Zoological Studies **61**:16 (2022)
doi:10.6620/ZS.2022.61-16

Zoological Studies

Open Access

Correlated Expression of the Opsin Retrogene LWS-R and its Host Gene in Two Poeciliid Fishes

Chia-Hao Chang[1,*]

[1]*Department of Science Education, National Taipei University of Education, No.134, Sec.2, Heping E. Rd., Da'an District, Taipei City 10671, Taiwan.*
**Correspondence: E-mail: chiahao0928@gmail.com (Chang). Tel.: +886-4-23590121#32453.*

Received 4 August 2021 / Accepted 22 February 2022 / Published 10 May 2022
Communicated by John Wang

The important role of retrogenes in genome evolution and species differentiation is becoming increasingly accepted. One synapomorphy among cyprinodontoid fish is a retrotransposed version of a long-wavelength sensitive (LWS) opsin gene, LWS-R, within an intron of the gephyrin (GPHN) gene. These two genes display opposing orientations. It had been speculated that LWS-R hijacks the cis-regulatory elements of GPHN for transcription, but whether their expression is correlated had remained unclear. Here, *in silico* predictions identified putative promoters upstream of the translation start site of LWS-R, indicating that its transcription is driven by its own promoter rather than by the GPHN promoter. However, consistent expression ratios of LWS-R:GPHN in the eyeball and brain of fishes indicate that the respective gene transcriptions are correlated. Co-expression is potentially modulated by histone exchange during GPHN transcription. Two isoforms were detected in this study, *i.e.*, intron-free and intron-retaining. Intron-free LWS-R was only expressed in the eyeball of fishes, whereas intron-retaining LWS-R occurred in both eyeball and brain. Expression of vision-associated LWS-R beyond the eyeball supports that it is co-expressed with more ubiquitous GPHN.

Key words: Retrotransposition, Opsin, Promoter, Gephyrin, Histone exchange, Intron retention.

BACKGROUND

Animals are equipped with various sensory modalities to constantly adjust their behavior in response to external stimuli. Among these sensory systems, vision is typically paramount. Apart from species living in extreme habitats, such as caves, most have evolved specialized cells to convert external light signals into internal nerve impulses. In vertebrates, this competence is attributable to specialized photoreceptors—cone cells (photopic or well-lit vision) and rod cells (scotopic or low-light vision)—in the retina of eyes. These photoreceptor cells express visual pigments comprising a light-absorbing protein (visual opsin) and a vitamin A-derived chromophore, on their outer segments. Visual opsins are membrane-bound G-protein-coupled receptors (GPCRs), and the vitamin A-derived chromophores include vitamin A1 (retinal) or A2 (3,4-dehydroretinal). The maximal absorbance wavelength (λ_{max}) of visual pigments depends on the type of visual opsin or chromophore (Yokoyama 2000). Most vertebrates only express one kind of rhodopsin (RH) in rod cells (Musilova et al. 2019), and the other four types of visual opsins are associated with cone cells. These cone-cell-specific visual opsins are short-wavelength sensitive 1 (SWS1, λ_{max} ranges from UV to violet), short-wavelength sensitive 2 (SWS2, λ_{max} ranges from violet to blue), medium-wavelength sensitive (RH2, λ_{max} green), and long-wavelength sensitive (LWS, λ_{max} red) (Yokoyama 2000).

Unlike other vertebrates, fish feature numerous cone opsin genes in their genome. For example,

Citation: Chang CH. 2022. Correlated expression of the opsin retrogene LWS-R and its host gene in two poeciliid fishes. Zool Stud **61**:16. doi:10.6620/ZS.2022.61-16.

zebrafish (*Danio rerio*) possess one SWS1 gene, one SWS2 gene, four RH2 genes, and two LWS genes (Chinen et al. 2003). The fish cone opsin repertoires is believed to have diversified due to variability in aquatic light levels (Hofmann et al. 2012; Escobar-Camacho et al. 2017; Terai et al. 2017). Whole genome duplication has played an important role in increasing cone opsin numbers during the evolution of ray-finned fishes. For example, lineage-specific whole genome duplication events have generated additional cone opsins in some cyprinids, such as the common carp (*Cyprinus carpio*) and in salmonids (Lin et al. 2017). Moreover, other gene duplication events mediated by transposable elements (TEs) can add even more cone opsin gene copies to a genome. Depending on their transposition intermediates, RNA or DNA respectively, TEs can be classified as class I (retrotransposons) or class II (DNA transposons) (Bourque et al. 2018). Class II TEs may lead to unequal crossovers during meiosis, with a DNA fragment moving from one chromosome to its homolog, generating new offspring genes. Such offspring genes inherit the entire gene structure from their parents, and they are often localized close to their parental genes (Shen 2019). The tandemly-duplicated opsin genes constitute an opsin gene array in humans and lancelets, both representing well-documented cases of gene duplication based on DNA transposons (Dulai et al. 1999; Pantzartzi et al. 2018). Retrotransposons can reverse-transcribe mRNA molecules into cDNA, so offspring genes may be generated when these cDNA copies are integrated back into the genome. Genes arising in this way are called retrogenes, and they are usually intron-less. Unlike the offspring genes arising from unequal crossovers, these intron-less offspring genes are usually located far from the parental genes (Kaessmann et al. 2009). A classical example of a visual opsin retrogene is the RH gene of teleost fish, which is intron-less, but its parental gene (EXO-RH) has four introns (Bellingham et al. 2003; Fujiyabu et al. 2019). The above-mentioned mechanisms all operate to increase the number of cone opsins in fish. However, as with all duplicated genes, functional redundancy drives selection to remove these duplicates, unless they acquire functional specialization (Kuzmin et al. 2022). For instance, whereas the Japanese eel (*Anguilla japonica*) and American eel (*A. rostrata*) have three RH2 genes, the European eel (*A. Anguilla*) only has two (Lin et al. 2017). Overall, opsin gene gain or loss is a common feature of fish evolution (Cortesiet al. 2015; Lin et al. 2017).

mRNA generally does not contain *cis*-regulatory elements, so retrogenes were classically viewed as evolutionary dead-ends with limited biological functions. However, some retrogenes are expressed, either by using inherent *cis*-regulatory elements, by acquiring *de novo* promoters, or by hijacking the regulatory mechanism of nearby genes (Kaessmann et al. 2009; Carelli et al. 2016). The zebrafish transcriptome has revealed that retrogene expression is correlated with that of respective parental genes, implying that *cis*-regulatory elements are shared (Zhong et al. 2016). Human genomics data have also revealed that transcribed retrogenes lie close to other genes or even within introns, indicating that they could take advantage of the regulatory elements of neighboring genes (Vinckenbosch et al. 2006). Transcriptomics analysis of zebrafish has also revealed that expression of some retrogenes is more tissue-specific relative to the parental genes, potentially representing subfunctionalization of the duplicated gene (Zhong et al. 2016). In fact, a genome-wide survey of European sea bass (*Dicentrarchus labrax*) demonstrated that most retrogenes in its genome are functional, with some retrogene-parental gene pairs experiencing positive selection (Tine et al. 2021). Accordingly, retrogenes are now regarded as facilitating new gene functions and their potential roles in genome evolution and interspecific differentiation are increasingly appreciated (Kabza et al. 2014; Carelli et al. 2016).

Cyprinodontoid fishes encompass many popular aquarium species such as bluefin killifish (*Lucania goodei*) and guppy (*Poecilia reticulata*), and Pohl et al. (2015) reported on the molecular phylogenetic relationships among cyprinodontoids. One dramatic feature of the visual system of cyprinodontoids is the high numbers of visual cone opsin genes in their genomes. For instance, sheepshead minnow (*Cyprinodon variegatus*) has 11 cone opsin genes, including 2 SWS1, 2 SWS2, 3 RH2, and 4 LWS (Lin et al. 2017). One of the LWS genes in this and other cyprinodontoids is particularly interesting as it does not occur in a genetic array comprising SWS2 and LWS genes, termed the SWS2-LWS synteny. Instead, it is sited within an intron of the gephyrin (GPHN) gene, albeit in an opposing orientation (Lin et al. 2017). GPHN is a postsynaptic protein that cooperates with inhibitory neurotransmitter receptors to mediate neural signal transduction at -aminobutyric acid (GABA) synapses (Sassoè-Pognetto and Fritschy 2000). This LWS gene has been named LWS-R as it originated from a retrotransposition event (Sandkam et al. 2018). Phylogenetic analysis also suggests that it is evolutionarily closer to LWS-2 than other LWS paralogs (Chang et al. 2021), though the detailed evolutionary history of these four cyprinodontoid LWS paralogs has not yet been studied systematically. Although cyprinodontoid LWS-R is a retrogene, it is not intron-less. Apart from in bluefin killifish (in which LWS-R is

intronless), cyprinodontoid LWS-R paralogs retain one intron (Watson et al. 2010 2011; Rennison et al. 2011; Lin et al. 2017; Chang et al. 2020 2021).

LWS-R transcripts have been detected in cDNA samples from eyeball tissue of bluefin killifish, guppy, green swordtail (*Xiphophorus hellerii*), and western mosquitofish (*Gambusia affinis*) (Watson et al. 2010; Laver and Taylor 2011; Ehlman et al. 2015; Kawamura et al. 2016; Chang et al. 2020 2021). However, the biological function of this retrogene remains unclear. Duplicated genes display functional redundancy, so pseudogenization is their destined fate unless they can retain functionality by means of neofunctionalization or subfunctionalization or under circumstances where increased levels of gene product are advantageous (Iñiguez and Hernández 2017). Harada et al. (2019) identified mutant medaka (*Oryzias latipes* and *O. sakaizumii*) with only one of two original LWS genes, but they exhibited no difference in gene expression or red-light sensitivity to the wild-type, supporting the functional redundancy of LWS genes. In bluefin killifish and guppy, the expression level of LWS-R is much lower than for other cone opsin genes and LWS-R has the same amino acid composition as LWS-1 at five key residues crucial for predicting the maximal absorbance wavelength value of visual opsins (Laver and Taylor 2011; Ehlman et al. 2015; Kawamura et al. 2016; Chang et al. 2020 2021). Therefore, the biological meaning of LWS-R is likely negligible and it may display an evolutionary trajectory toward pseudogenization. The finding that LWS-R transcripts in western mosquitofish not only are intron-retaining but also exhibit a 46-basepair deletion in the coding region supports pseudogenization of LWS-R (Chang et al. 2020). However, ontogenetically variable LWS-R expression in bluefin killifish and guppy, as well as significant differences in proportional expression of LWS-R upon exposing guppy to various turbidity conditions, indicate that LWS-R could play a role in vision (Laver and Taylor 2011; Ehlman et al. 2015; Chang et al. 2021). Thus, the true function of LWS-R remains to be determined.

Moreover, how LWS-R expression is regulated has not been established. Two proximal highly conserved segments located between the SWS2 and LWS genes in the SWS2-LWS synteny have been identified in many teleost fishes and they are believed to be the locus-control region (LCR) of LWS genes (Watson et al. 2010; O'Quin et al. 2011). In zebrafish and green swordtail, the LCR plays an enhancer role in LWS gene expression (Tsujimura et al. 2010; Tam et al. 2011). However, as a retrogene, LWS-R may not have inherent *cis*-regulatory elements. So, how is it transcribed? One hypothesis is that it hijacks the regulatory sequences of GPHN (Watson et al. 2010 2011; Laver and Taylor 2011); a plausible supposition given that retinal tissue of zebrafish and green swordtail has been demonstrated to express GPHN (Yazulla and Studholme 2001; Watson et al. 2010). If LWS-R does hijack the GPHN regulatory mechanism, then LWS-R would also be expressed beyond cone cells, and expression of both genes would be correlated. However, GPHN and LWS-R display opposing orientations, meaning that LWS-R transcription is unlikely to be driven by the GPHN promoter. Accordingly, LWS-R should have its own promoter located in its upstream region.

There were two goals of this study. First, by quantifying transcript levels of LWS-R and GPHN in eyeball and brain tissue of two poeciliid fishes, I wanted to determine if their expression levels are correlated. Second, I employed *in silico* promoter prediction software to identify candidate promoters for LWS-R.

MATERIALS AND METHODS

Subjects

Guppies (*Poecilia reticulata*), ranging from 35–44 mm in standard length (SL, the measurement from the most anterior tip of the body to the midlateral posterior edge of the hypural plate), were purchased from an aquarium store in Taichung City and green swordtails (*Xiphophorus hellerii*) (45–63 mm SL) were collected from an invasive population in Yilan County, Taiwan. About 30 individuals of each species were brought back to Tunghai University, and specimens were housed in a concrete tank (125 cm in length × 80 cm in width × 50 cm in height) equipped with an air-powered filter for water circulation located in a climate-controlled greenhouse for three weeks before being sacrificed. Fish were fed artificial fish food (Otohime B2, Marubeni Nisshin Feed) *ad libitum* twice a day. Experimental protocols and specimen handling were performed with approval (109-24) from the Institutional Animal Care and Use Committee (IACUC) of Tunghai University.

Gene annotation, primer design, and promoter prediction

The published LWS-R gene sequence of guppy (KX768568) and GPHN sequences from guppy and green swordtail (guppy: XM_008397472; green swordtail: XM_032584608) were used as references to query and annotate the protein-coding region of the NCBI database sequences (guppy: NC_024351 and green swordtail: GQ999833) using BLASTN 2.8.1+

(Zhang et al. 2000). Specific primers (see Table 1) were used to amplify a segment of the mitochondrial D-loop, β-actin gene, GPHN, and LWS-R. Primer pairs for amplifying LWS-R and LWS-R intronic segments were obtained from Watson et al. (2010) and Kawamura et al. (2016), or were designed based on available database sequences. The forward primers for amplifying LWS-R and quantifying its expression were the same and annealed in the 5' untranslated region (UTR). Primer annealing sites are shown in figure 1. The primer pair for amplifying GPHN was designed based on published gene transcripts from guppy and green swordtail.

Putative LWS-R promoters were probed using two different programs: 1) Promoter 2.0 was designed to recognize candidate RNA polymerase II promoter positions in eukaryotic DNA sequences (Knudsen 1999); and 2) PromPredict utilizes DNA duplex stability as an index of promoters (Rangannan and Bansal 2009) and has been demonstrated to display ~80% recall value in identifying gene promoters in fish sequences (Yella et al. 2018).

DNA and RNA extraction, reverse transcription-PCR, and gene cloning

In order to minimize the effects of circadian rhythms on gene expression, specimens were sampled between 2:00 PM and 3:00 PM. Fish were anesthetized using 0.025% buffered MS-222 (Ethyl 3-aminobenzoate, methanesulfonic acid salt) solution. After the fish were rendered comatose, the body weight and SL of each specimen were measured. DNA samples were extracted from muscle tissues using a Geneaid DNA extraction kit (Cat No./ID: GS100). An RNeasy Plus Universal Mini Kit (Cat No./ID: 73404, QIAGEN) was used to isolate total RNA according to the manufacturer's protocol. The eyeballs and brains from two guppy individuals were combined into a single tissue sample, respectively, but one green swordtail specimen was used for each sample. The eyeball and the brain of each sample was collected and placed in separate 2 ml microcentrifuge tubes containing two stainless steel beads, and then homogenized using a TissueLyser II apparatus (QIAGEN). Total RNA content and quality

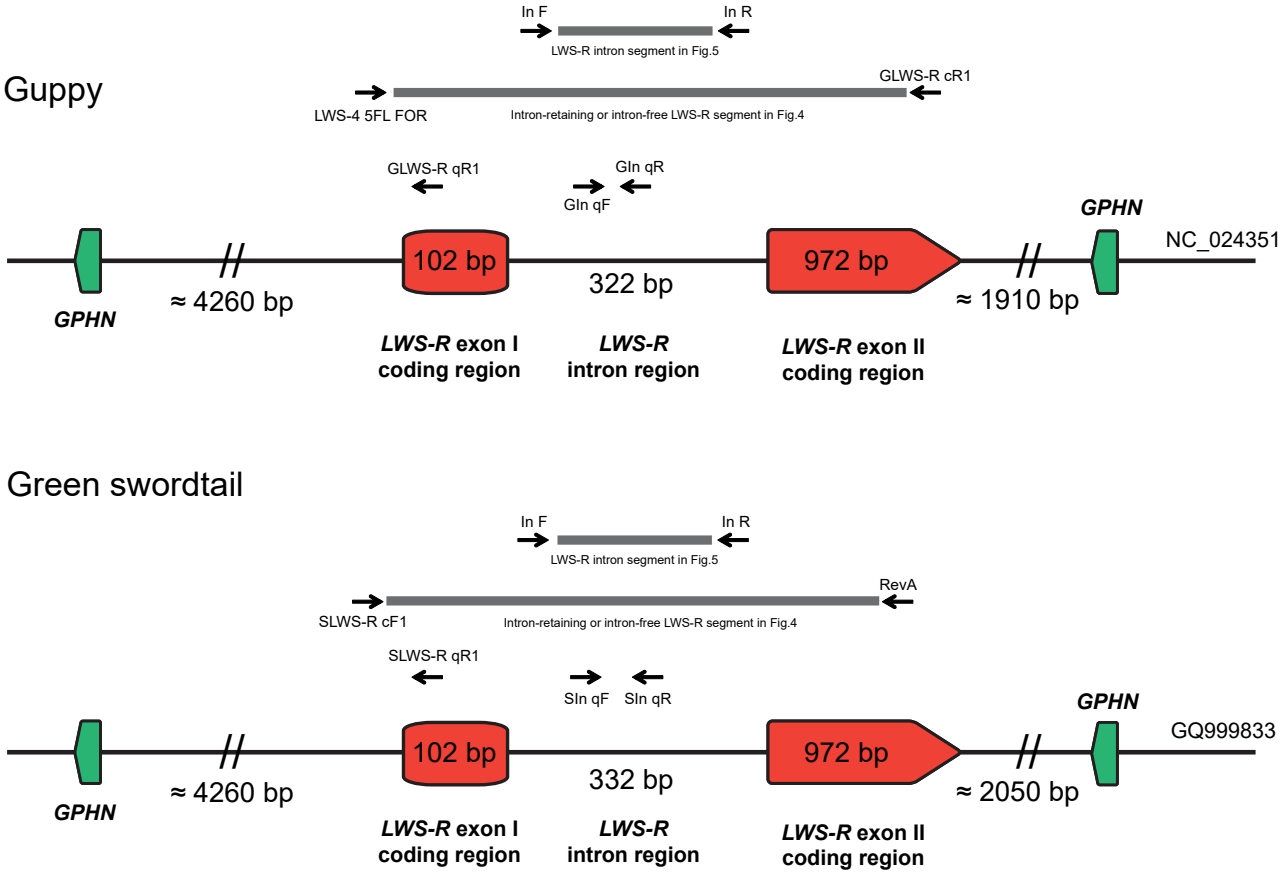

Fig. 1. Genomic organization of the LWS-R and GPHN genes and annealing sites of primers used in this study. Details on the primer sequences are presented in table 1.

were measured using a NanoDrop 1000 system (Thermo Fisher Scientific). The extracted RNA samples were first treated with TURBOTM DNase (Thermo Fisher Scientific) per the manufacturer's protocol, and then they were preserved at -80°C upon adding RNase inhibitor (E0126-40D6, Lucigen).

Before conducting reverse transcription-PCR, the stored RNA samples were treated with recombinant DNase I (04716728001, Roche). Then polymerase chain reaction (PCR) was conducted using double DNase-treated RNA samples as a template with mitochondrial D-loop primers to confirm an absence of contaminating DNA. Reverse transcription-PCR was conducted using approximately 1 μg of total RNA, which was reverse-transcribed using a Verso cDNA Synthesis Kit (Cat No. 00764129, Thermo Fisher Scientific) and Anchored Oligo-dT in a final volume of 20 μl. An RT Enhancer in the Verso cDNA Synthesis Kit degrades double-stranded DNA during the reverse-transcription reaction. PCR using cDNA samples as templates with *β*-actin primers was conducted to check the cDNA quality.

PCR amplification of target genes was performed in a final reaction volume of 25 μL containing 2 ng cDNA, 6 μmol each of forward and reverse primers, 12.5 μL of Fast-Run^{TM} Advanced Taq Master Mix (ProTech, Taipei, Taiwan), and distilled water. The thermal cycling protocol was as follows: one cycle at 94°C for 4 min; 35 cycles of denaturation at 94°C for 30 sec, 50–62°C for 30 sec, and 72°C for 1–4 min; one final single extension step at 72°C for 10 min. The PCR products were then purified using a Qiagen purification kit, subcloned into the pGEM-T Easy vector (Promega), and the clones were forward- and reverse-sequenced using M13 primers. Since LWS-R amplification of cDNA samples from fish eyeball resulted in PCR products of two different sizes, both were cloned. Sequencing was performed using an ABI 3730 version 3.2 analyzer (Applied Biosystems) and by following ABI PRISM BigDye Sequencing Kit protocols (Applied Biosystems). Cloning and sequencing were conducted by Mission Biotech Inc., Taipei, Taiwan. Contig sequences were constructed using the program CodonCode Aligner 9.0, and the results were identified by BLAST analysis against the National Centre for Biotechnology Information (NCBI) database.

Real-time quantitative PCR (qPCR)

The amplification efficiency and melting curve for each qPCR primer pair was tested by five-fold serial dilutions of the templates, with three replicates for each gene and sample. A qPCR primer pair was adopted only when its amplification efficiency was between 90% and 110% and the melting curve analysis revealed it only

Table 1. Sequences of primers used to amplify and quantify expression of the LWS-R, GPHN, and *β*-actin genes in guppy (*Poecilia reticulata*) and green swordtail (*Xiphophorus hellerii*)

Gene	Forward (5'-3')	Reverse (5'-3')
Gene amplification		
Guppy		
LWS-R	LWS-4 5FL FOR: AGCTCAGATCGTCTTTCCAA	GLWS-R qR1: GTCTTYGATGTGGACACTTC
Green swordtail		
LWS-R	SLWS-R cF1: AGATCAAGCAGCTCAGATCG	RevA: CATCCTAGATACTTCCTTCTGGG
Guppy & green swordtail		
LWS-R intron	In F: GTTAAAGTTAGTGTTATCAGAGA	In R: CTGGAAGCACAAGTTCACAT
GPHN	TCATCACGCTCAAGTCAAGG	TCCTTCTCTCCCATGGACAC
β-actin	*β*-actin Forward: CCTGTACGCTTCTGGTCGTA	*β*-actin Reverse: CCTCCAATCCAGACAGAGTA
Mitochondrial D-loop	AAGAGACCACCATCAGTTGA	ATGGTGGGTAACGAGGAGTA
qPCR		
Guppy		
LWS-R (*Ei* =94.90 %)	LWS-4 5FL FOR: AGCTCAGATCGTCTTTCCAA	GLWS-R qR1: CTTGTATGATTGCTGTTTGTGTA
LWS-R intron (E_i =100.09 %)	GIn qF: TGTCTGCTGGCACAAGTAAA	GIn qR: ACTGGCATGCAAGGAAAATA
Green swordtail		
LWS-R (E_i =98.90 %)	SLWS-R cF1: AGATCAAGCAGCTCAGATCG	SLWS-R qR1 ATGATTGCTGTTTGTGTATGTGA
LWS-R intron (E_i =100.05 %)	SIn qF: GCTTGCTGGCACAAAGTAAAC	SIn qR: GCATGTAGGCACGGAAAATA
Guppy & green swordtail		
GPHN (E_i = 99.55 % in Guppy; 99.40% in green swordtail)	CTGGACCCTCGTCCTGAATA	AGTCTGCTGCTCACCTGGTT

Efficiency (E_i) for qPCR primers. LWS-4 5FL FOR was adopted from Kawamura et al. (2016). RevA was adopted from Ward et al. (2008). *β*-actin Forward and Reverse were adopted form Chang et al. (2020).

generated a single product. Expression of opsin genes was determined by qPCR in a Roche LightCycler480 system (Roche). Each reaction contained 10 µl of Roche LightCycler480 SYBR Green I Master mix (Roche Applied Science), 50 ng of cDNA, and 1 µl of each primer (10 µM) (see Table 1) to a final volume of 20 µl. The qPCR reactions were performed in a LightCycler 480 Multiwell Plate system (Roche) with optical adhesive film (Applied Biosystems Ref. 4360954). The following thermal cycles were performed: one cycle of 50°C for 2 min and 95°C for 10 min; followed by 45 cycles of 95°C for 10 sec, 60°C for 10 sec, and 72°C for 10 sec; and then one cycle of 95°C for 5 sec and 65°C for 1 min. A melting-curve analysis was used to verify that only a pure, single amplicon was generated by qPCR. Additionally, representative samples were also electrophoresed to verify that only a single product (band) was present. RNA-free water was used as a template in the control reactions to determine non-specific primer amplification background levels. Three replicates were performed for each target gene for each specimen. Raw data have been submitted to Figshare (https://doi.org/10.6084/m9.figshare.14742708).

The LWS-R:GPHN expression ratio was calculated using the following equation:

$$\frac{T_{\text{LWS-R}}}{T_{\text{GPHN}}} = \frac{1/(1+E_{i(\text{LWS-R})})^{C_{t(\text{LWS-R})}}}{1/(1+E_{i(\text{GPHN})})^{C_{t(\text{GPHN})}}}$$

$T_{\text{LWS-R}}$ represents the expression level of LWS-R and T_{GPHN} is the expression level of the GPHN gene. $E_{i(\text{LWS-R})}$ and $E_{i(\text{GPHN})}$ are the amplification efficiencies for the LWS-R and GPHN primer pairs, respectively. $C_{t(\text{LWS-R})}$ and $C_{t(\text{GPHN})}$ are the average critical cycle numbers for the LWS-R and GPHN genes, respectively.

The LWS-R:GPHN expression ratio for eyeball and brain tissue were compared using a paired *t*-test. Statistical tests were performed in R version 3.6.0 (R Foundation for Statistical Computing, Vienna, Austria).

To determine the expression levels of intron-free and intron-retaining LWS-R transcripts, quantification of plasmid extracted from an intron-retaining LWS-R transcript clone was used as a standard. Then, the *Ct* values of LWS-R and intron-retaining LWS-R in various serially diluted plasmid concentrations were measured to generate linear regression equations. The linear regression for guppy LWS-R transcript was $\text{Log}(T_{\text{LWS-R}}) = -0.4404(Ct_{(\text{LWS-R})}) + 10.559$, and that for guppy intron-retaining LWS-R was $\text{Log}(T_{\text{LWS-R intron}}) = -0.4528(Ct_{(\text{LWS-R intron})}) + 10.517$. The linear regression for green swordtail LWS-R was $\text{Log}(T_{\text{LWS-R}}) = -0.2601(Ct_{(\text{LWS-R})}) + 6.3803$, and that for green swordtail intron-retaining LWS-R was $\text{Log}(T_{\text{LWS-R intron}}) = -0.25578(Ct_{(\text{LWS-R intron})}) + 6.2422$. $T_{\text{LWS-R intron}}$ represents the expression level of intron-retaining LWS-R transcript.

The ratio of intron-free LWS-R to total LWS-R was calculated using the following equation:

$$\frac{T_{\text{intron-free LWS-R}}}{T_{\text{LWS-R}}} = 1 - \frac{T_{\text{LWS-R intron}}}{T_{\text{LWS-R}}}$$

The intron-free LWS-R:LWS-R expression ratios for eyeball in guppy and green swordtail were compared using *t*-test. Statistical tests were performed in R version 3.6.0 (R Foundation for Statistical Computing, Vienna, Austria).

RESULTS

Gene annotation in this study disclosed in greater detail the organization of GPHN and LWS-R in these two poeciliids, and the distances between LWS-R and the upstream and downstream GPHN exons. It showed that LWS-R is located in the intronic region between exons X and XI of the GPHN gene, with the two genes displaying opposing orientations. The distance between the translation start site (TLS) of LWS-R and the upstream GPHN exon is ~4,260 basepairs (bp), whereas it is ~2,000 bp between the LWS-R translation terminus and its downstream GPHN exon (Fig. 1). Promoter2.0 identified one possible promoter in the region spanning the LWS-R TLS and GPHN exon XI for both guppy and green swordtail, whereas PromPredict highlighted four and seven putative promoters for guppy and green swordtail, respectively, in the same region (Supplementary data at https://doi.org/10.6084/m9.figshare.16830730).

GPHN transcripts were successfully amplified from both brain and eyeball cDNA samples (Fig. S1, https://doi.org/10.6084/m9.figshare.14742708), as were LWS-R transcripts (Fig. 2A, and 2B). However, LWS-R transcripts displayed two fragment sizes upon gel electrophoresis. The larger LWS-R transcript was amplified from genomic DNA and both brain and eyeball cDNA samples, whereas the smaller transcript was solely detected in eyeball cDNA samples (Fig. 2A, and 2B). Subsequent sequencing demonstrated that the large LWS-R transcript is intron-retaining and the smaller one is intron-free (Supplementary data at https://doi.org/10.6084/m9.figshare.1683074). Intron-retaining transcripts would encounter a stop codon early in the intronic region. Relatively less PCR product of intron-retaining than intron-free LWS-R was found in guppy (Fig. 2A), but the opposite was the case for green swordtail (Fig. 2B). PCR amplification using a pair of primers specific to the intronic region of LWS-R

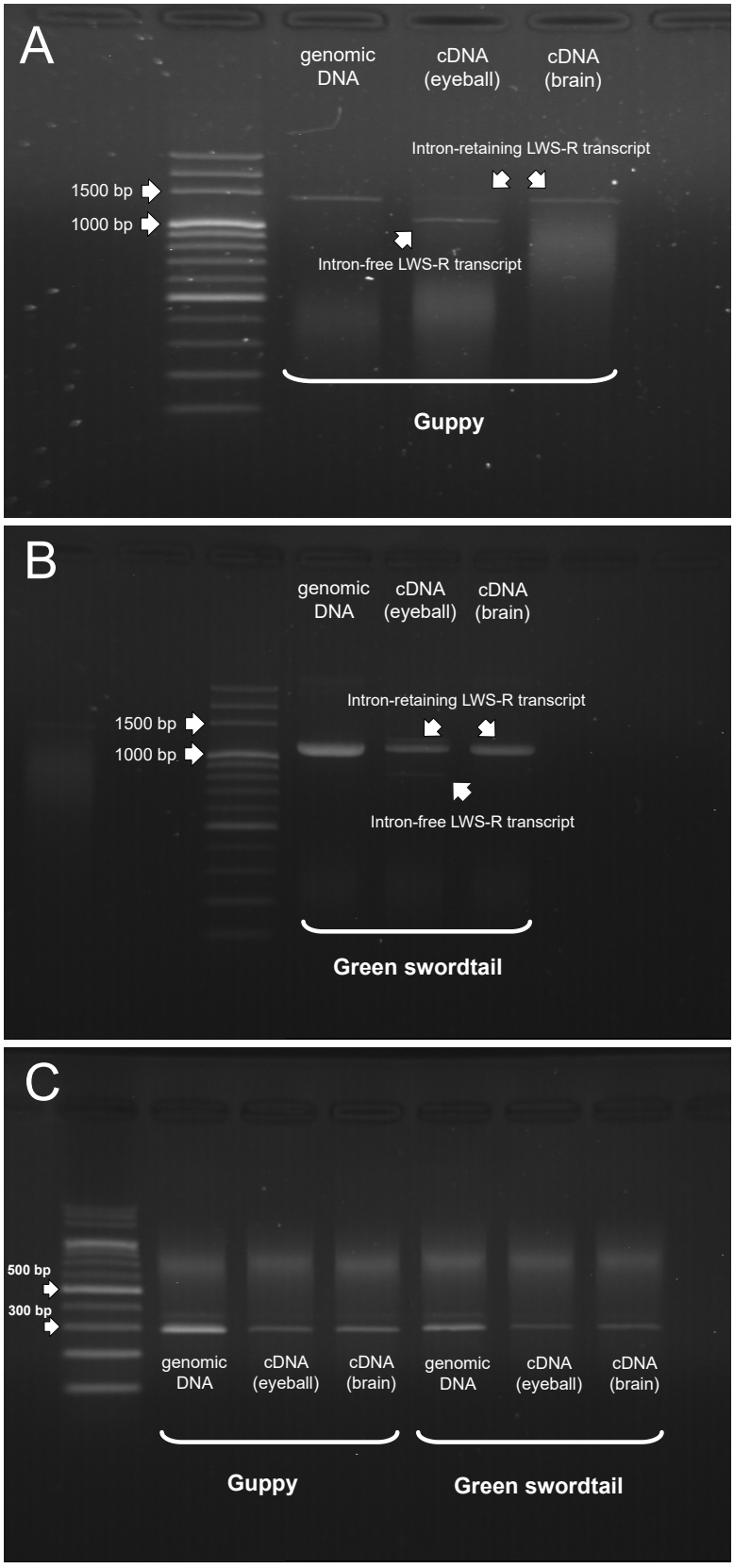

Fig. 2. PCR-amplified target segments from genomic DNA, eyeball cDNA, and brain cDNA. (A) LWS-R transcripts of guppy, (B) LWS-R transcripts of green swordtail, and (C) LWS-R intron of guppy and green swordtail.

intron region was successful regardless of whether the template used was genomic DNA or cDNA (Fig. 2C).

To test if LWS-R and GPHN are co-expressed, a paired *t*-test was used to compare the LWS-R:GPHN gene expression ratio between eyeball and brain, which revealed that the values did not differ significantly for guppy (N = 8, $t = 1.1171$, $p = 0.3008$) or green swordtail (N = 8, $t = -1.5154$, $p = 0.1734$). The mean gene expression ratio of intron-free LWS-R to total LWS-R is 65.05% in guppy and 51.93% in green swordtail (Fig. 3B), and a *t*-test demonstrated these two poeciliids did not have significantly different ratios ($t = 1.5382$, $p = 0.1463$).

DISCUSSION

A particularly interesting finding of this study is that the intron-retaining LWS-R transcript was detected in both poeciliid fishes. Through the increasing body of transcriptomic data, intron retention (IR) is now recognized as a relatively common phenomenon. In humans, up to 80% of protein-coding genes are estimated to display IR, with biological functions for IR, such as regulating gene expression, also having been verified (see Grabski et al. 2020 for a review). However, two factors may confound the detection of IR, *i.e.*, DNA contamination and unspliced nascent pre-mRNA. DNA contamination is almost inevitable during RNA extraction so, in this study, extracted RNA samples were double-treated with DNase before undergoing cDNA preparation, and the negative PCR results using mitochondrial D-loop primers on treated RNA template demonstrated an absence of DNA. Moreover, the RT Enhancer (Verso cDNA Synthesis Kit) also degrades double stranded DNA during reverse transcription. The other factor potentially leading to false detection of intron-retaining LWS-R transcripts is the occurrence of pre-mRNA. Random primers are often used in reverse transcription reactions to significantly improve cDNA synthesis. To avoid pre-mRNA interference, reverse transcription was conducted in this study using oligo-dT primer rather than a cocktail of oligo-dT and random primers. Therefore, without the post-transcriptional modification to acquire a 3' polyadenylated tail, pre-RNA molecules cannot be turned into cDNA.

Gene fusion may occur between a retrogene and its host gene. In the case of LWS-R, gene fusion is unlikely since LWS-R and its host gene display opposing orientations. However, Watson et al. (2010) obtained conflicting PCR results using different combinations of GPHN and LWS-R primers. In the current study, the forward primers located in the 5' UTR region both successfully amplified LWS-R transcript from guppy and green swordtail (Fig. 2). Moreover, based on the annotation analysis of this study, LWS-R lies in the intron between exons X and XI of GPHN. Given that the forward primer for amplifying GPHN anneals between exons VIII and IX, and the reverse primer anneals in exon XIX, if there had been any gene fusion transcripts, they would have been detected by PCR amplification. Since no such transcripts were observed, these experiments confirm that gene fusion between GPHN and LWS-R does not occur in these poeciliids.

In silico analyses using Promoter2.0 and PromPredict software revealed that LWS-R very likely has its own promoter. A 5' rapid amplification of cDNA ends (5' RACE) experiment on LWS-R transcripts would help to precisely locate the promoter. Although

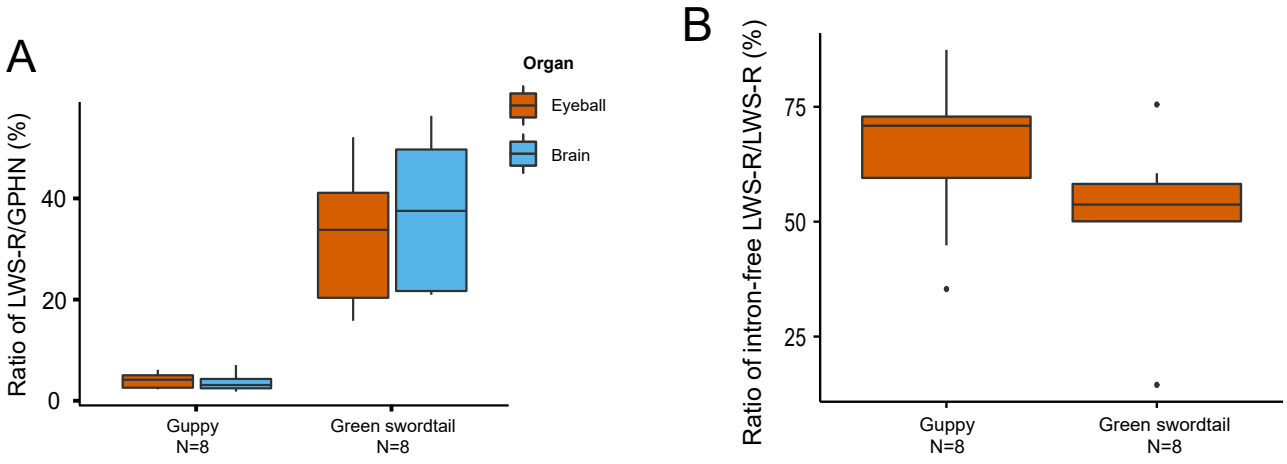

Fig. 3. qPCR-determined expression ratios of (A) LWS-R:GPHN for eyeball and brain samples and (B) intron-free LWS-R to total LWS-R for eyeball samples from guppy and green swordtail. The box boundaries indicate the median (—), 25th and 75th percentiles (boxes), 95% range (|), and outliers (•).

the LWS-R:GPHN expression ratio differs between the two species of poeciliid fishes assessed herein, there was no significant difference between eyeball and brain samples within species, indicating that expression of these two genes is correlated, at least in these two tissues (Fig. 3A). The identification of putative LWS-R promoters counters the hypothesis that LWS-R hijacks the GPHN promoter for transcription. However, why then are the expressions of LWS-R and GPHN correlated if each gene has its own promoters? LWS-R is a visual opsin and GPHN is a postsynaptic protein, so there is no direct functional link between them. A simple explanation is that LWS-R is still controlled by the GPHN enhancer, so LWS-R is transcribed along with GPHN. Alternatively, tight wrapping of DNA around histones naturally curbs transcription and dynamic changes in chromatin structure represent a critical mechanism contributing to gene regulation (Li et al. 2007; Venkatesh and Workman 2015), but LWS-R may not initiate histone exchange by itself. Instead, its transcription may take advantage of the histone exchange arising from production of GPHN transcripts. Accordingly, LWS-R expression becomes a by-product of GPHN transcription. This "hitchhiking" hypothesis may explain why LWS-R transcripts also occur in brain tissue because GPHN is a postsynaptic protein that anchors the $GABA_A$ postsynaptic receptor (Rs), which plays a critical role in the central nervous system (Watanabe et al. 2002; Pizzarelli et al. 2020). A selection pressure test on an additional collection of cyprinodontoid LWS-R sequences would verify whether LWS-R is still a functionally important visual opsin or if the gene is destined for pseudogenization.

Two distinct isoforms of LWS-R transcript were identified in this study, intron-free and intron-retaining. Intron-retaining transcripts were present in both eyeball and brain samples, but intron-free LWS-R was only detected in the eyeball (Fig. 2A and 2B). In the retina, not only cone cells, but also horizontal cells and Müller cells, express $GABA_A$ Rs (Picaud et al. 1998; Yang 2004; Rao et al. 2017). Therefore, LWS-R transcript should be found in many different retinal cells. The detection of intron-retaining LWS-R transcripts in both eyeball and brain supports the aforementioned "hitchhiking" hypothesis. A double fluorescence *in situ* hybridization experiment to detect co-expression of these two genes could help verify that even though LWS-R transcripts are not translated into functional protein, they are still generated where the GPHN gene is expressed. In this study, the proportion of intron-free LWS-R transcripts in the eyeball did not differ between guppy and green swordtail. However, details of the RNA splicing mechanism of LWS-R have not yet been established. If RNA splicing of LWS-R is achieved with *trans*-acting factors of other mature cone opsin mRNAs, the proportion of intron-free LWS-R may fluctuate through time along with the diurnal variation in cone opsin expression (Johnson et al. 2013; Yourick et al. 2019).

The occurrence of intron-retaining LWS-R transcript prompts two questions. Firstly, where does it localize in the cell, *i.e.*, in the nucleus or cytoplasm? If the intron-retaining transcripts are restricted to the nucleus, they may participate in gene regulation or are degraded there. Secondly, are these transcripts translated? If intron-retaining transcripts are transported into the cytoplasm, then the in-frame stop codon in the intronic region may induce nonsense-mediated mRNA decay or translation into a truncated polypeptide (Grabski et al. 2020). Further experiments, such as cytoplasmic RNA extraction and ribosomal profiling, will be needed to verify if intron-retaining LWS-R transcripts are translated.

CONCLUSIONS

This study supports the model that the expressions of LWS-R and GPHN are correlated, at least in brain and eyeball tissue. *In silico* identification of putative promoters and expression levels of these two genes do not support that LWS-R hijacks GPHN for transcription. Instead, LWS-R may hitchhike on chromatin changes during GPHN transcription. Consequently, neurons that express GPHN will also transcribe LWS-R, which is supported by the amplification of LWS-R transcripts from brain cDNA samples in this study. However, since intron-free LWS-R transcripts were only detected in eyeball tissue, only in cone cells do LWS-R transcripts subsequently proceed through the RNA splicing process. Although intron retention functions in gene regulation, transcriptome plasticity, and proteome diversity (Baralle and Giudice 2017; Grabski et al. 2020), the role of the intron-retaining LWS-R transcript remains unknown. It may simply represent a by-product of GPHN transcription that is subsequently degraded, or it exerts biological functions as mRNA or truncated polypeptides. Further detailed studies that incorporate additional cyprinodontoid fishes could reveal how LWS-R transcription is initiated, what post-transcriptional modifications the LWS-R transcripts are subjected to, whether both LWS-R isoforms are translated, and the phylogenetic and gene expression relationships between LWS-R and the other LWS genes, thereby unraveling the true biological function and evolutionary history of LWS-R.

Acknowledgments: The author was supported by

grants from the Ministry of Science and Technology, Taiwan (MOST 109-2621-B-029-006, 110-2621-B-029-005, 110-2621-B-152-001). The author also thanks Dr. Aditya Kumar at the Department of Molecular Biology and Biotechnology (Tezpur University, India) and Dr. Manju Bansal at the Molecular Biophysics Unit (Indian Institute of Science, India) for helping conduct PromPredict analysis. Dr. Yi Ta Shao at the Institute of Marine Biology (National Taiwan Ocean University) provided technical assistance. Dr. Da-Wei Liu's suggestion to quantify the expression levels of the two LWS-R isoforms is greatly appreciated. Dr. John O'Brien provided editing assistance.

Authors' contributions: CC conceived the project, was responsible for experimental animal care, data collection and analysis, and prepared the manuscript.

Competing interests: The author declares that he has no competing interests.

Availability of data and materials: Raw data and supplemental information have been submitted to Figshare (https://doi.org/10.6084/m9.figshare.14742708; https://doi.org/10.6084/m9.figshare.16830745; https://doi.org/10.6084/m9.figshare.16830730).

Consent for publication: Not applicable.

Ethics approval consent to participate: Not applicable.

REFERENCES

Baralle FE, Giudice J. 2017. Alternative splicing as a regulator of development and tissue identity. Nat Rev Mol Cell Biol **18:**437–451. doi:10.1038/nrm.2017.27.

Bellingham J, Tarttelin EE, Foster RG, Wells DJ. 2003. Structure and evolution of the Teleost extraretinal rod-like opsin (errlo) and ocular rod opsin (rho) genes: Is teleost rho a retrogene? J Exp Zool Part B **297B:**1–10. doi:10.1002/jez.b.18.

Bourque G, Burns KH, Gehring M, Gorbunova V, Seluanov A, Hammell M, Imbeault M, Izsvak Z, Levin HL, Macfarlan TS, Mager DL, Feschotte C. 2018. Ten things you should know about transposable elements. Genome Biol **19:**199. doi:10.1186/s13059-018-1577-z.

Carelli FN, Hayakawa T, Go Y, Imai H, Warnefors M, Kaessmann H. 2016. The life history of retrocopies illuminates the evolution of new mammalian genes. Genome Res **26:**301–314. doi:10.1101/gr.198473.115.

Chang C-H, Catchen J, Moran RL, Rivera-Colón AG, Wang Y-C, Fuller RC. 2021. Sequence analysis and ontogenetic expression patterns of cone opsin genes in the bluefin killifish (*Lucania goodei*). J Hered **112:**357–366. doi:10.1093/jhered/esab017.

Chang C-H, Wang Y-C, Shao YT, Liu S-H. 2020. Phylogenetic analysis and ontogenetic changes in the cone opsins of the western mosquitofish (*Gambusia affinis*). PLoS ONE **15:**e0240313. doi:10.1371/journal.pone.0240313.

Chinen A, Hamaoka T, Yamada Y, Kawamura S. 2003. Gene duplication and spectral diversification of cone visual pigments of zebrafish. Genetics **163:**663–675. doi:10.1093/genetics/163.2.663.

Cortesi F, Musilová Z, Stieb SM, Hart NS, Siebeck UE, Malmstrøm M, Tørresen OK, Jentoft S, Cheney KL, Marshall NJ, Carleton KL, Salzburger W. 2015. Ancestral duplications and highly dynamic opsin gene evolution in percomorph fishes. Proc Natl Acad Sci U S A **112:**1493–1498. doi:10.1073/pnas.1417803112.

Dulai KS, von Dornum M, Mollon JD, Hunt DM. 1999. The evolution of trichromatic color vision by opsin gene duplication in new world and old world primates. Genome Res **9:**629–638.

Ehlman SM, Sandkam BA, Breden F, Sih A. 2015. Developmental plasticity in vision and behavior may help guppies overcome increased turbidity. J Comp Physiol A **201:**1125–1135. doi:10.1007/s00359-015-1041-4.

Escobar-Camacho D, Ramos E, Martins C, Carleton KL. 2017. The opsin genes of amazonian cichlids. Mol Ecol **26:**1343–1356. doi:10.1111/mec.13957.

Fujiyabu C, Sato K, Utari NML, Ohuchi H, Shichida Y, Yamashita T. 2019. Evolutionary history of teleost intron-containing and intron-less rhodopsin genes. Sci Rep **9:**10653. doi:10.1038/s41598-019-47028-4.

Grabski DF, Broseus L, Kumari B, Rekosh D, Hammarskjold ML, Ritchie W. 2020. Intron retention and its impact on gene expression and protein diversity: a review and a practical guide. Wiley Interdiscip Rev-RNA **12:**e1631. doi:10.1002/wrna.1631.

Harada Y, Matsuo M, Kamei Y, Goto M, Fukamachi S. 2019. Evolutionary history of the medaka *long-wavelength sensitive* genes and effects of artificial regression by gene loss on behavioural photosensitivity. Sci Rep **9:**2726. doi:10.1038/s41598-019-39978-6.

Hofmann CM, Marshall NJ, Abdilleh K, Patel Z, Siebeck UE, Carleton KL. 2012. Opsin evolution in damselfish: convergence, reversal, and parallel evolution across tuning sites. J Mol Evol **75:**79–91. doi:10.1007/s00239-012-9525-0.

Iñiguez LP, Hernández G. 2017. The evolutionary relationship between alternative splicing and gene duplication. Front Genet **8:**14. doi:10.3389/fgene.2017.00014.

Johnson AM, Stanis S, Fuller RC. 2013. Diurnal lighting patterns and habitat alter opsin expression and colour preferences in a killifish. Proc Biol Sci **280:**20130796. doi:10.1098/rspb.2013.0796.

Kabza M, Ciomborowska J, Makałowska I. 2014. RetrogeneDB—A database of animal retrogenes. Mol Biol Evol **31:**1646–1648. doi:10.1093/molbev/msu139.

Kaessmann H, Vinckenbosch N, Long M. 2009. RNA-based gene duplication: mechanistic and evolutionary insights. Nat Rev Genet **10:**19–31. doi:10.1038/nrg2487.

Kawamura S, Kasagi S, Kasai D, Tezuka A, Shoji A, Takahashi A, Imai H, Kawata M. 2016. Spectral sensitivity of guppy visual pigments reconstituted *in vitro* to resolve association of opsins with cone cell types. Vision Res **127:**67–73. doi:10.1016/j.visres.2016.06.013.

Knudsen S. 1999. Promoter2.0: for the recognition of PolII promoter sequences. Bioinformatics **15:**356–361. doi:10.1093/bioinformatics/15.5.356.

Kuzmin E, Taylor JS, Boone C. 2022. Retention of duplicated genes in evolution. Trends Genet **38:**59–72. doi:10.1016/j.tig.2021.06.016.

Laver CRJ, Taylor JS. 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. doi:10.1186/1471-2148-11-81.

Li B, Carey M, Workman JL. 2007. The role of chromatin during transcription. Cell **128:**707–719. doi:10.1016/j.cell.2007.01.015.

Lin J-J, Wang F-Y, Li W-H, Wang T-Y. 2017. The rises and falls of opsin genes in 59 ray-finned fish genomes and their implications for environmental adaptation. Sci Rep **7:**15568. doi:10.1038/s41598-017-15868-7.

Musilova Z, Cortesi F, Matschiner M, Davies WIL, Patel JS, Stieb SM, de Busserolles F, Malmstrøm M, Tørresen OK, Brown CJ, Mountford JK, Hanel R, Stenkamp DL, Jakobsen KS, Carleton KL, Jentoft S, Marshall J, Salzburger W. 2019. Vision using multiple distinct rod opsins in deep-sea fishes. Science **364:**588–592. doi:10.1126/science.aav4632.

O'Quin KE, Smith D, Naseer Z, Schulte J, Engel SD, Loh YHE, Streelman JT, Boore JL, Carleton KL. 2011. Divergence in *cis*-regulatory sequences surrounding the opsin gene arrays of African cichlid fishes. BMC Evol Biol **11:**120. doi:10.1186/1471-2148-11-120.

Pantzartzi CN, Pergner J, Kozmik Z. 2018. The role of transposable elements in functional evolution of amphioxus genome: the case of opsin gene family. Sci Rep **8:**2506. doi:10.1038/s41598-018-20683-9.

Picaud S, Pattnaik B, Hicks D, Forster V, Fontaine V, Sahel J, Dreyfus H. 1998. $GABA_A$ and $GABA_C$ receptors in adult porcine cones: evidence from a photoreceptor-glia co-culture model. J Physiol **513:**33–42. doi:10.1111/j.1469-7793.1998.033by.x.

Pizzarelli R, Griguoli M, Zacchi P, Petrini EM, Barberis A, Cattaneo A, Cherubini E. 2020. Tuning GABAergic inhibition: Gephyrin molecular organization and functions. Neuroscience **439:**125–136. doi:10.1016/j.neuroscience.2019.07.036.

Pohl M, Milvertz FC, Mayer A, Vences M. 2015. Multigene phylogeny of cyprinodontiform fishes suggests continental radiations and a rogue taxon position of *Pantanodon*. Vertebr Zool **65:**37–44.

Rangannan V, Bansal M. 2009. Relative stability of DNA as a generic criterion for promoter prediction: whole genome annotation of microbial genomes with varying nucleotide base composition. Mol Bio Syst **5:**1758–1769. doi:10.1039/B906535K.

Rao MB, Didiano D, Patton JG. 2017. Neurotransmitter-regulated regeneration in the zebrafish retina. Stem Cell Rep **8:**831–842. doi:10.1016/j.stemcr.2017.02.007.

Rennison DJ, Owens GL, Allison WT, Taylor JS. 2011. Intra-retinal variation of opsin gene expression in the guppy (*Poecilia reticulata*). J Exp Biol **214:**3248–3254. doi:10.1242/jeb.057836.

Sandkam B, Dalton B, Breden F, Carleton K. 2018. Reviewing guppy color vision: integrating the molecular and physiological variation in visual tuning of a classic system for sensory drive. Curr Zool **64:**535–545. doi:10.1093/cz/zoy047.

Sassoè-Pognetto M, Fritschy J-M. 2000. Mini-review: gephyrin, a major postsynaptic protein of GABAergic synapses. Eur J Neurosci **12:**2205–2210. doi:10.1046/j.1460-9568.2000.00106.x.

Shen C-H. 2019. Diagnostic Molecular Biology. Academic Press, San Diego, CA, USA.

Tam KJ, Watson CT, Massah S, Kolybaba AM, Breden F, Prefontaine GG, Beischlag TV. 2011. Regulatory function of conserved sequences upstream of the long-wave sensitive opsin genes in teleost fishes. Vision Res **51:**2295–2303. doi:10.1016/j.visres.2011.09.010.

Terai Y, Miyagi R, Aibara M, Mizoiri S, Imai H, Okitsu T, Wada A, Takahashi-Kariyazono S, Sato A, Tichy H, Mrosso HDJ, Mzighani SI, Okada N. 2017. Visual adaptation in Lake Victoria cichlid fishes: depth-related variation of color and scotopic opsins in species from sand/mud bottoms. BMC Evol Biol **17:**200. doi:10.1186/s12862-017-1040-x.

Tine M, Kuhl H, Teske PR, Reinhardt R. 2021. Genome-wide analysis of European sea bass provides insights into the evolution and functions of single-exon genes. Ecol Evol **11:**6546–6557. doi:10.1002/ece3.7507.

Tsujimura T, Hosoya T, Kawamura S. 2010. A single enhancer regulating the differential expression of duplicated red-sensitive opsin genes in zebrafish. PLoS Genetics **6:**e1001245. doi:10.1371/journal.pgen.1001245.

Venkatesh S, Workman JL. 2015. Histone exchange, chromatin structure and the regulation of transcription. Nat Rev Mol Cell Biol **16:**178–189. doi:10.1038/nrm3941.

Vinckenbosch N, Dupanloup I, Kaessmann H. 2006. Evolutionary fate of retroposed gene copies in the human genome. Proc Natl Acad Sci USA **103:**3220–3225. doi:10.1073/pnas.0511307103.

Ward MN, Churcher AM, Dick KJ, Laver CRJ, Owens GL, Polack MD, Ward PR, Breden F, Taylor JS. 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. doi:10.1186/1471-2148-8-210.

Watanabe M, Maemura K, Kanbara K, Tamayama T, Hayasaki H. 2002. GABA and GABA receptors in the central nervous system and other organs. Int Rev Cytol **213:**1–47. doi:10.1016/s0074-7696(02)13011-7.

Watson CT, Gray SM, Hoffmann M, Lubieniecki KP, Joy JB, Sandkam BA, Weigel D, Loew E, Dreyer C, Davidson WS, Breden F. 2011. Gene duplication and divergence of long wavelength-sensitive opsin genes in the guppy, *Poecilia reticulata*. J Mol Evol **72:**40–222. doi:10.1007/s00239-010-9426-z.

Watson CT, Lubieniecki KP, Loew E, Davidson WS, Breden F. 2010. Genomic organization of duplicated short wave-sensitive and long wave-sensitive opsin genes in the green swordtail, *Xiphophorus helleri*. BMC Evol Biol **10:**87. doi:10.1186/1471-2148-10-87.

Yang X-L. 2004. Characterization of receptors for glutamate and GABA in retinal neurons. Prog Neurobiol **73:**127–150. doi:10.1016/j.pneurobio.2004.04.002.

Yazulla S, Studholme KM. 2001. Neurochemical anatomy of the zebrafish retina as determined by immunocytochemistry. J Neurocytol **30:**551–592. doi:10.1023/a:1016512617484.

Yella VR, Kumar A, Bansal M. 2018. Identification of putative promoters in 48 eukaryotic genomes on the basis of DNA free energy. Sci Rep **8:**4520. doi:10.1038/s41598-018-22129-8.

Yokoyama S. 2000. Molecular evolution of vertebrate visual pigments. Prog Retin Eye Res **19:**385–491. doi:10.1016/s1350-9462(00)00002-1.

Yourick MR, Sandkam BA, Gammerdinger WJ, Escobar-Camacho D, Nandamuri SP, Clark FE, Joyce B, Conte MA, Kocher TD, Carleton KL. 2019. Diurnal variaiton in opsin expression and common housekeeping genes necessitates comprehensive normalization methods for quantitative real-time PCR analyses. Mol Ecol Resour **19:**1447–1460. doi:10.1111/1755-0998.13062.

Zhong Z, Yang L, Zhang YE, Xue Y, He S. 2016. Correlated expression of retrocopies andparental genes in zebrafish. Mol Genet Genomics **291:**723–737. doi:10.1007/s00438-015-1140-5.

Supplementary materials

Fig. S1. PCR-amplified GPHN transcripts from eyeball and brain cDNA samples of guppy and green swordtail. (download)