

Contents lists available at [SciVerse ScienceDirect](http://www.sciencedirect.com)

## Molecular Phylogenetics and Evolution

journal homepage: [www.elsevier.com/locate/ympev](http://www.elsevier.com/locate/ympev)

## Review

## Opsin gene duplication and divergence in ray-finned fish

Diana J. Rennison<sup>1</sup>, Gregory L. Owens<sup>2</sup>, John S. Taylor\*

University of Victoria, Department of Biology, PO Box 3020, Station CSC, Victoria, BC, Canada V8W 3N5

## ARTICLE INFO

## Article history:

Received 10 May 2011

Revised 18 November 2011

Accepted 21 November 2011

Available online xxxx

## Keywords:

Ray-finned fish

Vision

Gene duplication

Opsins

Phylogenies

## ABSTRACT

Opsin gene sequences were first reported in the 1980s. The goal of that research was to test the hypothesis that human opsins were members of a single gene family and that variation in human color vision was mediated by mutations in these genes. While the new data supported both hypotheses, the greatest contribution of this work was, arguably, that it provided the data necessary for PCR-based surveys in a diversity of other species. Such studies, and recent whole genome sequencing projects, have uncovered exceptionally large opsin gene repertoires in ray-finned fishes (taxon, Actinopterygii). Guppies and zebrafish, for example, have 10 visual opsin genes each. Here we review the duplication and divergence events that have generated these gene collections. Phylogenetic analyses revealed that large opsin gene repertoires in fish have been generated by gene duplication and divergence events that span the age of the ray-finned fishes. Data from whole genome sequencing projects and from large-insert clones show that tandem duplication is the primary mode of opsin gene family expansion in fishes. In some instances gene conversion between tandem duplicates has obscured evolutionary relationships among genes and generated unique key-site haplotypes. We mapped amino acid substitutions at so-called key-sites onto phylogenies and this exposed many examples of convergence. We found that  $dN/dS$  values were higher on the branches of our trees that followed gene duplication than on branches that followed speciation events, suggesting that duplication relaxes constraints on opsin sequence evolution. Though the focus of the review is opsin sequence evolution, we also note that there are few clear connections between opsin gene repertoires and variation in spectral environment, morphological traits, or life history traits.

© 2011 Elsevier Inc. All rights reserved.

## Contents

1. Introduction .....	00
2. Methods .....	00
2.1. Database survey and phylogenetic analysis .....	00
2.2. Mapping duplication events onto a species tree .....	00
2.3. Gene orientation .....	00
2.4. Positive selection and $dN/dS$ ratios .....	00
2.5. Key-sites .....	00
2.6. Sequence divergence in green LWS opsins .....	00
3. Results .....	00
3.1. All opsins phylogeny .....	00
3.2. Opsin subfamily phylogenies .....	00
3.2.1. SWS1 .....	00
3.2.2. SWS2 .....	00
3.2.3. RH1 .....	00
3.2.4. RH2 .....	00
3.2.5. LWS .....	00
3.3. Gene loss and pseudogenization .....	00
3.4. Mechanisms of gene duplication .....	00

\* Corresponding author. Fax: +1 250 721 7120.

E-mail address: [taylorjs@uvic.ca](mailto:taylorjs@uvic.ca) (J.S. Taylor).<sup>1</sup> Present address: University of British Columbia, Department of Zoology, #2370-6270 University Blvd., Vancouver, BC, Canada V6T 1Z4.<sup>2</sup> Present address: University of British Columbia, Department of Botany, #3529-6270 University Blvd., Vancouver, BC, Canada V6T 1Z4.

3.5.	Gene conversion .....	00
3.6.	Non-synonymous and synonymous substitutions. ....	00
3.7.	Key-sites .....	00
4.	Discussion. ....	00
4.1.	Opsin gene duplication events span the age of the ray-finned fish taxon. ....	00
4.2.	Tandem duplication is the most common form of opsin duplication in fish. ....	00
4.3.	Opsin gene duplication is most prevalent in the RH2 and LWS subfamilies. ....	00
4.4.	Gene duplication and divergence. ....	00
4.5.	Opsins duplication and adaptation. ....	00
4.6.	Future directions of opsin gene research in ray-finned fish. ....	00
5.	Conclusions. ....	00
	Acknowledgements .....	00
	Appendix A. Supplementary material. ....	00
	References .....	00

## 1. Introduction

The rhodopsin-like G protein-coupled receptor (GPCR) family is a large group of membrane-bound proteins that includes opsins, olfactory receptors, neurotransmitter receptors, hormone, and opioid receptors (Fredriksson et al., 2003; Fredriksson and Schiöth, 2005). Opsin genes form a monophyletic clade within this group that can be divided into two major lineages. Genes from one group, the c-opsins, are expressed primarily in ciliary photoreceptor cells (e.g., rods and cones of vertebrate retinas and the extraocular ocelli of annelids); they are also expressed in a diversity of tissues in cnidarians (e.g., Kozmik et al., 2008). Genes from the second major group of opsins, the rhabdomeric- or r-opsins, are expressed primarily in rhabdomeric photoreceptor cells including those that make up the compound eyes of insects, the light-sensitive Joseph cells in amphioxus, and vertebrate retinal ganglion cells (Eakin, 1965; Falk and Applebury, 1988; Arendt et al., 2004; Plachetzki et al., 2005; Alvares, 2008; Suga et al., 2008).

Opsin proteins are bound to a vitamin A-derived chromophore at a lysine residue at position 296 (Palczewski et al., 2000). When it absorbs light, the chromophore isomerizes and induces a conformational change in the opsin. This leads to intracellular G-protein-mediated signal transduction culminating in membrane hyperpolarization (ciliary cells) or depolarization (rhabdomeric cells). There are two chromophores and several species of fish appear to tune their vision by switching from one to the other (reviewed in Levine and MacNichol (1979)). Intracellular oil droplets also influence vision by narrowing the range of wavelengths available to the opsin-expressing photoreceptor (Bowmaker and Knowles, 1977), although these are primarily found in birds and reptiles. However, neither chromophore-based tuning nor oil droplets will be discussed further in this review.

Retinal pigments (opsin protein plus chromophore) were isolated more than 100 years ago (Kühne, 1879; Tansley, 1931). However, the first opsin gene sequence, bovine (*Bos taurus*) rhodopsin (RH1), was obtained in 1983 (Nathans and Hogness, 1983). Shortly thereafter the human (*Homo sapiens*) RH1 gene was characterized (Nathans and Hogness, 1984). RH1 genes are expressed in rod cells, which are used for dim light, or scotopic, vision. Cone cells are used for bright light or photopic vision and cones cells in humans express one of three additional opsins (a short wavelength sensitive (SWS) and two long wavelength sensitive genes), all of which were first sequenced in 1986 (Nathans et al., 1986).

In the 1990s, studies in non-mammalian vertebrates uncovered additional visual opsins and all have now been sorted into five subfamilies: Two short wave- or blue-sensitive opsin subfamilies (SWS1 and SWS2), two middle wave- or green-sensitive opsin subfamilies (RH1 and RH2), and the long wave- or red-sensitive (LWS) opsin subfamily (Okano et al., 1992; Johnson et al., 1993; Yokoyama,

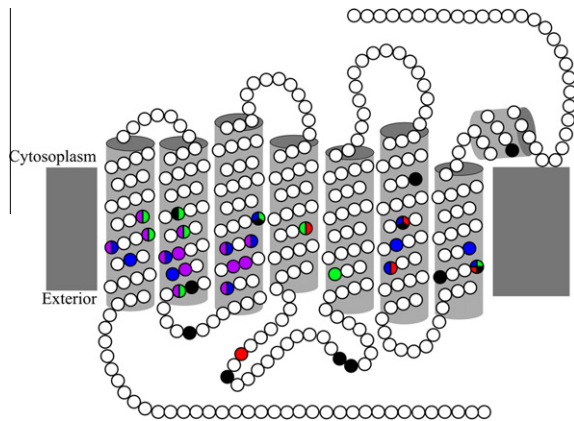
1994, 2000; Kawamura and Yokoyama, 1995, 1996). The mammalian SWS opsin turned out to be an ortholog of the SWS1 genes. All five of the opsin subclasses can be found in the lamprey, *Geotria australis* (Collin and Trezise, 2004; Davies et al., 2007), indicating that a five-gene repertoire is the ancestral state for vertebrates.

These five opsin subfamilies appear to be products of a tandem duplication that produced the LWS opsin and a second gene that, via two whole genome duplication events, gave rise to the SWS1, SWS2, RH1 and RH2 opsins. This one-tandem-plus-two-whole-genome duplication hypothesis requires a large number of opsin gene losses (see Fig. 3 in Larhammar et al., 2009) and is not well supported by phylogenetic analysis (Okano et al., 1992). However, synteny data indicate that the regions of human chromosomes 1, 3, 7 and X that carry opsin genes are paralogs (Larhammar et al., 2009), a finding which is consistent with the hypothesis that either chromosome or whole genome duplication played a role in the expansion of the opsin family.

Data from a diversity of species show that mammals have fewer opsins than their early vertebrate ancestors (having lost the SWS2 and RH2 genes). Although mammals have been extensively surveyed, only three lineages have increased their opsin repertoire, bats (*Haplonycteris fischeri*) (Wang et al., 2004), great apes (Ibbotson et al., 1992) and fat tailed dunnart (*Sminthopsis crassicaudata*) (Cowing et al., 2008). On the other hand, PCR-based surveys and whole genome sequences have shown that ray-finned fish (actinopterygians) often possess many more opsins than were present in the vertebrate ancestor but until now fish data have not been assembled in a single analysis.

For this review we reconstructed phylogenetic trees of fish visual opsins to provide an indication of the ages of gene duplication events and the number of species expected to possess said duplicates. Some new fish opsin sequences are reported as part of this tree reconstruction work. For many species we have also been able to determine what types of duplication events generated different opsin gene repertoires. Additionally, by comparing the number of non-synonymous substitutions per non-synonymous site ( $dN$ ) to the number of synonymous substitutions per synonymous site ( $dS$ ) it was possible to gain insight into the selective pressures that influence opsin gene evolution. We used PAML (Yang, 2007) to estimate  $dN/dS$  ( $\omega$ ). We compared  $\omega$  values among opsin gene subfamilies, and we compared  $\omega$  for branches that followed gene duplication events to those that followed speciation events. We also mapped key-site substitutions onto the opsin subfamily phylogenies in an effort to correlate sequence variation and function. Key-sites are a subset of the amino acid positions within an opsin that have been shown to influence the wavelength of maximal absorption ( $\lambda_{max}$ ) (see Fig. 1).

Five major conclusions can be drawn from our survey and analysis: Large opsin gene repertoires in ray-finned fishes are a



**Fig. 1.** Opsin protein structure. Schematic diagram of the seven transmembrane domain structure of the opsin protein. Key-sites known to affect spectral sensitivity are highlighted by subfamily: Violet, SWS1; Blue, SWS2; Black, RH1; Green, RH2; Red, LWS. Structure adapted from Palczewski et al., 2000. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

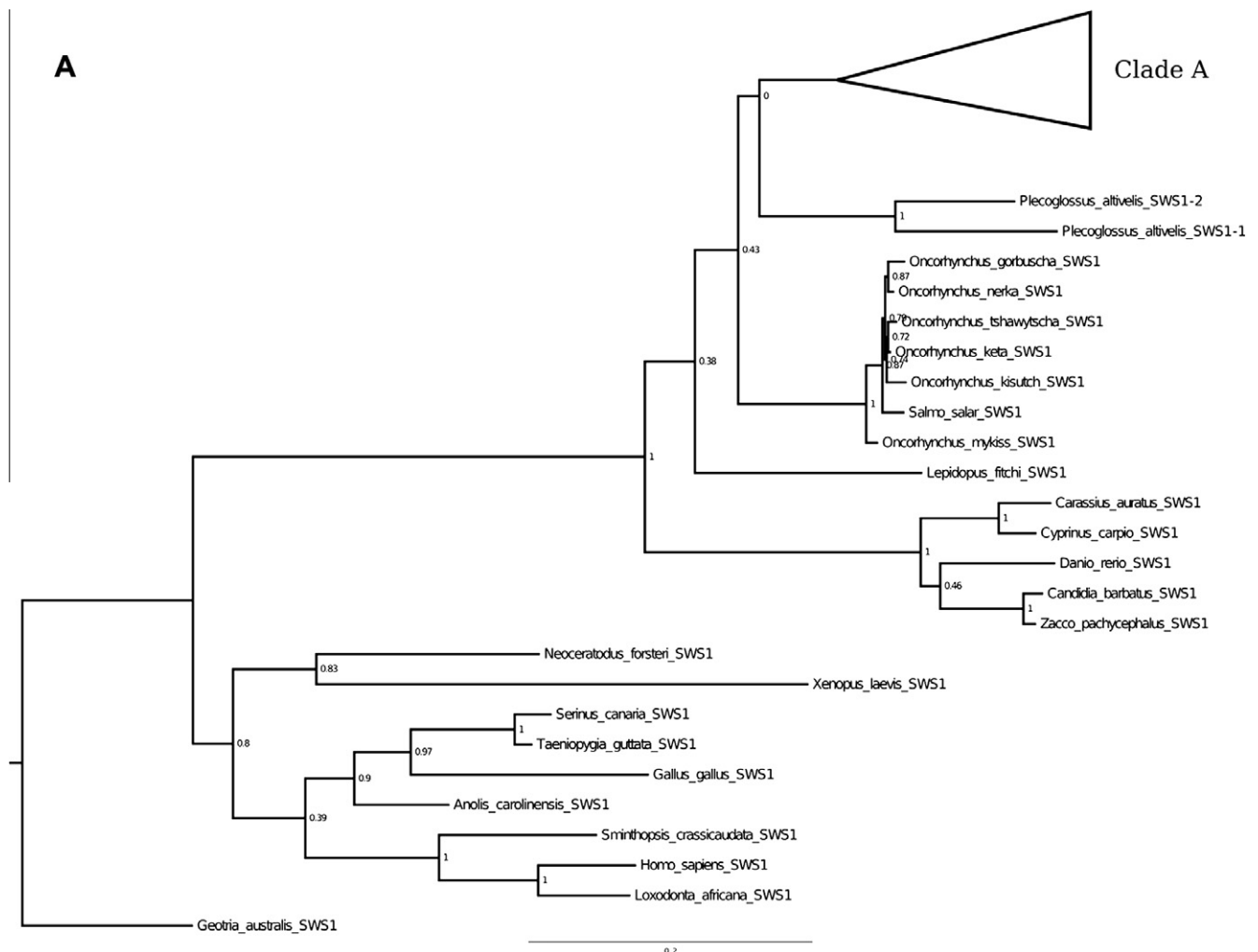
consequence of gene duplication events spanning the age of the taxon. Tandem duplication produces more opsin gene duplicates in fish than any other mode of duplication and surprisingly none

of the duplication nodes appear to coincide with '3R', a whole genome duplication event that occurred in the ancestor of teleosts (Amores et al., 1998; Taylor et al., 2003; Hoegg et al., 2004). Gene duplicates are most prevalent in the RH2 and LWS opsin subfamilies, which tend to be expressed in double cones. We observed increased  $\omega$  following opsin gene duplication, which suggests that duplication facilitates evolutionary change at the amino acid level in this family of proteins. Finally, we found no support for the hypothesis that repertoire variation among fish is correlated with life history, behavior (e.g., sexual selection) or morphology (e.g., coloration); it appears that phylogenetic history is a much more important consideration when interpreting opsin repertoire size in ray-finned fish.

## 2. Methods

### 2.1. Database survey and phylogenetic analysis

The majority of opsin gene sequences were obtained using BLASTn with default parameters (Altschul et al., 1990). The databases surveyed were the NCBI nucleotide database and Ensembl genome sequence databases. Opsin genes from *Danio rerio* and *Oryzias latipes* were employed as queries. Most, but not all ray-finned fish hits to these query sequences were used in our



**Fig. 2.** Phylogenetic tree of SWS1 opsins in fish. The tree was created using the maximum-likelihood method. Accession numbers are listed in Supplementary Table 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 and Gascuel, 2003; Posada and Crandall, 1998). Tree topology was tested using the best of NNI and SPR. Numbers at nodes represent aLRT values (Anisimova and Gascuel, 2006). The model of evolution was determined to be HKY85 + I + G (I = 0.2299, G = 1.1920). Clade A encompasses Neoteleostei.

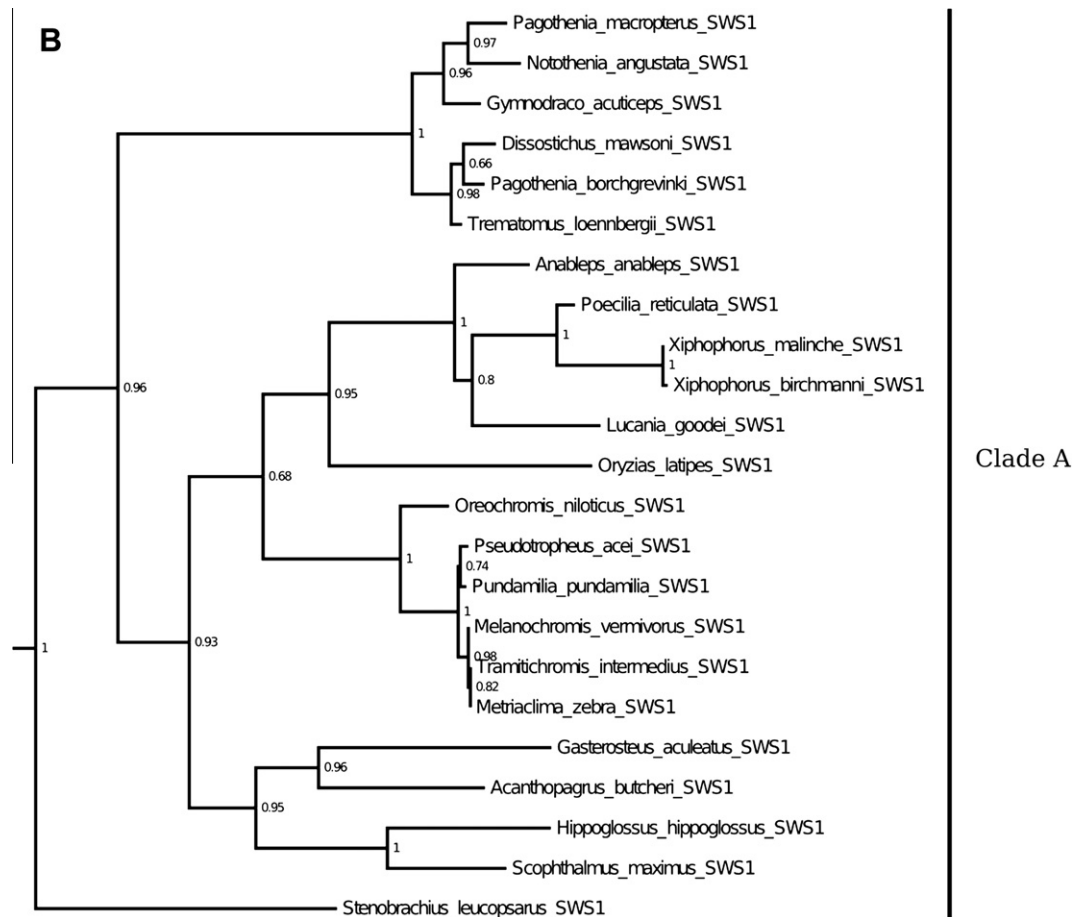


Fig. 2 (continued)

analyses: Some lineages have been the subjects of extensive surveys that have generated an enormous number of very similar sequences (e.g., family Cichlidae) and in these cases, we selected a representative subset of the available data. The seabream (*Pagrus* and *Acanthopagrus*) opsins in this study were obtained from Dr. F.Y. Wang (Personal correspondence) and new sequence data for pencilfish (*Nannostomus beckfordi*) and American flag fish (*Jordanella floridae*) from our lab were also included. All of the included nucleotide sequences were translated and aligned by hand using BioEdit (Hall, 1999).

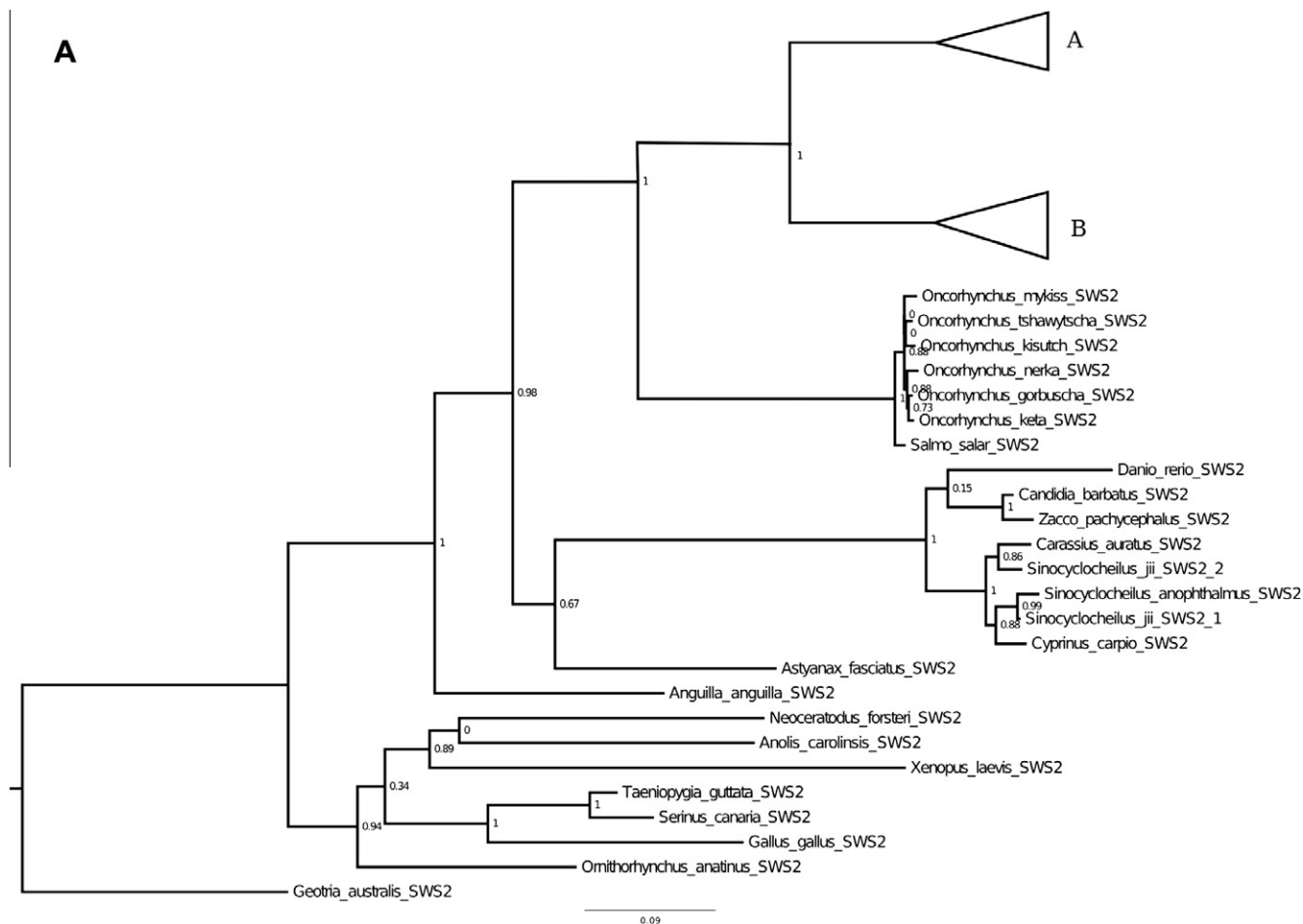
All phylogenetic trees were constructed from nucleotide alignments. A single “all-opsin” multiple sequence alignment was used in the first analysis. This included pinopsins, vertebrate ancient (VA) opsins, amphiopsins, a c-opsin from the annelid worm (*Platynereis drumerilii*), and an r-opsin from *Drosophila melanogaster*. We included the non-visual opsins because recent reports have suggested that pinopsins might be nested among the visual opsins (e.g., Max et al., 1995). Mega4 (Tamura et al., 2007) was used to generate a neighbor-joining (NJ) phylogenetic tree based upon Log Det distances (Tamura and Kumar, 2002). This analysis helped us to confirm that some of the especially divergent genes had been assigned to subfamilies correctly.

We used PAUP\* 4.8B and Modeltest (Posada and Crandall, 1998; Swofford, 2002) to select models of sequence evolution suitable for the data from each opsin subfamily and then maximum likelihood (ML) trees were reconstructed using PhyML (Guindon and Gascuel, 2003). Tree improvement was done using the best of nearest neighbor interchange (NNI) and subtree pruning regrafting (SPR) (Hordijk and Gascuel, 2005). Support for nodes on ML trees was

estimated using an approximate likelihood ratio test (Anisimova and Gascuel, 2006). Neighbor joining trees using Tamura–Nei distances with bootstrap (1000 replicates) were also reconstructed for subfamily alignments using PAUP (Felsenstein, 1985; Saitou and Nei, 1987; Tamura and Nei, 1993; Swofford, 2002). Lastly, for each subfamily a strict consensus maximum parsimony tree was created, also using PAUP. Pair-wise deletion was used for instances of missing nucleotides in all analyses.

## 2.2. Mapping duplication events onto a species tree

We inferred relative timing of gene duplication events from gene phylogenies (Figs. 2–6). Nodes (bifurcations) in our phylogenetic trees demark either speciation or gene duplication events. The duplication nodes on each subfamily ML tree were numbered based upon the amount of sequence divergence between genes found in the two post-duplication/paralogous clades. These divergence estimates between paralogous clades were based upon average Tamura–Nei (TN) distances (Tamura and Nei, 1993; Tamura et al., 2007) for all pairs of genes in the paralogous clades. Where one paralog was the sister to a large clade of homologs, e.g., zebrafish RH1-2, which was the sister group to all other RH1 opsins (including its paralog, RH1-1), we based the node label on the TN distance between that gene and its paralog only (i.e., not the average TN distance between it and all genes in the sister clade) because we suspected that the position of one gene might be disrupted by long branch attraction. Paralogs that we suspected were modified by gene conversion were excluded from this analysis.



**Fig. 3.** Phylogenetic tree of SWS2 opsins in fish. The tree was created using the maximum-likelihood method. Accession numbers are listed in [Supplementary Table 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 and Gascuel, 2003; Posada and Crandall, 1998). Tree topology was tested using the best of NNI and SPR. Numbers at nodes represent aLRT values (Anisimova and 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.

As a summary, opsin gene duplication events were displayed on a species tree (Fig. 7). The topology of this tree was generated from a maximum likelihood analysis of RH1 sequences with terminal branches added for *J. floridae*, *Scophthalmus maximus*, *Zacco pachycephalus*, *Candidia barbatus*, *Clupea harengus* and *N. beckfordi*, (species that lacked RH1 gene sequences), in a manner that was consistent with fish taxonomy (Nelson, 2006).

### 2.3. Gene orientation

In order to discriminate among several possible modes of gene duplication we recorded the location, either on chromosomes or on long-insert clones, of opsin genes for six fish species. Data for three spine stickleback (*Gasterosteus aculeatus*), fugu (*Takifugu rubripes*), the green spotted pufferfish (*Tetraodon nigroviridis*), the Japanese rice fish (*O. latipes*), and the zebrafish (*D. rerio*) were obtained from the Ensembl genome browser. Data for tilapia (*Oreochromis niloticus*) was reported by in Hofmann and Carleton, 2009 and O'Quin et al., 2011. BAC clone sequence data for the swordtail (*Xiphophorus helleri*) was reported by (Watson et al., 2010a).

### 2.4. Positive selection and dN/dS ratios

Purifying selection occurs when the number of non-synonymous substitutions per non-synonymous site divided by the num-

ber of synonymous substitutions per synonymous site ( $\omega$ ) is much less than 1.0. Purifying selection is most common because amino acid changes (i.e., non-synonymous substitutions) are usually detrimental to protein function (Fay and Wu, 2003). When  $\omega$  is close to 1.0, purifying selection has been relaxed and the sequence is considered to be evolving in a neutral manner. This is expected to be temporary or to generate pseudogenes. Positive selection (selection favouring amino acid-level change) is believed to have occurred when  $\omega$  is greater than 1.0. In this study PAML (Yang, 2007) was used to compare average  $\omega$  among opsin subfamilies, between paralogous clades within two opsin subfamilies, SWS2 and RH2, and, for post-duplication and post-speciation branches within subfamilies. Phylogenetic trees used for these analyses are shown in [Supplementary Figs. 1–5](#) and the post-duplication branches are indicated. These analyses used only full-length opsin gene sequences (58 species, see [Supplementary Table 1](#) for accession numbers).

To identify codons under positive selection, we used two tests in the PAML package: M1a vs. M2a and M8 vs. M8a (Nielsen and Yang, 1998; Wong et al., 2004; Yang et al., 2005). M1a divides codons into two categories, those under neutral selection ( $\omega = 1$ ) and those experiencing negative selection ( $\omega < 1$ ), while M2a adds a third category, codons under positive selection ( $\omega > 1$ ). M8 assumes a beta distribution, from 0 to 1, of  $\omega$  for sites and an additional class of sites under positive selection ( $\omega > 1$ ), while M8a



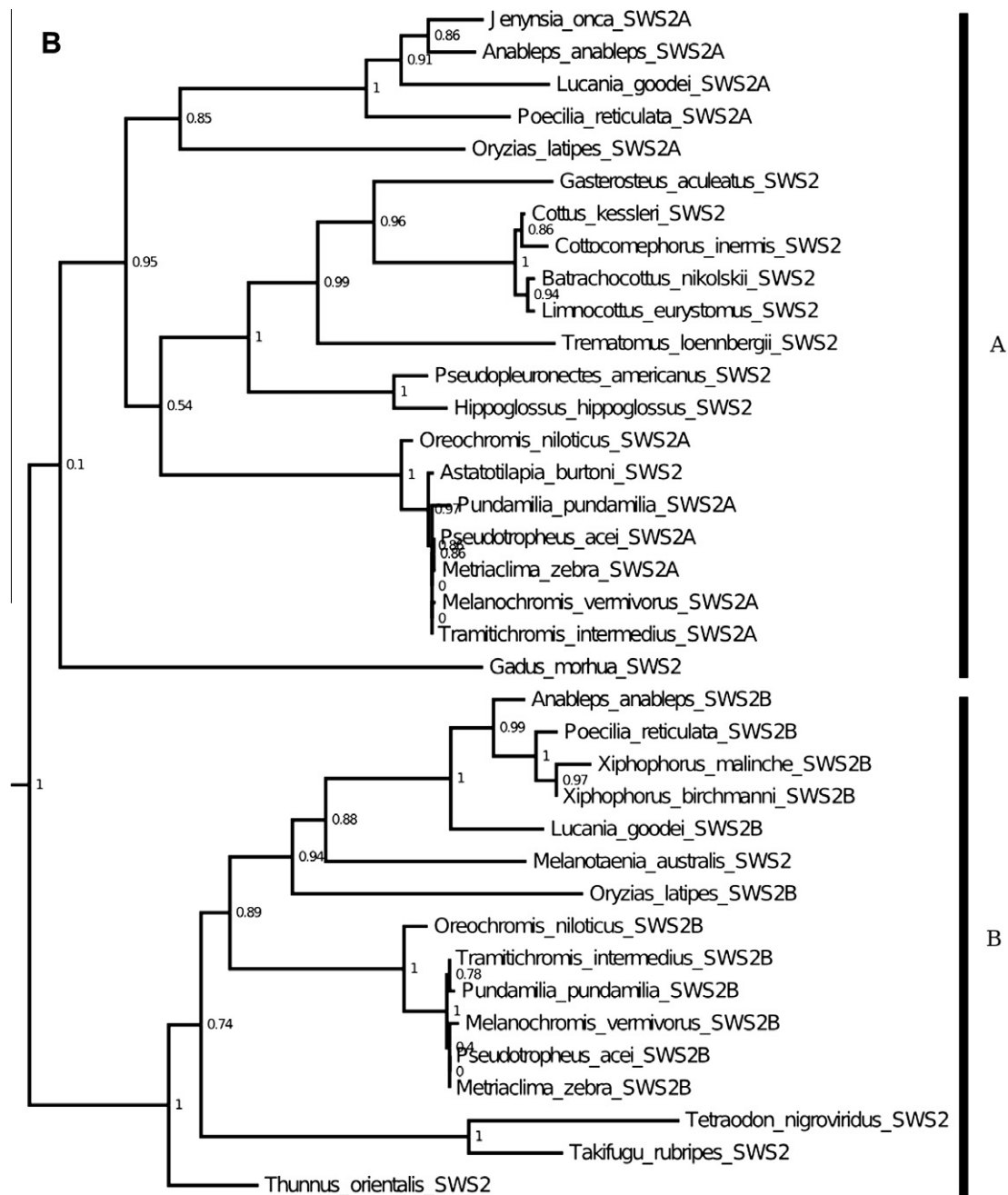


Fig. 3 (continued)

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., 2000; Anisimova et al., 2001). To ameliorate this issue, starting  $\omega$  values of less than and greater than one were used.

To uncover codon-level positive selection only on post-duplication branches, we used 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 highlighted in Supplementary Figs. 1–5. Codons are then divided into categories, which allow a subset of foreground 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 codons to be under

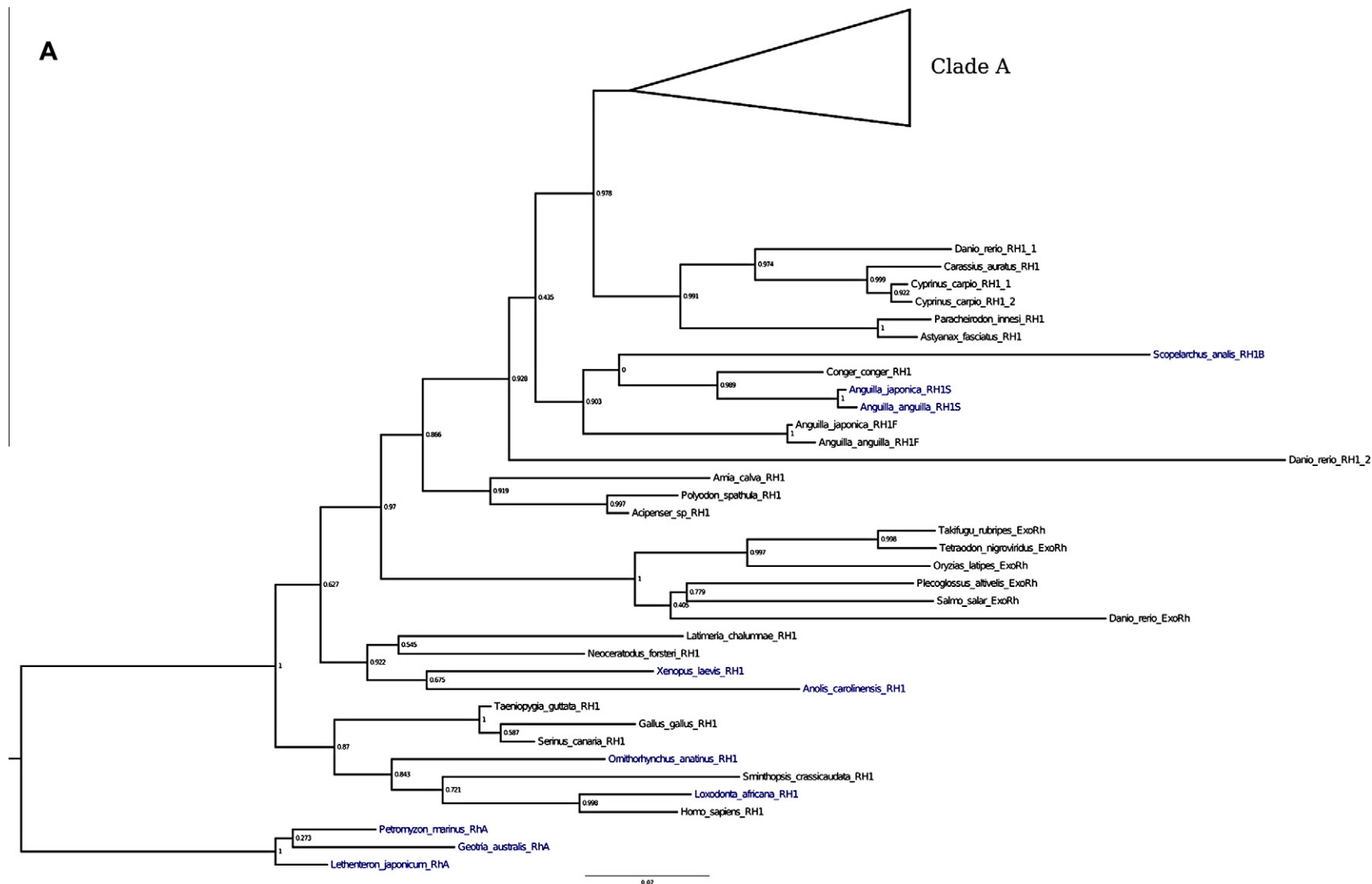
positive selection using a likelihood ratio test. To account for multiple testing on each subfamily, Bonferroni's correction was applied (Miller, 1981; Anisimova and Yang, 2007).

## 2.5. Key-sites

Key-sites are positions in opsin proteins where amino acid substitutions result in a shift (from 1 nm to greater than 60 nm) in wavelength sensitivity of the opsin-chromophore complex (Fig. 1). Three examples of convergent key-site substitution were mapped onto the RH1, RH2 and LWS subfamily trees (Figs. 4–6).

## 2.6. Sequence divergence in green LWS opsins

During our analyses we discovered an unusual pattern of sequence divergence between the green LWS genes from Mexican



**Fig. 4.** Phylogenetic tree of RH1 opsins in fish. The tree was created using the maximum-likelihood method. Accession numbers are listed in [Supplementary Table 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 and Gascuel, 2003; Posada and Crandall, 1998). Tree topology was tested using the best of NNI and SPR. Numbers at nodes represent aLRT values (Anisimova and Gascuel, 2006). The model of evolution was determined to be GTR + I + G (I = 0.2746, G = 1.1396). Taxa names are color coded to show amino acid at position 83 (based on bovine rhodopsin numbering): black, aspartic acid; blue, asparagine. Clade A encompasses Euteleostei. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

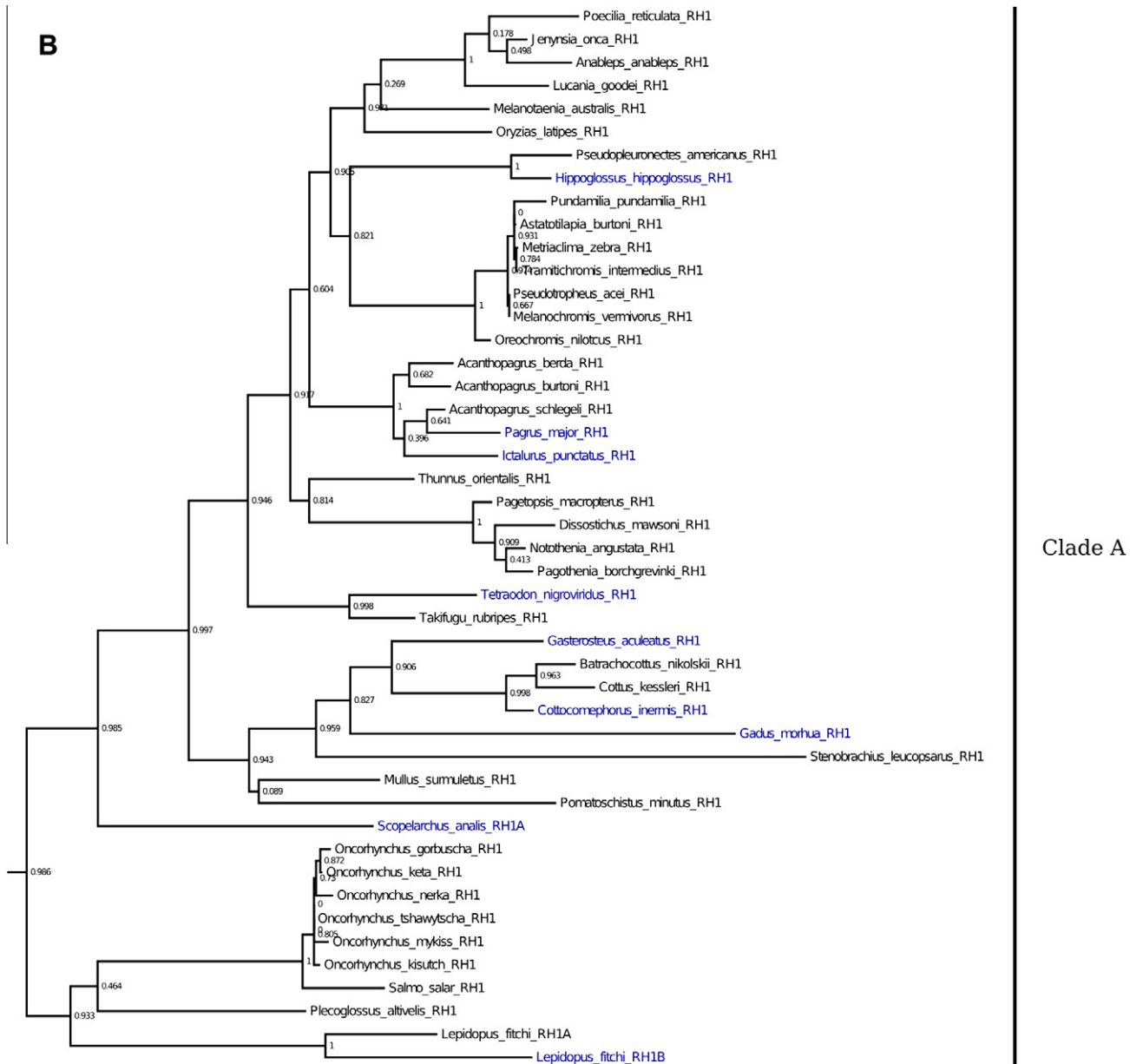


Fig. 4 (continued)

cavefish (*Astyanax fasciatus*), neon tetras (*Paracheirodon innesi*), and pencilfish (*N. beckfordi*) and other LWS opsins. A sliding window analysis (30-residue window) was performed using Swaap 1.0.3 to highlight this pattern (Pride, 2000).

### 3. Results

#### 3.1. All opsins phylogeny

Our genetic distance-based phylogenetic analysis of all opsin sequences generated a topology with five well-supported visual opsin clades (LWS, SWS1, SWS2, RH1 & RH2), a pinopsin clade, and a vertebrate ancient (VA) opsin clade (Supplementary Fig. 6). The visual opsins were not monophyletic. LWS opsins and the pinopsins formed a monophyletic group that was sister to all other visual opsins, although the node supporting this hypothesis had low bootstrap support. We looked for visual opsin synapomorphies

(shared derived amino acid residues) by using the VA opsins and the *P. drumerii* c-opsin as out-groups (i.e., to polarize character state changes), but we found none. Interestingly, two derived character states, a valine (V) at position 70 and a tryptophan (W) at position 188, were consistent with the distance-based hypothesis that LWS opsins and pinopsins are monophyletic. Though not the focus of this study, we note that the relationships inferred for LWS opsins from sarcopterygians (lobed-finned fishes and tetrapods) were not consistent with taxonomy. Similar observations have been made in other studies (e.g., Max et al., 1995). The green opsins from cavefish, neon tetras and the pencilfish formed a well-supported clade at the base of the LWS tree.

#### 3.2. Opsin subfamily phylogenies

##### 3.2.1. SWS1

We analyzed SWS1 gene sequences from 37 fish species. The topology of the SWS1 gene tree was largely consistent with fish





**Fig. 5.** Phylogenetic tree of RH2 opsins in fish. The tree was created using the maximum-likelihood method. Accession numbers are listed in [Supplementary Table 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 and Gascuel, 2003](#); [Posada and Crandall, 1998](#)). Tree topology was tested using the best of NNI and SPR. Numbers at nodes represent aLRT values ([Anisimova and Gascuel, 2006](#)). The model of evolution was determined to be GTR + I + G (I = 0.2485, G = 0.8147). Taxa names are color coded to show amino acid at position 112 (based on bovine rhodopsin numbering): black, glutamic acid; blue, glutamine. The letters A and B indicate RH2A and RH2B clades respectively.

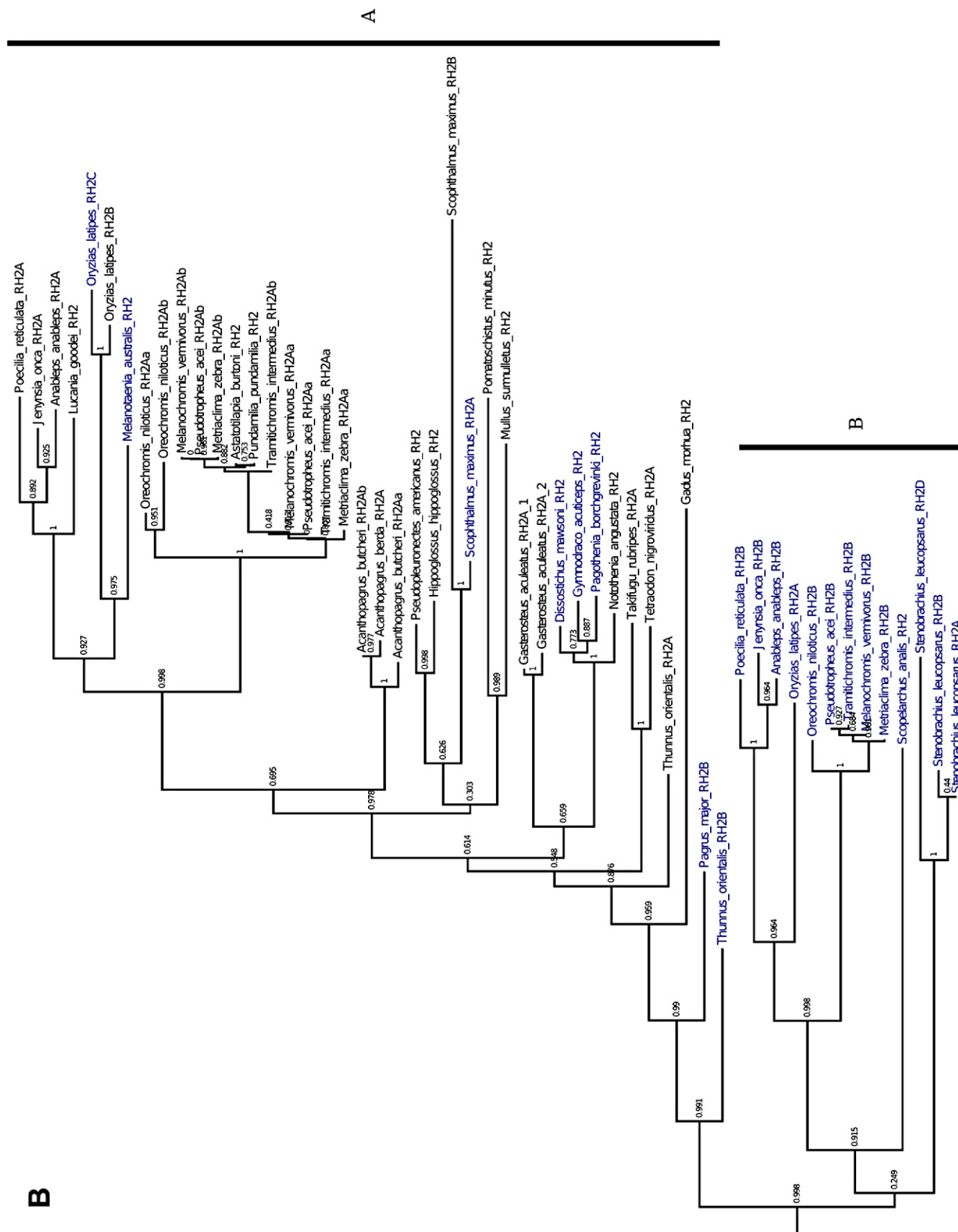
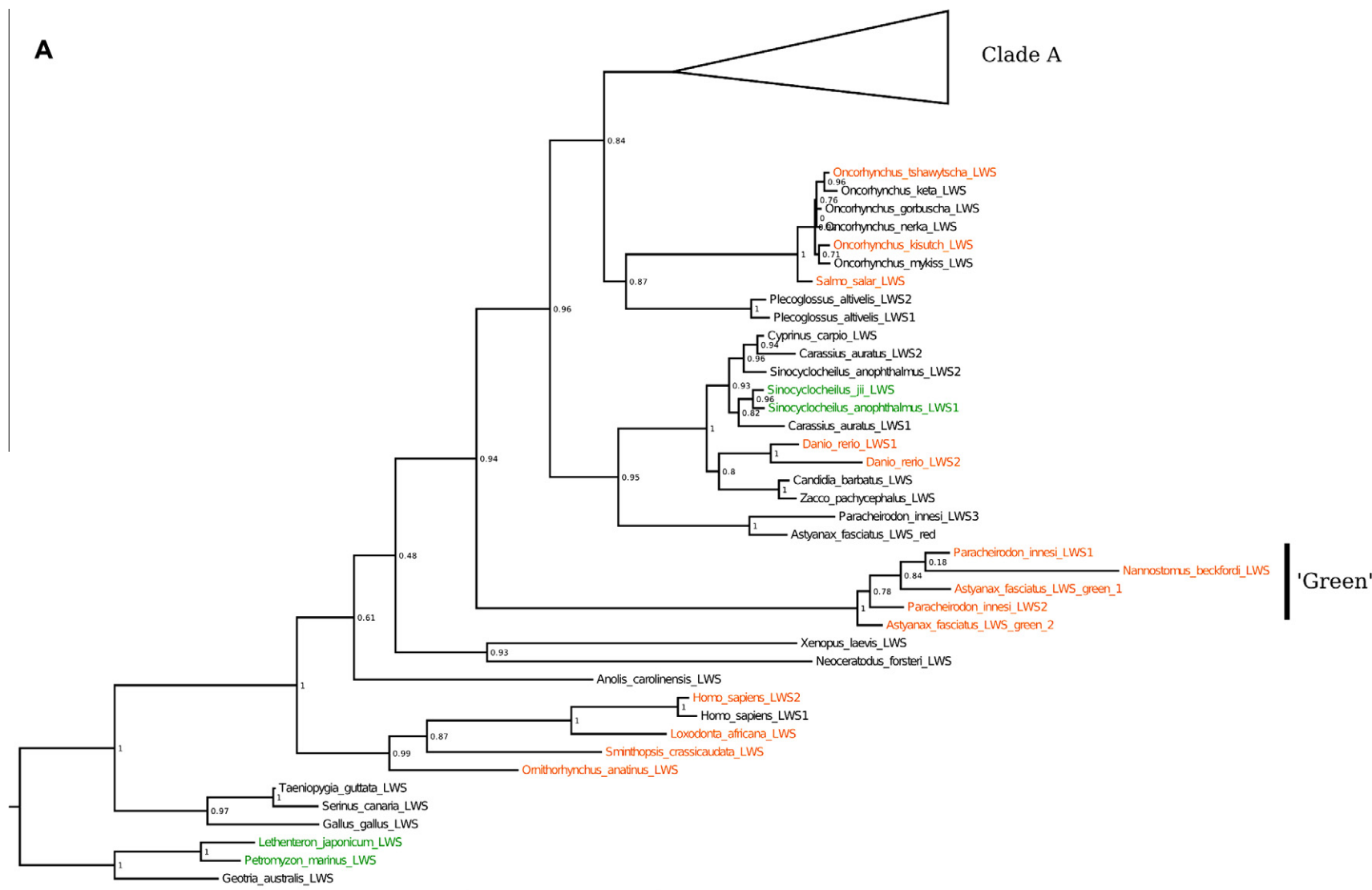


Fig. 5 (continued)



**Fig. 6.** Phylogenetic tree of LWS opsins in fish. The tree was created using the maximum-likelihood method. Accession numbers are listed in [Supplementary Table 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 and Gascuel, 2003; Posada and Crandall, 1998). Tree topology was tested using the best of NNI and SPR. Numbers at nodes represent aLRT values (Anisimova and 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). Taxa names are color coded to show amino acid at position 164 (based on bovine rhodopsin numbering): black, serine; orange, alanine; green, proline. Clade A encompasses Acanthopterygii.

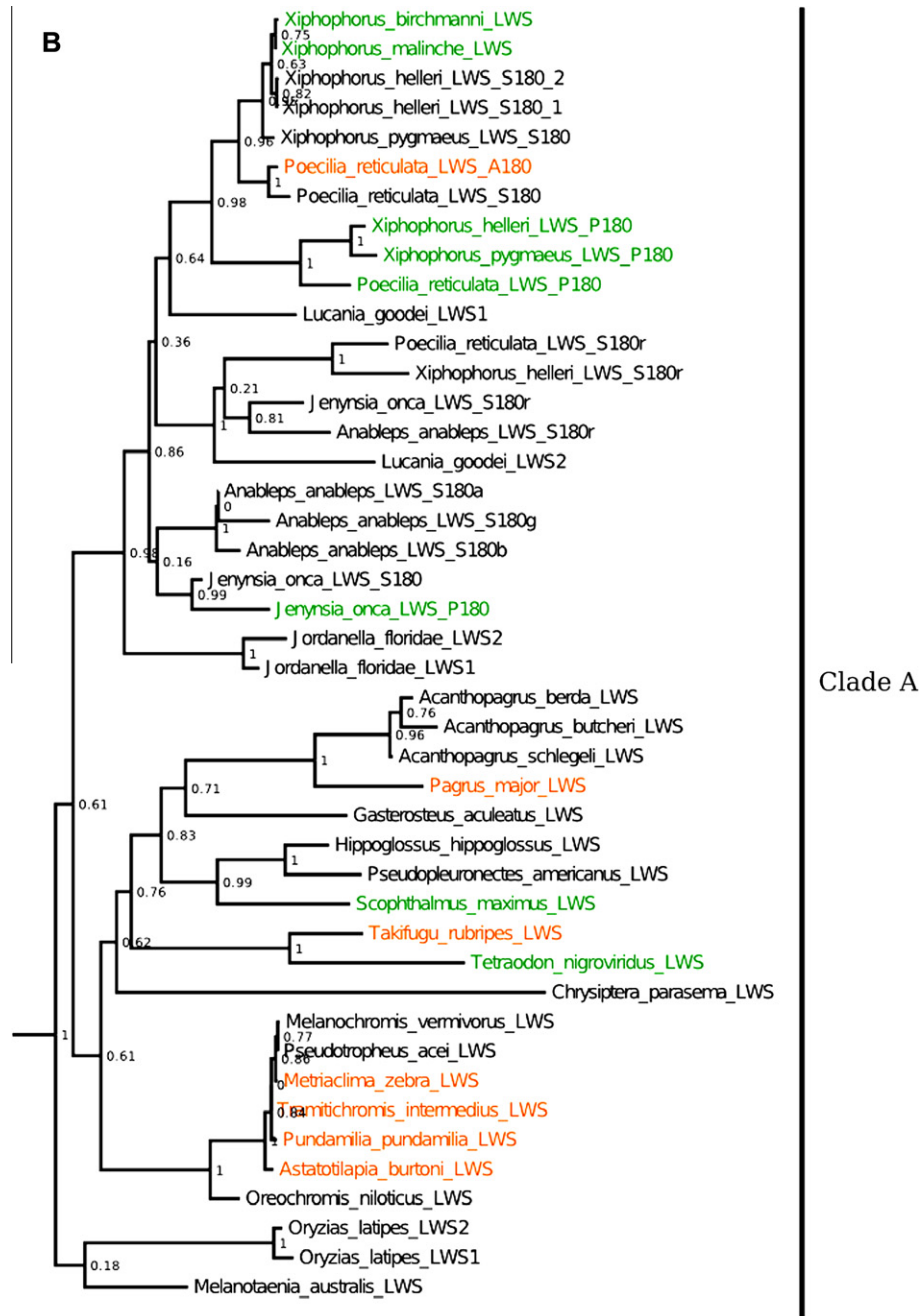


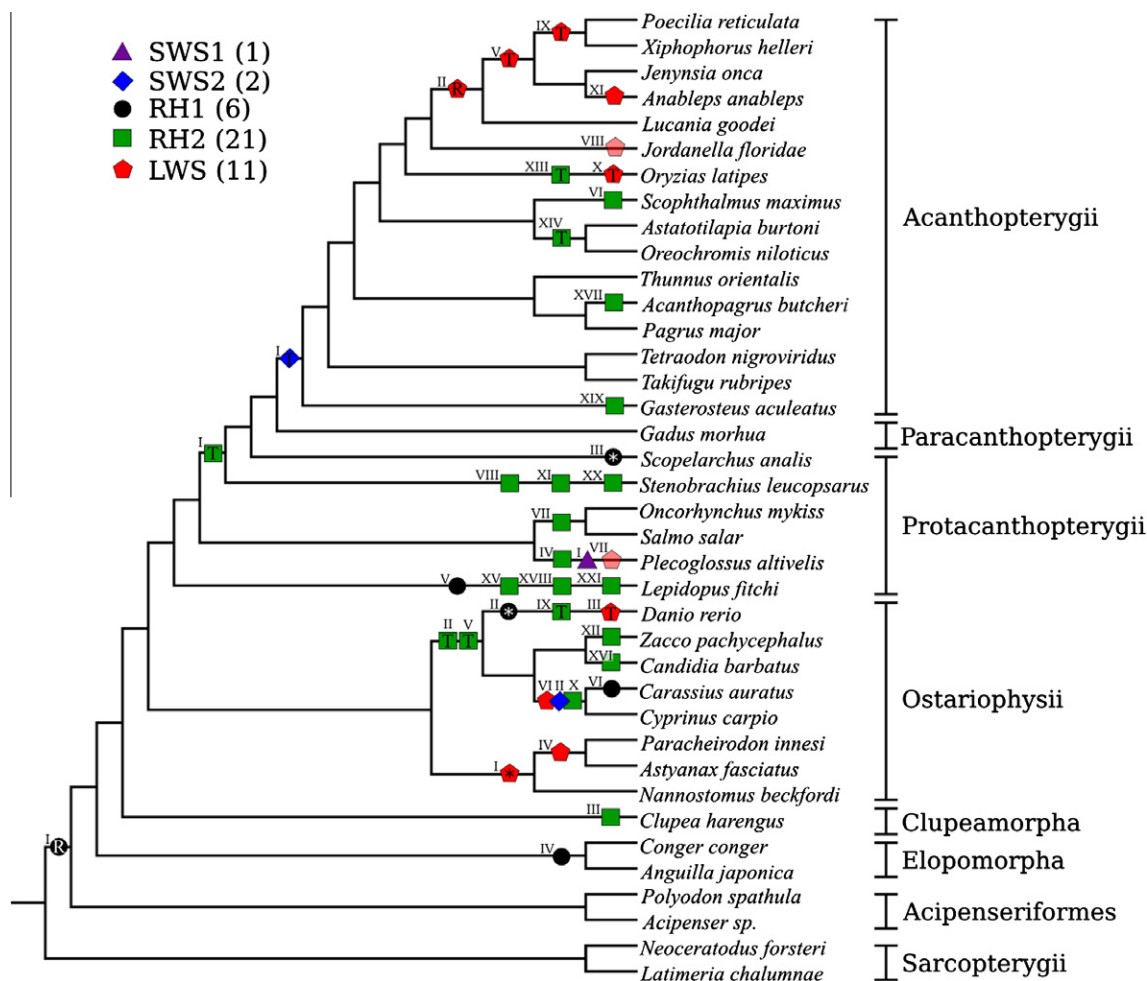
Fig. 6 (continued)

taxonomy with the exception of the position of the scabbardfish (*Lepidopus fitchi*) gene, which appears more closely related to SWS1 sequences from salmonids and smelt (Salmoniformes and Osmeriformes) and the cyprinids (Cypriniformes) than to those from euteleosts. Only one of the 41 opsin gene duplication nodes observed in the study occurred on the SWS1 tree (Fig. 2 and summarized in Fig. 7): Ayu smelt (*Plecoglossus altivelis*) have two SWS1 genes (AYU-UV1 and AYU-UV2). Although only SWS1-2 is expressed in the eye of ayu smelt, both genes have intact open reading frames (Minamoto and Shimizu, 2005). These SWS1 duplicates are 85% identical over a 1025 bp alignment (BLASTn alignment, not shown) and are, therefore, almost as different from one another as they are from single-copy SWS1 genes possessed by species in the family Salmonidae (data not shown). This suggests that the duplication event occurred very early during the evolution of osmerids

and that many other species in this family are likely to possess a pair of SWS1 genes.

### 3.2.2. SWS2

SWS2 opsin genes from 39 species were analyzed. The gene tree was largely consistent with fish taxonomy. Two gene duplication events are represented on the SWS2 subfamily tree (Fig. 3). The first, producing SWS2A and SWS2B genes (SWS2-dup1), occurred in either the ancestor of the clade Holacanthopterygii, a taxonomic group that includes Paracanthopterygii (represented by *Gadus morhua*) and Acanthopterygii (e.g., cichlids, livebearers and stickleback), or in the ancestor of Acanthopterygii. The ML tree indicates that the *Gadus morhua* sequence occurs at the base of the SWS2A clade, but with poor support (0.105 aLRT). The NJ and MP analyses placed it as the first out-group to the duplication



**Fig. 7.** Fish opsin duplication 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. Roman numerals correspond to duplication numbers in [Supplementary Table 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 indicate possible allelic variation.

node. The second duplication (SWS2-dupII) occurred in the common ancestor of three cyprinids, the goldline fish (*Sinocyclocheilus*), the carp (*Cyprinus carpio*), and the goldfish (*Carassius auratus*).

### 3.2.3. RH1

RH1 opsin gene sequences from seventy species of ray-finned fish were analyzed. Sequence relationships were largely consistent with taxonomy: Exceptions included the observation that the Acipenseriformes (sturgeon and paddlefish) and Amiiformes (bowfin) RH1 genes formed a monophyletic group. If the gene tree matched the species tree, then the bowfin RH1 sequence would be more closely related to orthologs from teleosts than to those from Acipenseriformes. Also, sarcopterygian RH1 genes did not form a monophyletic clade (Fig. 4).

Forty-one opsin gene duplication events occur on the opsin subfamily trees, six of these are found on the RH1 opsin tree (summarized in Fig. 4). The first (RH1-dupI) generated the non-visual "exo-rhodopsins" and a clade of genes simply called RH1s. The post-duplication RH1 genes have no introns, supporting the hypothesis that this node reflects retro-duplication (Venkatesh et al., 1999). Venkatesh et al.'s (1999) analysis of intron presence or absence and our ML analysis suggest that the retro-duplication occurred near the base of Actinopterygii, although our NJ and MP analysis place the duplication node before actinopterygians and sarcopterygians diverged.

Zebrafish (*D. rerio*) have two single-exon RH1 genes in addition to exo-rhodopsin, which are both expressed in the eye (Morrow et al., 2011). One of these RH1 paralogs (RH1-2) is the sister sequence to most other teleost RH1 genes. A similar pattern was observed for RH1 duplicates in the pearl eye (*Scopelarchus analis*), where one of the paralogs is sister group to the RH1 genes from eels (Elopomorpha). Pairwise distances for the zebrafish RH1 paralogs and the pearleye RH1 paralogs (Supplementary Table 2) show that they are both products of old duplication events, however, we suspect that their positions in the tree are incorrect because an enormous number of gene losses are inferred by this topology. We show both duplication nodes as species-specific events (node labels RH1-dupII & III) on our summary tree (Fig. 7) and look forward to data from additional species to help resolve the positions of these events.

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 eel clade, RH1 was duplicated and produced the freshwater and deep-sea paralogs (RH1-dupIV) (Hope et al., 1998) before the genera *Anguilla* and *Conger* diverged.

Two additional gene duplication events are present in the RH1 tree: RH1-dupVI occurred in the carp (genus *Cyprinus*), after it diverged from goldfish (genus *Carassius*) and RH1-dupV occurred in the scabbard fish (*L. fitchi*). The scabbard fish RH1 sequences



grouped with salmonid sequences contrary to expectations based upon taxonomy (we observed the same unexpected pattern in SWS1 gene tree).

Another anomaly in the RH1 phylogeny was the location of the catfish (*Ictalurus punctatus*) gene. It was expected to occur in a clade with goldfish and zebrafish, as all three species occur in the taxon Ostariophysi, within Otocephala, but instead it formed a monophyletic clade with RH1 sequences from the breams (e.g., genera *Pagrus* and *Acanthopagrus*). We suspect the sample was misidentified in the original study (Blackshaw and Snyder, 1997).

### 3.2.4. RH2

Forty-seven species are represented in the RH2 tree (Fig. 5). Contrasting the SWS1, SWS2, and RH1 opsin subfamilies, where duplication appears to be rare, the RH2 opsin subfamily has 21 of the 41 duplication nodes (summarized in Fig. 7).

We focused first on RH2 duplication events within the order Cypriniformes. There were six; two shared by all cypriniforms (RH2-dupII and RH2-dupV), one in the ancestor of carp (*Cyprinus*) and goldfish (*Carassius*) (RH2-dupX), and one at the base of each of the following three lineages, *Danio* (RH2-dupIX), *Zacco* (RH2-dupXII) and *Candidia* (RH2-dupXVI).

The Atlantic herring (*C. harengus*) also has two RH2 genes. Although herring occur in the taxon Otocephala, one of the duplications (RH2\_2) is sister to the Euteleostei fish sequences and the other (RH2\_1) is sister sequence to a clade of four RH2 genes from the scabbardfish, *L. fitchi*. Ayu (*P. altivelis*) also has two RH2 genes that are not sister sequences. Tree reconciliation would attribute both the herring and ayu paralogs to ancestral duplication events and infer the subsequent loss of one paralog in all other species. We suspect the timing of these duplication events were not accurately reflected in this phylogenetic analysis. In the MP trees (not shown), 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 and this is the pattern we present in the summary tree (Fig. 7).

The scabbardfish has three independent RH2 duplication events (RH2-dupXV, RH2-dupXVIII and RH2-dupXXI) and all are in positions inconsistent with taxonomy (see Supplement). Coho salmon (*Oncorhynchus kisutch*) has two RH2 paralogs stemming from a gene duplication event (RH2-dupVII) at the base of the salmonid clade.

The oldest RH2 duplication node (RH2-dupI) 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, we have adopted the more commonly used RH2A–RH2B notation. This tandem duplication (see below) occurred in the ancestor of fish in the taxon Neoteleostei and loss of one of the two paralogs appears to have occurred independently in several lineages. There have been RH2A gene duplications in stickleback (*G. aculeatus*) (RH2-dupXIX), seabreams (genus: *Acanthopagrus*) (RH2-dupXVII), medaka (*O. latipes*) (RH2-dupXIII), turbot (*S. maximus*) (RH2-dupVI) and cichlids (family: Cichlidae) (RH2-dupXIV). Our RH2 tree also shows independent RH2A duplications in Nile tilapia (*O. niloticus*) and the African Great Lake cichlids (e.g., *Pseudotropheus acei*), but a single event is inferred in our all-opsin tree (Supplementary Fig. 6) and in several other analyses (Spady et al., 2006; Shand et al., 2008).

The lanternfish (*Stenobrachius leucopsarus*) has undergone three independent RH2B duplication events (RH2-dupVIII, RH2-dupXI and RH2-dupXX) to produce four RH2B genes (Yokoyama and Tada, 2010). One of these genes is a pseudogene and was not included in our analysis. RH2B from the tuna (*Thunnus orientalis*) and red seabream (*Pagrus major*) are likely to be orthologs of the RH2B genes from other species and their inclusion at the base of the RH2A clade might be explained by long-branch attraction or gene conversion (see online Supplementary material). In the all-opsin anal-

ysis, the tuna RH2B occurred in a monophyletic group with all other RH2B genes, though bootstrap support for this clade was weak (Fig. 5B).

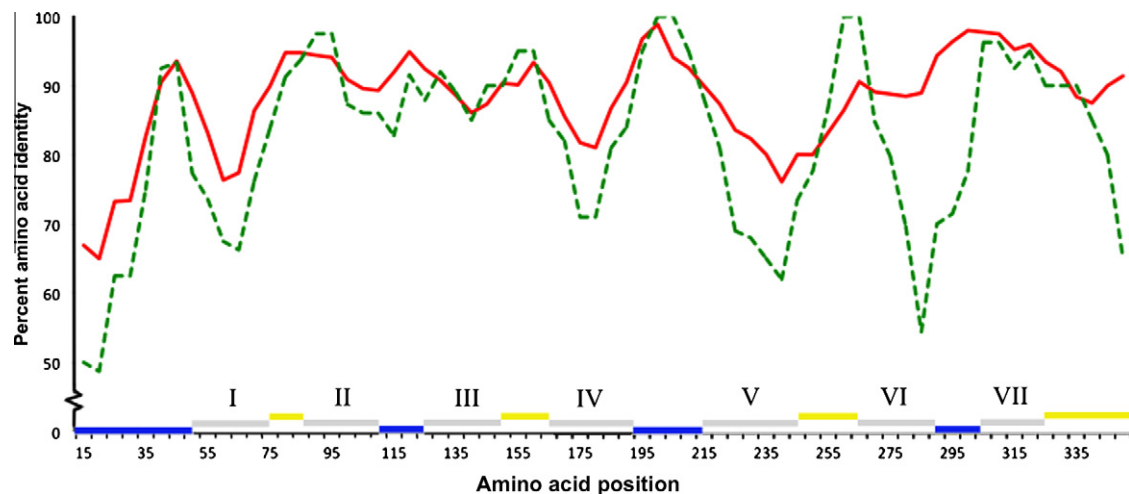
### 3.2.5. LWS

Phylogenetic analysis of fish LWS opsins (sequences from fifty-eight species) produced a tree with a topology that was largely consistent with species-level relationships among teleosts (Fig. 6). Although, there was one especially interesting deviation, which was also noted in the all-opsin analysis: Mexican cave fish (*A. fasciatus*) and neon tetra (*P. innesi*), which are members of the taxon Ostariophysi, possess a gene that is similar to LWS opsins from other ostariophysians (e.g., zebrafish) in the survey. However, both species also have a pair of genes that differ from the LWS opsins found in all other fish. In this study an ortholog of these green opsins was sequenced from golden pencilfish (*N. beckfordi*) cDNA. These sequences have been called green opsins (Register et al., 1994) because they possess the same five key-site haplotype (AHFAA) as one of the human LWS duplicates, which is most commonly called the MWS or green opsin (Nathans et al., 1986). The sister group relationship between this five-gene clade and the LWS opsin genes in other fish places a duplication node (LWS-dupI) at the base of the fish LWS tree, however, we did not find orthologs of these genes in a BLASTn search of any of the whole genome sequences available for ray-finned fish. By including the data from pencilfish, *N. beckfordi*, our LWS opsin gene analysis suggested that the cavefish and neon tetra green LWS opsin duplicates were produced by an event that occurred after the families Lebiasinidae (pencilfish) and Characidae (cavefish and neon tetras) diverged. All five of these genes have diverged from other LWS opsins in the transmembrane 6 (TM6) and extracellular 3 (E3) domains, two regions highly conserved in LWS opsins from other fish (Fig. 8).

Within the family Cypriniformes, there are two LWS opsin duplication events: One produced duplicates in zebrafish (*D. rerio*) (LWS-dupIII), while the other appears to have occurred in the common ancestor of the genera *Cyprinus*, *Carassius*, and *Sinocyclocheilus* (LWS-dupVI). The zebrafish duplicates have a pattern of divergence suggesting that they have been modified by gene conversion (see below). The smelt, *P. altivelis*, has two LWS genes (98% identical) that were sequenced in separate studies and may represent alleles or gene duplicates (LWS-dupVII). The medaka (*O. latipes*) also has a recent LWS opsin duplication event (LWS-dupX). These sequences are 99% identical but are derived from distinct loci (Matsumoto et al., 2006).

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 retro-duplication giving rise to the LWS S180r clade (LWS-dupII) (Ward et al., 2008; Watson et al., 2010a). This occurred after the American flagfish, *J. floridae*, lineage (Family: Cyprinodontidae) diverged from the other species surveyed from Cyprinidontoidei. The next event was a tandem duplication producing the LWS P180 opsins (*J. onca* LWS P180, *A. anableps* S180γ and Poeciliid LWS P180) and LWS S180. Relationships among the paralogs produced by this tandem duplication have been obscured by gene conversion, but are discernable in sequence comparisons involving only the 3' region of the genes within each of the two clades (Owens et al., 2009; Windsor and Owens, 2009). The LWS S180 opsin duplicated independently to produce LWS S180α/LWS S180β in *A. anableps* (LWS-dupXI) and the LWS S180/LWS A180 gene pair in the poeciliids, *Poecilia reticulata* and *X. helleri*. There was also an independent LWS duplication in *J. floridae* (the flagfish).

In summary, 41 visual opsin gene duplication nodes were identified. Duplication events were most prevalent in the RH2 and LWS



**Fig. 8.** Sliding window analysis of 'green' LWS amino acid divergence. The dotted 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, gray 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. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

subfamilies (Fig. 7). In several cases we suspected that duplication nodes were incorrectly localized. In these instances, marked with an asterisk, duplication events were repositioned on our summary tree (Fig. 7). Average TN distances, across all positions, for genes in paralogs clades ranged from 0.011 to 0.346, though the majority were less than 0.1 (Supplementary Table 2).

### 3.3. Gene loss and pseudogenization

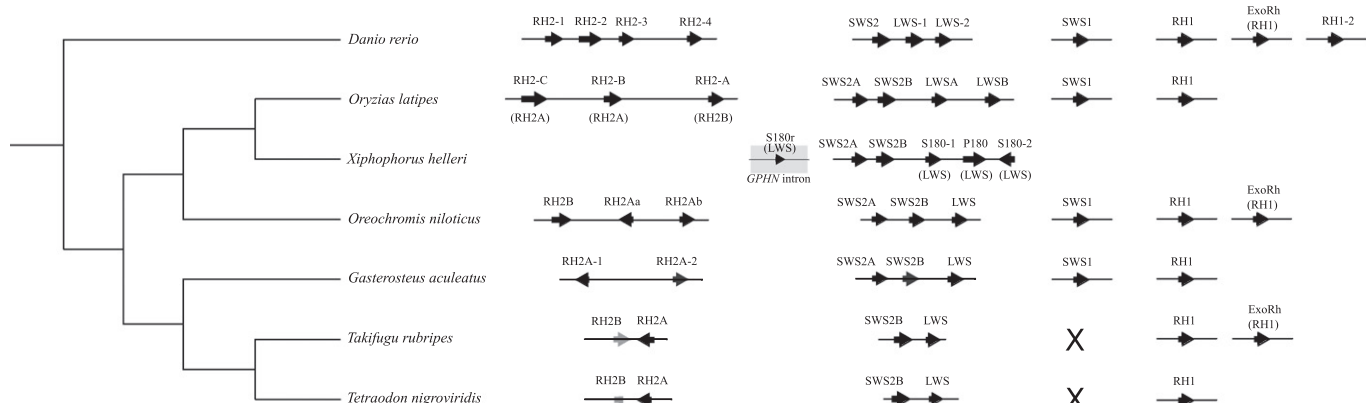
Our phylogenetic analysis indicates that many species are missing genes that were present in their ancestors. Gene loss or an incomplete survey, are equally valid explanations in most cases. However, for a few species opsins have not been detected despite significant effort to locate them (i.e., whole genome sequencing, a PCR survey and/or southern blotting). For example SWS1 and SWS2A genes have not been detected in either of the puffer fish genomes (Neafsey and Hartl, 2005). We looked for, but could not find RH2B in the three-spine stickleback (*G. aculeatus*) genome. Genes, including RH2A, that are adjacent to RH2B in other species were located but RH2B was not. In these three cases, gene loss has occurred long after the duplication events that produced them.

Pseudogenes are genes with disruptive mutations (e.g., internal stop codons, frame shifts, and deletions) that retain enough se-

quence similarity to be detectable by BLAST or by motif-based bioinformatics methods. They have been identified in some species; for example, RH2B is a pseudogene in the pufferfishes, *T. nigroviridis* and *T. rubripes*. Approximately two thirds of RH2B is missing in *T. nigroviridis*. In *T. rubripes*, RH2B is disrupted by the insertion of a long interspersed nuclear element (LINE) and by a deletion (Neafsey and Hartl, 2005). The differences in the types of disruptive mutations that the two pufferfish pseudogenes exhibit and the observation that five other species in the genus *Takifugu* have functional RH2A and RH2B opsin genes, indicate that the RH2B losses in *T. rubripes* and *T. nigroviridis* were recent and independent events (Neafsey and Hartl, 2005). Additionally, we uncovered only exon II of a SWS2B pseudogene in the three-spine stickleback (*G. aculeatus*) genome by searching the region of its genome downstream from SWS2A, a region that contains SWS2B in other species.

### 3.4. Mechanisms of gene duplication

Tandem duplication appears to be the most common mode of opsin gene family expansion in fishes. Of the 15 duplication events in species with genomic resources, 12 are tandem duplication events (Fig. 9). The SWS2 paralogs, produced in the ancestor of Neoteleostei (SWS2-dupl) occur next to one another in a head to



**Fig. 9.** 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. Gray arrows are pseudogenes. X represents gene not present.

tail orientation in the Nile tilapia (*O. niloticus*), medaka (*O. latipes*), green swordtail (*X. helleri*) and the three-spine stickleback (*G. aculeatus*). RH2A and RH2B occur in an inverted (tail to tail) orientation in the two pufferfishes and the cichlid, *Oreochromis niloticus*. Several additional tandem duplication events have taken place at this locus. After the first tandem duplication event that produced RH2A and RH2B, RH2A was again tandemly duplicated in *O. niloticus* (RH2-dupXIV), leaving this lineage with three linked RH2 genes (RH2B, RH2Aa and RH2Ab). In stickleback, RH2B was lost (see above) and RH2A was tandemly duplicated. These stickleback RH2A paralogs (RH2A-1 and RH2A-2) are oriented in a head to head pattern (Fig. 9). They are 99% identical in coding region, suggesting that this second tandem duplication event was very recent. However, 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-labeled. 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 led to the three-gene repertoire and orientation present in medaka (Fig. 9). The single copy RH2 gene in the ancestor of zebrafish experienced three tandem duplication events producing the four-gene array present in this species (Fig. 9).

Tandem duplication is also the major contributor to LWS opsin gene subfamily amplification. Between-gene PCR and sequencing, as well as, screening of large BAC clones indicates that the LWS S180-2 (LWS S180 in guppies) and LWS P180 genes in livebearers are in an inverted position (tail to tail orientation) and that LWS S180-1 (A180 in guppies) is upstream of LWS P180 (Ward et al., 2008; Watson et al., 2010a,b). Zebrafish and medaka also possess LWS opsin duplicates linked in a head to tail orientation (Fig. 9).

Whole genome duplication can also cause gene family expansion. It has been suggested that the LWS-dupVI and SWS2-dupII duplications in Cyprininae are a consequence of tetraploidy (Li et al., 2009). An RH2 gene duplication event in salmonids (RH2-dupVII) may also be a result of tetraploidy, however there are currently no linkage data to confirm this hypothesis (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 retro-duplication. However, this form of duplication overall appears to be relatively rare.

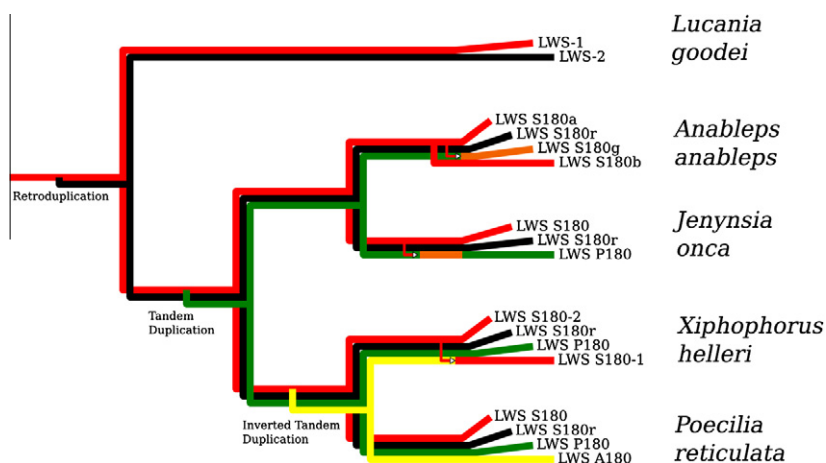
### 3.5. Gene conversion

There appear to be at least three examples of gene conversion between tandem duplicates in this study, all of which occur in the LWS opsin subfamily. Although we proposed that a tandem duplication event produced the P180 and S180 LWS opsins in the ancestor of anablepids and poeciliids, the phylogenetic analysis did not produce a tree consistent with this hypothesis. Sequence similarity among P180 orthologs is limited to a region at the 3' end of the gene that is approximately 243 bp long. When full-length sequences are analyzed a tree is generated that infers the independent evolution of P180 genes in guppies, one-sided livebearers and in the four-eyed fish (see Windsor and Owens, 2009). Independent gene conversion events in the two anablepids, where the S180 gene has over-written most of the P180 gene, have disrupted the phylogenetic signal (see Fig. 10). Also, Watson et al. (2010b) argued that an LWS A180 gene in *X. helleri* was over-written by an adjacent LWS S180. Finally, *D. rerio* LWS gene paralogs exhibit high sequence similarity (95%) except in exon I where LWS-1 and LWS-2 are only 57% identical. It is possible that the pair is a product of a tandem duplication that is older than the tree indicates; a tree produced using only LWS first exon of both sequences places the *D. rerio* LWS-2 outside the cyprinid clade (data not shown).

### 3.6. Non-synonymous and synonymous substitutions

There were far fewer non-synonymous substitutions per non-synonymous site than there are synonymous substitutions per synonymous site in all opsin subfamily trees indicating that opsins are under strong purifying selection (Table 1). However, average values can conceal interesting patterns of DNA sequence substitution along specific branches or in regions within a gene. To investigate these possibilities we compared  $\omega$  values among subsets of branches within opsin gene subfamilies and among codons. When examining large gene duplication clades in the RH2 and SWS2 subfamilies, we detected a significant increase in the  $\omega$  in only the RH2B clade ( $p = 1.07E-12$ ) (Supplementary Table 3).

We then tested the hypothesis that genes diverged faster after duplication events than they did after speciation events by allowing post-duplication branches and post-speciation branches to have an independent  $\omega$  values. This produced significant likelihood gains in the SWS1 ( $p = 1.49E-4$ ), LWS ( $p = 9.61E-3$ ) and RH2



**Fig. 10.** Hypothesized relationships of livebearer LWS opsin duplicates. Each path represents one opsin gene. Duplication events are represented by additional paths originating from their progenitor gene. They are labeled with duplication event type when available. White arrows represent gene conversion events. Orange bars are partially converted genes, in the case of *J. onca* LWS P180, the P180 haplotype was regained. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



**Table 1**

Average  $\omega$  for opsin subfamilies, subclades and branch types.  $\omega$  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  $\omega$  from all other branches (Yang, 2007).

Gene	Overall	Branch type	Subclades	
SWS1	0.113	Gene duplication	<b>0.232</b>	–
		Speciation	<b>0.113</b>	–
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	<b>0.211</b>	RH2A 0.126
		Speciation	<b>0.114</b>	RH2B <b>0.173</b>
LWS	0.118	Gene duplication	<b>0.153</b>	–
		Speciation	<b>0.111</b>	–

( $p = 1.07E-12$ ) subfamilies. In each case,  $\omega$  was higher for the gene duplication branches (Table 1 and Supplementary Table 3).

We also looked for positive selection on specific branches. We used this to test the hypothesis that following gene duplication, opsin genes have codons that are under positive selection for spectral divergence. A total of 65 post-duplication branches were tested individually and eleven (17%) were found to have evidence for codons evolving under  $\omega > 1$  (Supplementary Table 3). Of those, two indicated that a key spectral tuning site was under positive selection; sites 46 and 86 on branch SWS1-dupla and site 195 on branch RH1-duplla. Selected branches are indicated in Supplementary Figs. 1–5.

### 3.7. Key-sites

Key-sites are residue positions that have been demonstrated through site-directed mutagenesis to have a significant effect on spectral sensitivity (e.g., Yokoyama and Radlwimmer, 1998) (Fig. 1). At many of these sites there are a few amino acids that 'toggle' back and forth on the phylogeny.

We found three compelling examples of convergent evolution at opsin key-sites. The first is in the RH2 opsin subfamily, where the key-site substitution E112Q has occurred 15 times in ray-finned fish (Fig. 5) and results in a 20 nm blue shift (Sakmar et al., 1989). In the RH1 subclass a D83N substitution has occurred 12 times among ray-finned fish (Fig. 4). This substitution shifts lambda max 6 nm toward blue region of the spectrum (Nathans, 1990). These substitutions are reversions to what appears to be the ancestral residue, based on its occurrence in *G. australis* RhA, *Petromyzon marinus* RhA, *Lethenteron japonicum* RhA, and homologs in *Xenopus laevis* and *Anolis carolinensis*. Within the LWS opsin subfamily, three of the five key-sites are polymorphic in ray-finned fish (Fig. 6). Six S180P mutations have occurred in fish, including one in *J. onca* that followed a gene conversion event, which had replaced a proline with a serine from the adjacent donor locus (Windsor and Owens, 2009). An S180P substitution has also been reported in the lamprey LWS opsin. S180P shifts lambda max –19 nm (Davies et al., 2009a). S180A substitutions have occurred nine times in ray-finned fish. This mutation also occurred after LWS gene duplication in hominids (Nathans et al., 1986) and it shifts lambda max by –7 nm (Yokoyama and Radlwimmer, 1998), which is why the second LWS gene in humans is often referred to as MWS.

## 4. Discussion

Our phylogenetic analysis of vertebrate visual opsin genes produced a tree with five monophyletic clades, each with representatives from lamprey (Petromyzontiformes), tetrapods (Sarco-

pterygii) and ray-finned fish (Actinopterygii). This is consistent with the hypothesis that gene duplication events in the common ancestor of vertebrates produced the five-subfamilies of visual opsins. Kuraku et al. (2009) pointed out that these observations support the hypothesis that two rounds of whole genome duplication (the 2R hypothesis) occurred prior to the divergence of the jawed and jawless fishes. It is clear that pinopsins are very similar to the visual opsins, indeed, our analyses uncovered some support (two synapomorphies) for the hypothesis that pinopsins and LWS opsins are sister groups. The all-opsins analysis also showed that all fish genes, including the green LWS genes from cavefish, neon tetras and pencilfish, were correctly labeled. While confirming gene identity might seem overly precautionary, these so-called green LWS genes had not previously been analyzed with such a large set of homologs. The relationships among fish opsin sequences in our analysis were largely consistent with fish taxonomy. Exceptions were reported in Section 3 and are discussed briefly in the online Supplementary materials.

Forty-one opsin gene duplication events were identified on the opsin subfamily trees. We make five major conclusions from our survey and analysis: (1) the large number of opsins in many ray-finned fishes was produced by duplication events spanning the age of the taxon and the retention of old duplicates contributes to large repertoires as much as the generation of new duplicates. (2) Tandem duplication produced more opsin gene duplicates in fish than any other mode of duplication. Tandem duplication followed by gene conversion appears to facilitate homogenization in some instances, and diversification in other instances. Retro-duplication also contributed to opsin gene amplification and diversification, while the ancient whole genome duplication in fishes (3R) (Taylor et al., 2003) appears to have had no contribution. (3) Gene duplication is most prevalent in the RH2 and LWS opsin subfamilies. (4) Opsin gene duplication appears to enhance amino-acid level evolutionary change, however, sequence variability is constrained at many key-sites. This conclusion is based upon the observation that the  $dN/dS$  ratio increases on branches following opsin gene duplication but that there are also a large number of parallel substitutions in each of the opsin subfamily clades. (5) While opsin gene loss and sequence evolution appears to be correlated with changes in the spectral environment, especially water depth, connections between opsin repertoire size and variation in habitat or life history are not obvious.

### 4.1. Opsin gene duplication events span the age of the ray-finned fish taxon

Those fish with large opsin gene repertoires have retained duplicates derived from events that span the evolution of the ray-finned fish. We emphasize this observation by comparing repertoire evolution in guppies, cichlids and zebrafish. The guppy (*P. reticulata*) has 10 visual opsins and cichlids have eight. Both species have an RH2 opsin gene pair derived from a duplication event that occurred approximately 229 ( $\pm 29$ ) MYA (Spady, 2006) in a fish that gave rise to all acanthopterygians. Guppies and cichlids also have two SWS2 opsins derived from genes produced in the common ancestor of all species in Acanthomorpha. This fish lived about 198 ( $\pm 8$ ) million years ago (MYA) (Spady, 2006). The four LWS opsins in guppy are products of relatively recent duplication events. The first, which occurred about 72 MYA (Spady, 2006), was a retro-duplication in the ancestor of Cyprinodontoidae, a taxon that includes killifish and livebearers, but not cichlids. The second was a tandem duplication in the fish that gave rise to species in the sister families Poeciliidae and Anablepidae and the third (another tandem duplication event) produced the LWS A180 and LWS S180 opsins in Poeciliidae before the divergence of the genera, *Poecilia* and

*Xiphophorus*. Cichlids have only one LWS opsin, but the repertoire of this taxon was enlarged after it diverged from the ancestor of guppies by an additional RH2 gene duplication. Zebrafish is another species with a large number of opsin genes. This 10-gene repertoire is a product of a different set of duplication events in the Otocephala lineage across a similar time span to those in Acanthopterygii. Zebrafish have four RH2 genes; the first of the duplication events that produced this RH2 clade occurred in the common ancestor of all cypriniforms. The second RH2 duplication event produced a gene pair so far only detected in *D. rerio* and *Z. pachycephalus* and the last RH2 duplication event generated genes that have been found only in *D. rerio*. Zebrafish possess two LWS genes, the products of a duplication event that appears to have taken place ~14 MY ago (Spady, 2006), although higher levels of divergence in exon I hint at an older date (see Section 3). Zebrafish also have two single-exon RH1 genes. The timing of this duplication is difficult to determine due to a lack of paralogs in other species but recent analysis has suggested it occurred around 140 MYA (Morrow et al., 2011).

The pufferfish, *T. rubripes* and *T. nigroviridis* have comparatively small opsin gene repertoires. The topology of our trees (and an analysis by Neafsey and Hartl, 2005) indicates that this is more a consequence of gene loss, than failure to duplicate: The ancestor of all puffer fish had an SWS1, two SWS2, an RH1, two RH2 and an LWS opsin. SWS1 and SWS2A were lost in their common ancestor and RH2B became nonfunctional in each species independently (Neafsey and Hartl, 2005; Hurley et al., 2007). Gene loss of RH2B has also been observed in three-spine stickleback (*G. aculeatus*) and many Antarctic ice fish have lost their LWS genes (Pointer et al., 2005).

The oldest LWS duplication and one of the more interesting events is LWS-dupl. Paralogs from this duplication, known as 'green' LWS, have only been found in the three species in the family Characiformes, yet the level of divergence between paralogs suggests the duplication occurred long before the emergence of this lineage. There are two possible explanations for this observation; the 'green' LWS evolve at a faster rate than other opsins and the duplication is not as old as it seems or the duplication is ancient and orthologs have been lost, or not yet detected, in all other fish lineages.

Interestingly, genes in the LWS 'green' clade differ most from other opsins in the TM6 and E3 domains (Fig. 8), two regions that are typically highly conserved. TM6 plays a key role in G-protein binding and provides an opening for retinal to enter the binding pocket (Park et al., 2008; Scheerer et al., 2008). This hints that these LWS opsins might interact with a different G-protein to coincide with their unusual skin expression pattern (Kasai and Oshima, 2006).

#### 4.2. Tandem duplication is the most common form of opsin duplication in fish

The completion of several whole genome and large-insert BAC sequencing projects, allowed us to examine not only gene sequences but also location and orientation of opsins in seven fish genomes. This provided insight into the mechanisms of opsin gene duplication and rearrangement. For seven species with genomic data, there were 15 duplication events, 12 of which were tandem. Preferential survival of tandem duplicates might reflect the apparent ability of opsin gene regulatory modules or locus control regions (LCRs) to exert long-range influence. In humans, one LCR regulates the expression of many downstream LWS opsins and expression is negatively correlated with the distance between the LCR and the genes (Winderickx et al., 1992). This trend has also been observed in zebrafish and guppies where one LCR appears to drive expression of SWS2 and LWS duplicates and does so in a

manor that is correlated with distance (Tsujimura et al., 2007; Laver and Taylor, 2011); however, this is not the case for medaka (Matsumoto et al., 2006).

A consequence of tandem duplication, in addition to the potential for co-regulation, is that it facilitates conversion between paralogs. Gene conversion occurs when one gene is over-written by a similar, and often neighboring, 'donor'. If the so-called donor is functional, conversion reduces genetic variation by generating two copies of one functional gene where two distinct genes formerly existed. If the donor is a pseudogene, conversion can eliminate function; though there are examples where conversion between pseudogenes and functional genes generates functional diversity (e.g., chicken immunoglobulins) (McCormack et al., 1991). Homogenization and diversification have been observed in human opsins (Reyniers et al., 1995; Winderickx et al., 1993). Many of the gene conversion events we observed in fish appear to be incomplete: In zebrafish, exon I in one of the LWS paralogs was not overwritten. In the livebearers, the 3' ends of the LWS P180 opsins have been spared. However, Watson et al. (2010b) recently argued, using synteny data, that the LWS A180 opsin was completely overwritten by the LWS S180 gene in the green swordtail. It is also possible, indeed likely, that other opsin duplicates that appear to have been produced independently in different lineages are the products of older (shared) events that have been disguised by gene conversion. In *A. anableps*, the four-eyed fish, conversion has generated a unique LWS opsin five key-site haplotype, SHYAA, which is a chimera generated from SHYTA and AHFAA.

The opsin gene repertoires of fish have also been expanded by retro-duplication. This occurs when the mRNA of a 'parental' gene is reverse transcribed into cDNA and is then inserted into chromosomal DNA at a distinct location (Brosius, 1999). Retro-duplication typically creates an intron-less paralog of the parental gene. Although opsin retro-duplication appears to be rare, the prevalence of this mode of duplication cannot be estimated precisely because most opsin sequences in NCBI are derived from processed mRNA, and therefore, few data on the presence or absence of introns are available.

Successful duplication requires the maintenance or acquisition of gene regulatory modules and this has generally been considered unlikely with retro-duplication. However, it is now clear that proximal promoters can be retained or acquired at the insertion site by adopting those used by neighboring genes (Okamura and Nakai, 2008). The opsin retro-genes RH1 and LWS-S180r are expressed in photoreceptors (Raymond et al., 1993; Rennison et al., 2011), thus they have retained or acquired eye-specific enhancers. For LWS-S180r, eye-specific regulatory modules might have been carried in the first intron, which survived retro-transposition (Watson et al., 2010a). It is also possible that the insertion of LWS S180r into intron X1 of the gephyrin gene (Watson et al., 2010a) contributes to its expression pattern. While retro-duplication is possible in any tissue, only events that take place in germ cells (sperm and eggs) will produce retro-genes that are passed from one generation to the next. This suggests that germ cells may be among the many non-ocular tissue recently shown to express visual opsins (e.g., Kasai and Oshima, 2006).

It appears that whole genome duplication events have had only a minor role in the expansion of the opsin repertoires of ray-finned fish. We find this surprising given that almost all species in our survey are descended from a tetraploid ancestor (Taylor et al., 2003; Hoegg et al., 2004). Opsin duplicates in the LWS & SWS2 subfamilies in members of the subfamily Cyprininae and RH2 in Salmonids might have been produced by whole genome duplication. This speculation is based upon the position of these gene duplication events relative to known genome duplications events (Li et al., 2009; Temple et al., 2008).



#### 4.3. Opsin gene duplication is most prevalent in the RH2 and LWS subfamilies

Opsin gene duplicates are not evenly distributed among subfamilies. Gojobori and Innan (2009) hypothesized that duplicates of 'boundary opsins', those opsins encoding proteins sensitive to wavelengths at either end of the visible spectrum (i.e., SWS1 and LWS), are retained less often than middle wave opsins (RH2 and SWS2). We did not observe this pattern. While a large number of fish possess SWS2 duplicates, almost all of these genes were produced by a single event. By counting duplication nodes rather than genes, we concluded that half of all opsin duplications in fish occurred in the RH2 subfamily and that the majority of the remaining duplication events occurred in the LWS subfamily. The retention of RH2 and LWS duplicates might be a consequence of the fact that they are expressed in double cones (Hisatomi et al., 1997; Vihtelic et al., 1999). Although double cones appear to be involved in a number of visual processes including wavelength discrimination, as well as, the detection of polarized light, luminance and movement (Pignatelli et al., 2010; Wagner, 1990; Cameron and Pugh, 1991; Osorio and Vorobyev, 2005), their role in background matching is well established (Loew and Lythgoe, 1978). Background light in most aquatic environments ranges from 470 to 600 nm (shorter wavelengths are absorbed by organic and inorganic material) and this, coincidentally, overlaps with the region of the spectrum that RH2 and LWS opsins are most sensitive to. Given that the transmission of light through water is highly dependent on water quality, having a diversity of opsins in these two subfamilies might allow a fish to better match background illumination in a heterogeneous spectral environment (e.g., over time and/or space). SWS1 and SWS2 opsins, which are expressed in single cones, do not appear to be used for background matching. Consequently, we suspect that the duplication and divergence of SWS1 and SWS2 genes is less likely to be favored by natural selection. It will be interesting to see if other aquatic vertebrates, that have double cones, have also retained RH2 and LWS opsin duplicates. The only full opsin repertoire of a cartilaginous fish, *Callorhynchus milii*, shows that there has been duplication in the LWS subfamily of this species (Davies et al., 2009b).

#### 4.4. Gene duplication and divergence

The potential for gene duplication to provide raw material for protein-level innovation has been discussed for almost 100 years (reviewed by Taylor and Raes, 2005). Several methods have been developed to detect changes in the rate of amino acid evolution. We used PAML (Yang, 2007) to calculate  $\omega$ , the ratio of non-synonymous per non-synonymous site, to synonymous substitutions per synonymous site. We also characterized amino acids substitutions (among orthologs and paralogs) at sites known to influence spectral sensitivity.

Previous attempts to study visual opsins using PAML have produced little evidence for positive selection (Yokoyama et al., 2008; Nozawa et al., 2009). In these cases, sites predicted to be under positive selection were not known to cause changes in  $\lambda_{\max}$  and those known to cause  $\lambda_{\max}$  shifts were not identified. We found that the average  $\omega$  for opsin subfamilies ranged from 0.074 (RH1) to 0.167 (SWS2) indicating that opsins are under strong purifying selection. Having estimated  $\omega$  for all opsins within each subfamily, we looked for an increase in  $\omega$  associated with ancient gene duplications in the SWS2 and RH2 subfamilies. The average  $\omega$  for RH2B genes was 0.173, whereas mean  $\omega$  for the RH2A and pre-duplication RH2 genes was 0.123. This increase in  $\omega$  appears to be associated with changes in wavelength sensitivity, as RH2B genes tend to be sensitive to shorter wavelengths ( $\lambda_{\max}$  = 452–488 nm) than RH2A genes ( $\lambda_{\max}$  = 492–555 nm). We also compared  $\omega$  values

along branches that follow duplication to average values for branches that follow speciation. In the SWS1, RH2, and LWS subfamilies there appears to be a relaxation of purifying selection following gene duplication. This trend has been seen in a variety of other genes and organisms (Kondrashov et al., 2002).

As mentioned, an increase in  $\omega$  might reflect a change from purifying selection to neutral evolution or a change to positive selection. However, there is less ambiguity when  $\omega$  exceeds 1.0. Most authors accept that positive selection is occurring when  $\omega$  exceeds 1.0. Such values have only rarely been observed for whole genes or even domains within genes, but when  $\omega$  is estimated for individual codons,  $\omega$  values >1.0 are not uncommon. We surveyed fish opsins for codons with  $\omega$  values >1.0 using the branch-site models in PAML, as in most cases, codon-level positive selection does not occur in all branches of a phylogeny (Bielawski and Yang, 2001; Raes and Van de Peer, 2003).

The branch-site model searches for codons under positive selection on specific branches (Zhang et al., 2005). We examined post-duplication branches and found eleven (17%) had codons evolving under positive selection. Five of the branches identified had codons with  $\omega$  = 999 (the maximum value). These are likely statistical errors stemming from a lack of synonymous substitutions (Hughes and Friedman, 2008). For the other seven branches,  $\omega$  estimates varied from 9.1 to 242.9 for a small subset of the codons (<5%). In six 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 rule out the hypothesis that there are adaptive advantages to these substitutions. In cases where these branches lead to a spectrally novel duplicate (particularly RH2-dupVII in salmonids and LWS-dupV in livebearers), these sites may provide targets for future mutagenesis experiments.

Another branch of interest is the one connecting the SWS1-dupla node and the ayu SWS1-1 gene, which has two key-site codons that appear to have been under the influence of positive selection. Changes at these two codons lead to the following amino acid substitutions; F46T and F86A. In mammals, F46T, in conjunction with several other SWS1 mutations, causes a shift in maximal absorption from the UV to the violet region of the spectrum (Shi et al., 2001). While F86A has not been empirically tested, F86L has been found to contribute to a red shift in mammals (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  $\omega$  value of 0.09. This hypothesis needs to be confirmed by *in vitro* reconstitution and mutagenesis experiments to measure the lambda 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 positive selection for a shift in lambda max seen in a vertebrate opsin.

Comparisons across species show many examples of key-site 'toggling' and this suggests that only a few mutations are accepted at these positions. Though many of these substitutions have been shown to alter spectral sensitivity, this does not necessarily mean they are adaptive. It is possible, for example, that a fish with an LWS opsin with the SHYTA five key-site haplotype sees food, predators, or mates no better than a fish with the AHYTA haplotype in a spectrally heterogeneous environment. Alternatively, these recurring substitutions might represent examples of independent

adaptation to the same ecological (*i.e.*, spectral) challenges (*e.g.*, Simpson, 1953; Endler, 1986). In this sense, key-site toggling could be evidence of the colonization of similar spectral environments by multiple independent species (*i.e.*, a signature of convergent evolution). Interestingly, opsin gene duplication is almost always associated with changes at key spectral tuning sites among paralogs. Furthermore, there are several instances where the same, or very similar, opsin gene pairs have evolved independently after duplication. For example, the key site haplotype AHFAA has evolved from SHYTA twice, once in the howler monkey and once in great apes (Jacobs et al., 1996) and PHFAA has evolved from SHYTA after gene duplication in guppies (Ward et al., 2008). Thus, the pattern of recurrent key-site substitution among paralogs and orthologs in the ray-finned phylogeny could represent functional constraint or a signature of the action of natural selection. However, direct testing of these hypotheses is required before any conclusions can be made.

#### 4.5. Opsins duplication and adaptation

In a recent review, Osorio and Vorobyev (2008) expressed astonishment that most birds have the same opsin gene repertoires. It is hard to imagine, they remarked, that a sea bird such as a shearwater might use color vision that, at the receptor level, is comparable to a peacock. Bees and ants also have the same set of opsin genes. The implications of this, commented Osorio and Vorobyev, is that opsin gene repertoires in birds and insects have not evolved in response to changes in life history. At first glance, the story seems to be similar in fishes; opsin gene repertoires vary to a much greater extent in fishes, but in only a few instances (see below) are correlations between this genetic variation and variation in habitat, life history, or behavior clear.

An observation that has been made by several authors is that gene repertoires in deep-water fishes differ from those living closer to the surface. Several deep-water species (those living 200 m below sea level) have lost their LWS genes. In addition, the opsins that have been retained in these fishes are often more sensitive to blue light than their orthologs in other species (Douglas and Partridge, 1997; Partridge et al., 1989). For example, both gene loss and blue shifts have been seen in the cottoid fish of Lake Baikal, which live at depths of up to 1000 m (Hunt et al., 1997; Cowing et al., 2002). Another putative example of a correlation between opsin repertoire and environment occurs in Antarctic fish. Short and long wavelength light is filtered by sea-ice and this generates a spectrum similar to that which is experienced by deep-water species (Littlepage, 1965; Pankhurst and Montgomery, 1989). As in some deep-water species, the LWS opsins in many Antarctic species have been lost (Pankhurst and Montgomery, 1989; Perovich et al., 1998; Pointer et al., 2005). The scabbard fish appears to be an exception. This species generally lives between 100 m and 250 m below the surface (Nakamura and Parin, 1993), yet it has a surprisingly large opsin repertoire given the narrow range of wavelengths that reach to that depth. Regular migrations toward the surface (Nakamura, 1995) and/or the detection of bioluminescence (Douglas et al., 1998) may have played a role in the evolution of a large opsin gene repertoire in this deep-water species.

Hoffmann et al. (2007), Weadick and Chang (2007) and Ward et al. (2008) all suspected that the large repertoire in guppies might play a role in color-based sexual selection. However, we now know there are almost as many opsin genes in guppy relatives that are not colorful, including the one-sided livebearer and four-eyed fish. Additionally, the more distantly related zebrafish, not typically thought of as a colorful species, possesses more opsins than the guppy.

With these observations in mind, and following our hypothesis that opsin diversity within species (*i.e.*, when it occurs among par-

alogs) is adaptive, we propose that opsin gene duplication and divergence is driven by environmental heterogeneity. This heterogeneity can occur at different levels: in the aquatic environment, up-welling light often differs in spectral properties from down-welling light due to the filtering effects of water and several studies have reported data showing that opsin gene expression differs among regions of the retina (*e.g.*, Takechi and Kawamura, 2005; Owens et al., 2011; Rennison et al., 2011). Also, as mentioned above, light transmission in water is influenced by organic and inorganic material in the water. Thus, heavy rain, which occurs on a daily basis in many tropical environments, is likely to have a dramatic effect on the spectral environment for fishes. Differences in opsin gene expression have been associated with the spectral environment in killifish (Fuller et al., 2005). Additionally, some species migrate between aquatic environments with very different spectral properties (*e.g.*, eels (Archer et al., 1995) and salmon (Temple et al., 2008). The loss of opsins in species that live in comparatively homogeneous environments (deep sea and under ice) is consistent with this idea. These observations also indicate that it is important to take opsin expression patterns into consideration when considering adaptive explanations of large opsin repertoires in fish.

#### 4.6. Future directions of opsin gene research in ray-finned fish

**Repertoires:** The future of opsin research is bright, as new sequencing technology and new genome projects seem poised to increase the opsin repertoires of public databases (Haussler et al., 2009). Whole genome sequencing has the advantage of finding pseudogenes, non-expressed genes, and cryptic paralogs, often missed by PCR-based surveys (see Watson et al., 2010a).

**Mechanisms of duplication:** Whole genomes also provide gene structure and orientation data, which will allow us to test the hypothesis that tandem duplication is the most prevalent method of opsin duplication, in a much larger dataset. These new genomes, if obtained from a diversity of species, will improve our ability to reconstruct opsin gene trees and therefore, characterize duplication events and patterns of sequence evolution.

**Expression (transcriptomics):** As we have discussed above, opsin repertoire tells us only half the story and expression information is essential for understanding how that repertoire is used. The traditional method of opsin quantification is RT-QPCR and has been used to great effect in several species (*e.g.*, O'Quin et al., 2010; Laver and Taylor, 2011). In the future, RT-QPCR may be supplanted by sequencing based methods such as retinal or even single-cell transcriptomics (Tang et al., 2009), which allow for absolute quantification of all opsin genes as well as every other mRNA present. Opsin expression patterns in ray-finned fish have also been found to exhibit intra-retinal variability (*e.g.*, Takechi and Kawamura, 2005), in the future we may be able to determine what factors contribute to this variation and whether it is plastic. Behavioral assays will also be important in determining correlations between wavelength discrimination and sensitivity and intra-retinal opsin expression variability.

When these new data become available, we will be poised to answer important biological questions that have only been studied with small and species-restricted datasets. To what extent are opsin number and sequence correlated with the spectral environment? Intriguing trends have been found in lake Baikal cottoid fishes, but do these trends also occur in the ocean with taxonomically diverse fish groups (Cowing et al., 2002)? Are opsin duplication events associated with increased diversification or life history shifts? While two of the largest fish orders (Perciformes and Cypriniformes) have gene duplications at their base, the fact that overall sampling is both phylogenetically sparse and individually incomplete (due to biases from PCR surveys) does not allow

us to answer these questions. Do changes in opsin genes play a role in sexual selection or species recognition? This has been explored in cichlids and livebearers but many other taxa can be surveyed, for example stickleback, reef fish and in deep sea fish the role of bioluminescence could be examined.

## 5. Conclusions

Large opsin gene repertoires in ray-finned fish are not the result of a single duplication event (e.g., whole genome duplication) in their shared common ancestor, nor are they the product of many independent events near the tips of the fish phylogeny. Rather, large opsin repertoires reflect the gradual accumulation of new genes, generated largely by tandem duplications at fairly regular intervals over the past 250 million years. The distribution of tandem duplication events has not been even among the five opsin subfamilies; duplication events have been most numerous in the RH2 and LWS subfamilies, which tend to be expressed in double cones. Gene conversion among tandem duplicates has obscured some evolutionary relationships. However, these relationships can be exposed when phylogenetic analyses consider different regions of a gene separately (e.g., a sliding window approach) and/or by considering gene location when identifying orthologs and paralogs. Though largely a mechanism that reduces variation, gene conversion can and has generated unique opsin gene sequences (with novel key-site haplotypes). We found that  $dN/dS$  values were higher on the branches of our trees that followed gene duplication than on branches that followed speciation events, suggesting that duplication relaxes constraints on opsin sequence evolution. However, we also show that key-sites have limited variability and toggle through their amino acid options across the phylogeny; this could be neutral variation or be a signature of the action of natural selection. The observation that within species, opsins tend to diverge at key sites suggests that many of these key site substitutions are adaptive when they occur among paralogs. When all available opsin data is considered together for ray-finned fish, there are surprisingly few clear connections between opsin gene repertoires and variation in spectral environment, morphological traits, or life history traits that emerge. We speculate that the expanded and phenotypically diversified opsin repertoires of many ray-finned fish reflect the spatial and temporal environmental heterogeneity that these species experience.

## Acknowledgements

The authors thank referees for helpful comments and acknowledge support from an NSERC Discovery Grant (JST) and two University of Victoria graduate fellowships (DJR and GLO).

## Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.ympev.2011.11.030](https://doi.org/10.1016/j.ympev.2011.11.030).

## References

- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D., 1990. Basic Local alignment search tool. *J. Mol. Biol.* 215, 403–410.
- Alvares, C.E., 2008. On the origin of arrestin and rhodopsin. *BMC Evol. Biol.* 8, 222.
- Amores, A., Force, A., Yan, Y.L., Joly, L., Amemiya, C., Fritz, A., Ho, R.K., Langeland, J., Prince, V., Wang, Y.-L., Westerfield, M., Ekker, M., Postlethwait, J.H., 1998. Zebrafish hox clusters and vertebrate genome evolution. *Science* 282, 1711–1714.
- Anisimova, M., Gascuel, O., 2006. Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. *Syst. Biol.* 55, 539–552.
- Anisimova, M., Yang, Z., 2007. Multiple hypothesis testing to detect lineages under positive selection that affects only a few sites. *Mol. Biol. Evol.* 24, 1219–1228.
- Anisimova, M., Bielawski, J.P., Yang, Z., 2001. Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Mol. Biol. Evol.* 18, 1585–1592.
- Archer, S., Hope, A., Partridge, J.C., 1995. The molecular basis for the green-blue sensitivity shift in the rod visual pigments of the European eel. *Proc. Biol. Sci.* 262, 289–295.
- Arendt, D., Tessmar-Raible, K., Snyman, H., Adriaan, W.D., Wittbrodt, J., 2004. Ciliary photoreceptors with a vertebrate-type opsin in an invertebrate brain. *Science* 306, 869–871.
- Bielawski, J.P., Yang, Z., 2001. Positive and negative selection in the DAZ gene family. *Mol. Biol. Evol.* 18, 523–529.
- Blackshaw, S., Snyder, S.H., 1997. Parapinopsin, a novel catfish opsin localized to the parapineal organ, defines a new gene family. *J. Neurosci.* 17, 8083–8092.
- Bowmaker, J.K., Knowles, A., 1977. The visual pigments and oil droplets of the chicken, *Gallus gallus*. *Vis. Res.* 17, 755–764.
- Brosius, J., 1999. RNAs from all categories generate retrosequences that may be exapted as novel genes or regulatory elements. *Gene* 238, 115–134.
- 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.
- Collin, S.P., Trezise, A.E., 2004. The origins of colour vision in vertebrates. *Clin. Exp. Optom.* 87, 217–223.
- Cowing, J.A., Poopalasundaram, S., Wilkie, S.E., Bowmaker, J.K., Hunt, D.M., 2002. Spectral tuning and evolution of short wave-sensitive cone pigments in cottoid fish from Lake Baikal. *Biochemistry* 41, 6019–6025.
- Cowing, J.A., Arrese, C.A., Davies, W.L., Beazley, L.D., Hunt, D.M., 2008. Cone visual pigments in two marsupial species: the fat-tailed dunnart (*Sminthopsis crassicaudata*) and the honey possum (*Tarsipes rostratus*).
- Davies, W.L., Cowing, J.A., Carvalho, L.S., Potter, I.C., Trezise, A.E., Hunt, D.M., Collin, S.P., 2007. Functional characterization, tuning, and regulation of visual pigment gene expression in an anadromous lamprey. *FASEB J.* 21, 2713–2724.
- Davies, W.L., Collin, S.P., Hunt, D.M., 2009a. Adaptive gene loss reflects differences in the visual ecology of basal vertebrates. *Mol. Biol. Evol.* 26, 1803–1809.
- Davies, W.L., Carvalho, L.S., Tay, B.H., Brenner, S., Hunt, D.M., Venkatesh, B., 2009b. Into the blue: gene duplication and loss underlie color vision adaptations in a deep-sea chimaera, the elephant shark *Callorhynchus milii*. *Genome Res.* 19, 415–426.
- Douglas, R.H., Partridge, J.C., 1997. On the visual pigments of deep-sea fish. *J. Fish. Biol.* 50, 68–85.
- Douglas, R.H., Partridge, J.C., Marshall, N.J., 1998. The eyes of deep-sea fish I. Lens pigmentation, tapeta and visual pigments. *Prog. Ret. Eye Res.* 17, 597–636.
- Eakin, R.M., 1965. Evolution of photoreceptors. *Cold. SH. Q. B.* 30, 363–370.
- Endler, J.A., 1986. Natural Selection in the Wild. Princeton Univ. Press, Princeton, NJ.
- Falk, J.D., Applebury, M.L., 1988. The molecular genetics of photoreceptor cells. In: Osborne, N., Chader, G.J. (Eds.), *Progress in Retinal Research*, vol. 7, pp. 89–112.
- Fay, J.C., Wu, C.-I., 2003. Sequence divergence, functional constraint, and selection in protein evolution. *Annu. Rev. Genomics. Hum. Genet.* 4, 213–235.
- Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using bootstrap. *Evolution* 39, 783–791.
- Fredriksson, R., Schiöth, H.B., 2005. The repertoire of G-protein-coupled receptors in fully sequenced genomes. *Mol. Pharmacol.* 67, 1414–1425.
- Fredriksson, R., Lagerström, M.C., Lundin, L.-G., Schiöth, H.B., 2003. The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. *Mol. Pharmacol.* 63, 1256–1272.
- 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*. *J. Evol. Biol.* 18, 516–523.
- Gojobori, J., Innan, H., 2009. Potential of fish opsin gene duplications to evolve new adaptive functions. *Trends Genet.* 25, 198–202.
- Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 52, 696–704.
- Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids Symp.* 41, 95–98.
- Haussler, D., O'Brien, S.J., Ryder, O.A., Barker, F.K., Clamp, M., Crawford, A.J., Hanner, R., Hanotte, O., Johnson, W.E., McGuire, J.A., et al., 2009. Genome 10K: a proposal to obtain whole-genome sequence for 10,000 vertebrate species. *J. Hered.* 100, 659–674.
- Hisatomi, O., Satoh, T., Tokunaga, F., 1997. The primary structure and distribution of killifish visual pigments. *Vis. Res.* 37, 3089–3096.
- Hoegg, S., Brinkmann, H., Taylor, J.S., Meyer, A., 2004. Phylogenetic timing of the fish-specific genome duplication correlates with the diversification of teleost fish. *J. Mol. Evol.* 59, 190–203.
- 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. *Proc. Roy. Soc. B* 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. *Integr. Comp. Biol.* 49, 630–643.
- Hope, A.J., Partridge, J.C., Hayes, P.K., 1998. Switch in rod opsin gene expression in the European eel, *Anguilla anguilla* (L.). *Proc. Roy. Soc. B* 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.



- Hunt, D.M., Fitzgibbon, J., Slobodyanyuk, S.J., Bowmaker, J.K., Dulai, K.S., 1997. Molecular evolution of the cottoid fish endemic to lake Baikal deduced from nuclear DNA evidence. *Mol. Phylogenet. Evol.* 8, 415–422.
- 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. *Proc. Roy. Soc. B* 274, 489–498.
- Ibbotson, R.E., Hunt, D.M., Bowmaker, J.K., Mollon, J.D., 1992. Sequence divergence and copy number of the middle- and long-wave photopigment genes in old world monkeys. *Proc. Roy. Soc. B* 247, 145–154.
- Jacobs, G.H., Neitz, M., Deegan, J.F., Neitz, J., 1996. Trichromatic color vision in new world monkeys. *Nature* 382, 156–158.
- Johnson, R.L., Grant, K.B., Zankel, T.C., Boehm, M.F., Merbs, S.L., Nathans, J., Nakanishi, K., 1993. Cloning and expression of goldfish opsin sequences. *Biochemistry* 32, 208–214.
- Kasai, A., Oshima, N., 2006. Light-sensitive motile iridophores and visual pigments in the neon tetra, *Paracheirodon innesi*. *Zool. Sci.* 23, 815–819.
- Kawamura, S., Yokoyama, S., 1995. Paralogous origin of the rhodopsin-like opsin genes in lizards. *J. Mol. Evol.* 40, 594–600.
- Kawamura, S., Yokoyama, S., 1996. Phylogenetic relationships among short wavelength-sensitive opsins of American Chameleon (*Anolis carolinensis*) and other vertebrates. *Vis. Res.* 36, 2797–2804.
- Kondrashov, F.A., Rogozin, I.B., Wolf, Y.I., Koonin, E.V., 2002. Selection in the evolution of gene duplicates. *Genome Biol.* 3, 0008.1–0008.9.
- Kozmik, Z., Ruzickova, J., Jonasova, K., Matsumoto, Y., Volapensky, P., Kozmikova, I., Strnad, H., Kawamura, S., Piatigorsky, J., Paces, V., Vlcek, C., 2008. Assembly of cnidarian camera-type eye from vertebrate-like components. *Proc. Natl. Acad. Sci. USA* 105, 8989–8993.
- Kühne, W., 1979. Chemische Vorgänge in der Netzhaut. In: Hermann, L., Handbuch der Physiologie, vol. 3, pp. 235–342.
- Kuraku, S., Meyer, A., Kuratani, S., 2009. Timing of genome duplications relative to the origin of the vertebrates: did cyclostomes diverge before or after? *Mol. Biol. Evol.* 26, 47–59.
- Larhammar, D., Nordström, K., Larsson, T.A., 2009. Evolution of vertebrate rod and cone phototransduction genes. *Proc. Roy. Soc. B* 364, 2867–2880.
- Laver, C.R.J., Taylor, J.S., 2011. RT-qPCR reveals opsin gene upregulation associated with age and sex in guppies (*Poecilia reticulata*) – a species with color-based sexual selection and 11 visual-opsin genes. *BMC Evol. Biol.* 11, 81.
- Levine, J.S., MacNichol Jr., E.F., 1979. Visual pigments in teleost fishes: effects of habitat, microhabitat, and behaviour on visual system evolution. *Sens. Process.* 3, 95–131.
- Li, Z., Gan, X., He, S., 2009. Distinct evolutionary patterns between two duplicated color vision genes within cyprinid fishes. *J. Mol. Evol.* 69, 346–359.
- Littlepage, J.L., 1965. Oceanographic investigations in McMurdo Sound, Antarctica. In: Llano, G.A. (Ed.), *Biology of the Antarctic Seas II*. Antarct. Res. Ser., vol. 5. American Geophysical Union, Washington, pp. 1–37.
- Loew, E.R., Lythgoe, J.N., 1978. The ecology of cone pigments in teleost fishes. *Vis. Res.* 18, 715–722.
- Matsumoto, Y., Fukamachi, S., Mitani, H., Kawamura, S., 2006. Functional characterization of visual opsin repertoire in Medaka (*Oryzias latipes*). *Gene* 371, 268–278.
- Max, M., McKinnon, P.J., Seidenman, K.J., Barrett, R.K., Applebury, M.L., Takahashi, J.S., Margolskee, R.F., 1995. Pineal opsin: a nonvisual opsin expressed in chick pineal. *Science* 267, 1502–1506.
- McCormack, W.T., Tjoelker, L.W., Thompson, C.B., 1991. Avian B-cell development: generation of an immunoglobulin repertoire by gene conversion. *Annu. Rev. Immunol.* 9, 219–241.
- 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). *Comp. Biochem. Phys.* 140, 197–205.
- Morrow, J.M., Lazic, S., Chang, B.S.W., 2011. A novel rhodopsin-like gene expressed in zebrafish retina. *Visual Neurosci.* 28, 325–335.
- Nakamura, I., 1995. Trichiuridae. *Peces sables, cintillas*. In: Fischer, W., Krupp, F., Schneider, W., Sommer, C., Carpenter, K.E., Niem, V. (Eds.), *Guía FAO para Identificación de Especies para lo Fines de la Pesca Pacifico Centro-Oriental*. FAO, Rome, pp. 1638–1642.
- Nakamura, I., Parin, N.V., 1993. FAO Species Catalogue. Snake Mackerels and Cutlassfishes of the World (families Gempylidae and Trichiuridae). An annotated and illustrated Catalogue of the Snake Mackerels, Snoeks, Escalors, Gemfishes, Sackfishes, Domine, Oilfish, Cutlassfishes, Scabbardfishes, Hairtails, and Frostfishes Known to Date, vol. 15. FAO, Rome, pp. 125–136.
- Nathans, J., 1990. Determinants of visual pigment absorbance: role of charged amino acids in the putative transmembrane segments. *Biochemistry* 29, 937–942.
- 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., Hogness, D.S., 1984. Isolation and nucleotide sequence of the gene encoding human rhodopsin. *Proc. Natl. Acad. Sci. USA* 81, 4851–4855.
- Nathans, J., Thomas, D., Hogness, D.S., 1986. Molecular genetics of human color vision: the genes encoding blue, green, and red pigments. *Science* 232, 193–202.
- Neafsey, D.E., Hartl, D.L., 2005. Convergent loss of an anciently duplicated, functionally divergent RH2 opsin gene in the fugu and *Tetraodon* pufferfish lineages. *Gene* 350, 161–171.
- Nelson, J.S., 2006. *Fishes of the World*, fourth ed. Wiley, New York.
- Nielsen, R., Yang, Z., 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* 148, 929–936.
- Nozawa, M., Suzuki, Y., Nei, M., 2009. Reliabilities of identifying positive selection by the branch-site and the site-prediction methods. *Proc. Natl. Acad. Sci. USA* 106, 6700–6705.
- Okamura, K., Nakai, K., 2008. Retrotransposition as a source of new promoters. *Mol. Biol. Evol.* 25, 1231–1238.
- Okano, T., Kojima, D., Fukada, Y., Shichida, Y., Yoshizawa, T., 1992. Primary structures of chicken cone visual pigments: vertebrate rhodopsins have evolved out of cone visual pigments. *Proc. Natl. Acad. Sci. USA* 89, 5932–5936.
- O'Quin, K.E., Hofmann, C.M., Hofmann, H.A., Carleton, K.L., 2010. Parallel evolution of opsin gene expression in African cichlid fishes. *Mol. Biol. Evol.* 27, 2839–2854.
- O'Quin, K.E., Smith, D., Naseer, Z., Schulte, J., Engel, S.D., Loh, Y.-H., Streelman, J.T., Boore, J.L., Carleton, K.L., 2011. Divergence in cis-regulatory sequences surrounding the opsin gene arrays of African cichlid fishes. *BMC Evol. Biol.* 11, 120.
- Osorio, D., Vorobyev, M., 2005. Photoreceptor spectral sensitivities in terrestrial animals: adaptations for luminance and colour vision. *Proc. Roy. Soc. B* 272, 1745–1752.
- Osorio, D., Vorobyev, M., 2008. A review of the evolution of animal colour vision and visual communication signals. *Vis. Res.* 48, 2042–2051.
- 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.
- Owens, G.L., Rennison, D.J., Allison, W.T., Taylor, J.S., 2011. In the four-eyed fish (*Anableps anableps*), the regions of the retina exposed to aquatic and aerial light do not express the same set of opsin genes. *Biol. Lett.*
- Palczewski, K., Kumasaka, T., Hori, T., Behnke, C.A., Motoshima, H., Fox, B.A., Le Trong, I., Teller, D.C., Okada, T., Stenkamp, R.E., Yamamoto, M., Miyano, M., 2000. Crystal structure of rhodopsin: a G protein-coupled receptor. *Science* 289, 739–745.
- Pankhurst, N.W., Montgomery, J.C., 1989. Visual function in four Antarctic nototheniid fishes. *J. Exp. Biol.* 142, 311–324.
- 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.
- 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. *J. Comp. Phys.* 164, 513–529.
- Perovich, D.K., Longacre, J., Barber, D.G., Maffione, R.A., Cota, G.F., Mobley, C.D., Gow, A.J., Onstott, R.G., Grenfell, T.C., Pegau, W.S., Laundry, M., Roesler, C.S., 1998. Field observations of the electromagnetic properties of first-year sea ice. *IEEE Trans. Geosci. Remote* 36, 1705–1715.
- Pignatelli, V., Champ, C., Marshall, J., Vorobyev, M., 2010. Double cones are used for colour discrimination in the reef fish, *Rhinecanthus aculeatus*. *Biol. Lett.* 6, 537–539.
- Plachetzki, D.C., Serb, J.M., Oakley, T.H., 2005. New insights into the evolutionary history of photoreceptor cells. *TREE* 20, 465–467.
- 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 nototheniid fish. *J. Exp. Biol.* 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. *Appl. Bioinform.* 2, 91–101.
- Raymond, P.A., Barthel, L.K., Rounsifer, M.E., Sullivan, S.A., Knight, J.K., 1993. Expression of rod and cone visual pigments in goldfish and zebrafish: a rhodopsin-like gene is expressed in cones. *Vis. Res.* 10, 1161–1174.
- Register, A.R., Yokoyama, R., Yokoyama, S., 1994. Multiple origins of green-sensitive opsins in fish. *J. Mol. Evol.* 39, 268–273.
- Rennison, D.J., Owens, G.L., Allison, W.T., Taylor, J.S., 2011. Intra-retinal variation of opsin gene expression in the guppy (*Poecilia reticulata*). *J. Exp. Biol.* 214, 3248–3254.
- Reyniers, E., van Thienen, M.N., Meire, F., de Boule, K., Devries, K., Kestelijn, P., Willems, P.J., 1995. Gene conversion between red and defective green opsin gene in blue cone monochromacy. *Genomics* 29, 323–328.
- Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406.
- Sakmar, T.P., Franke, R.R., Khorana, H.G., 1989. Glutamic acid-113 serves as the retinylidene Schiff base counterion in bovine rhodopsin. *Proc. Natl. Acad. Sci. USA* 86, 8309–8313.
- 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.
- 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*. *J. Exp. Biol.* 211, 1495–1503.
- Shi, Y., Radlwimmer, F.B., Yokoyama, S., 2001. Molecular genetics and the evolution of ultraviolet vision in vertebrates. *Proc. Natl. Acad. Sci. USA* 98, 11731–11736.
- Simpson, G.G., 1953. *The Major Features of Evolution*. Columbia Univ. Press, New York.
- Spady, T.C., 2006. *Cichlids as a Model for the Evolution of Visual Sensitivity*. University of New Hampshire, Durham, NH.

- 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. *Mol. Biol. Evol.* 23, 1538–1547.
- Suga, H., Schmid, V., Gehring, W.J., 2008. Evolution and functional diversity of jellyfish opsins. *Curr. Biol.* 18, 51–55.
- Swofford, D.L., 2002. PAUP: Phylogenetic Analysis Using Parsimony, Version 4.0 b10. Sinauer Associates, Sunderland, MA.
- Takechi, M., Kawamura, S., 2005. Temporal and spatial changes in the expression pattern of multiple red and green subtype opsin genes during zebrafish development. *J. Exp. Biol.* 208, 1337–1345.
- Tamura, K., Kumar, S., 2002. Evolutionary distance estimation under heterogeneous substitution pattern among lineages. *Mol. Biol. Evol.* 19, 1727–1736.
- Tamura, K., Nei, M., 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* 10, 512–526.
- Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24, 1596–1599.
- Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodea, J., Tuch, B.B., Siddiqui, A., Lao, K., Surani, M.A., 2009. MRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods* 6, 377–382.
- Tansley, K., 1931. The regeneration of visual purple: its relation to dark adaptation and night blindness. *J. Physiol.* 71, 442–458.
- Taylor, J.S., Raes, J., 2005. Small-scale gene duplication. In: Gregory, T.R. (Ed.), *The Evolution of the Genome*. Elsevier Academic Press, London.
- Taylor, J.S., Braasch, I., Frickey, T., Meyer, A., Van der Peer, Y., 2003. Genome duplication, a trait shared by 22,000 species of ray-finned fish. *Genome Res.* 13, 382–390.
- 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). *J. Exp. Biol.* 211, 3879–3888.
- Tsujimura, T., Chinen, A., Kawamura, S., 2007. Identification of a locus control region for quadruplicated green-sensitive opsin genes in zebrafish. *Proc. Natl. Acad. Sci. USA* 104, 12813–12818.
- Venkatesh, B., Ning, Y., Brenner, S., 1999. Late changes in spliceosomal introns define clades in vertebrate evolution. *Proc. Natl. Acad. Sci. USA* 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 Neurosci.* 16, 571–585.
- Wagner, H.J., 1990. Retinal structure of fishes. In: Djamgoz, M. (Ed.), *The Visual System of Fish*. Chapman and Hall Ltd., London, pp. 109–157.
- Wang, D., Oakley, T., Mower, J., Shimmin, L.C., Yim, S., Honeycutt, R.L., Tsao, H., Li, W.H., 2004. Molecular evolution of bat color vision genes. *Mol. Biol. Evol.* 21, 295–302.
- 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 Evol. Biol.* 8, 210.
- Watson, C.T., Lubieniecki, K.P., Loew, E., Davidson, W.S., Breden, F., 2010a. Genomic organization of duplicated short wave-sensitive and long wave-sensitive opsin genes in the green swordtail, *Xiphophorus helleri*. *BMC Evol. Biol.* 10, 87.
- Watson, C.T., Gray, S.M., Hoffmann, M., Lubieniecki, K.P., Joy, J.B., Sandkam, B.A., Weigel, D., Loew, E., Davidson, W.S., Breden, F., 2010b. Gene duplication and divergence of long wavelength-sensitive opsin genes in the guppy, *Poecilia reticulata*. *J. Mol. Evol.* 72, 240–252.
- 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 Evol. Biol.* 7, S11.
- Winderickx, J., Battisti, L., Motulsky, A.G., Deeb, S.S., 1992. Selective expression of human X chromosome-linked green opsin genes. *Proc. Natl. Acad. Sci. USA* 89, 9710–9714.
- Winderickx, J., Battisti, L., Hibiya, Y., Motulsky, A.G., Deeb, S.S., 1993. Haplotype diversity in the human red and green opsin genes: evidence for frequent sequence exchange in exon 3. *Hum. Mol. Genet.* 2, 1413–1421.
- Windsor, D.J., Owens, G.L., 2009. The opsin repertoire of *Jenynsia onca*: a new perspective on gene duplication and divergence in livebearers. *BMC Res. 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.
- Yang, Z., 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24, 1586–1591.
- Yang, Z., Nielsen, R., Goldman, N., Pedersen, A.M., 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155, 431–449.
- Yang, Z., Wong, W.S., Nielsen, R., 2005. Bayes empirical bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.* 22, 1107–1118.
- Yokoyama, S., 1994. Gene duplications and evolution of the short wavelength-sensitive visual pigments in vertebrates. *Mol. Biol. Evol.* 11, 32–39.
- Yokoyama, S., 2000. Molecular evolution of vertebrate visual pigments. *Prog. Retin. Eye Res.* 19, 385–420.
- Yokoyama, S., Radlwimmer, F.B., 1998. The “five-sites” rule and the evolution of red and green color vision in mammals. *Mol. Biol. Evol.* 15, 560–567.
- Yokoyama, S., Tada, T., 2010. Evolutionary dynamics of rhodopsin type 2 opsins in vertebrates. *Mol. Biol. Evol.* 27, 133–141.
- Yokoyama, S., Tada, T., Zhang, H., Britt, L., 2008. Elucidation of phenotypic adaptations: molecular analyses of dim-light vision proteins in vertebrates. *Proc. Natl. Acad. Sci. USA* 105, 13480–13485.
- Zhang, J., Nielsen, R., Yang, Z., 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol. Biol. Evol.* 22, 2472–2479.