Supporting Information

Cortesi et al. 10.1073/pnas.1417803112

SI Materials and Methods

Study Species. This study included 97 fish species, of which molecular data for 44 species were available from public databases (Table S1), and 53 species were sequenced specifically for this study; 38 species were part of the teleost/Acanthomorpha whole-genome sequencing project at the Centre for Ecological and Evolutionary Synthesis, and 12 samples were obtained from the aquarium trade to be newly sequenced. Fin clips of seven species from the aquarium trade were preserved in 95% ethanol (95:5, ethanol:ddH_2O) until total DNA was extracted using a QiaGen DNeasy Tissue commercial kit (www.qiagen.com), and of the remaining five species, retinas were preserved in RNAlater (www.lifetechnologies.com) for subsequent transcriptome sequencing (Table S1). Three dottyback species (Pseudochromidae) were caught at Lizard Island (14°40′ S, 145°27′ E), Great Barrier Reef, Australia between 2007 and 2013 (Table S1). Dottybacks were collected on snorkel from shallow reefs (depth of 2–5 m) surrounding the island using an anesthetic clove oil solution (10% clove oil, 40% ethanol, and 50% seawater) and hand nets. A fin clip was preserved in 95% ethanol (95:5, ethanol:ddH_2O) until total DNA was extracted using a standard salt precipitation protocol (1). In addition, several dusky dottyback (*Pseudochromis fuscus*) tissues were preserved on RNAlater for subsequent transcriptome sequencing and gene expression analysis. Larval dusky dottybacks were caught overnight using light traps during the summer recruitment pulses in November of 2007 and October and November of 2013 and either directly used for MSP measurements or kept on RNAlater for subsequent gene expression analysis.

SWS2 Gene Synteny.

Transcriptome sequencing and SWS2 reference mapping. Total RNA from various dusky dottyback tissues was extracted using a QiaGen RNeasy Plus Universal Kit (QiaGen): skin, liver, eyes, and gonads from a brown male individual [total length (TL) = 69 mm]; skin, brain, anal fin, caudal fin, and gonads from a yellow female individual (TL = 71 mm); and one entire small immature individual (TL = 22 mm). BioAnalyzer (www.genomics.agilent.com) was used to measure the initial concentration of the different extracts, after which they were diluted to the same concentration and pooled. The pool was then used to prepare a library for high-throughput sequencing using the Dynabeads mRNA Purification Kit (LifeTechnologies) for mRNA selection and the Ion Total RNA-Seq Kit (LifeTechnologies) for standard steps, such as RNA-to-cDNA transcription and size selection. Transcriptome sequencing was performed on an IonTorrent PGM platform (LifeTechnologies) using a 316 chip, standard run conditions, and a 120-bp length restriction. The run produced >3.4 million unique reads (70% efficiency) with a mean length of 113 bp, equaling a total number of 389.06 Mbp, of which 330.67 Mbp had a Phred quality score of Q20 or higher (i.e., >99% base call accuracy). Subsequent quality filtering of reads was performed on the Galaxy online web server (usegalaxy.org). Data were initially trimmed using a sliding window approach with a window size of 20 and step size of 1, and reads were trimmed from both sides until reaching a base pair with a score of ≥Q20. Reads with a read length of zero were discarded, and the trimmed reads were filtered for quality so that 95% of a single read had an overall score of Q20 or higher (Q20/95). After quality filtering and removal of sequencing artifacts, the library contained >2.5 million reads with a mean length of 80 bp. Filtered reads were mapped against publicly available SWS2A and SWS2B coding sequences of the Nile tilapia (*Oreochromis niloticus*; Cichlidae) in Geneious v.6.0.2 (www.geneious.com) using customized sensitivity settings (index word length = 11; maximal gap size = 2,000 bp). Assembled reads with an average depth of 16× per gene were manually assigned to the different copies before generating their consensus. The resulting sequences were scored for similarity to publicly available genes using BLASTN (www.ncbi.nlm.nih.gov/BLAST). This approach produced three distinct gene products, which were thereafter used as references for mapping of orthologous genes (see below). To verify the synteny of the dusky dottyback SWS2 copies, we furthermore sequenced the genomic region containing the three genes using a combination of long-amplicon sequencing on IonTorrent and Sanger sequencing (see below). The dusky dottyback transcriptome is made available on the short-read archive database in GenBank (www.ncbi.nlm.nih.gov/genbank) (Table S1).

Additionally, we used a HiSeq 2000 DNA sequencer from Illumina (www.illumina.com) to generate retina-specific transcriptomes for five species of labrids (Labridae) and cardinalfishes (Apogonidae) (Table S1). Raw reads from these approaches were then mapped against the dusky dottyback SWS2 genes in Geneious v.6.0.2 (average depth of 250–2,500× per SWS2 gene), and genes were extracted as previously described for the dusky dottyback.

Public data mining. Whole-genome sequences of 24 species and the transcriptome sequences of 1 species (*Tripterygion delaisi*; average depth of 360–620× per SWS2 gene) were accessed from the Ensembl Genome browser (www.ensembl.org) or the Assembly (assembled contigs or scaffolds) and the short-read archive databases in GenBank (Table S1). Initially, the raw reads from unassembled datasets were mapped against SWS2 exons from the three dusky dottyback SWS2 genes in Geneious v.6.0.2 using medium-sensitivity settings (70% identity threshold for mapping) to efficiently recover all SWS2 copies. Matching reads were then manually split by copies (if more than one gene copy was present in the species) and de novo assembled, and their consensus was used as a species-specific reference for subsequent low-sensitivity mapping (only reads over 90% sequence identity map) in Geneious v.6.0.2. During this cyclic mapping, unassembled reads were mapped repeatedly against the prolonging reference (originally single exons) until the mapped regions would overlap and could be connected into an entire gene. The cyclic mapping continued the same way until the genes could not be prolonged anymore or could be connected into a genomic region (~30 kbp) that contained the highly conserved up- and downstream neighboring genes HCFC1 and LWS or GNL3L (in case of LWS loss), respectively. Alignments were continuously inspected visually to exclude ambiguous mapping of genes. In species that retained all three SWS2 paralogs, genes were interspaced by around 1,500 bp with an upstream SWS2Aβ, middle SWS2Aα, and downstream SWS2B copy (Fig. 1).

Sequencing of the SWS2 target region. The synteny of SWS2 genes in nine species was investigated by sequencing the SWS2 target region between HCFC1 and LWS. The region was initially separated into three overlapping stretches, and universal primers were designed to amplify each stretch separately (Table S2). Long PCR was used to amplify the 5- to 13-kbp-long products using the TaKaRa LA polymerase (TaKaRa Bio Inc.; program: 35× 98 °C for 10 s, 60–68 °C for 1 min, and 68 °C for 20 min), and the QIAquick PCR Purification Kit (QiaGen) was used to purify the products cut from the electrophoresis gel. After purified, products were used to prepare a long-amplicon library following the

genomic DNA library preparation protocol (Ion Xpress Plus gDNA and Amplicon Library Preparation; LifeTechnologies) and sequenced on IonTorrent PGM using a 316 v2 ChIP combined with the Ion PGM Sequencing 400 Kit (LifeTechnologies). Reads were quality filtered (same as for the dusky dottyback transcriptome; see above) and de novo assembled in Geneious v.6.0.2. In several species, the consensus sequences would not cover the entire genomic region, and we, therefore, designed specific primers to sequence the missing parts by Sanger on an Applied Biosystems 3130*xl* Genetic Analyzer (www.appliedbiosystems.com) (Table S2). Additionally, the genomic raw reads and scaffolds of 38 species that were part of the whole-genome sequencing project at the Centre for Ecological and Evolutionary Synthesis were used to BLAST search and assemble the target SWS2 genomic region (HCFC1 upstream and LWS or GNL3L downstream) (Table S1).

SWS2 presence and synteny were assessed by mapping single exons from the dusky dottyback against the target region in Geneious v.6.0.2 using high-sensitivity settings (see above). Coding regions of SWS2 copies were subsequently extracted from the region and used for phylogenetic analyses.

Phylogenetic analyses. SWS2 coding sequences from 67 species were cut from genomic regions and combined with the transcriptome sequences from 7 species and publicly available single-gene coding sequences of 23 species (Table S1). Sequences were aligned using MAFFT v.6.8 (2), and the most appropriate model of sequence evolution was estimated in jModeltest v.2 (3) using the Akaike information criterion as the criterion for model selection. Subsequent Bayesian inference was conducted on the CIPRES platform (4) using the GTR+I+Γ model in MrBayes v.3.2.1 (5) and a Markov chain Monte Carlo search with two independent runs and four chains each. Each run was set to 10 million generations, with trees sampled every 1,000 generations (i.e., 10,000 trees per run) with 25% of burn in after the sampling. SWS2 sequences from eel (*Anguilla anguilla*), zebrafish (*Danio rerio*), carp (*Cyprinus carpio*), and salmons (*Salmo salar* and *Oncorhcynchus keta*) were used as outgroups to reconstruct phylogenetic relationships between SWS2 copies. This approach produced a partly unresolved gene tree with low phylogenetic support for SWS2Aα genes in particular (Fig. S5). Consequently, to increase the phylogenetic signal, we repeated the analysis after removing the genetic regions that were affected by gene conversion (see below for gene conversion approaches) (Fig. S1).

Gene Conversion.

Single-exon phylogenies. To investigate which SWS2 copies and what genetic regions would be affected by gene conversion, we ran additional MrBayes analyses under the same conditions as mentioned above but for each exon separately (five in total) (Fig. S2).

Sliding window analysis of gene conversion. To measure the divergence between SWS2 genes, we calculated the dS (neutral process) along the coding sequences of gene copies using a sliding window strategy with a step size of 1 and a window of 30 in DNAsp v.5.10.1 (6). To avoid a bias toward clades with more representatives, we calculated the rates for one fish species per family that possesses more than one SWS2 copy. Converted regions were identified based on a sharp drop in dS between genes, which is equivalent to high sequence similarities (Fig. S3). These regions were subsequently removed from the coding sequence alignment to generate the final SWS2 gene tree (see above) (Fig. S1).

Gene resurrection: gene conversion from pseudogenes. Pseudogenes with the potential to be resurrected by converting with functional paralogs were identified in five species/lineages (Fig. 1 and Fig. S3). To test for potential gene resurrection, we ran phylogenetic analyses with the aforementioned dataset (i.e., genes without converted regions) and additionally included the converted region of the pseudogene and its functional paralog of the species of interest (Fig. S4). Analyses were run for each species separately (i.e., a total of five analyses) in MrBayes under the same conditions as described above. Two out of five analyses were found to support the proposed gene resurrection scenario (Fig. S4).

Functional Analysis.

Neofunctionalization of SWS2 genes. Putative amino acid substitutions of importance for spectral tuning were searched for by comparing amino acid alignments of known (7) and potential key tuning sites (i.e., retinal binding pocket sites) of SWS2 genes from one fish species per family (based on alignments in ref. 8). Initially, sites were extracted based on differences in clade consensus (majority rule applied after removal of converted regions of sequences) between paralogs (Fig. S5). To identify those sites with clade specificity, we reconstructed their ancestral state (under maximal parsimony) in Mesquite v.3.0 (9). Additionally, all species were screened for the specific substitution of A269T, which is known to cause a positive shift in spectral sensitivity of 6 nm (7).

Function of the percomorph-specific SWS2A paralogs. A functional analysis of the percomorph-specific SWS2A duplication was conducted in the dusky dottyback using a combination of MSP and quantitative real-time RT-PCR (qRT-PCR) approaches.

For MSP, adult (n = 3) and larval (n = 1) dusky dottybacks were dark-adapted overnight and euthanized with an overdose of MS222 (1:2,000). Eyes were removed and dissected under IR illumination with the aid of an IR-sensitive image converter. Small pieces (~1–3 mm^2) of retinal tissue were mounted on a no. 1 glass coverslip in a drop of PBS (410 mOsm kg^1, pH 7.2) containing 4% dextran (molecular weight of 282,000; D-7265; Sigma). This preparation was covered with a smaller no. 0 coverslip, and the edges of the top coverslip were sealed with nail varnish to prevent dehydration. Absorbance spectra of individual photoreceptor outer segments were measured using a single-beam wavelength-scanning microspectrophotometer and analyzed as described in detail elsewhere (10, 11).

For qRT-PCR, RNA was extracted from retina tissues (adults) or the whole head (larvae) using TriZol following the protocol of the manufacturer (LifeTechnologies). To remove possible genomic contamination, we treated the RNA extract with DNase according to the DNA Free protocol from the manufacturer (LifeTechnologies). RNA was subsequently reverse-transcribed to cDNA using the High-Capacity RNA-to-cDNA Kit (Applied Biosystems), and the resulting concentration was measured on a NanoDrop1000 Spectrophotometer (ThermoScientific).

The relative expression of the SWS opsin genes (SWS1 and SWS2s) was quantified by qRT-PCR on a StepOnePlus Real-Time PCR System (LifeTechnologies). A 20-mL reaction volume was prepared using SYBR Green Master (Rox) dye (www.lifescience.roche.com) with a final cDNA concentration of 10 ng/μL and a final primer concentration of 200 nM. The qRT-PCR was then performed under the following cycling conditions: 95 °C for 10 min, 40 cycles at 95 °C for 15 s, and 61 °C for 60 s. All qRT-PCR amplifications included a melt curve step after cycling. Unique primers for each opsin gene were designed with a primer specificity of ≥90% and either the forward or the reverse primer spanning an exon–exon boundary to ensure cDNA-specific amplification of the product (60–100 bp) (Table S2). Products were furthermore sequenced by Sanger to assure accuracy of the reaction.

All primers were initially validated on a dilution series of factor 5 of a species-specific pool containing equal ratios of fragments (molarity measured on BioAnalyzer) of each of the gene copies with a starting concentration of 0.1–0.5 nM/μL. qRT-PCR efficiencies (Es) were calculated for each reaction from the slope of the standard curve using the equation $E = 10^{(-1/slope)}$ as implemented in the StepOnePlus software (LifeTechnologies), with an efficiency of 2 being equal to 100% ($E\% = [10^{(-1/slope)} - 1] - 100$) and an indicator of a robust assay. All experiments were

carried out with three technical replicates, and the opsin pool was added to each plate as an internal reference. The relative expression of each gene was then calculated as described in detail elsewhere (12).

We used *t* and Wilcoxon rank sum tests to examine whether the expression of SWS2 genes between larval and adult fish differed (SWS1 was found not to be expressed). Expression data were initially ln-transformed and assessed for normality and homogeneity of variance using histograms, residuals plots, and quantile–quantile plots. Because SWS2A copies did not conform to normality, we used Wilcoxon tests to compare their expression between larval and adult dottybacks. To account for multiple comparisons of tests, we used Bonferroni corrections (13) to adjust *P* values. All statistical analyses were performed in SPSS v.15.0 (SPSS).

1. Miller SA, Dykes DD, Polesky HF (1988) A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* 16(3):1215.
2. Katoh K, Toh H (2008) Recent developments in the MAFFT multiple sequence alignment program. *Brief Bioinform* 9(4):286–298.
3. Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: More models, new heuristics and parallel computing. *Nat Methods* 9(8):772–772.
4. Miller MA, Pfeiffer W, Schwartz T (2010) Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *Proceedings of the Gateway Computing Environments Workshop (GCE)* (IEEE, New Orleans), pp 1–8.
5. Ronquist F, et al. (2012) MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* 61(3):539–542.
6. Librado P, Rozas J (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25(11):1451–1452.
7. Yokoyama S (2008) Evolution of dim-light and color vision pigments. *Annu Rev Genomics Hum Genet* 9:259–282.
8. Carleton KL, Spady TC, Cote RH (2005) Rod and cone opsin families differ in spectral tuning domains but not signal transducing domains as judged by saturated evolutionary trace analysis. *J Mol Evol* 61(1):75–89.
9. Maddison WP, Maddison DR (2014) *Mesquite: A Modular System for Evolutionary Analysis, Version 3.0*. Available at mesquiteproject.org. Accessed September 7, 2014.
10. Hart NS, Theiss SM, Harahush BK, Collin SP (2011) Microspectrophotometric evidence for cone monochromacy in sharks. *Naturwissenschaften* 98(3):193–201.
11. Hart NS, Coimbra JP, Collin SP, Westhoff G (2012) Photoreceptor types, visual pigments, and topographic specializations in the retinas of hydrophiid sea snakes. *J Comp Neurol* 520(6):1246–1261.
12. Carleton KL, Kocher TD (2001) Cone opsin genes of african cichlid fishes: Tuning spectral sensitivity by differential gene expression. *Mol Biol Evol* 18(8):1540–1550.
13. Rice WR (1989) Analyzing tables of statistical tests. *Evolution* 43(1):223–225.

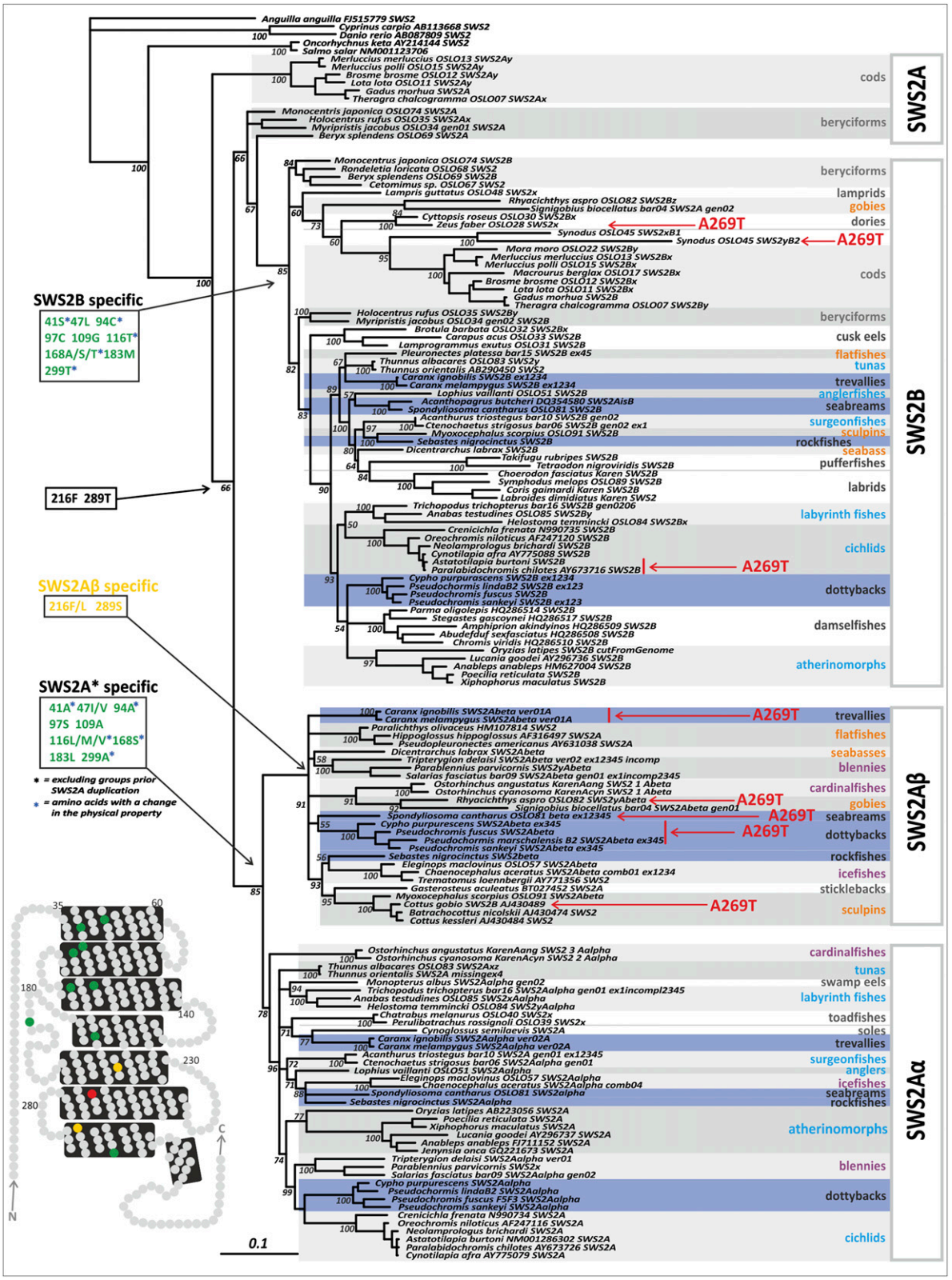

Fig. S1. SWS2 gene phylogeny and potential key amino acid substitutions. Consensus phylogeny based on the coding region of SWS2 genes without converted parts. Bayesian posterior probabilities are shown for deeper nodes (i.e., maximum one value per lineage). Lineages that retained all three SWS2 copies are shaded in blue; those with two copies are highlighted by two shades of gray. No shading means one copy retained. Common names are color-coded according to the identity of retained genes based on currently available data: blue, SWS2Aα + SWS2B; orange, SWS2Aβ + SWS2B; violet, SWS2Aα + SWS2Aβ. The boxes show clade-specific amino acid substitutions (as per ancestral state reconstruction; standardized to bovine rhodopsin) between SWS2A and SWS2B (green) and for SWS2Aβ (yellow). Amino acid substitutions that vary in physical properties are marked with blue asterisks. Species with a substitution of A269T,

Legend continued on following page

which is likely to cause a positive shift in spectral sensitivity of 6 nm, are indicated with red arrows (1). *Lower Left* shows a schematic drawing of the bovine rhodopsin (based on ref. 2), with potentially important amino acid substitutions marked accordingly.

1. Yokoyama S (2008) Evolution of dim-light and color vision pigments. *Annu Rev Genomics Hum Genet* 9:259–282.
2. Palczewski K, et al. (2000) Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* 289(5480):739–745.

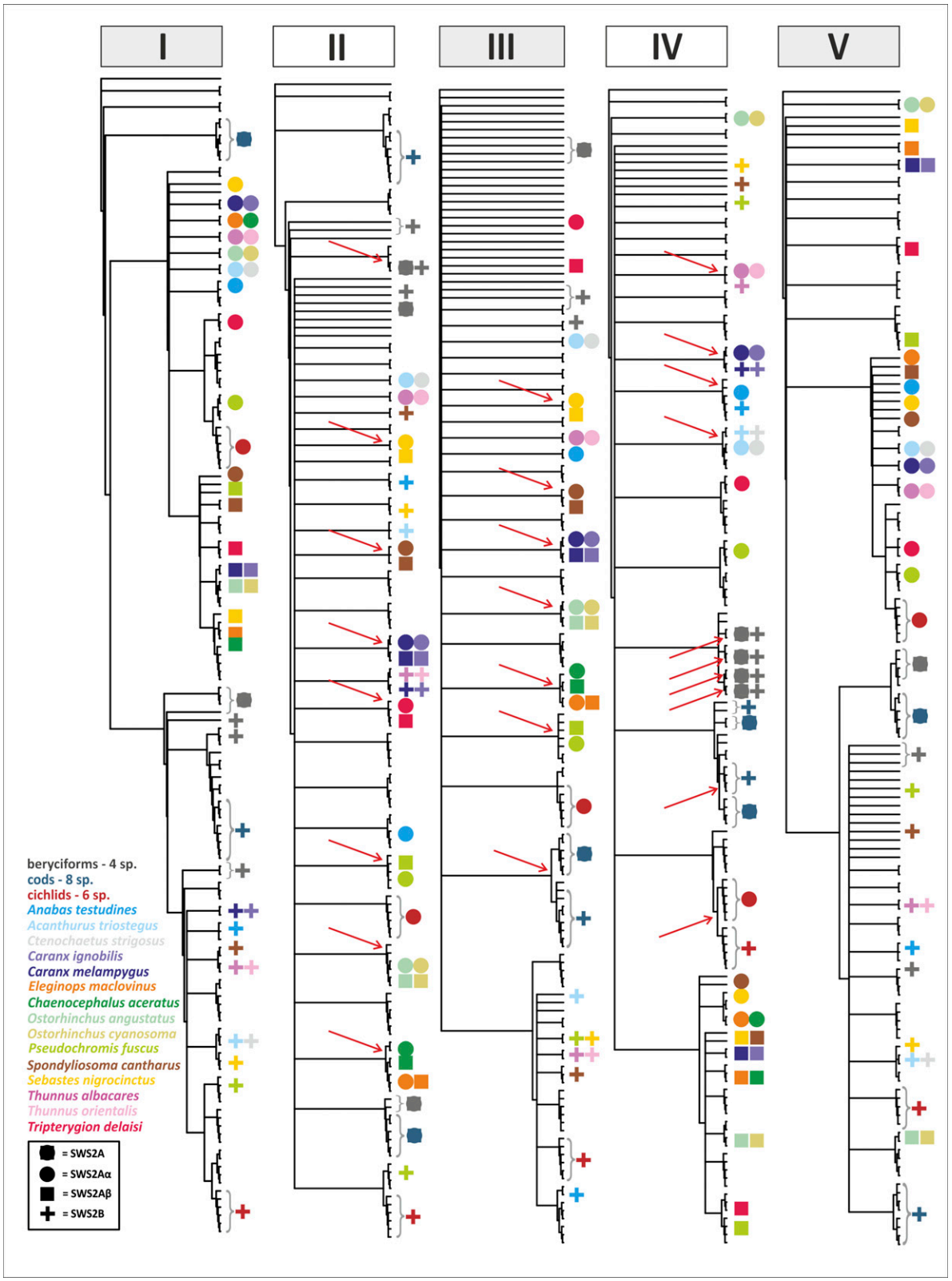

Fig. S2. Single-exon phylogenies used to identify regions under gene conversion. SWS2 copies are clustered as sisters if exons were affected by gene conversion (red arrows). Most of the conversions in exons 2 and 3 happened between SWS2Aα and SWS2Aβ, whereas the conversions of exon 4 happened mostly between SWS2B and one of the SWS2A copies. Gene conversions of different phylogenetic age can be detected based on the tree [e.g., family-specific gene conversion in cichlids (Cichlidae; exon 4) and cods (Gadiformes; exon 3), genus-specific conversion in *Caranx* (Carangiformes; exons 2–4) and *Ostorhinchus* (Apogonidae; exons 2 and 3), or species-specific conversion in icefishes (Notothenioidei; *E. maclovinus* and *C. aceratus*; exons 2 and 3) and beryciforms (Beryciformes; exon 4)]. Note that, for beryciforms, the exon 4 of SWS2A and SWS2B always cluster as sisters within species, indicating that conversion occurred independently multiple times over in this lineage.

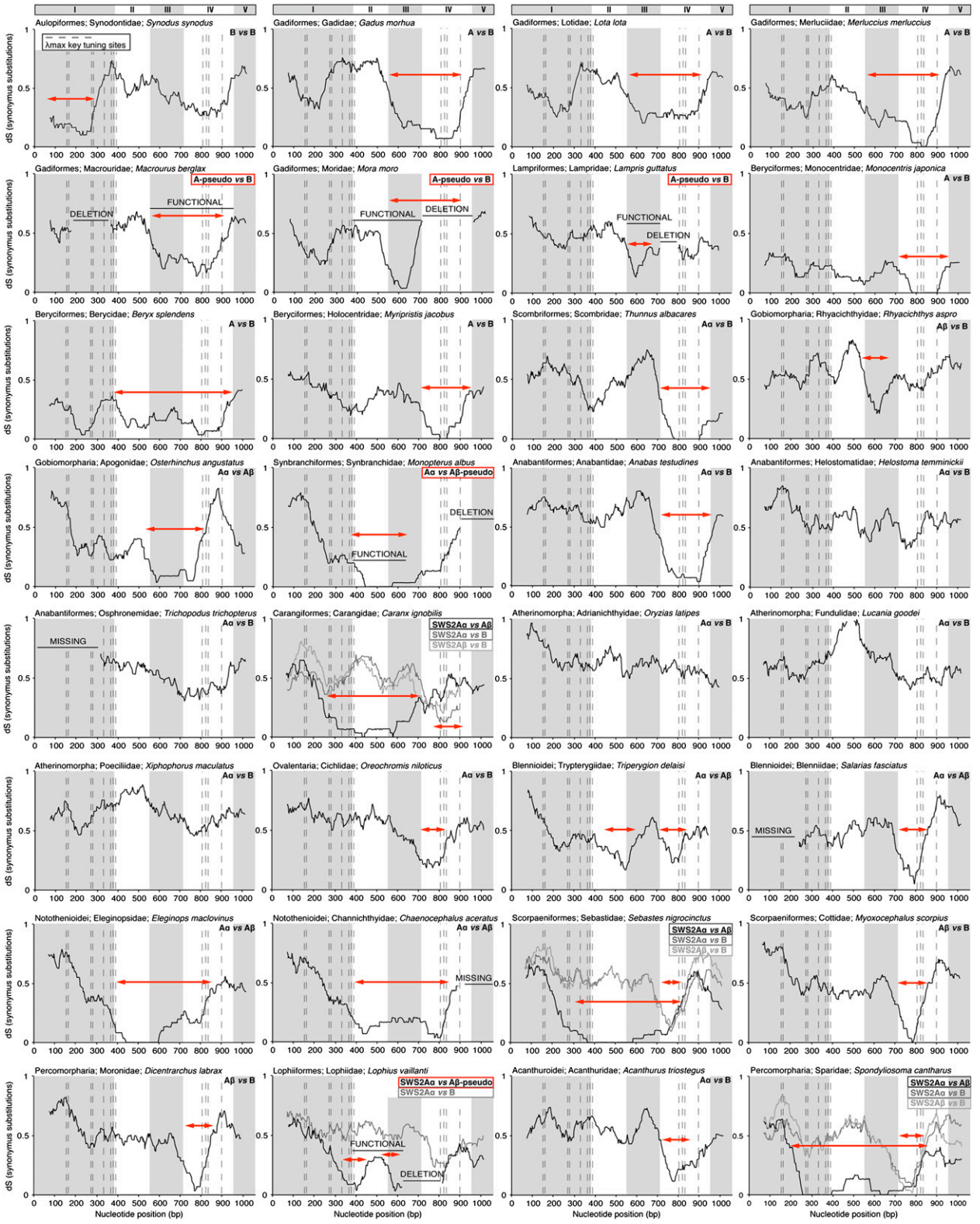

Fig. S3. Sequence comparison of SWS2 genes from one representative fish species per family. Synonymous substitution rates (dS) along gene sequences were assessed by sliding windows (size of 30 bp; step size = 1). Probable regions under gene conversion, which were removed to generate the final gene phylogeny (Fig. S1), are identified by a drop in dS and marked by red arrows. Note that, in several families, pseudogenized genes are likely to have contributed to gene conversion (framed in red) (also see Fig. S4).

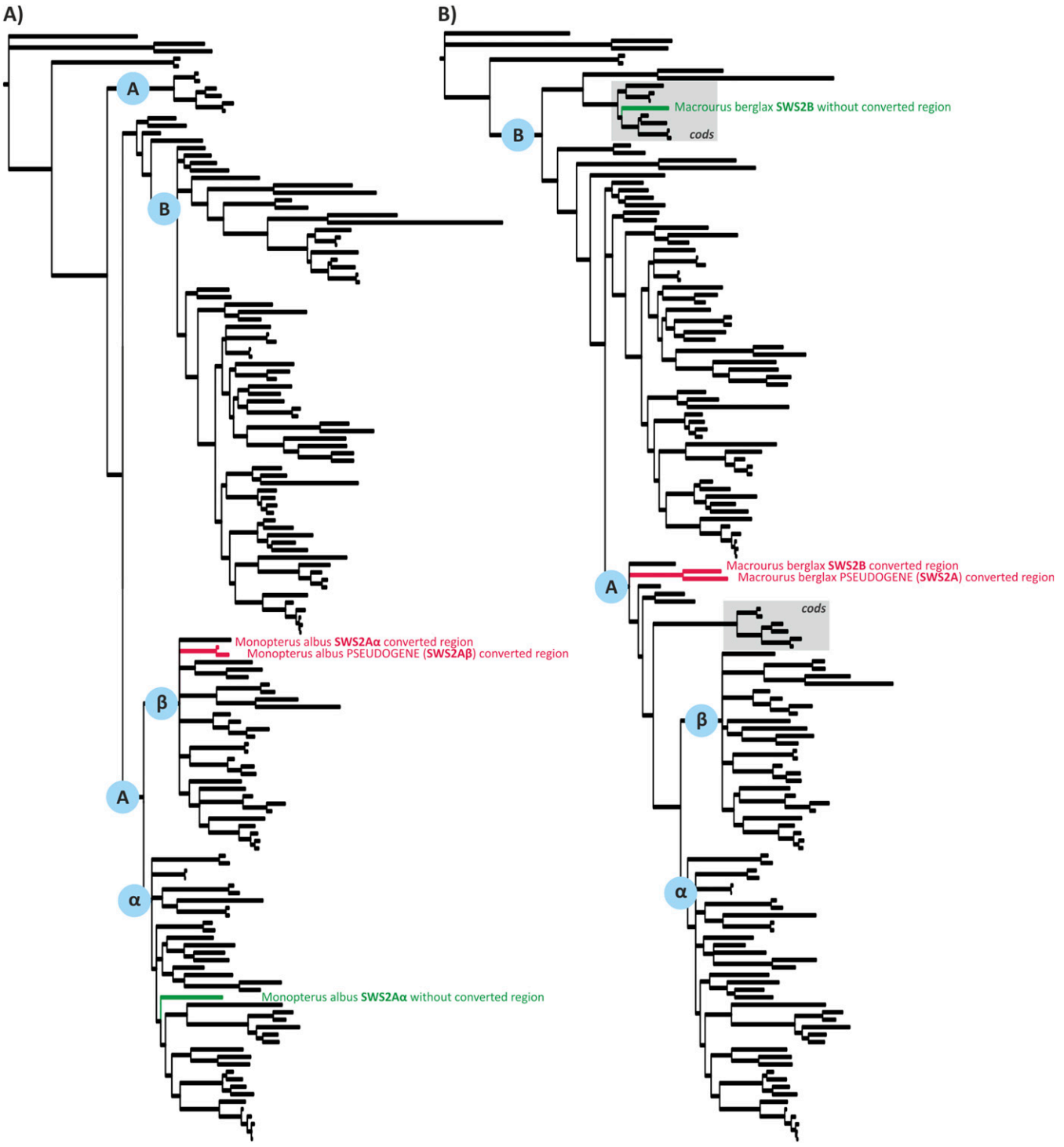

Fig. S4. Conversion from pseudogene to functional paralog potentially leading to gene resurrection. Phylogenies based on the coding genes without converted regions (all species) (similar to Fig. S1) and the converted region of the pseudogene and the target paralog (of a tested species) identified potential gene resurrection in two cases. (*A*) Asian swamp eel (*Monopterus albus*; Synbranchiformes). Red shows the converted regions of SWS2Aα and the SWS2Aβ pseudogene; green shows SWS2Aα without the converted region. The phylogenetic position of SWS2Aα converted region in the SWS2Aβ clade clearly suggests the origin of the converted region in the SWS2Aβ pseudogene. (*B*) Roughhead grenadier (*Macrourus berglax*; Gadiformes). Red shows the converted regions of SWS2B and the SWSA pseudogene; green shows SWS2B without the converted region. Shaded in gray is the position of other cod species (Gadiformes).

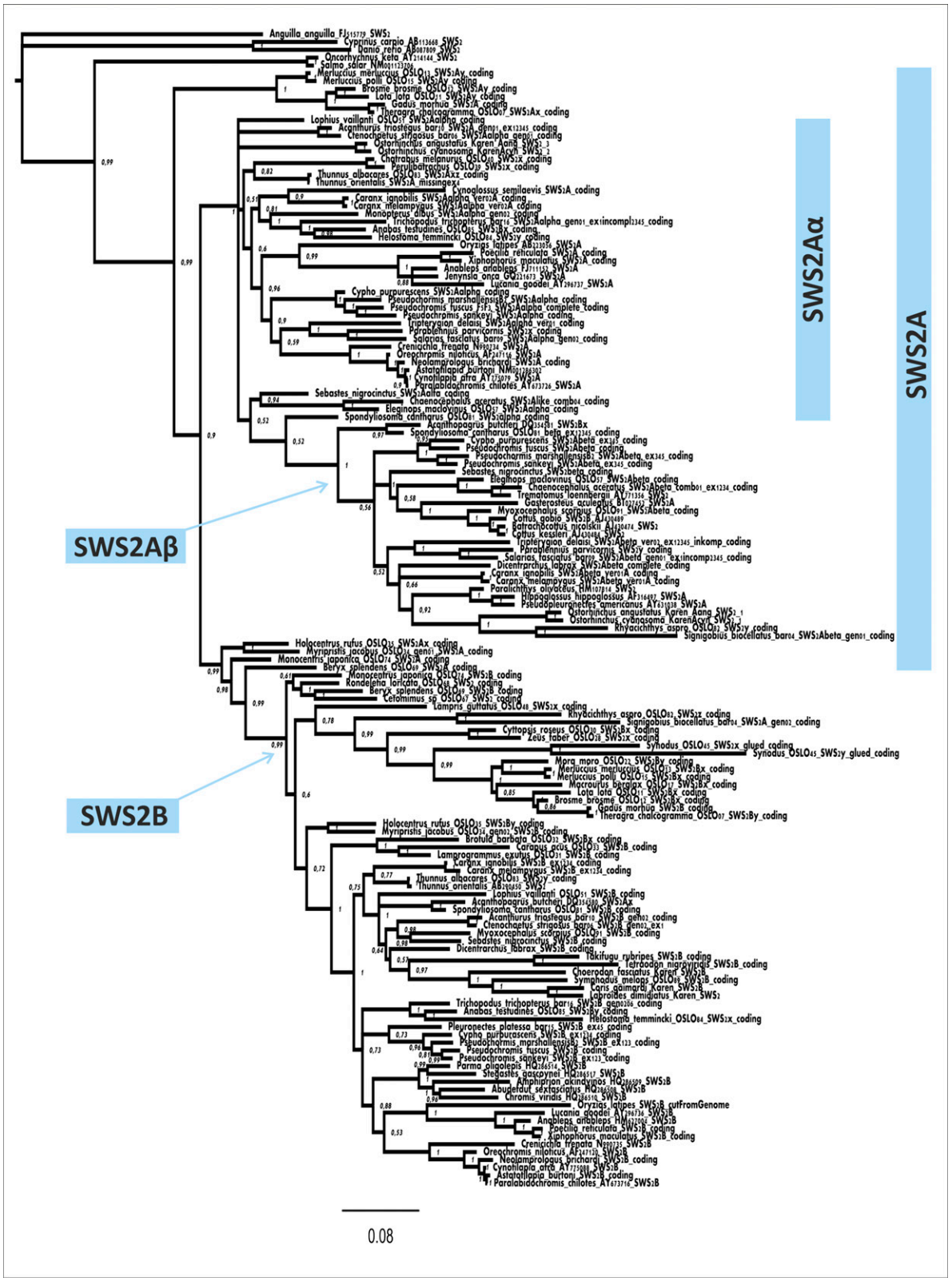

Fig. S5. Consensus phylogeny based on the full coding region of SWS2 genes (i.e., including converted regions). Only Bayesian support values >0.5 are shown. Note that, indicative of conversion, SWS2Aα genes especially do not resolve properly.

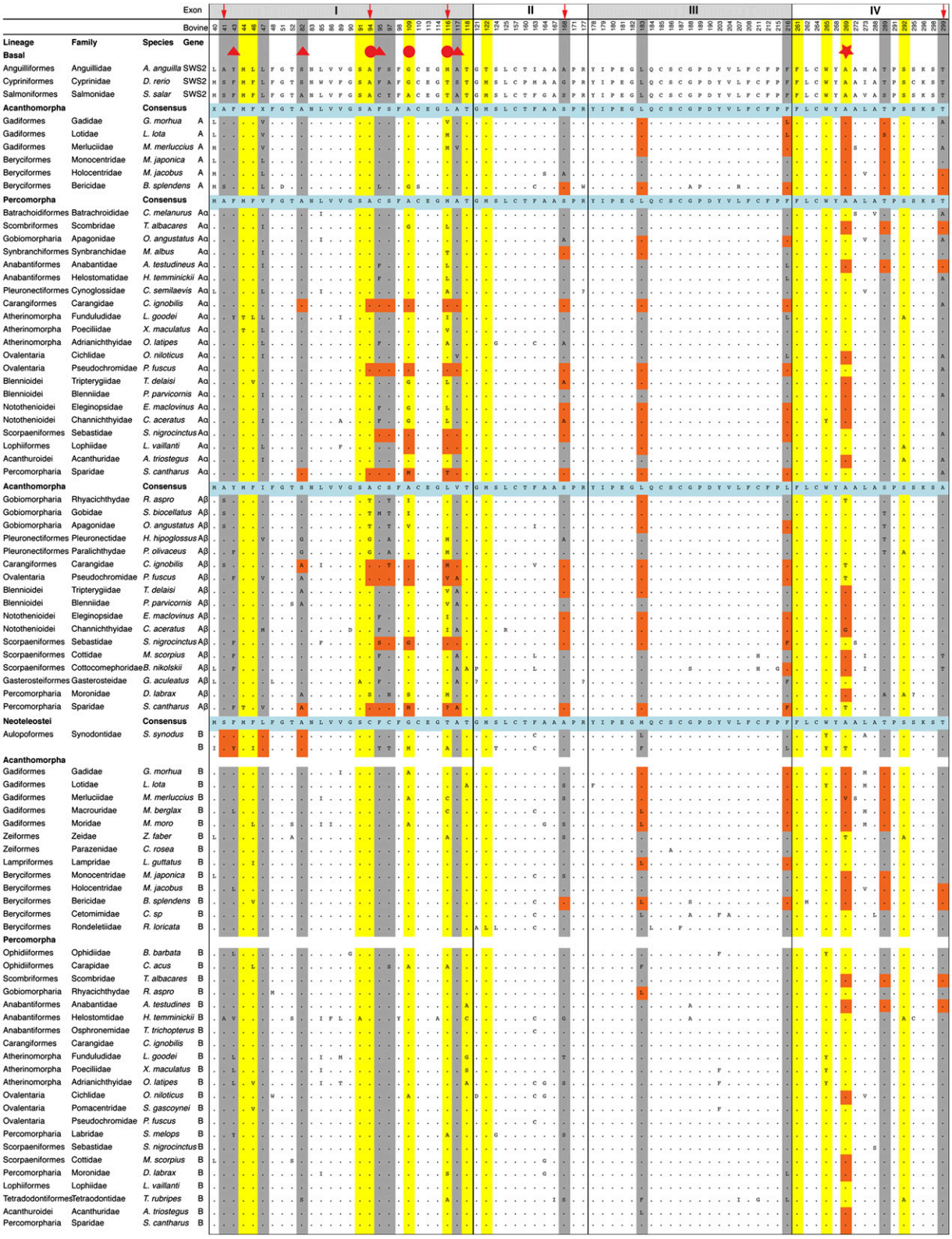

Fig. S6. SWS2 amino acid alignments (standardized to bovine rhodopsin) of known key tuning (yellow) (1) and retinal binging pocket sites. Pictured is one representative fish species per family. Highlighted in gray or marked by a red sphere are potentially functional amino acid substitutions that were identified based on clade consensus (after removing amino acids affected by conversion; orange). The red asterisk marks site 269, at which a substitution of A269T is known to cause a positive shift of 6 nm. Red triangles mark sites that did not confer to clade specificity based on an ancestral state reconstruction (after maximum parsimony). Arrows indicate those potential key substitutions that also vary in physical properties between SWS2 genes. Additional information is in Fig. S1.

1. Yokoyama S (2008) Evolution of dim-light and color vision pigments. *Annu Rev Genomics Hum Genet* 9:259–282.

Table S1. Species list including accession numbers and type of data

Order	Family	Species	Source/locality	Type of data	Database	Accession no.	Blue opsins
Outgroups							
Anguilliformes	Anguillidae	*Anguilla anguilla*	Public database	Single-gene sequence(s)	GenBank	FJ515779	SWS2
Cypriniformes	Cyprinidae	*Danio rerio*	Public database	Genome assembly	Ensembl	Zv9.76	SWS2
Cypriniformes	Cyprinidae	*Cyprinus carpio*	Public database	Single-gene sequence(s)	GenBank	AB113668	SWS2
Salmoniformes	Salmonidae	*Salmo salar*	Public database	Genome assembly	GenBank	AGKD00000000.3	SWS2
Salmoniformes	Salmonidae	*Oncorhynchus keta*	Public database	Single-gene sequence(s)	GenBank	AY214144	SWS2
Esociformes	Esocidae	*Esox lucius*	Public database	Genome assembly	GenBank	AZJR00000000.1	Pseudogene
Euteleostei—basal							
Osmeriformes	Osmeridae	*Osmerus eperlanus*	Teleost Genome Project	Genomic region	GenBank*	KP004260, KP004256	No
Stomiiformes	Stomiidae	*Borostomias antarcticus*	Teleost Genome Project	Genomic region	GenBank*	KP004255	No
Ateleopodiformes	Ateleopodidae	*Guentherus altivela*	Teleost Genome Project	Genomic region	GenBank*	KP004324, KP004266	No
Aulopiformes	Synodontidae	*Synodus synodus*	Teleost Genome Project	Genomic region	GenBank*	KP004275, KP004251, KP004258, KP004248, KP004312, KP004319	SWS2B1 + SWS2B2
Myctophiformes	Myctophidae	*Benthosema glaciale*	Teleost Genome Project	Genomic region	GenBank*	KP004273, KP004309	No
Acantomorpha—basal							
Percopsiformes	Percopsidae	*Percopsis transmontana*	Teleost Genome Project	Genomic region	GenBank*	KP004278, KP004247	No
Polymixiiformes	Polymixiidae	*Polymixia japonica*	Teleost Genome Project	Genomic region	GenBank*	KP004289, KP004300	No
Gadiiformes	Gadidae	*Theragra chalcogramma*	Teleost Genome Project	Genomic region	GenBank*	KP004279, KP004272, KP004276	SWS2A + SWS2B
Gadiiformes	Gadidae	*Gadus morhua*	Public database	Genome assembly	Ensembl	GadMor1.76	SWS2A + SWS2B
Gadiiformes	Lotidae	*Lota lota*	Teleost Genome Project	Genomic region	GenBank*	KP004274	SWS2A + SWS2B
Gadiiformes	Lotidae	*Brosme brosme*	Teleost Genome Project	Genomic region	GenBank*	KP004269, KP004326, KP004295, KP004321	SWS2A + SWS2B
Gadiiformes	Merluciidae	*Merluccius merluccius*	Teleost Genome Project	Genomic region	GenBank*	KP004323, KP004253, KP004317	SWS2A + SWS2B
Gadiiformes	Merluciidae	*Merluccius polli*	Teleost Genome Project	Genomic region	GenBank*	KP004310, KP004280	SWS2A + SWS2B
Gadiiformes	Macrouridae	*Macrourus berglax*	Teleost Genome Project	Genomic region	GenBank*	KP004264, KP004261	Pseudogene + SWS2B
Gadiiformes	Moridae	*Mora moro*	Teleost Genome Project	Genomic region	GenBank*	KP004293, KP004294, KP004298	Pseudogene + SWS2B
Zeiformes	Zeidae	*Zeus faber*	Teleost Genome Project	Genomic region	GenBank*	KP004286, KP004322	Pseudogene + SWS2B
Zeiformes	Parazenidae	*Cyttopsis roseus*	Teleost Genome Project	Genomic region	GenBank*	KP004302, KP004270	Pseudogene + SWS2B
Lampriformes	Lampridae	*Lampris guttatus*	Teleost Genome Project	Genomic region	GenBank*	KP004285, KP004288, KP004252	Pseudogene + SWS2B
Beryciformes	Holocentridae	*Myripristis jacobus*	Teleost Genome Project	Genomic region	GenBank*	KP004257, KP004271, KP004262	SWS2A + SWS2B
Beryciformes	Holocentridae	*Holocentrus rufus*	Teleost Genome Project	Genomic region	GenBank*	KP004299, KP004318	SWS2A + SWS2B
Beryciformes	Berycidae	*Beryx splendens*	Teleost Genome Project	Genomic region	GenBank*	KP004297, KP004314	SWS2A + SWS2B
Beryciformes	Monocentridae	*Monocentris japonica*	Teleost Genome Project	Genomic region	GenBank*	KP004305, KP004290	SWS2A + SWS2B
Beryciformes	Rondeletiidae	*Rondeletia loricata*	Teleost Genome Project	Genomic region	GenBank*	KP004307, KP004313	SWS2B
Beryciformes	Cetomimidae	*Cetomimus sp.*	Teleost Genome Project	Genomic region	GenBank*	KP004259, KP004311, KP004283	SWS2B
Percomorpha—basal							
Ophidiiformes	Ophidiidae	*Brotula barbata*	Teleost Genome Project	Genomic region	GenBank*	KP004277	SWS2B
Ophidiiformes	Ophidiidae	*Lamprogrammus exutus*	Teleost Genome Project	Genomic region	GenBank*	KP004325	SWS2B
Ophidiiformes	Carapidae	*Carapus acus*	Teleost Genome Project	Genomic region	GenBank*	KP004282	SWS2B

Table S1. Cont.

Order	Family	Species	Source/locality	Type of data	Database	Accession no.	Blue opsins
Batrachoidiformes	Batrachoididae	*Chatrabus melanurus*	Teleost Genome Project	Genomic region	GenBank*	KP004287, KP004296, KP004265, KP004267	SWS2Aα + pseudogene
Batrachoidiformes	Batrachoididae	*Perulibatrachus rossignoli*	Teleost Genome Project	Genomic region	GenBank*	KP004254, KP004308, KP004263	SWS2Aα
Scomberiformes	Scombridae	*Thunnus albacares*	Teleost Genome Project	Genomic region	GenBank*	KP004306, KP004291	SWS2Aα + SWS2B
Scomberiformes	Scombridae	*Thunnus orientalis*	Public database	Single-gene sequence(s); genome assembly	GenBank	AB290450, BADN01000000	SWS2Aα + SWS2B
Gobiomorpharia							
Former Perciformes	Gobiidae	*Signigobius biocellatus*	Aquarium trade	Long PCR genomic region	GenBank*	KP004338, KP004332	SWS2Aα + SWS2B
Former Perciformes	Gobiidae	*Rhyacichthys aspro*	Teleost Genome Project	Genomic region	GenBank*	KP004301	SWS2Aα + SWS2B
Former Perciformes	Apogonidae	*Ostrorhinchus cyanosoma*	Aquarium trade	Transcriptome	GenBank*	KP004342, KP004343	SWS2Aα + SWS2Aβ
Former Perciformes	Apogonidae	*Ostrorhinchus angustatus*	Aquarium trade	Transcriptome	GenBank*	KP004345, KP004341	SWS2Aα + SWS2Aβ
Carangomorphariae and anabantomorphariae							
Anabantiformes	Anabantidae	*Anabas testudineus*	Teleost Genome Project	Genomic region	GenBank*	KP004316, KP004284	SWS2Aα + SWS2B + pseudogene
Anabantiformes	Osphronemidae	*Trichopodus trichopterus*	Aquarium trade	Long PCR genomic region	GenBank*	KP004337	SWS2Aα, SWS2B
Anabantiformes	Helostomatidae	*Helostoma temmincki*	Teleost Genome Project	Genomic region	GenBank*	KP004249	SWS2Aα + SWS2B + pseudogene
Carangiformes	Carangidae	*Caranx melampygus*	Public database	Genomic raw reads	SRA GenBank	SRX360285	SWS2Aα + SWS2Aβ + SWS2B
Carangiformes	Carangidae	*Caranx ignobilis*	Public database	Genomic raw reads	SRA GenBank	SRX360276	SWS2Aα + SWS2Aβ + SWS2B
Pleuronectiformes	Cynoglossidae	*Cynoglossus semilaevis*	Public database	Genome assembly; raw reads	GenBank	AGRG00000000.1, SRX100168	SWS2Aα
Pleuronectiformes	Pleuronectidae	*Hippoglossus hippoglossus*	Public database	Single-gene sequence(s)	GenBank	AF316497	SWS2Aβ
Pleuronectiformes	Pleuronectidae	*Pseudopleuronectes americanus*	Public database	Single-gene sequence(s)	GenBank	AY631038	SWS2Aβ
Pleuronectiformes	Paralichthyidae	*Paralichthys olivaceus*	Public database	Single-gene sequence(s)	GenBank	HM107814	SWS2Aβ
Pleuronectiformes	Pleuronectidae	*Pleuronectes platessa*	Trade	Long PCR genomic region	GenBank*	KP004328	SWS2B (partial)
Synbranchiformes	Synbranchidae	*Monopterus albus*	Public database	Genomic raw reads	SRA GenBank	SRX218061	SWS2Aα + pseudogene
Ovalentaria							
Former Perciformes	Cichlidae	*Oreochromis niloticus*	Public database	Genome assembly	Ensembl	Orenil1.0.76	SWS2Aα + SWS2B + pseudogene
Former Perciformes	Cichlidae	*Astatotilapia burtoni*	Public database	Genome assembly	GenBank	AFNZ00000000.1	SWS2Aα + SWS2B + pseudogene
Former Perciformes	Cichlidae	*Neolamprologus brichardi*	Public database	Genome assembly	GenBank	AFNY00000000.1	SWS2Aα + SWS2B + pseudogene
Former Perciformes	Cichlidae	*Cynotilapia afra*	Public database	Single-gene sequence(s)	GenBank	AY775088, AY775079	SWS2Aα + SWS2B
Former Perciformes	Cichlidae	*Paralabidochromis chilotes*	Public database	Single-gene sequence(s)	GenBank	AY673716, AY673726	SWS2Aα + SWS2B
Former Perciformes	Cichlidae	*Crenicihla frenata*	Public database	Single-gene sequence(s)	GenBank	JN990735, JN990734	SWS2Aα + SWS2B
Cyprinodontiformes	Poeciliidae	*Poecilia reticulata*	Public database	Single-gene sequence(s)	GenBank	AZHG00000000.1	SWS2Aα + SWS2B
Cyprinodontiformes	Poeciliidae	*Xiphophorus maculatus*	Public database	Genome assembly	GenBank	AGAJ00000000.1	SWS2Aα + SWS2B
Cyprinodontiformes	Fundulidae	*Lucania goodei*	Public database	Single-gene sequence(s)	GenBank	AY296736, AY296737	SWS2Aα + SWS2B
Cyprinodontiformes	Anablepidae	*Anableps anableps*	Public database	Single-gene sequence(s)	GenBank	HM627004, FJ711152	SWS2Aα + SWS2B
Beloniformes	Adrianichthyidae	*Oryzias latipes*	Public database	Genome assembly	Ensembl	MEDAKA1.76	SWS2Aα + SWS2B

Table S1. Cont.

Order	Family	Species	Source/locality	Type of data	Database	Accession no.	Blue opsins
Former Perciformes	Pomacentridae	*Stegastes partitus*	Public database	Genome assembly; raw reads	GenBank	JMKM00000000.1, SRX526491	SWS2B
Former Perciformes	Pomacentridae	*Stegastes gascoynei*	Public database	Single-gene sequence(s)	GenBank	HQ286517	SWS2B
Former Perciformes	Pomacentridae	*Parma oligolepis*	Public database	Single-gene sequence(s)	GenBank	HQ286514	SWS2B
Former Perciformes	Pomacentridae	*Amphiprion akindynos*	Public database	Single-gene sequence(s)	GenBank	HQ286509	SWS2B
Former Perciformes	Pomacentridae	*Abudefduf sexfasciatus*	Public database	Single-gene sequence(s)	GenBank	HQ286508	SWS2B
Former Perciformes	Pomacentridae	*Chromis viridis*	Public database	Single-gene sequence(s)	GenBank	HQ286510	SWS2B
Blennioidei (former Perciformes)	Tripterygiidae	*Tripterygion delaisi*	Public database	Transcriptomic raw reads	SRA GenBank	SRX237746	SWS2Aα + SWS2Aβ
Blennioidei (former Perciformes)	Blenniidae	*Salarias fasciatus*	Aquarium trade	Long PCR genomic region	GenBank*	KP004330	SWS2Aα + SWS2Aβ
Blennioidei (former Perciformes)	Blenniidae	*Parablennius parvicornis*	Teleost Genome Project	Genomic region	GenBank*	KP004303, KP004304, KP004315	SWS2Aα + SWS2Aβ
Former Perciformes	Pseudochromidae	*Pseudochromis fuscus*	Lizard Island, Australia	Long PCR genomic region; transcriptome	GenBank,* SRA GenBank	KP004335, SRX736911, KP017246, KP017247	SWS2Aα + SWS2Aβ + SWS2B
Former Perciformes	Pseudochromidae	*Pseudochromis sankeyi*	Aquarium trade	Long PCR genomic region	GenBank*	KP004334	SWS2Aα + SWS2Aβ + SWS2B
Former Perciformes	Pseudochromidae	*Pseudochromis marshallensis*	Lizard Island, Australia	Long PCR genomic region	GenBank*	KP004331	SWS2Aα + SWS2Aβ + SWS2B
Former Perciformes	Pseudochromidae	*Cypho purpurascens*	Lizard Island, Australia	Long PCR genomic region	GenBank*	KP004327	SWS2Aα + SWS2Aβ + SWS2B
Percomorpharia							
Former Perciformes	Acanthuridae	*Acanthurus triostegus*	Aquarium trade	Long PCR genomic region	GenBank*	KP004336	SWS2Aα + SWS2B
Former Perciformes	Acanthuridae	*Ctenochaetus strigosus*	Aquarium trade	Long PCR genomic region	GenBank*	KP004329, KP004333	SWS2Aα + SWS2B
Former Perciformes	Labridae	*Symphodus melops*	Teleost Genome Project	Genomic region	GenBank*	KP004320, KP004281, KP004250	SWS2B
Former Perciformes	Labridae	*Choerodon fasciatus*	Aquarium trade	Transcriptome	GenBank*	KP004344	SWS2B
Former Perciformes	Labridae	*Coris gaimardi*	Aquarium trade	Transcriptome	GenBank*	KP004340	SWS2B
Former Perciformes	Labridae	*Labroides dimidiatus*	Aquarium trade	Transcriptome	GenBank*	KP004339	SWS2B
Lophiiformes	Lophiidae	*Lophius vaillanti*	Teleost Genome Project	Genomic region	GenBank*	KM978047	SWS2Aα + SWS2B + pseudogene
Gasterosteiformes	Gasterosteidae	*Gasterosteus aculeatus*	Public database	Genome assembly	Ensembl	BROADS1.76	SWS2A + SWS2B
Former Perciformes	Moronidae	*Dicentrarchus labrax*	Public database	Genome assembly	GenBank	FQ310506.3	Pseudogene + SWS2Aβ + SWS2B
Tetraodontiformes	Tetraodontidae	*Tetraodon nigriviridis*	Public database	Genome assembly	Ensembl	TETRAODON8.75	SWS2B
Tetraodontiformes	Tetraodontidae	*Takifugu rubripes*	Public database	Genome assembly	Ensembl	FUGU4.76	SWS2B
Notothenioidei	Eleginopsidae	*Eleginops maclovinus*	Teleost Genome Project	Genomic region	GenBank*	KP004292, KP004268	SWS2Aα + SWS2Aβ
Notothenioidei	Channichthyidae	*Chaenocephalus aceratus*	Public database	Genomic raw reads	SRA GenBank	SRX272123	SWS2Aα + SWS2Aβ
Notothenioidei	Nototheniidae	*Trematomus loennbergii*	Public database	Single-gene sequence(s)	GenBank	AY771356	SWS2Aβ
Scorpaeniformes	Sebastidae	*Sebastes nigrocinctus*	Public database	Genome assembly	GenBank	AUPR00000000.1	SWS2Aα + SWS2Aβ + SWS2B
Scorpaeniformes	Cottidae	*Myoxocephalus scorpius*	Teleost Genome Project	Genomic region	GenBank*	KM978045, KM978046	Pseudogene + SWS2Aβ + SWS2B
Scorpaeniformes	Cottidae	*Cottus gobio*	Public database	Single-gene sequence(s)	GenBank	AJ430489	SWS2Aβ
Scorpaeniformes	Cottocomephoridae	*Batrachocottus nikolskii*	Public database	Single-gene sequence(s)	GenBank	AJ430474	SWS2Aβ
Scorpaeniformes	Cottocomephoridae	*Cottus kessleri*	Public database	Single-gene sequence(s)	GenBank	AJ430484	SWS2Aβ

Table S1. Cont.

Order	Family	Species	Source/locality	Type of data	Database	Accession no.	Blue opsins
Former Perciformes	Sparidae	*Spondyliosoma cantharus*	Teleost Genome Project	Genomic region	GenBank*	KM978043, KM978044	SWS2Aα + SWS2Aβ + SWS2B
Former Perciformes	Sparidae	*Acanthopagrus butcheri*	Public database	Single-gene sequence(s)	GenBank	DQ354581, DQ354580	SWS2Aβ, SWS2B

SRA, short-read archive.
*Data obtained within this study.

Table S2. Primer list for this study

Method and targeted region: gene/exon or intron	Primer name	Orientation	Primer sequence	Species
Long PCR				
HCFC1/ex1	*D1_DUP_HCFC1_F1*	Forward	CTCCTTTATAGCCACAGCTCTGTGTCC	*P. fuscus*
SWS2Aβ/intron1	*D10_fus_SWS2Abet_exintr1_R1*	Reverse	GTACCAAACTCATCTTACCTCCAAGTGTTG	
Long PCR				
SWS2B/ex4	*D6_DUP_SWS2Abet_F1*	Forward	GAGCGGGAGGTGACCAGGATGGTGG	*P. fuscus*, *T. trichopterus*
LWS/ex2	*D9_DUP_LWS_ex2_R2*	Reverse	CCAGTTTAGAGGRTGACGGAGTTTCTTG	*C. strigosus*
Long PCR				
SWS2Aβ/ex3	*ENDF1*	Forward	CCTATGTGATRTTTCTCTTCTGCTTCTGCTTCG	*P. fuscus*, *P. sankeyi*, *P. marshallensis*
SWS2B/ex3	*BEGR1*	Reverse	GCAGTGCTCCTGTGGACCAGACTGGTACACCAC	*Cypho purpurescens*
Long PCR				
HCFC1/ex1	*D1_DUP_HCFC1_F1*	Forward	Sequence above	*A. triostegus*
SWS2B/ex3	*D16_DUP_SWS2B_ex3_R1*	Reverse	GTTTCATTGTTAAACTTGTTGCCTGTTG	
Long PCR				
HCFC1/ex1	*D1_DUP_HCFC1_F1*	Forward	Sequence above	*C. strigosus*
SWS2B/ex3	*D18_DUP_SWS2B_ex1_R3*	Reverse	TGTATCTGAAGGCAAAGCAGTAGAAGCAG	
Long PCR				
SWS2Aα/ex2	*D6_DUP_SWS2Abet_F1*	Forward	Sequence above	*S. biocellatus*
SWS2B/ex5	*D5_DUP_SWS2B_R2*	Reverse	GCAAGATTGAAGGATTTACAGCAAC	
Long PCR				
SWS2Aα/ex3	*ENDF1*	Forward	Sequence above	*T. trichopterus*
SWS2B/ex3	*BEGR1*	Reverse	Sequence above	
Long PCR				
SWS2B/ex4	*D6_DUP_SWS2Abet_F1*	Forward	Sequence above	*S. biocellatus*, *P. platessa*
LWS/ex2	*D8_DUP_LWS_ex2_R1*	Reverse	CTGGTTGCAYACACTGATGGTGCTGGC	*A. triostegus*
Long PCR				
SWS2Aβ/ex2	*D6_DUP_SWS2Abet_F1*	Forward	Sequence above	*S. fasciatus*
SWS2Aα/ex5	*D5_DUP_SWS2B_R2*	Reverse	Sequence above	
Long PCR				
SWS2Aβ/SWS2Aα/ex1	*D22_DUP_beta_ex1_F1*	Forward	ATGAAGCACGGCCGTGTCACRGAGC	*S. fasciatus*
LWS/ex2	*D8_DUP_LWS_ex2_R1*	Reverse	Sequence above	
Long PCR				
SWS2Aα/ex3	*D17_DUP_SWS2A_ex3_F1*	Forward	GACTGGTACACCACAAACAACAAATAC	*S. fasciatus*
LWS/ex2	*D9_DUP_LWS_ex2_R2*	Reverse	Sequence above	
Sanger sequencing				
SWS2Aβ/ex1	*SWS2A_betF6*	Forward	CATCAATGCGCTTACCG	*P. fuscus*
SWS2Aβ/ex4	*SWS2A_betR1*	Reverse	GAAGGAGGTGTAGGGGG	
Sanger sequencing				
SWS2B/intron3	*BF5_intron34*	Forward	CACATCTAAACTTCACCAGG	*P. fuscus*
SWS2B/ex5	*ABbetR6*	Reverse	CCCACTTTGGAGACTTC	
Sanger sequencing				
SWS2B/ex1	*D27_Ctenoch_SWS2ex1_F*	Forward	GCGCTCTTTTATTCAATGTCAGC	*A. triostegus*
SWS2B/ex4	*D26_Acanth_ex4_R2*	Reverse	GTAGATAACAGGGTTGTAGAC	
Sanger sequencing				
SWS2B/ex1	*D27_Ctenoch_SWS2ex1_F*	Forward	Sequence above	*C. strigosus*
SWS2B/ex4	*D28_Ctenoch_SWS2ex4_R*	Reverse	GATAACAGGGTTATAGACGGTG	
Sanger sequencing				
SWS2B/ex2	*D29_Tricho_ex2_F*	Forward	TACAGCGTAATCATCGTCAGTC	*T. trichopterus*
SWS2B/ex4	*D30_Tricho_ex4_R*	Reverse	CCACCTGTTTATTGAGGAGTATG	
Sanger sequencing				
SWS1	*POOL_Pfus_SWS1_F*	Forward	CTGTGTGCCATGGAGTCTGCC	*P. fuscus*
Pool for quantitative PCR reference	*SWS1_R2d_dam*	Reverse	TCGTTGTGGGTGTACCAGTC	
Sanger sequencing				
SWS2B	*POOL_Pfus_SWS2B_F*	Forward	GTGACTGGTACTGCCATCAATATC	*P. fuscus*
Pool for quantitative PCR reference	*POOL_Pfus_SWS2B_R*	Reverse	AACGATGGTGAAGAAGGGGATGGAA	
Sanger sequencing				
SWS2Aα	*POOL_Pfus_SWS2Aalfa_F*	Forward	CTCACTATTGCATGCACCGCC	*P. fuscus*
Pool for quantitative PCR reference	*POOL_Pfus_SWS2Aalfa_R*	Reverse	GCCCATGCCCAGCATCGCT	

Table S2. Cont.

Method and targeted region: gene/exon or intron	Primer name	Orientation	Primer sequence	Species
Sanger sequencing				
SWS2Aβ	*POOL_Pfus_SWS2Abeta_F*	Forward	CTTACCGTTGCATGCACCGTG	*P. fuscus*
Pool for quantitative PCR reference	*POOL_Pfus_SWS2Abeta_R*	Reverse	TCCACTCATCCCCAGCATCTTC	
qRT-PCR				
SWS1 (efficiency: 90%)	*Pfus_SWS1_2F*	Forward	TTTTGGAGCCTTCAAGTTCACCAG	*P. fuscus*
SWS1 (efficiency: 90%)	*Pfus_SWS1_23R*	Reverse	GATGTACCTGCTCCAGCCAAAG	
qRT-PCR				
SWS2B (efficiency: 94%)	*Pfus_SWS2B_1F1*	Forward	CCGTGGGCTCCTTCACCTG	*P. fuscus*
SWS2B (efficiency: 94%)	*Pfus_SWS2B_12R1*	Reverse	GGCTCACCATGCCTCCAATC	
qRT-PCR				
SWS2Aα (efficiency: 96%)	*Pfus_SWS2Aalfa_12F1*	Forward	CATGGCAACACTCGGGGGTATG	*P. fuscus*
SWS2Aα (efficiency: 96%)	*Pfus_SWS2Aalfa_2R1*	Reverse	CGCAAACACCCAGGTGAACC	
qRT-PCR				
SWS2Aβ (efficiency: 96%)	*Pfus_SWS2Abeta_1F2*	Forward	GGTGAACTTGGCTGCCGCG	*P. fuscus*
SWS2Aβ (efficiency: 96%)	*Pfus_SWS2Abeta_12R1*	Reverse	CCATACCTCCAAGTGTTGCTAC	