fish

The clear relationship from opsin gene sequence and expression to vision make this gene family ideal for studying the evolution of sensory systems. Gene duplication provided vertebrates with the potential for colour vision. In mammals, this repertoire was reduced through an extended nocturnal period of evolution but in fish the opposite appears to have occurred (Bowmaker, 2008). Opsin repertoires of ten genes are not uncommon in fish. The reason for frequent opsin repertoire expansion in fish has never been fully elucidated. In this thesis, I surveyed all opsin gene duplication events in fish. I also examined opsin gene expression in the four-eyed fish, *A. anableps*, to understand how opsin duplication contributes to morphological and spectral specialization. What we learn from *A. anableps* provides insight into the flexibility of the retina with respect to opsin repertoire utilization.

Opsin repertoire expansion

Opsin gene repertoires in fish have been explored since the 1980's but it is only in the last decade, with the rapid increase in whole genome sequencing, that the exceptionality of fish opsin repertoires was fully appreciated. As many duplication events are young and in tandem arrays, they are easy to overlook when using PCR and southern blotting. Initial efforts found four and five cone opsin genes in medaka and zebrafish respectively, while later genomic library screening experiments found eight in each (Hisatomi et al., 1997; Vihtelic et al., 1999; Matsumoto et al, 2006; Chinen et al., 2003).

With this in mind, it is likely that the number of opsin genes in many species have been underestimated. Even with this underestimation, it is clear that fish have retained many opsin gene duplications while tetrapods have retained very few. My work has shown that opsin gene duplications occur and are fixed primarily in the RH2 (50% of duplications) and LWS (29%) subfamilies. This general trend has not been noted before, nor has it been explained.

It is interesting to note that RH2 and LWS genes are typically the only opsins expressed in double cones (Hisatomi et al., 1997; Vihtelic et al., 1999). Double cones,

found in a variety of vertebrate species, are thought to be linked to polarized light, luminance and movement detection (Wagner, 1990; Osorio & Vorobyev, 2005; Cameron & Pugh, 1991). Their sensitivity is often matched to background illumination (Loew & Lythgoe, 1978; Bowmaker, 1995). Background illumination is highly dependent on water quality. Although aerial vision is broad spectrum, aquatic vision is spectrally limited by the absorption of particular wavelengths by water and dissolved solute. Maximal transmission in water varies between 470 nm and 600 nm, coincidentally the range of maximal absorptions provided by the RH2 and LWS subfamilies. Having multiple spectrally diverged duplicates in these subfamilies would allow double cones to be tuned to the maximal absorption of the environment. As water turbidity can vary seasonally, geographically and over evolutionary time scales, this tuning may occur through individual expression changes (as has been observed under differing light conditions (Fuller et al., 2005; Shand et al., 2008)) or species level changes during evolution (such as in the cichlids of lake Malawi (Hofmann et al., 2009)). Since the single cones do not appear to match background spectral illumination, there is less pressure for the retention of spectrally diverse duplicates in the SWS1 and SWS2 subfamilies. As double cones are common across vertebrates, I would expect the RH2 and LWS subfamilies to duplicate in other aquatic vertebrates, such as sharks. Indeed, the only full opsin repertoire for a cartilaginous fish, *Callorhinchus milii,* showed a duplication in the LWS subfamily (Davies et al., 2009).

Opsin expression variation

Compared to the relatively simple and unorganized expression pattern of opsins in the human eye, fish eyes present a more opaque but fascinating picture. In most fish, cone photoreceptors are one of three morphological types: short single cones expressing SWS1, long single cones expressing SWS2, and double cones expressing RH2 or LWS. This picture is complicated by gene duplications, e.g., if there are two SWS2 genes where does each go?

With regards to opsin gene expression, the most well studied fish clade is the African cichlids. This group has undergone an adaptive radiation in the rift lakes and consequently there are many, very closely related species. All cichlids have seven cone opsins, but even closely related species can have very different expression patterns.

Generally three opsins are expressed predominantly, two of the LWS and RH2 subfamilies and one of the SWS1 and SWS2 subfamilies. Although this work allows for general sensitivity of the entire retina to be estimated, it ignores the intraretinal variation.

My work demonstrates that examining cellular level expression patterns provides a much more intriguing picture than is gained from a homogenized measure of eye opsin expression. The total expression of RH2A and LWS in *A. anableps* would likely be similar from whole eye qPCR, but *in situ* hybridization revealed starkly different expression patterns between retinal hemispheres. Although *A. anableps*' eye is highly specialized to its unique visual niche, it represents merely an exaggeration of the standard fish eye, not an entirely novel creation. Together with non-uniform expression patterns seen in *Danio rerio*, *J. onca* and *Poecilia reticulata,* it is apparent that the fish retina is not a homogenous space (Takechi & Kawamura, 2005; Rennison & Owens, unpublished). It is highly specialized to a complex light environment that varies in every axis (Munz & McFarland, 1977).

Opsin duplicates in the retina are utilized in different ways. Some are expressed during different developmental periods, some in different cone types and others in different areas of the retina (Takechi & Kawamura, 2005; Minamoto & Shimizu, 2005; Spady et al., 2006). My work suggests that in Anablepids, SWS2B is expressed to the exclusion of SWS2A. Both RH2A and RH2B can be expressed in the same area of the retina, although without dual labelling it is impossible to know if they are expressed in the same cone cells.

The LWS subfamily has undergone several duplication events in Anablepids. *A. anableps* has four LWS genes, while *J. onca* has three. This study did not differentiate the expression of these duplicates. As key site analysis indicates that in both cases one duplicate has a different spectral sensitivity; LWS S180γ in *A. anableps* and LWS P180 in *J. onca*. Thus, differences in the expression pattern of LWS duplicates would have functional implications. It is possible that the homogenous expression patches shown with my degenerate LWS probe are actually multiple neighbouring expression domains. Future work should use shorter, more specific probes to explore this possibility.

Seeing beyond model species

The expansion of full genomic sequences is going to vastly increase the ease of molecular analysis of non-model species. It is the responsibility of future vision researchers to use these data to correlate opsin sequence information with both the environmental and phylogenetic context. Perhaps certain lineages or environmental conditions favour the retention or rapid evolution of opsin duplicates. Beyond this, expression analysis provides irreplaceable context to this information and effectively bridges the gap between sequence and environment.

Conclusions

Over the course of my work, I have fully characterized the opsin repertoires of two previously unstudied fish, *A. anableps* and *J. onca,* and placed them in a complete gene family phylogeny (Owens et al., 2009; Windsor & Owens, 2009). Using that information, I demonstrated a unique opsin expression pattern and provided a hypothesis on how it evolved and its function in nature.

Bibliography

Albensi, B. C., & Powell, J. H. (1998). The differential optomotor response of the four-eyed fish *Anableps anableps*. Perception, 27, 1475-1483.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research, 25, 3389.

Anisimova, M., & Gascuel, O. (2006). Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. Systematic Biology, 55, 539-552.

Anisimova, M., & Yang, Z. (2007). Multiple hypothesis testing to detect lineages under positive selection that affects only a few sites. Molecular Biology and Evolution, 24, 1219-1228.

Anisimova, M., Bielawski, J. P., & Yang, Z. (2001). Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. Molecular Biology and Evolution, 18, 1585-1592.

Applebury, M. L., Antoch, M. P., Baxter, L. C., Chun, L. L. Y., Falk, J. D., Farhangfar, F., Kage, K., Krzystolik, M. G., Lyass, L. A., & Robbins, J. T. (2000). The murine cone photoreceptor: A single cone type expresses both S and M opsins with retinal spatial patterning. Neuron, 27, 513-523.

Archer, S. N., & Lythgoe, J. N. (1990). The visual pigment basis for cone polymorphism in the guppy, *Poecilia reticulata*. Vision Research, 30, 225 - 233.

Avery, J. A., & Bowmaker, J. K. (1982). Visual pigments in the four-eyed fish, *Anableps anableps*. Nature, 298, 62-63.

Bailey, J. A., Gu, Z., Clark, R. A., Reinert, K., Samonte, R. V., Schwartz, S., Adams, M. D., Myers, E. W., Li, P. W., & Eichler, E. E. (2002). Recent segmental duplications in the human genome. Science, 297, 1003-1007.

Beatty, D. D. (1975). Visual pigments of the American eel *Anguilla rostrata*. Vision Research, 15, 771-776.

Bergsten, J. (2005). A review of long-branch attraction. Cladistics, 21, 163-193.

Bielawski, J. P., & Yang, Z. (2001). Positive and negative selection in the DAZ gene family. Molecular Biology and Evolution, 18, 523-529.

Björn, L. O. (2008). Photobiology: the science of life and light, second ed. Springer Netherlands.

Bloch, B., Popovici, T., Le Guellec, D., Normand, E., Chouham, S., Guitteny, A. F., & Bohlen, P. (1986). *In situ* hybridization histochemistry for the analysis of gene expression in the endocrine and central nervous system tissues: a 3-year experience. Journal of Neuroscience Research, 16, 183-200.

Borwein, B., & Hollenberg, M. J. (1973). The photoreceptors of the "four-eyed" fish, *Anableps anableps L.* Journal of Morphology, 140, 405-441.

Bowmaker, J. K. (1995). The visual pigments of fish. Progress in Retinal and Eye Research, 15, 1-31.

Bowmaker, J. K. (2008). Evolution of vertebrate visual pigments. Vision Research, 48, 2022-2041.

Bowmaker, J. K., Dartnall, H. J. A., & Herring, P. J. (1988). Longwave-sensitive visual pigments in some deep-sea fishes: segregation of 'paired' rhodopsins and porphyropsins. Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology, 163, 685-698.

Bowmaker, J. K., Govardovskii, V. I., Shukolyukov, S. A., Zueva, J. L., Hunt, D. M., Sideleva, V. G., & Smirnova, O. G. (1994). Visual pigments and the photic environment: The cottoid fish of Lake Baikal. Vision Research, 34, 591 - 605.

Braasch, D. A., & Corey, D. R. (2001). Locked nucleic acid (LNA): fine-tuning the recognition of DNA and RNA. Chemistry and Biology, 8, 1-7.

Brann, M. R., & Young, W. S. (1986). Localization and quantitation of opsin and transducin mRNAs in bovine retina by *in situ* hybridization histochemistry. FEBS Letters, 200, 275-278.

Brauer, A. (1902). Uber den Bau des Augen einiger Tiefseefische. Verhandlungen der Deutschen Zoologischen Gesellschaft (Leipzig), 12, 42-57.

Brenner, M., & Krumme, U. (2007). Tidal migration and patterns in feeding of the four-eyed fish *Anableps anableps L.* in a north Brazilian mangrove. Journal of Fish Biology, 70, 406-427.

Brett, J. R. (1957). The sense organs: The eye, in: Varley, M. E. (Eds.) The physiology of fishes, volume 2. Academic press inc. New York.

Brigati, D. J., Myerson, D., Leary, J. J., Spalholz, B., Travis, S. Z., Fong, C. K. Y., Hsiung, G. D., & Ward, D. C. (1983). Detection of viral genomes in cultured cells and paraffin-embedded tissue sections using biotin-labeled hybridization probes. Virology, 126, 32-50.

Briscoe, A. D., Bybee, S. M., Bernard, G. D., Yuan, F., Sison-Mangus, M. P., Reed, R. D., Warren, A. D., Llorente-Bousquets, J., & Chiao, C. C. (2010). Positive selection of a duplicated UV-sensitive visual pigment coincides with wing pigment evolution in Heliconius butterflies. Proceedings of the National Academy of Sciences of the United States of America, 107, 3628-3633.

Bruhn, S. L., & Cepko, C. L. (1996). Development of the pattern of photoreceptors in the chick retina. Journal of neuroscience, 16, 1430.

Bumsted, K., Jasoni, C., Szél, A., & Hendrickson, A. (1997). Spatial and temporal expression of cone opsins during monkey retinal development. The Journal of Comparative Neurology, 378, 117-134.

Cameron, D. A., & Pugh, E. N. (1991). Double cones as a basis for a new type of polarization vision in vertebrates. Nature, 353, 161-164.

Carlisle, D. B., & Denton, E. J. (1959). On the metamorphosis of the visual pigments of *Anguilla anguilla* (L.). Journal of the Marine Biological Association of the UK, 38, 97-102.

Chang, B. S. W., Crandall, K. A., Carulli, J. P., & Hartl, D. L. (1995). Opsin phylogeny and evolution: a model for blue shifts in wavelength regulation. Molecular Phylogenetics and Evolution, 4, 31-43.

Cheng, C. L., & Flamarique, I. N. (2007). Chromatic organization of cone photoreceptors in the retina of rainbow trout: single cones irreversibly switch from UV (SWS1) to blue (SWS2) light sensitive opsin during natural development. The Journal of Experimental Biology, 210, 4123-4135.

Cheng, C. L., Flamarique, I. N., Hárosi, F. I., Rickers-Haunerland, J., & Haunerland, N. H. (2006). Photoreceptor layer of salmonid fishes: transformation and loss of single cones in juvenile fish. The Journal of Comparative Neurology, 495, 213-235.

Chinen, A., Hamaoka, T., Yamada, Y., & Kawamura, S. (2003). Gene duplication

and spectral diversification of cone visual pigments of zebrafish. Genetics, 163, 663-675.

Chinen, A., Matsumoto, Y., & Kawamura, S. (2005a). Reconstitution of ancestral green visual pigments of zebrafish and molecular mechanism of their spectral differentiation. Molecular Biology and Evolution, 22, 1001-1010.

Chinen, A., Matsumoto, Y., & Kawamura, S. (2005b). Spectral differentiation of blue opsins between phylogenetically close but ecologically distant goldfish and zebrafish. The Journal of Biological Chemistry, 280, 9460-9466.

Collin, S. P., & Trezise, A. E. O. (2004). The origins of colour vision in vertebrates. Clinical and Experimental Optometry, 87, 217-223.

Crescitelli, F. (1990). Adaptations of visual pigments to the photic environment of the deep sea. Journal of Experimental Zoology, 256, 66-75.

Davies, W. L., Carvalho, L. S., Tay, B. H., Brenner, S., Hunt, D. M., & Venkatesh, B. (2009). Into the blue: gene duplication and loss underlie color vision adaptations in a deep-sea chimaera, the elephant shark *Callorhinchus milii*. Genome Research, 19, 415-426.

Davies, W. L., Hankins, M. W., & Foster, R. G. (2010). Vertebrate ancient opsin and melanopsin: divergent irradiance detectors. Photochemical & Photobiological Sciences, Epub ahead of print: DOI: 10.1039/c0pp00203h.

Davies-Colley, R. J., Vant, W. N., & Wilcock, R. J. (1988). Lake water color: comparison of direct observations with underwater spectral irradiance. JAWRA Journal of the American Water Resources Association, 24, 11-18.

Denton, E. J., Gilpin-Brown, J. B., & Wright, P. G. (1970). On the 'filters' in the photophores of mesopelagic fish and on a fish emitting red light and especially sensitive to red light. The Journal of Physiology, 208, 72P-73P.

Denton, E. J., Herring, P. J., Widder, E. A., Latz, M. F., & Case, J. F. (1985). The roles of filters in the photophores of oceanic animals and their relation to vision in the oceanic environment. Proceedings of the Royal Society of London. Series B, Biological Sciences, 225, 63-97.

Farmer, C. T., Moore, C. A., Zika, R. G., & Sikorski, R. J. (1993). Effects of low and high Orinoco River flow on the underwater light field of the Eastern Caribbean Basin. Journal of Geophysical Research, 98, 2279-2288.

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using bootstrap. Evolution, 39 783-791

Force, A., Lynch, M., Pickett, F. B., Amores, A., Yan, Y. L., & Postlethwait, J. (1999). Preservation of duplicate genes by complementary, degenerative mutations. Genetics, 151, 1531-1545.

Frederiksen, R. D. (1973). On the retinal diverticula in the tubular-eyed opisthoproctid deep-sea fishes *Macropinna microstoma* and *Dolichopteryx longipes*. Videnskabelige Meddelelser fra Dansk Naturhistorisk Forening, 136, 233-244.

Fuller, R. C., Carleton, K. L., Fadool, J. M., Spady, T. C., & Travis, J. (2004). Population variation in opsin expression in the bluefin killifish, *Lucania goodei*: a real-time PCR study. Journal of Comparative Physiology. A, Neuroethology, Sensory, Neural, and Behavioral Physiology, 190, 147-154.

Fuller, R. C., Carleton, K. L., Fadool, J. M., Spady, T. C., & Travis, J. (2005). Genetic and environmental variation in the visual properties of bluefin killifish, *Lucania goodei*. Journal of Evolutionary Biology, 18, 516-523.

Gojobori, J., & Innan, H. (2009). Potential of fish opsin gene duplications to evolve new adaptive functions. Trends in Genetics, 25, 198-202.

Guindon, S., & Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Systematic Biology, 52, 696-704.

Hall, T. A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series, 41, 95-98.

Helvik, J. V., Drivenes Ø, Harboe, T., & Seo, H. C. (2001). Topography of different photoreceptor cell types in the larval retina of Atlantic halibut (*Hippoglossus hippoglossus*). The Journal of Experimental Biology, 204, 2553-2559.

Herring, P. J. (1996). Light, colour, and vision in the ocean, in: Summerhayes, C. P., Thorpe, S. A. Oceanography: An illustrated guide, Manson publishing, London, pp. 212-227.

Hisatomi, O., Satoh, T., & Tokunaga, F. (1997). The primary structure and distribution of killifish visual pigments. Vision Research, 37, 3089-3096.

Hoffmann, M., Tripathi, N., Henz, S. R., Lindholm, A. K., Weigel, D., Breden, F.,

& Dreyer, C. (2007). Opsin gene duplication and diversification in the guppy, a model for sexual selection. Proceedings of the Royal Society of London. Series B, Biological Sciences, 274, 33-42.

Hofmann, C. M., & Carleton, K. L. (2009). Gene duplication and differential gene expression play an important role in the diversification of visual pigments in fish. Integrative and Comparative Biology, 49, 630-643.

Hofmann, C. M., O'Quin, K. E., Marshall, N. J., Cronin, T. W., Seehausen, O., & Carleton, K. L. (2009). The eyes have it: regulatory and structural changes both underlie cichlid visual pigment diversity. PLoS biology, 7, e1000266.

Hoke, K. L., Evans, B. I., & Fernald, R. D. (2006). Remodeling of the cone photoreceptor mosaic during metamorphosis of flounder (*Pseudopleuronectes americanus*). Brain, Behavior and Evolution, 68, 241-254.

Hope, A. J., Partridge, J. C., & Hayes, P. K. (1998). Switch in rod opsin gene expression in the European eel, *Anguilla anguilla* (L.). Proceedings of the Royal Society of London. Series B, Biological Sciences, 265, 869-874.

Hordijk, W., & Gascuel, O. (2005). Improving the efficiency of SPR moves in phylogenetic tree search methods based on maximum likelihood. Bioinformatics, 21, 4338-4347.

Hughes, A. L., & Friedman, R. (2008). Codon-based tests of positive selection, branch lengths, and the evolution of mammalian immune system genes. Immunogenetics, 60, 495-506.

Hughes, A. L., & Nei, M. (1988). Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature, 335, 167-170.

Hunt, D. M., Fitzgibbon, J., Slobodyanyuk, S. J., & Bowmaker, J. K. (1996). Spectral tuning and molecular evolution of rod visual pigments in the species flock of cottoid fish in Lake Baikal. Vision Research, 36, 1217-1224.

Hurley, I. A., Mueller, R. L., Dunn, K. A., Schmidt, E. J., Friedman, M., Ho, R. K., Prince, V. E., Yang, Z., Thomas, M. G., & Coates, M. I. (2007). A new time-scale for ray-finned fish evolution. Proceedings of the Royal Society of London. Series B, Biological Sciences, 274, 489.

Jerlov, N. G. (1968). Optical oceanography. Elsevier, Amsterdam.

Jondeung, A., & Karinthanyakit, W. (2010). The complete mitochondrial DNA sequence of the short mackerel (*Rastrelliger brachysoma*), and its phylogenetic position within Scombroidei, Perciformes. Mitochondrial DNA, 21, 36-47.

Jukes, T. H., & Cantor, C. R. (1969). Evolution of protein molecules. Mammalian Protein Metabolism, 3, 21-132.

Kampa, E. M. (1970). Underwater daylight and moonlight measurements in the eastern North Atlantic. Journal of the Marine Biological Association of the UK, 50, 397-420.

Kasai, A., & Oshima, N. (2006). Light-sensitive motile iridophores and visual pigments in the neon tetra, *Paracheirodon innesi*. Zoological Science, 23, 815-819.

Klein, J., & Horejsi, V. (1997) Immunology, second ed. Wiley-Blackwell, Oxford.

Kloosterman, W. P., Wienholds, E., De Bruijn, E., Kauppinen, S., & Plasterk, R. H. A. (2005). *In situ* detection of miRNAs in animal embryos using LNA-modified oligonucleotide probes. Nature Methods, 3, 27-29.

Knowles, A., & Dartnall, H. J. A. (1977). The photobiology of vision. Academic press, New York.

Kochendoerfer, G. G., Lin, S. W., Sakmar, T. P., & Mathies, R. A. (1999). How color visual pigments are tuned. Trends in Biochemical Sciences, 24, 300-305.

Kolb, H., Nelson, R., Ahnelt, P., & Cuenca, N. (2001). Cellular organization of the vertebrate retina. Progress in Brain Research, 131, 3-26.

Kondrashov, F. A., Rogozin, I. B., Wolf, Y. I., & Koonin, E. V. (2002). Selection in the evolution of gene duplications. Genome Biology, 3, 8-1.

Kuffler, S. W. (1953). Discharge patterns and functional organization of mammalian retina. Journal of Neurophysiology, 16, 37-68.

Land, M. F. (2000). On the functions of double eyes in midwater animals. Proceedings of the Royal Society of London. Series B, Biological Sciences, 355, 1147-1150.

Lee, Y. H., & Vacquier, V. D. (1992). The divergence of species-specific abalone sperm lysins is promoted by positive Darwinian selection. The Biological Bulletin, 182,

97.

Lee, Y. H., Ota, T., & Vacquier, V. D. (1995). Positive selection is a general phenomenon in the evolution of abalone sperm lysin. Molecular Biology and Evolution, 12, 231-238.

Leonard, D. W., & Meek, K. M. (1997). Refractive indices of the collagen fibrils and extrafibrillar material of the corneal stroma. Biophysical Journal, 72, 1382-1387.

Levine, J. S., MacNichol, E. F., Kraft, T., & Collins, B. A. (1979). Intraretinal distribution of cone pigments in certain teleost fishes. Science, 204, 523-526.

Levinson, G., & Gutman, G. A. (1987). Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Molecular Biology and Evolution, 4, 203-221.

Li, Z., Gan, X., & He, S. (2009). Distinct evolutionary patterns between two duplicated color vision genes within cyprinid fishes. Journal of Molecular Evolution, 69, 346-359.

Locket, N. A. (1977). Adaptations to the deep-sea environment, in: Crescitelli, F. (Eds.) Handbook of sensory physiology. Sperlig-Verlag, Berlin, pp. 67-192.

Loew, E. R., & Lythgoe, J. N. (1978). The ecology of cone pigments in teleost fishes. Vision Research, 18, 715-722.

Lynch, M., & Conery, J. S. (2003). The evolutionary demography of duplicate genes. Journal of Structural and Functional Genomics, 3, 35-44.

Lynch, M., & Force, A. (2000). The probability of duplicate gene preservation by subfunctionalization. Genetics, 154, 459-473.

Lythgoe, J. N. (1972). The adaptation of visual pigments to the photic environment, in: Crescitelli, F. (Eds.) Handbook of Sensory Physiology. Sperlig-Verlag, Berlin, pp. 193-274.

Lythgoe, J. N. (1988). Light and vision in the aquatic environment, in: Atema, J., Fay, R. R., Popper, A. N., Tavolga, W. N. (Eds.) Sensory Biology of Aquatic Animals, Springer-Verlag, New York, pp. 57-82.

Lythgoe, J. N., & Partridge, J. C. (1989). Visual pigments and the acquisition of visual information. The Journal of Experimental Biology, 146, 1-20.

Mano, H., Kojima, D., & Fukada, Y. (1999). Exo-rhodopsin: a novel rhodopsin expressed in the zebrafish pineal gland. Brain research. Molecular Brain Research, 73,

110-118.

Matsumoto, Y., Fukamachi, S., Mitani, H., & Kawamura, S. (2006). Functional characterization of visual opsin repertoire in Medaka (*Oryzias latipes*). Gene, 371, 268-278.

Miller, R. G. (1981). Simultaneous statistical inference. Springer-Verlag, New York.

Minamoto, T., & Shimizu, I. (2005). Molecular cloning of cone opsin genes and their expression in the retina of a smelt, Ayu (*Plecoglossus altivelis*, Teleostei). Comparative Biochemistry and Physiology. Part B, Biochemistry & Molecular Biology, 140, 197-205.

Miyata, T., & Yasunaga, T. (1980). Molecular evolution of mRNA: a method for estimating evolutionary rates of synonymous and amino acid substitutions from homologous nucleotide sequences and its application. Journal of Molecular Evolution, 16, 23-36.

Miyazaki, T., Kohbara, J., Takii, K., Ishibashi, Y., & Kumai, H. (2008). Three cone opsin genes and cone cell arrangement in retina of juvenile Pacific bluefin tuna *Thunnus orientalis*. Fisheries Science, 74, 314-321.

Morris, B. J., Haarmann, I., Kempter, B., Höllt, V., & Herz, A. (1986). Localization of prodynorphin messenger RNA in rat brain by *in situ* hybridization using a synthetic oligonucleotide probe. Neuroscience Letters, 69, 104-108.

Munk, O. (1966). Ocular anatomy of some deep-sea teleosts. Dana Rep. 70, 1-62.

Munz, F. W., & McFarland, W. N. (1977). Evolutionary adaptations of fishes to the photic environment. Handbook of Sensory Physiology, 7, 193-274.

Nathans, J., & Hogness, D. S. (1983). Isolation, sequence analysis, and intron-exon arrangement of the gene encoding bovine rhodopsin. Cell, 34, 807-814.

Nathans, J., Piantanida, T. P., Eddy, R. L., Shows, T. B., & Hogness, D. S. (1986a). Molecular genetics of inherited variation in human color vision. Science, 232, 203-203.

Nathans, J., Thomas, D., & Hogness, D. S. (1986b). Molecular genetics of human color vision: the genes encoding blue, green, and red pigments. Science, 232, 193.

Nelson, J. S. (2006). Fishes of the World, 4th Edition. Wiley, New York.

Nicol, J. A., & Somiya, H. (1996). The eyes of fishes. Oxford University Press, Oxford.

Nielsen, R., & Yang, Z. (1998). Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics, 148, 929.

Nilsson, D. E. (2009). The evolution of eyes and visually guided behaviour. Proceedings of the Royal Society of London. Series B, Biological Sciences, 364, 2833-2847.

Nozawa, M., Suzuki, Y., & Nei, M. (2009). Reliabilities of identifying positive selection by the branch-site and the site-prediction methods. Proceedings of the National Academy of Sciences of the United States of America, 106, 6700-6705.

O'Day, W. T., & Fernandez, H. R. (1974). *Aristostomias scintillans* (Malacosteidae): A deep-sea fish with visual pigments apparently adapted to its own bioluminescence. Vision Research, 14, 545 - 550.

Oliveria, F. G., Coimbra, J. P., Yamada, E. S., Montag, L. F. D. E. A., Nascimento, F. L., Oliverira, V. A., Mota, D. L. D. A., Bittencourt, A. M., Silva, V. L D. A., & Costa, B. L. D. A. (2006). Topographic analysis of the ganglion cell layer in the retina of the four-eyed fish *Anableps anableps*. Visual Neuroscience, 23, 879-886.

Oprian, D. D., Asenjo, A. B., Lee, N., & Pelletier, S. L. (1991). Design, chemical synthesis, and expression of genes for the three human color vision pigments. Biochemistry, 30, 11367-11372.

Osorio, D., & Vorobyev, M. (2005). Photoreceptor spectral sensitivities in terrestrial animals: adaptations for luminance and colour vision. Proceedings of the Royal Society of London. Series B, Biological Sciences, 272, 1745-1752.

Owens, G. L., Windsor, D. J., Mui, J., & Taylor, J. S. (2009). A fish eye out of water: ten visual opsins in the four-eyed fish, *Anableps anableps*. PloS one, 4, e5970.

Pardue, M. L., & Gall, J. G. (1969). Molecular hybridization of radioactive DNA to the DNA of cytological preparations. Proceedings of the National Academy of Sciences of the United States of America, 64, 600-604.

Park, J. H., Scheerer, P., Hofmann, K. P., Choe, H. W., & Ernst, O. P. (2008). Crystal structure of the ligand-free G-protein-coupled receptor opsin. Nature, 454, 183-

187.

Parry, J. W., Carleton, K. L., Spady, T., Carboo, A., Hunt, D. M., & Bowmaker, J. K. (2005). Mix and match color vision: tuning spectral sensitivity by differential opsin gene expression in Lake Malawi cichlids. Current Biology, 15, 1734-1739.

Partridge, J. C., Archer, S. N., & Lythgoe, J. N. (1988). Visual pigments in the individual rods of deep-sea fishes. Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology, 162, 543-550.

Partridge, J. C., Shand, J., Archer, S. N., Lythgoe, J. N., & van Groningen-Luyben, W. A. (1989). Interspecific variation in the visual pigments of deep-sea fishes. Journal of Comparative Physiology. A, Neuroethology, Sensory, Neural, and Behavioral Physiology, 164, 513-529.

Patel, S., Marshall, J., & Fitzke 3rd, F. W. (1995). Refractive index of the human corneal epithelium and stroma. Journal of Refractive Surgery, 11, 100-105.

Pearcy, W. G., Meyer, S. L., & Munk, O. (1965). A 'four-eyed' fish from the deep-sea: *Bathylychnops exilis* Cohen, 1958. Nature, 207, 1260-1262.

Plachetzki, D. C., Fong, C. R., & Oakley, T. H. (2010). The evolution of phototransduction from an ancestral cyclic nucleotide gated pathway. Proceedings of the Royal Society of London. Series B, Biological Sciences, 277, 1963-1969.

Pointer, M. A., Cheng, C. H., Bowmaker, J. K., Parry, J. W., Soto, N., Jeffery, G., Cowing, J. A., & Hunt, D. M. (2005). Adaptations to an extreme environment: retinal organisation and spectral properties of photoreceptors in Antarctic notothenioid fish. The Journal of Experimental Biology, 208, 2363-2376.

Posada, D., & Crandall, K. A. (1998). MODELTEST: testing the model of DNA substitution. Bioinformatics, 14, 817-818.

Pride, D. T., (2000) SWAAP Version 1.0.3-Sliding windows alignment analysis program: a tool for analyzing patterns of substitutions and similarity in multiple alignments.

Raes, J., & Van de Peer, Y. (2003). Gene duplication, the evolution of novel gene functions, and detecting functional divergence of duplicates in silico. Applied Bioinformatics, 2, 91-101.

Reuter, T. E., White, R. H., & Wald, G. (1971). Rhodopsin and porphyropsin

fields in the adult bullfrog retina. The Journal of General Physiology, 58, 351-371.

Saidel, W. M., & Fabiane, R. S. (1998). Optomotor response of *Anableps anableps* depends on the field of view. Vision Research, 38, 2001-2006.

Saitou, N., & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular Biology and Evolution, 4, 406.

Scheerer, P., Park, J. H., Hildebrand, P. W., Kim, Y. J., Krauss, N., Choe, H. W., Hofmann, K. P., & Ernst, O. P. (2008). Crystal structure of opsin in its G-protein-interacting conformation. Nature, 455, 497-502.

Schultz, L. P., & Stern, E. M. (1948). The ways of fishes. D. Van Nostrand Co, New York.

Schwab, I. R., Ho, V., Roth, A., Blankenship, T. N., & Fitzgerald, P. G. (2001). Evolutionary attempts at 4 eyes in vertebrates. Transactions of the American Ophthalmological Society, 99, 145-56.

Schwassmann, H. O., & Kruger, L. (1965). Experimental analysis of the visual system of the four-eyed fish *Anableps microlepis*. Vision Research, 5, 269-281.

Shand, J., Davies, W. L., Thomas, N., Balmer, L., Cowing, J. A., Pointer, M., Carvalho, L. S., Trezise, A. E., Collin, S. P., Beazley, L. D., & Hunt, D. M. (2008). The influence of ontogeny and light environment on the expression of visual pigment opsins in the retina of the black bream, *Acanthopagrus butcheri*. The Journal of Experimental Biology, 211, 1495-1503.

Shi, Y., Radlwimmer, F. B., & Yokoyama, S. (2001). Molecular genetics and the evolution of ultraviolet vision in vertebrates. Proceedings of the National Academy of Sciences of the United States of America, 98, 11731.

Sivak, J. G. (1976). Optics of the eye of the "four-eyed fish" (*Anableps anableps*). Vision Research, 16, 531-534.

Spady, T. C. (2006) Cichlids as a model for the evolution of visual sensitivity. Durham, NH: University of New Hampshire.

Spady, T. C., Parry, J. W., Robinson, P. R., Hunt, D. M., Bowmaker, J. K., & Carleton, K. L. (2006). Evolution of the cichlid visual palette through ontogenetic subfunctionalization of the opsin gene arrays. Molecular Biology and Evolution, 23, 1538-1547.

Stenkemp, D. L. Neurogenesis in the fish retina. International Review of Cytology, 259, 173-224

Stenkamp, D. L., Hisatomi, O., Barthel, L. K., Tokunaga, F., & Raymond, P. A. (1996). Temporal expression of rod and cone opsins in embryonic goldfish retina predicts the spatial organization of the cone mosaic. Investigative Ophthalmology & Visual Science, 37, 363-376.

Swamynathan, S. K., Crawford, M. A., Robison, W. G., Kanungo, J., & Piatigorsky, J. (2003). Adaptive differences in the structure and macromolecular compositions of the air and water corneas of the "four-eyed" fish (*Anableps anableps*). The FASEB Journal, 17, 1996-2005.

Swofford, D. L. (2002). PAUP: phylogenetic analysis using parsimony, version 4.0 b10. Sunderland, MA: Sinauer Associates.

Takechi, M., & Kawamura, S. (2005). Temporal and spatial changes in the expression pattern of multiple red and green subtype opsin genes during zebrafish development. The Journal of Experimental Biology, 208, 1337-1345.

Takenaka, N., & Yokoyama, S. (2007). Mechanisms of spectral tuning in the RH2 pigments of Tokay gecko and American chameleon. Gene, 399, 26-32.

Tamura, K., & Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Molecular Biology and Evolution, 10, 512-526.

Tamura, K., Dudley, J., Nei, M., & Kumar, S. (2007). MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Molecular Biology and Evolution, 24, 1596.

Temple, S., Hart, N. S., Marshall, N. J., & Collin, S. P. (2010). A spitting image: specializations in archerfish eyes for vision at the interface between air and water. Proceedings of the Royal Society of London. Series B, Biological Sciences, Epub ahead of print, DOI: 10.1098/rspb.2010.0345.

Temple, S. E., Veldhoen, K. M., Phelan, J. T., Veldhoen, N. J., & Hawryshyn, C. W. (2008). Ontogenetic changes in photoreceptor opsin gene expression in coho salmon (*Oncorhynchus kisutch*, Walbaum). The Journal of Experimental Biology, 211, 3879-3888.

Terakita, A. (2005). The opsins. Genome Biology, 6, 213.

Tesch, F. W. (1977). The eel: biology and management of anguillid eels. Halsted Press, London.

Unger, E. R., Budgeon, L. R., Brigati, D. J., & Myerson, D. (1986). Viral diagnosis by *in situ* hybridization: description of a rapid simplified colorimetric method. The American Journal of Surgical Pathology, 10, 1-8.

Venkatesh, B., Ning, Y., & Brenner, S. (1999). Late changes in spliceosomal introns define clades in vertebrate evolution. Proceedings of the National Academy of Sciences of the United States of America, 96, 10267-10271.

Vihtelic, T. S., Doro, C. J., & Hyde, D. R. (1999). Cloning and characterization of six zebrafish photoreceptor opsin cDNAs and immunolocalization of their corresponding proteins. Visual Neuroscience, 16, 571-585.

Vollrath, D., Nathans, J., & Davis, R. W. (1988). Tandem array of human visual pigment genes at Xq28. Science, 240, 1669.

Wagner, H. J. (1990). Retinal structure of fishes, in: Djamgoz, M., (Eds.) The Visual System of Fish. Chapman and Hall Ltd., London, 109-157.

Wagner, H. J., Douglas, R. H., Frank, T. M., Roberts, N. W., & Partridge, J. C. (2009). A novel vertebrate eye using both refractive and reflective optics. Current Biology, 19, 108-114.

Walls, G. L. (1967). The vertebrate eye and its adaptive radiation. Hafner, New York; London.

Ward, M. N., Churcher, A. M., Dick, K. J., Laver, C. R. J., Owens, G. L., Polack, M. D., Ward, P. R., Breden, F., & Taylor, J. S. (2008). The molecular basis of color vision in colorful fish: Four Long Wave-Sensitive (LWS) opsins in guppies (*Poecilia reticulata*) are defined by amino acid substitutions at key functional sites. BMC Evolutionary Biology, 8, 210.

Watson, C. T., Lubieniecki, K. P., Loew, E., Davidson, W. S., & Breden, F. (2010). Genomic organization of duplicated short wave-sensitive and long wave-sensitive opsin genes in the green swordtail, *Xiphophorus helleri*. BMC Evolutionary Biology, 10, 87.

Weadick, C. J., & Chang, B. S. (2007). Long-wavelength sensitive visual

pigments of the guppy (*Poecilia reticulata*): six opsins expressed in a single individual. BMC Evolutionary Biology, 7, S11.

Wheeler, T. G. (1982). Color vision and retinal chromatic information processing in teleost: a review. Brain Research Reviews, 4, 177-235.

Windsor, D. J., & Owens, G. L. (2009). The opsin repertoire of *Jenynsia onca*: a new perspective on gene duplication and divergence in livebearers. BMC Research Notes, 2, 159.

Wong, W. S., Yang, Z., Goldman, N., & Nielsen, R. (2004). Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. Genetics, 168, 1041-1051.

Wood, P., & Partridge, J. C. (1993). Opsin substitution induced in retinal rods of the eel (*Anguilla anguilla* (L.)): a model for G-protein-linked receptors. Proceedings of the Royal Society of London. Series B, Biological Sciences, 254, 227-232.

Yang, L., Mayden, R. L., Sado, T., He, S., Saitoh, K., & Miya, M. (2010). Molecular phylogeny of the fishes traditionally referred to Cyprinini *sensu stricto* (Teleostei: Cypriniformes). Zoologica Scripta, Epub ahead of print, DOI: 10.1111/j.1463-6409.2010.00443.x .

Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. Molecular Biology and Evolution, 24, 1586.

Yang, Z., & Nielsen, R. (1998). Synonymous and nonsynonymous rate variation in nuclear genes of mammals. Journal of Molecular Evolution, 46, 409-418.

Yang, Z., & Nielsen, R. (2002). Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. Molecular biology and evolution, 19, 908-917.

Yang, Z., Nielsen, R., Goldman, N., & Pedersen, A. M. (2000a). Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics, 155, 431-449.

Yang, Z., Swanson, W. J., & Vacquier, V. D. (2000b). Maximum-likelihood analysis of molecular adaptation in abalone sperm lysin reveals variable selective pressures among lineages and sites. Molecular Biology and Evolution, 17, 1446-1455.

Yang, Z., Wong, W. S., & Nielsen, R. (2005). Bayes empirical bayes inference of

amino acid sites under positive selection. Molecular Biology and Evolution, 22, 1107-1118.

Yokoyama, R., & Yokoyama, S. (1990). Convergent evolution of the red- and green-like visual pigment genes in fish, Astyanax fasciatus, and human. Proceedings of the National Academy of Sciences of the United States of America, 87, 9315-9318.

Yokoyama, S. (1994). Gene duplications and evolution of the short wavelength-sensitive visual pigments in vertebrates. Molecular Biology and Evolution, 11, 32-39.

Yokoyama, S. (2000). Molecular evolution of vertebrate visual pigments. Progress in Retinal and Eye Research, 19, 385-420.

Yokoyama, S. (2008). Evolution of dim-light and color vision pigments. Annual review of Genomics and Human Genetics, 9, 259-282.

Yokoyama, S., & Radlwimmer, F. B. (2001). The molecular genetics and evolution of red and green color vision in vertebrates. Genetics, 158, 1697.

Yokoyama, S., & Tada, T. (2003). The spectral tuning in the short wavelength-sensitive type 2 pigments. Gene, 306, 91-98.

Yokoyama, S., & Tada, T. (2010). Evolutionary dynamics of rhodopsin type 2 opsins in vertebrates. Molecular Biology and Evolution, 27, 133-141.

Yokoyama, S., Radlwimmer, F. B., & Kawamura, S. (1998). Regeneration of ultraviolet pigments of vertebrates. FEBS Letters, 423, 155-158.

Yokoyama, S., Tada, T., Zhang, H., & Britt, L. (2008). Elucidation of phenotypic adaptations: Molecular analyses of dim-light vision proteins in vertebrates. Proceedings of the National Academy of Sciences of the United States of America, 105, 13480-13485.

Yokoyama, S., Takenaka, N., & Blow, N. (2007). A novel spectral tuning in the short wavelength-sensitive (SWS1 and SWS2) pigments of bluefin killifish (*Lucania goodei*). Gene, 396, 196-202.

Young, J. M., & Trask, B. J. (2002). The sense of smell: genomics of vertebrate odorant receptors. Human Molecular Genetics, 11, 1153.

Yu, W. P., Brenner, S., & Venkatesh, B. (2003). Duplication, degeneration and subfunctionalization of the nested synapsin-Timp genes in Fugu. Trends in Genetics, 19, 180-183.

Zhang, J., Nielsen, R., & Yang, Z. (2005). Evaluation of an improved branch-site

likelihood method for detecting positive selection at the molecular level. Molecular Biology and Evolution, 22, 2472-2479.

Zhang, J., Rosenberg, H. F., & Nei, M. (1998). Positive Darwinian selection after gene duplication in primate ribonuclease genes. Proceedings of the National Academy of Sciences of the United States of America, 95, 3708-3713.

Zhao, H., Ru, B., Teeling, E. C., Faulkes, C. G., Zhang, S., & Rossiter, S. J. (2009). Rhodopsin molecular evolution in mammals inhabiting low light environments. PloS One, 4, e8326.

Scientific Name	Common Name	SWS1	SWS2	RH1	RH2	LWS
Actinopterygians						
Acanthopagrus berda	Dark-finned black porgy	X	X	*	*	*
Acanthopagrus butcheri	Southern black bream	DQ354579.1	X	DQ354577.1	**EU090914.1, EU090913.1**	DQ354578.1
Acanthopagrus schlegeli	Black porgy	X	X	*	X	*
Acipenser sp.	Sturgeon	X	X	AF137206.1	X	X
Amia calva	Bowfin	X	X	AF137208.1	X	X
Anableps anableps	Four-eyed fish	FJ711153.1	FJ711151.1, HM627004	FJ711149.1	FJ711149.1, FJ711150.1	FJ711158.1, FJ711157.1, FJ711155.1, FJ711154.1
Anguilla anguilla	European eel	X	**FJ515779.1**	**L78007.1, L78008.1**	**FJ515778.1**	X
Anguilla japonica	Japanese eel	X	X	**AJ249203.1, AJ249202.1**	X	X
Astatotilapia burtoni	Tanganyika cichlid	X	**AY660538.1**	**AF315354.1**	AY660539.1	**AY660540.1**
Astyanax fasciatus	Blind cave fish	X	**AH007939.1**	**U12328.1**	X	**M90075.1, U12025.1, U12024.1**
Batrachocottus nikolskii	Fat sculpin	X	**AJ430474.1**	U97268.1	X	X
Candidia barbatus	Lake Candidus dace	**EU410458.1**	**EU410459.1**	EU919559.1	**EU410461.1, EU410460.1**	**EU410457.1**
Carassius auratus	Goldfish	**D85863.1**	**L11864.1**	**L11863.1**	**L11866.1, L11865.1**	**L11867.1, GQ168789.1**
Chrysiptera parasema	Yellowtail damselfish	X	X	X	X	DQ073800.1
Conger conger	Conger eel	X	X	**S82619.1**	X	X
Cottocomephorus inermis	Longfin Baikal sculpin	X	**AJ430479.1**	U97266.1	X	X
Cottus kessleri	Kessler's sculpin	X	**AJ430484.1**	L42953.1	X	X
Cyprinus carpio	Common carp	**AB113669.1**	**AB113668.1**	**S74449.1, Z71999.1**	**AB110603.1, AB110602.2**	**AB055656.1**
Danio rerio	Zebrafish	**BC060894.1**	**NM_131192.1**	**BC045288.1, XM_001336009.2 (NM_131212.2)**	**NM_131253.2, NM_182891.2, NM_182892.2, NM_131254.1**	**NM_001002443.1, NM_131175.1**
Dissostichus mawsoni	Antarctic cod	**AY927651.1**	X	DQ498794.1	**AY771352.1**	X
Gadus morhua	Atlantic cod	X	**AF385822.1**	**AF385832.1**	**AF385824.1**	X
Gasterosteus aculeatus	Threespine stickleback	**BT028514.1**	**BT027452.1**	**BT028623.1**	**BT027598.1, ENSGACT00000001859**	**BT027981.1**
Girella punctata	Largescale blackfish	X	AB158256.1	X	X	AB158261.1
Gymnodraco acuticeps	Ploughfish	**AY927656.1**	X	X	**AY771355.1**	X
Hippoglossus hippoglossus	Atlantic halibut	**AF156264.1**	**AF316497.1**	**AF156265.1**	**AF156263.1**	**AF316498.1**
Ictalurus punctatus	Channel catfish	X	X	AF028016.1	X	X
Jenynsia onca	One-sided livebearer	X	GQ221673.1	GQ221670.1	GQ221668.1, GQ221669.1	GQ221677.1, GQ221675.1, GQ221671.1
Jordanella floridae	American flagfish	X	X	X	X	HM627007, HM627009
Lepidopus fitchi	Pacific scabbardfish	**FJ443126.1**	X	**EU407252.1, EU407253.1**	**GQ414752.1, GQ414753.1, GQ414754.1,** GQ414755.1	X

Scientific Name	Common Name	SWS1	SWS2	RH1	RH2	LWS
Limnocottus eurystomus	Deep-water sculpin	X	**AJ430469.1**	X	X	X
Lucania goodei	Bluefin killifish	**AY296735.1**	**AY296737.1, AY296736.1**	**AY296738.1**	**AY296739.1**	**AY296740.1**, AY296741.1
Melanochromis vermivorus	Purple Mbuna	**DQ088643.1**	**DQ088637.1, DQ088640.1**	**GQ422472.1**	**DQ088634.1, DQ088646.1, DQ088631**	**DQ088628.1**
Melanotaenia australis	Rainbowfish	X	FJ940705.1	FJ940704.1	FJ940703.1	FJ940702.1
Metriaclima zebra	Malawi cichlid	DQ088648.1	**AF247118.1, AF247114.1**	**AY775114.1**	**DQ088650.1, DQ088651.1, DQ088652**	**AF247126.1**
Mullus surmuletus	Striped red mullet	X	X	**Y18666.1**	**Y18680.1**	X
Nannostomus beckfordi	Golden pencilfish	X	X	X	X	HM627006
Notothenia angustata	Maori chief	**AY927654.1**	X	DQ498787.1	AY771354.1	X
Oncorhynchus gorbuscha	Pink salmon	**AY214153.1**	**AY214154.1**	**AY214151.1**	**AY214152.1**	**AY214150.1**
Oncorhynchus keta	Chum salmon	**AY214143.1**	**AY214144.1**	**AY214141.1**	**AY214142.1**	**AY214140.1**
Oncorhynchus kisutch	Coho salmon	**AY214148.1**	**AY214149.1**	**AY214146.1**	**AY214147.1, DQ309027**	**AY214145.1**
Oncorhynchus mykiss	Rainbow trout	**NM_001124321.1**	**NM_001124322.1**	**NM_001124319.1**	**NM_001124323.1**	**NM_001124320.1**
Oncorhynchus nerka	Sockeye salmon	**AY214158.1**	**AY214159.1**	**AY214156.1**	**AY214157.1**	**AY214155.1**
Oncorhynchus tshawytscha	Chinook salmon	**AY214138.1**	**AY214139.1**	**AY214136.1**	**AY214137.1**	**AY214135.1**
Oreochromis niloticus	Nile tilapia	**AF191221.1**	**AF247116.1, AF247120.1**	AB084938.1	**DQ235683.1, DQ235681.1, DQ235682.1**	**AF247128.1**
Oryzias latipes	Medaka	**AB223058.1**	**AB223056.1, AB223057.1**	**AB180742.1, (ENSORLG00000010979)**	**AB223053.1, AB223054.1. AB223055.1**	AB223052.1, AB223051.1
Pagetopsis macropterus	Crocodile icefish	**AY927655.1**	X	EU637990.1	X	X
Pagothenia borchgrevinki	Bald notothen	**AY927653.1**	X	DQ498791.1	**AY771353.1**	X
Pagrus major	Red seabream	X	X	*	*	*
Paracheirodon innesi	Neon tetra	X	X	**AB245433.1**	X	AB239250.1, AB239249.1, AB545021.1
Plecoglossus altivelis	Ayu	**AB098706.1, AB098705.1**	X	**AB074484.1 (AB089247.1)**	**AB098703.1 AB098704.1**	**AB098702.1, AB107771.1**
Poecilia reticulata	Endler's guppy	**DQ234861.1**	**FJ711159.1, DQ234860.1**	**DQ912024.1**	**DQ234859.1, DQ234858.1**	**EU329457.1, EU329453.1, HM627005, EU329433.1**
Polyodon spathula	American paddlefish	X	X	AF369050.1	X	X
Pomatoschistus minutus	Sand goby	X	X	**FN430594.1**	**Y18679.1**	X
Pseudopleuronectes americanus	Winter flounder	X	**AY631038.1**	**AY631036.1**	**AY631037.1**	**AY631039.1**
Pseudotropheus acei	Yellow-tail acei	**DQ088642.1**	**DQ088636.1, DQ088639.1**	**GQ422475.1**	**DQ088645.1, DQ088633, DQ088630**	**DQ088627.1**
Pundamilia pundamilia	Victoria cichlid	**AY673729.1**	**AY673719.1, AY673709.1**	**AY673739.1**	**AY673699.1**	AB448425.1
Salmo salar	Atlantic salmon	**NM_001123708.1**	**NM_001123706.1**	**AF201470 (NM_001123536.1)**	**NM_001123707.1**	**NM_001123705.1**
Scopelarchus analis	Short fin pearleye	X	X	**EF517404.1, EF517405.1**	**EF517406.1**	X
Scophthalmus maximus	Turbot	**AF385825.1**	X	X	AF385828.1, AF385827.1	**AF385826.1**

Scientific Name	Common Name	SWS1	SWS2	RH1	RH2	LWS
Sinocyclocheilus anophthalmus	Eyeless golden-line fish	X	GQ168759.1	EU606007.1	X	GQ168786.1 GQ168787.1
Sinocyclocheilus jii	Gara Jiova	X	GQ168748.1, GQ168749.1	EU605989.1	X	GQ168768.1
Stenobrachius leucopsarus	Northern lampfish	**FJ443127.1**	X	**EU407251.1**	**GQ414753.1, GQ414754.1, GQ414756.1**	X
Takifugu rubripes	Japanese pufferfish	X	**AY598947.1**	**NM_001078631.1 (NM_001033849.1)**	**NM_001033712.1**	**AY598942.1**
Tetraodon nigroviridus	Green spotted puffer	X	**AY598948.1**	**AJ293018.1, (ENSTNIG00000017925)**	**AY598944.1**	**AY598943.1**
Thunnus orientalis	Pacific bluefin tuna	X	**AB290450.1**	**AB290449.1**	AB290451.1	X
Tramitichromis intermedius	Malawi cichlid	DQ088644.1	**DQ088638.1, DQ088641.1**	**GQ422473.1**	**DQ088635.1, DQ088647, DQ088632.1**	**DQ088629.1**
Trematomus loennbergii	Deepwater notothn	**AY927652.1**	AY771356.1	X	X	X
Xiphophorus birchmanni	Sheepshead swordtail	EU825689.1	EU825682.1	X	X	EU825688.1
Xiphophorus malinche	Highland swordtail	EU825684.1	EU825683.1	X	X	EU825686.1
Xiphophorus pygmaeus	Pygmy swordtail	X	X	X	X	EU329480.1, EU329478.1
Xiphopohorus helleri	Green swordtail	X	GQ999832	X	X	**GQ999832, GQ999833**
Zacco pachycephalus	Thickhead chub	**EU410468.1**	**EU410467.1**	EU919560.1	**EU410463.1, EU410464.1,** EU410465.1	**EU410466.1**
Sarcopterygians						
Anolis carolinensis	American chameleon	AH007736.1	AF133907.1	L31503.1	AH007735.1	U08131.1
Gallus gallus	Chicken	NM_205438.1	NM_205517.1	D00702.1	NM_205490.1	X57490.1
Homo sapiens	Human	NM_001708.2	X	U49742.1	X	NM_020061.4, NM_001048181.2
Latimeria chalumnae	West Indian Ocean coelacanth	X	X	**AH007712.1**	**AH007713.1**	X
Loxodonta africana	African savanna elephant	AY686753.1	X	AY686752.1	X	AY686754.1
Neoceratodus forsteri	Queensland lungfish	**EF526298.1**	**EF526299.1**	**EF526295.1**	**EF526296.1**	**EF526297.1**
Omithorhynchus anatinus	Platypus	X	EU624412.1	NM_001127627.1	X	NM_001127625.1
Serinus canaria	Common canary	AJ277922.1	AJ277923.1	AJ277926.1	AJ277924.1	AJ277925.1
Sminthopsis crassicaudata	Fat-tailed dunnart	AY442173.1	X	AY159786	X	EU232013.1
Taeniopygia guttata	Zebrafinch	NM_001076704	NM_001076697	NM_001076695	NM_001076696.1	NM_001076702
Xenopus laevis	African clawed toad	U23463.1	BC080123.1	L07770.1	X	U90895.1
Lamprey						
Geotria australis	Pouched lamprey	**AY366495.1**	**AY366492.1**	**AY366493.1**	**AY366494.1**	**AY366491.1**
Lethenteron japonicum	Arctic lamprey	X	X	**AB116382.1**	X	**AB116381.1**
Petromyzon marinus	Sea lamprey	X	X	**AH005459**	X	**EU571209.1**

Non-visual opsins			
Carassius auratus	Goldfish	VA opsin	AB383149.1
Danio rerio	Zebrafish	VA opsin	NM_131586.2
Gallus gallus	Chicken	VA opsin	GQ280390.1
Oryzias latipes	Medaka	VA opsin	NM_001136515.1
Plecoglossus altivelis	Ayu	VA opsin	AB074483.1
Salmo salar	Atlantic salmon	VA opsin	NM_001123626.1
Xenopus laevis	African clawed toad	VA opsin	NM_001171892.1
Anolis carolinensis	American chameleon	Pinopsin	AH007737.1
Gallus gallus	Chicken	Pinopsin	NM_205409.1
Taeniopygia guttata	Zebrafinch	Pinopsin	XM_002198907.1
Invertebrate visual opsins			
Drosophila melanogaster	Common fruit fly	RH4	BT044485.1

Appendix 1: Accession numbers for all sequences used in analyses.

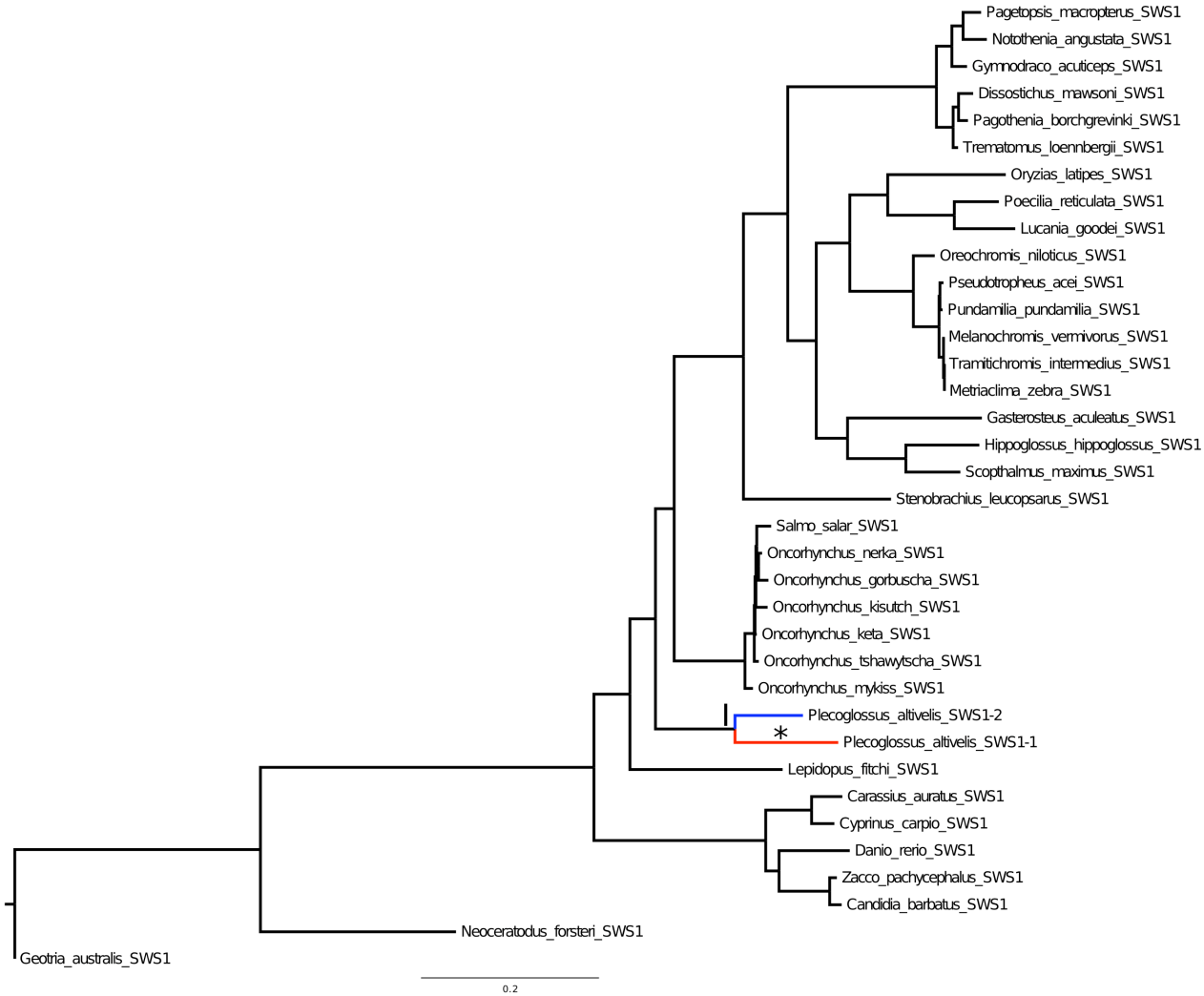
Bold numbers are those included in PAML analysis. Starred sequences are those from personal correspondence and lack accession numbers. Numbers in brackets are members of the exo-rhodopsin clade.

Subfamily	ID	Group A	Group B	All sites	1st & 2nd Sites	3rd Site
SWS1	I	PlAlSWS1-1	PlAlSWS1-2	0.166	0.108	0.305
SWS2	I	AnabSWS2A, AsBuSWS2, BaNiSWS2, CoInSWS2, CoKeSWS2, GaAcSWS2, HiHiSWS2, JeOnSWS2A, LiEuSWS2, LuGoSWS2A, MeVeSWS2A, MeZeSWS2A, OrLaSWS2A, OrNiSWS2A, PoReSWS2A, PsAcSWS2A, PsAmSWS2, PuPuSWS2A, TrInSWS2A, TrLoSWS2	AnabSWS2B, LuGoSWS2B, MeAuSWS2, MeVeSWS2B, MeZeSWS2B, OrLaSWS2B, OrNiSWS2B, PoReSWS2B, PsAcSWS2B, PuPuSWS2B, TaRuSWS2, TeNiSWS2, ThOrSWS2, TrInSWS2B, XiBiSWS2B, XiMaSWS2B	0.314	0.174	0.823
SWS2	II	CyCaSWS2, SiAnSWS2, SiJiSWS2_1	CaAuSWS2, SiJiSWS2_2	0.051	0.032	0.092
RH1	I	DaReExoRh, OrLaExoRh, PlAlExoRh, SaSaExoRh, TaRuExoRh, TeNiExoRh	AcBeRH1, AcBuRH1, AcScRH1, AcSpRH1, AmCaRH1, AnAnRH1F, AnAnRH1S, AnJaRH1F, AnJaRH1S, AnabRH1, AsBuRH1, AsFaRH1, BaNiRH1, CaAuRH1, CoCoRH1, CoInRH1, CoKeRH1, CyCaRH1_1, CyCaRH1_2, DaReRH1A, DaReRH1B, DiMaRH1, GaAcRH1, GaMoRH1, HiHiRH1, IcPuRH1, JeOnRH1, LeFiRH1A, LeFiRH1B, LuGoRH1, MeAuRH1, MeVeRH1, MeZeRH1, MuSuRH1, NoAnRH1, OnGoRH1, OnKeRH1, OnKiRH1, OnMyRH1, OnNeRH1, OnTsRH1, OrLaRH1, OrNiRH1, PaBoRH1, PaInRH1, PaMaRH1, PaMacRH1, PlAlRH1, PoMiRH1, PoReRH1, PoSpRH1, PsAcRH1, PsAmRH1, PuPuRH1, SaSaRH1, ScAnRH1A, ScAnRH1B, StLeRH1, TaRuRH1, TeNiRH1, ThOrRH1, TrInRH1	0.346	0.189	1.022
RH1	II	DaReRH1A	DaReRH1B	0.337	0.141	**n/c**
RH1	III	ScAnRH1A	ScAnRH1B	0.293	0.155	0.996
RH1	IV	AnAnRH1F, AnJaRH1F	AnAnRH1S, AnJaRH1S, CoCoRH1	0.194	0.078	0.645
RH1	V	LeFiRH1A	LeFiRH1B	0.107	0.069	0.206
RH1	VI	CyCaRH1_1	CyCaRH1_2	0.022	0.011	0.045
LWS	I	AsFaLWS1, PaInLWS3	AsFaLWS2, AsFaLWS3, NaBeLWS, PaInLWS1, PaInLWS2	0.316	0.163	0.830
LWS	II	AnabLWSa, AnabLWSb, AnabLWSg, JeOnLWSp, JeOnLWSs, LuGoLWS1, PoReLWSa, PoReLWSp, PoReLWSs, XiBiLWS, XiHeLWSp, XiHeLWSs1, XiHeLWSs2, XiMaLWS, XiPyLWSp, XiPyLWSs	AnabLWSr, JeOnLWSr, LuGoLWS2, PoReLWSr, XiHeLWSr	0.119	0.056	0.271
LWS	III	DaReLWS1	DaReLWS2	0.077	0.046	0.143
LWS	IV	AsFaLWS2, PaInLWS1	AsFaLWS3, PaInLWS2	0.072	0.043	0.138
LWS	V	PoReLWSp, XiHeLWSp, XiPyLWSp	PoReLWSa, PoReLWSs, XiBiLWS, XiHeLWSs1, XiHeLWSs2, XiMaLWS, XiPyLWSs	0.072	0.052	0.114
LWS	VI	CaAuLWS2, CyCaLWS, SiAnLWS2	CaAuLWS1, SiAnLWS1, SiJiLWS	0.053	0.029	0.105
LWS	VII	PlAlLWS1	PlAlLWS2	0.024	0.011	0.050
LWS	VIII	JoFlLWS1	JoFlLWS2	0.020	0.012	0.037
LWS	IX	PoReLWSa	PoReLWSs	0.013	0.008	0.023
LWS	X	OrLaLWS1	OrLaLWS2	0.012	0.007	0.023
LWS	XI	AnabLWSa	AnabLWSb	0.011	0.011	0.011
LWS	XII	XiHeLWSs1	XiHeLWSs2	0.001	0.001	0.000

Subfamily	ID	Group A	Group B	All sites	1st & 2nd Sites	3rd Site
RH2	I	AcBeRH2A, AcBuRH2Ab, AcBuRh2a, AnabRH2A, AsBuRH2, DiMaRH2, GaAcRH2A, GaAcRH2B, GaMoRH2, GyAcRH2, HiHiRH2, JeOnRH2A, LuGoRH2, MeAuRH2, MeVeRH2Aa, MeZeRH2Aa, MeZeRH2Ab, MuSuRH2, NoAnRH2, OrLaRH2B, OrLaRH2C, OrNiRH2Aa, OrNiRH2Ab, PaBoRH2, PoMiRH2, PoReRH2A, PsAcRH2Aa, PsAcRH2Ab, PsAmRH2, PuPuRH2, ScMaRH2A, ScMaRH2B, TaRuRH2A, TeNiRH2A, ThOrRH2A, TrInRH2Aa, TrInRH2Ab	AnabRH2B, JeOnRH2B, MeVeRH2B, MeZeRH2B, OrLaRH2A, OrNiRH2B, PoReRH2B, PsAcRH2B, ScAnRH2, StLeRH2A, StLeRH2B, StLeRH2D, TrInRH2B	0.298	0.169	0.706
RH2	II	DaReRH2_1, DaReRH2_2, ZaPaRH2C	CaAuRH2_1, CaAuRH2_2, CaBaRH2A, CaBaRH2B, CyCaRH2_1, CyCaRH2_2, DaReRH2_3, DaReRH2_4, ZaPaRH2A, ZaPaRH2B	0.272	0.115	0.772
RH2	III	ClHaRH2_1	ClHaRH2_2	0.250	0.107	0.690
RH2	IV	PlAlRH2A	PlAlRH2B	0.186	0.084	0.461
RH2	V	DaReRH2_1	DaReRH2_2, ZaPaRH2C	0.183	0.054	0.568
RH2	VI	ScMaRH2A	ScMaRH2B	0.147	0.099	0.269
RH2	VII	OnKiRH2B	OnGoRH2, OnKeRH2, OnKiRH2A, OnMyRH2, OnNeRH2, OnTsRH2, SaSaRH2	0.140	0.089	0.264
RH2	VIII	StLeRH2A, StLeRH2B	StLeRH2C, StLeRH2D	0.108	0.077	0.176
RH2	IX	DaReRH2_4	DaReRH2_3	0.078	0.044	0.152
RH2	X	CaAuRH2_2, CyCaRH2_2	CaAuRH2_1, CyCaRH2_1	0.076	0.032	0.173
RH2	XI	StLeRH2C	StLeRH2D	0.060	0.041	0.098
RH2	XII	ZaPaRH2A	ZaPaRH2B	0.051	0.029	0.098
RH2	XIII	OrLaRH2B	OrLaRH2C	0.039	0.032	0.054
RH2	XIV	MeVeRH2Aa, MeZeRH2Aa, OrNiRH2Aa, PsAcRH2Aa, TrInRH2Aa	AsBuRH2, MeVeRH2Ab, MeZeRH2Ab, OrNiRH2Ab, PsAcRH2Ab, PuPuRH2, TrInRH2Ab	0.037	0.024	0.063
RH2	XV	LeFiRH2A	LeFiRH2B, LeFiRH2C, LeFiRH2D	0.029	0.028	0.033
RH2	XVI	CaBaRH2A	CaBaRH2B	0.026	0.017	0.044
RH2	XVII	AcBuRH2Aa	AcBeRH2, AcBuRH2Ab	0.024	0.019	0.035
RH2	XVIII	LeFiRH2B	LeFiRH2C, LeFiRH2D	0.017	0.009	0.035
RH2	XIX	GaAcRH2A	GaAcRH2B	0.013	0.013	0.013
RH2	XX	StLeRH2A	StLeRH2B	0.012	0.012	0.012
RH2	XXI	LeFiRH2C	LeFiRH2D	0.011	0.011	0.009

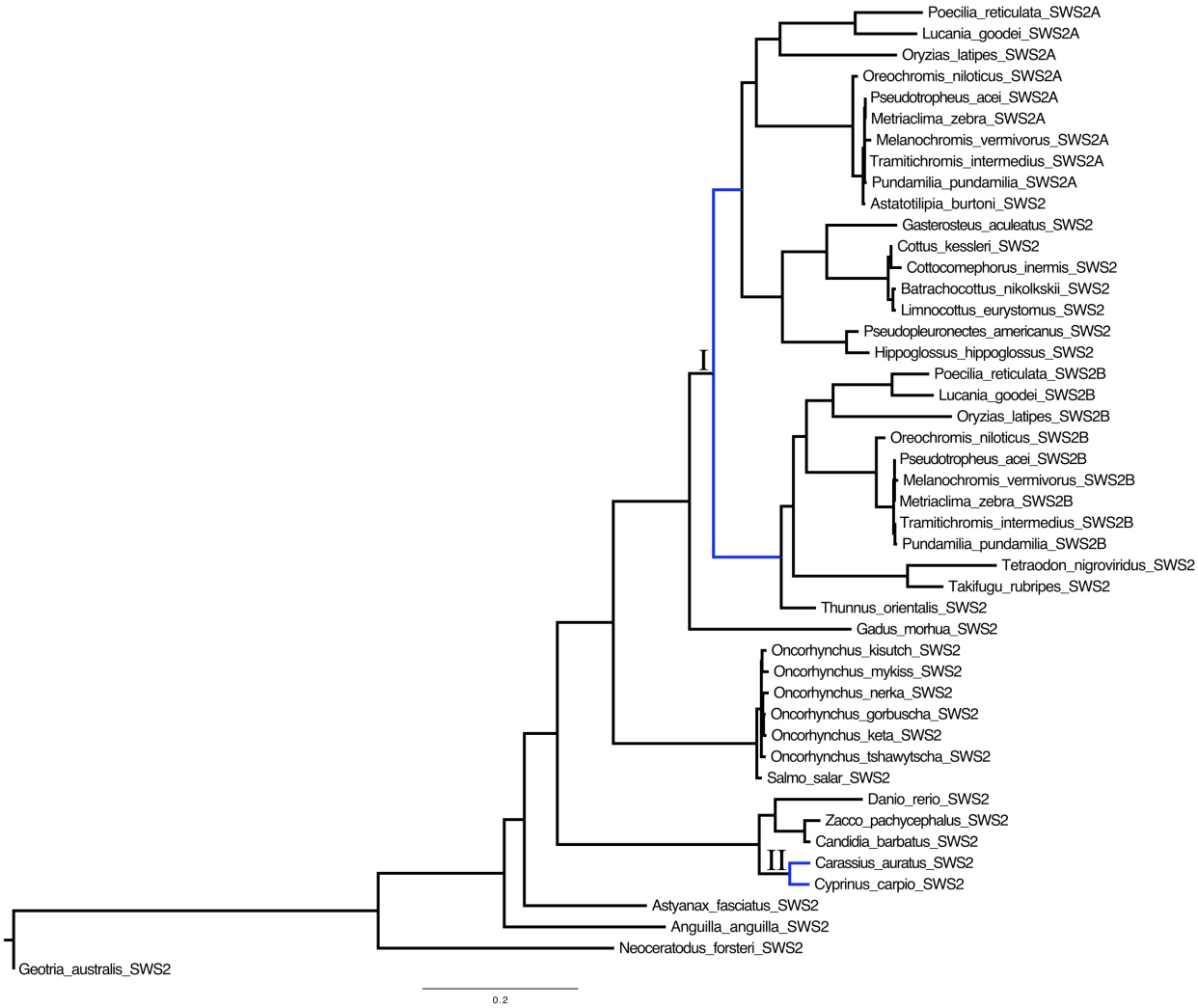
Appendix 2: Distance and ID of all identified opsin duplication events.

All opsin gene duplication events were given ID numbers from oldest to youngest using all sites distance. Distances were mean group distance measured using Tamura-Nei method in MEGA4 (Tamura et al., 2007; Tamura & Nei, 1993). Species names are abbreviated to the first two letters of genus and species, except for *A. anableps* which is abbreviated 'Anab'.



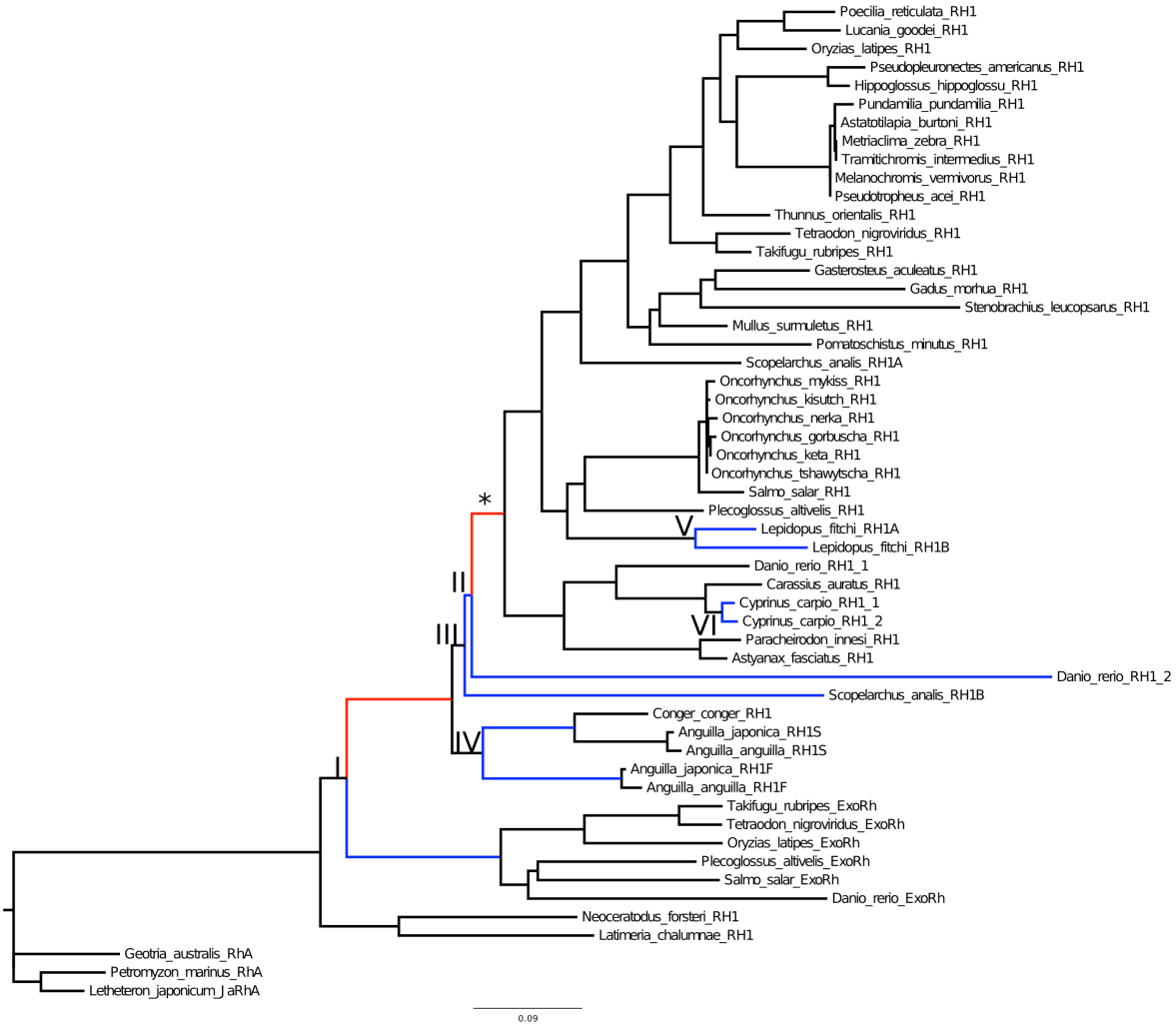
Appendix 3: Phylogenetic tree of SWS1 opsins in fish used for PAML analyses.

The tree was created using the maximum-likelihood method. Accession numbers are bolded and listed in Appendix 1. SWS1 opsin from lamprey (*G. australis*) was used as a root. PAUP was used to estimate genetic distances, based on modeltest's best-fit model of evolution, and complete phylogenetic analysis (Swofford, 2002; Posada & Crandall, 1998). Tree topology was tested using SPR. The model of evolution was determined to be HKY85+I+G (I = 0.2299, G = 1.1920). Gene duplication branches tested in PAML are labelled with roman numerals. Blue branches did not have sites evolving under positive selection, while red branches did. The star indicates that at least one site under positive selection is a key spectral tuning site on that branch.



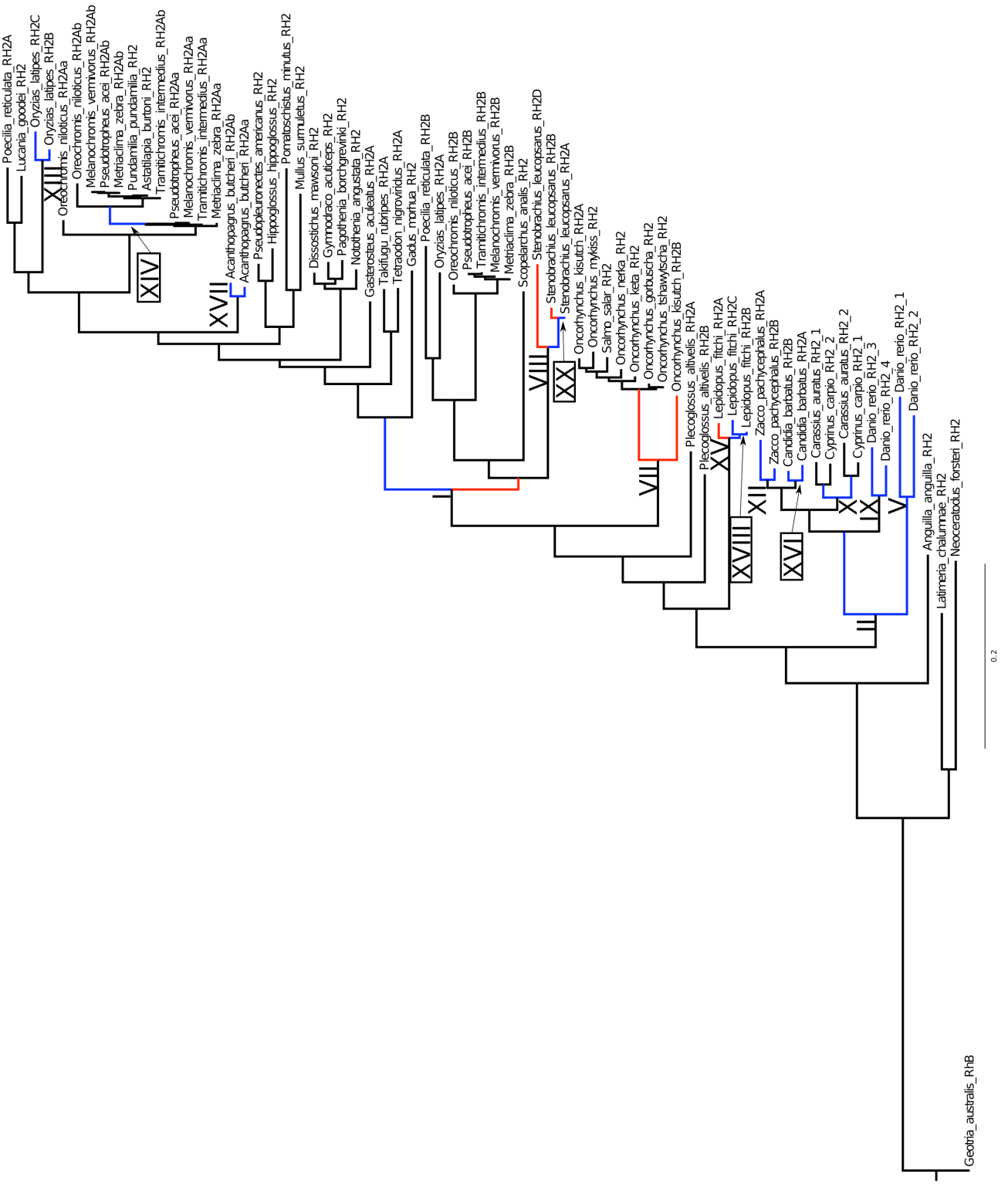
Appendix 4: Phylogenetic tree of SWS2 opsins in fish used for PAML analyses.

The tree was created using the maximum-likelihood method. Accession numbers are bolded and listed in Appendix 1. SWS2 opsin from lamprey (*G. australis*) was used as a root. PAUP was used to estimate genetic distances, based on modeltest's best-fit model of evolution, and complete phylogenetic analysis (Swofford, 2002; Posada & Crandall, 1998). Tree topology was tested using SPR. The model of evolution was determined to be HKY85+I+G (I = 0.2238, G = 1.1396). Gene duplication branches tested in PAML are labelled with roman numerals. Blue branches did not have sites evolving under positive selection.



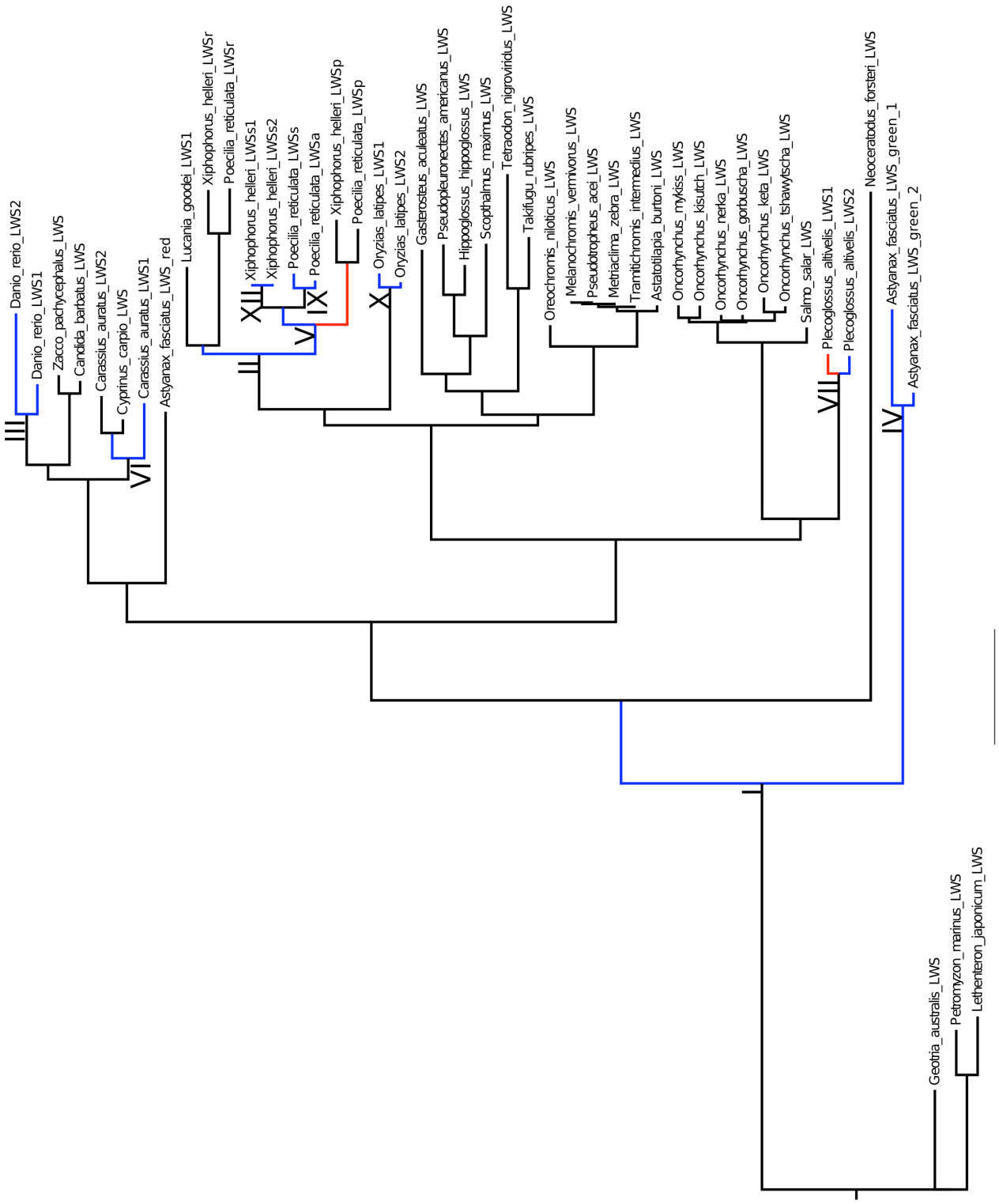
Appendix 5: Phylogenetic tree of RH1 opsins in fish used for PAML analyses.

The tree was created using the maximum-likelihood method. Accession numbers are bolded and listed in Appendix 1. RhA opsins from lamprey (*G. australis, P. marinus* and *L. japonicum*) were used as a root. PAUP was used to estimate genetic distances, based on modeltest's best-fit model of evolution, and complete phylogenetic analysis (Swofford, 2002; Posada & Crandall, 1998). Tree topology was tested using SPR. The model of evolution was determined to be GTR+I+G (I = 0.2746, G = 1.1396). Gene duplication branches tested in PAML are labelled with roman numerals. Blue branches did not have sites evolving under positive selection, while red branches did. The star indicates that at least one site under positive selection is a key spectral tuning site on that branch.


Poecilia_reticulata_RH2A
Lucania_goodei_RH2
Oryzias_latipes_RH2C
Oryzias_latipes_RH2B
Oreochromis_niloticus_RH2Aa
Oreochromis_niloticus_RH2Ab
Melanochromis_vermivorus_RH2Ab
Pseudotropheus_acei_RH2Ab
Metriaclima_zebra_RH2Ab
Pundamilia_pundamilia_RH2
Astatilapia_burtoni_RH2
Tramitichromis_intermedius_RH2Ab
Pseudotropheus_acei_RH2Aa
Melanochromis_vermivorus_RH2Aa
Tramitichromis_intermedius_RH2Aa
Metriaclima_zebra_RH2Aa
Acanthopagrus_butcheri_RH2Ab
Acanthopagrus_butcheri_RH2Aa
Pseudopleuronectes_americanus_RH2
Hippoglossus_hippoglossus_RH2
Pomatoschistus_minutus_RH2
Mullus_surmuletus_RH2
Dissostichus_mawsoni_RH2
Gymnodraco_acuticeps_RH2
Pagothenia_borchgrevinki_RH2
Notothenia_angustata_RH2
Gasterosteus_aculeatus_RH2A
Takifugu_rubripes_RH2A
Tetraodon_nigroviridus_RH2A
Gadus_morhua_RH2
Poecilia_reticulata_RH2B
Oryzias_latipes_RH2A
Oreochromis_niloticus_RH2B
Pseudotropheus_acei_RH2B
Tramitichromis_intermedius_RH2B
Melanochromis_vermivorus_RH2B
Metriaclima_zebra_RH2B
Scopelarchus_analis_RH2
Stenobrachius_leucopsarus_RH2D
Stenobrachius_leucopsarus_RH2B
Stenobrachius_leucopsarus_RH2A
Oncorhynchus_kisutch_RH2A
Salmo_salar_RH2
Oncorhynchus_mykiss_RH2
Oncorhynchus_nerka_RH2
Oncorhynchus_keta_RH2
Oncorhynchus_gorbuscha_RH2
Oncorhynchus_tshawytscha_RH2
Oncorhynchus_kisutch_RH2B
Plecoglossus_altivelis_RH2A
Plecoglossus_altivelis_RH2B
Lepidopus_fitchi_RH2A
Lepidopus_fitchi_RH2C
Lepidopus_fitchi_RH2B
Zacco_pachycephalus_RH2A
Zacco_pachycephalus_RH2B
Candidia_barbatus_RH2B
Candidia_barbatus_RH2A
Carassius_auratus_RH2_1
Cyprinus_carpio_RH2_2
Carassius_auratus_RH2_2
Cyprinus_carpio_RH2_1
Danio_rerio_RH2_3
Danio_rerio_RH2_4
Danio_rerio_RH2_1
Danio_rerio_RH2_2
Anguilla_anguilla_RH2
Latimeria_chalumnae_RH2
Neoceratodus_forsteri_RH2
Geotria_australis_RHB
0.2

Appendix 6: Phylogenetic tree of RH2 opsins in fish used for PAML analyses.

The tree was created using the maximum-likelihood method. Accession numbers are bolded and listed in Appendix 1. RhB opsin from lamprey (*G. australis*) was used as a root. PAUP was used to estimate genetic distances, based on modeltest's best-fit model of evolution, and complete phylogenetic analysis (Swofford, 2002; Posada & Crandall, 1998). Tree topology was tested using SPR. The model of evolution was determined to be GTR+I+G (I = 0.2485, G = 0.8147). Gene duplication branches tested in PAML are labelled with roman numerals. Blue branches did not have sites evolving under positive selection, while red branches did.


Danio_rerio_LWS2
Danio_rerio_LWS1
Zacco_pachycephalus_LWS
Candida_barbatus_LWS
Carassius_auratus_LWS2
Cyprinus_carpio_LWS
Carassius_auratus_LWS1
Astyanax_fasciatus_LWS_red
Lucania_goodei_LWS1
Xiphophorus_helleri_LWSr
Poecilia_reticulata_LWSr
Xiphophorus_helleri_LWSs1
Xiphophorus_helleri_LWSs2
Poecilia_reticulata_LWSs
Poecilia_reticulata_LWSa
Xiphophorus_helleri_LWSp
Poecilia_reticulata_LWSp
Oryzias_latipes_LWS1
Oryzias_latipes_LWS2
Gasterosteus_aculeatus_LWS
Pseudopleuronectes_americanus_LWS
Hippoglossus_hippoglossus_LWS
Scopthalmus_maximus_LWS
Tetraodon_nigroviridus_LWS
Takifugu_rubripes_LWS
Oreochromis_niloticus_LWS
Melanochromis_vermivorus_LWS
Pseudotropheus_acei_LWS
Metriaclima_zebra_LWS
Tramitichromis_intermedius_LWS
Astatotilapia_burtoni_LWS
Oncorhynchus_mykiss_LWS
Oncorhynchus_kisutch_LWS
Oncorhynchus_nerka_LWS
Oncorhynchus_gorbuscha_LWS
Oncorhynchus_keta_LWS
Oncorhynchus_tshawytscha_LWS
Salmo_salar_LWS
Plecoglossus_altivelis_LWS1
Plecoglossus_altivelis_LWS2
Neoceratodus_forsteri_LWS
Astyanax_fasciatus_LWS_green_1
Astyanax_fasciatus_LWS_green_2
Geotria_australis_LWS
Petromyzon_marinus_LWS
Lethenteron_japonicum_LWS
III
VI
XII
II
V
IX
X
VII
IV
0.08

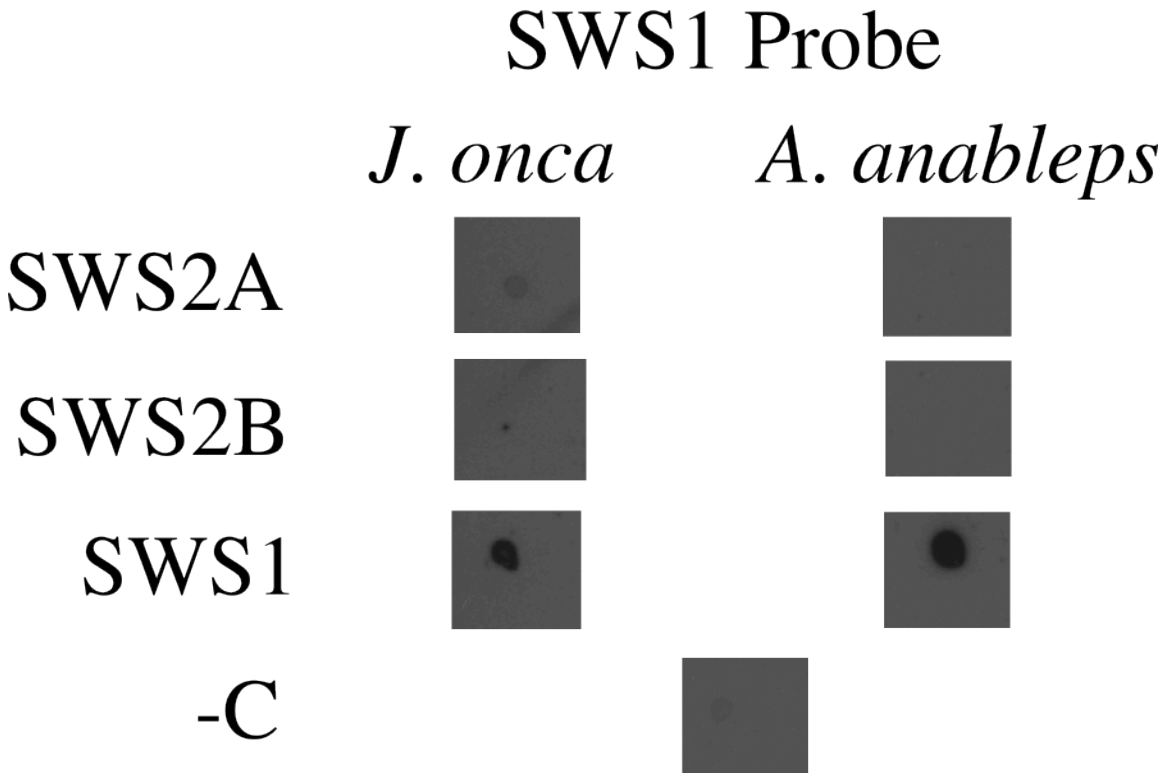
Appendix 7: Phylogenetic tree of LWS opsins in fish used for PAML analyses.

The tree was created using the maximum-likelihood method. Accession numbers are listed in Appendix 1. LWS opsins from lamprey (*G. australis, P. marinus* and *L. japonicum*) were used as a root. PAUP was used to estimate genetic distances, based on modeltest's best-fit model of evolution, and complete phylogenetic analysis (Swofford, 2002; Posada & Crandall, 1998). Tree topology was tested using SPR. The model of evolution was determined to be GTR+I+G (I = 0.3378, G = 1.2425). Gene duplication branches tested in PAML are labelled with roman numerals. Blue branches did not have sites evolving under positive selection, while red branches did.

		Probe					
		SWS1	SWS2A	SWS2B	RH2A	RH2B	LWS
Anableps anableps	SWS1	100*	-	-	-	-	-
	SWS2A	-	100*	61.9	-	-	-
	SWS2B	-	61.7	100*	-	-	-
	RH1	-	-	-	50.4	46.8	-
	RH2A	-	-	-	100*	65.1	-
	RH2B	-	-	-	64.2	100*	-
	LWS S180a	-	-	-	-	-	99.8*
	LWS S180b	-	-	-	-	-	98.7*
	LWS S180g	-	-	-	-	-	100*
	LWS S180r	-	-	-	-	-	92.7*
Jenynsia onca	SWS1	95.4*	-	-	-	-	-
	SWS2A	-	95.3*	60.6	-	-	-
	SWS2B	-	64.2	96.8*	-	-	-
	RH1	-	-	-	49.6	46.1	-
	RH2A	-	-	-	96.6*	67.1	-
	RH2B	-	-	-	63.8	95.9*	-
	LWS S180	-	-	-	-	-	95.8*
	LWS P180	-	-	-	-	-	95.3*
	LWS S180r	-	-	-	-	-	94.5*

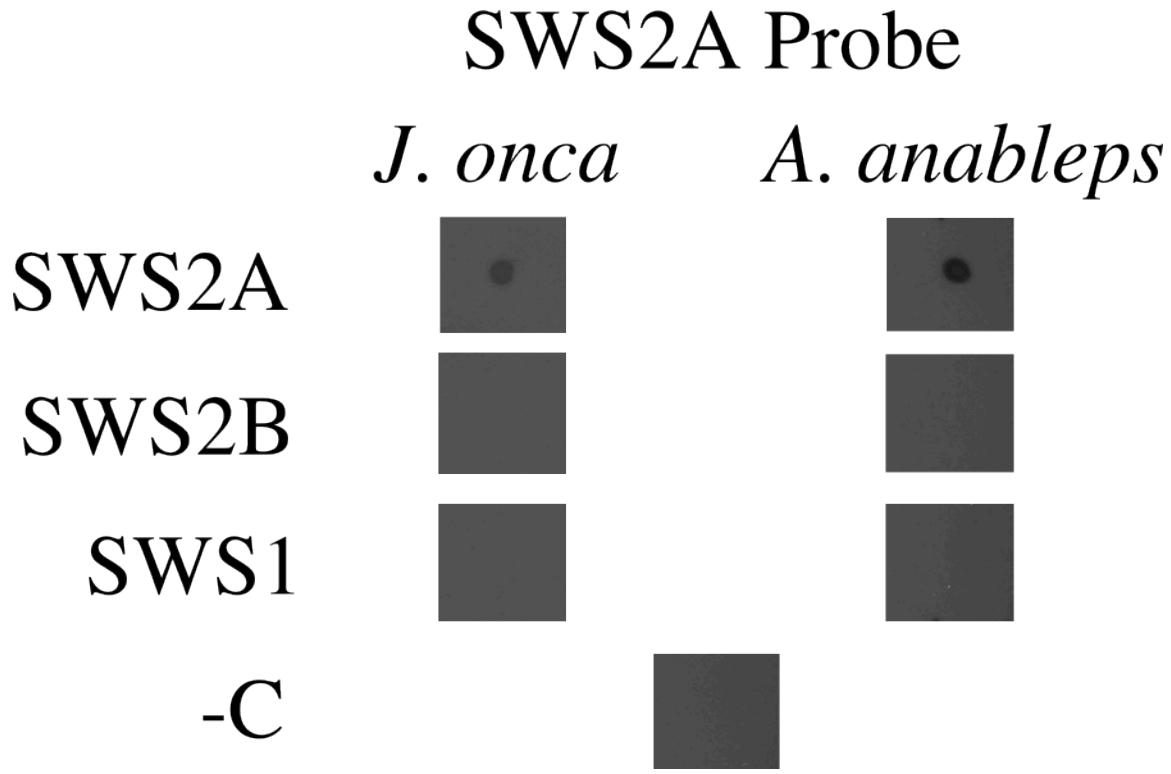
Appendix 8: Percent identity between probe and target genes.

Nucleotide identity between probe sequence and gene sequence for both *A. anableps* and *J. onca.* A star indicate genes targeted by the probe.



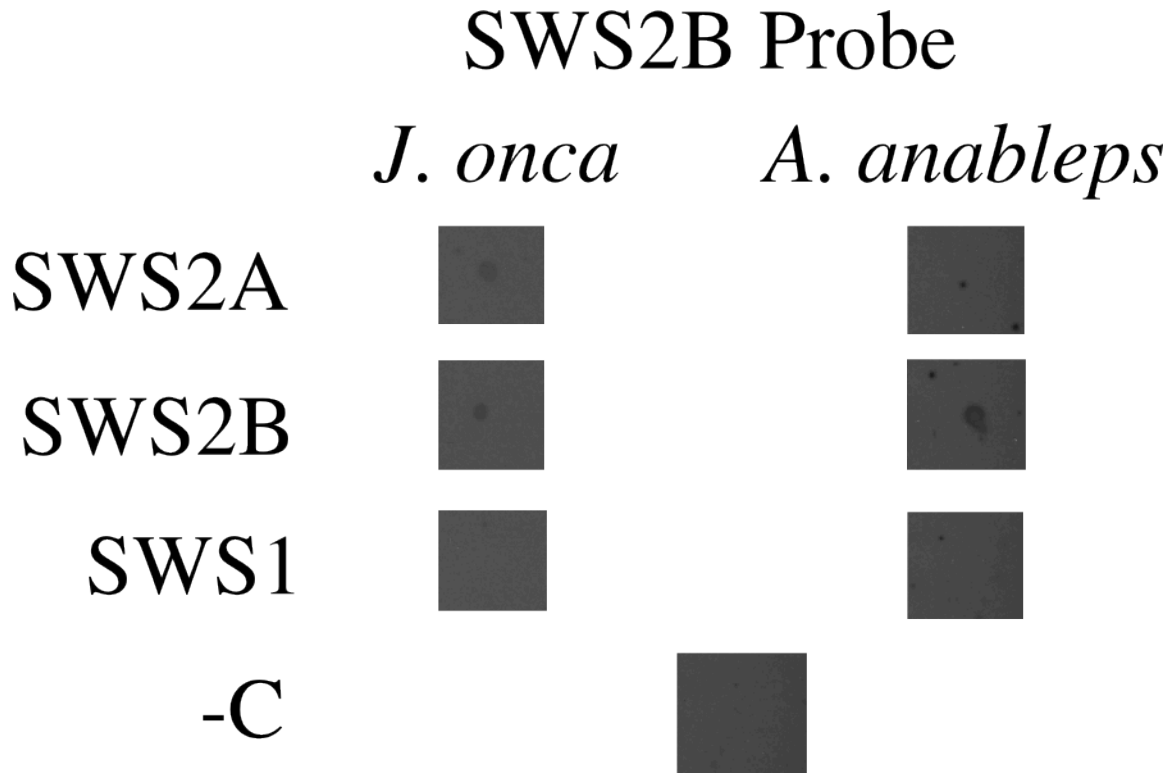
Appendix 9: Dot blot probe specificity test for SWS1 probe.

RNA template was diluted 10^{-4}. Hybridization and high stringency washes were completed at 65°C. The high stringency wash used 0.5X SSC with Tween-20. –C is antisense RH2A transcript.



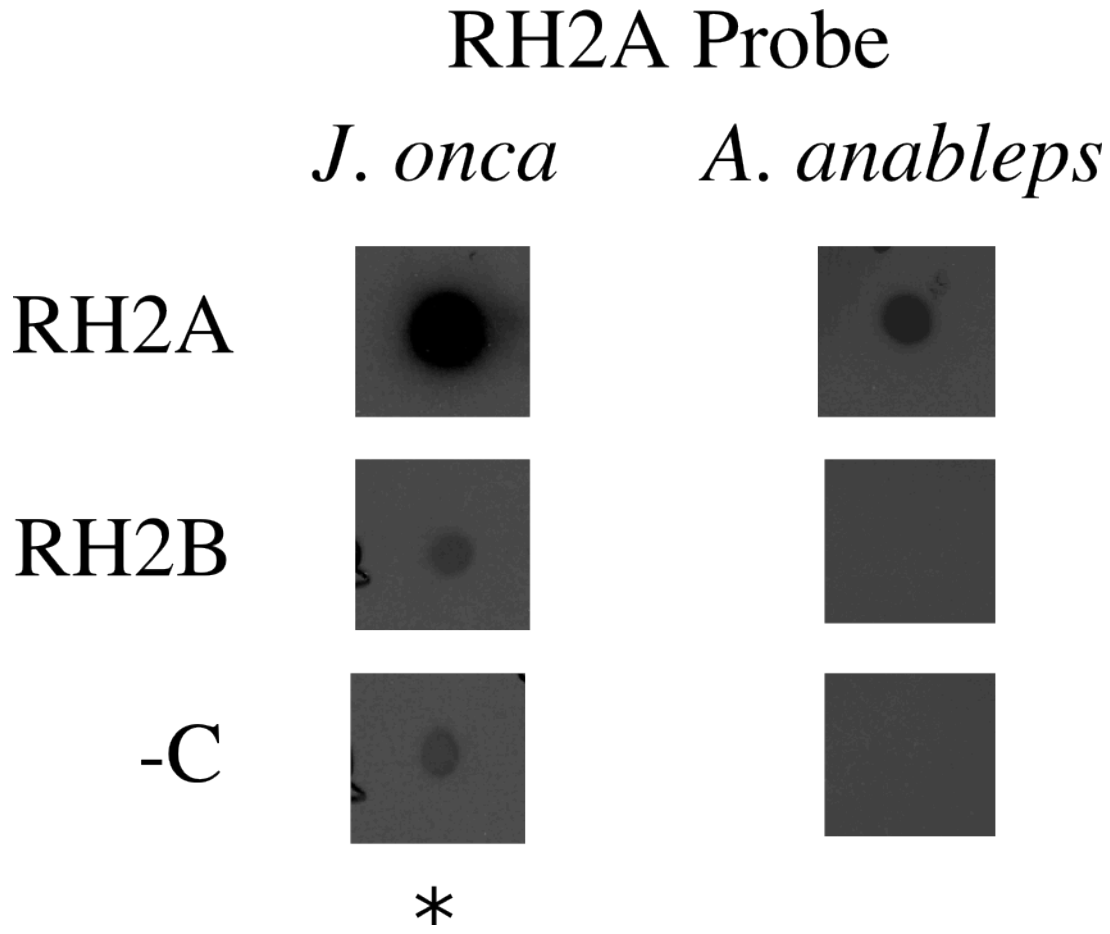
Appendix 10: Dot blot probe specificity test for SWS2A probe.

RNA template was diluted 10^{-4}. Hybridization and high stringency washes were completed at 65°C. The high stringency wash used 0.5X SSC with Tween-20. –C is antisense RH2A transcript.



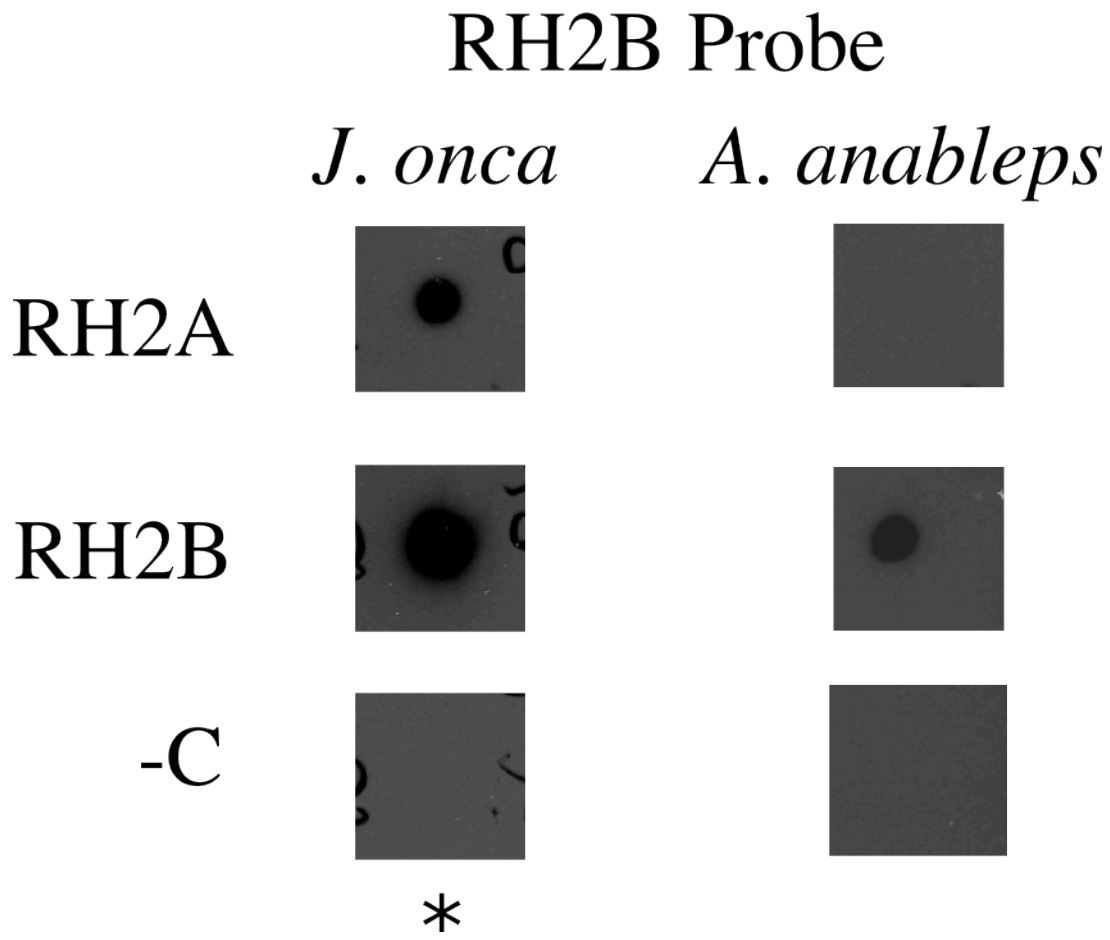
Appendix 11: Dot blot probe specificity test for SWS2B.

RNA template was diluted 10^{-4}. Hybridization and high stringency washes were completed at 65°C. The high stringency wash used 0.5X SSC with Tween-20. –C is antisense RH2A transcript.



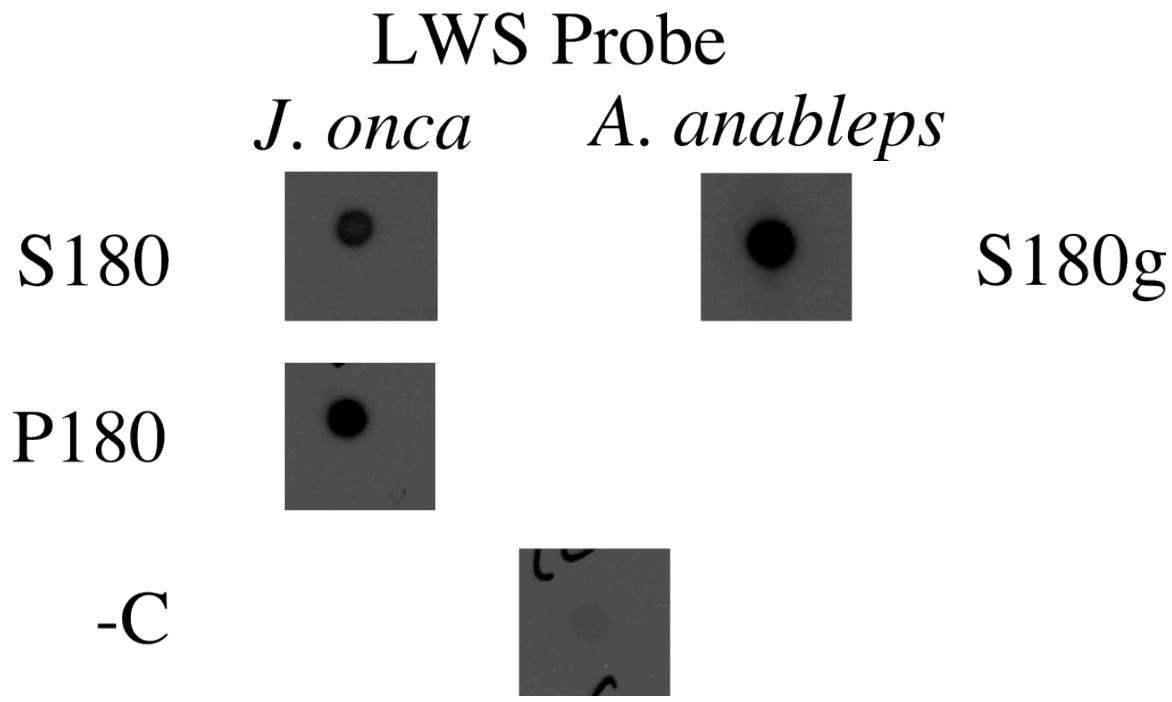
Appendix 12: Dot blot probe specificity test for RH2A.

RNA template was diluted 10^{-4} (*10^{-2}). Hybridization and high stringency washes were completed at 65°C (*68°C). The high stringency wash used 0.5X SSC (*0.2X SSC) with Tween-20. –C is antisense RH2A transcript.



Appendix 13: Dot blot probe specificity test for RH2B.

RNA template was diluted 10^{-4} (*10^{-2}). Hybridization and high stringency washes were completed at 65°C (*68°C). The high stringency wash used 0.5X SSC (*0.2X SSC) with Tween-20. –C is antisense RH2A transcript.



Appendix 14: Dot blot probe specificity test for LWS.

RNA template was diluted 10^{-2}. Hybridization and high stringency washes were completed at 68°C. The high stringency wash used 0.2X SSC with Tween-20. –C is antisense RH2A transcript.