

# Convergent loss of an anciently duplicated, functionally divergent RH2 opsin gene in the fugu and *Tetraodon* pufferfish lineages

Daniel E. Neafsey\*, Daniel L. Hartl

Department of Organismic and Evolutionary Biology, Harvard University, 16 Divinity Avenue, Cambridge, MA 02138, USA

Received 10 December 2004; received in revised form 21 January 2005; accepted 22 February 2005

Available online 7 April 2005

## Abstract

We describe the complete opsin gene families from the sequenced fugu and *Tetraodon* pufferfish genomes. We report the convergent loss of function of an anciently duplicated, functionally divergent RH2 or “green-sensitive” opsin gene in both pufferfish lineages, designated *RH2-2*. In fugu, *RH2-2* apparently ceased to function very recently following a transposon-induced deletion that truncated the N-terminal 115 amino acids from the translated protein. Although a lack of frameshift or nonsense mutations in the fugu *RH2-2* pseudogene suggests that the gene was lost very recently in this lineage, we were unable to detect any evidence of a selective sweep associated with the fixation of the truncated allele from population data. Interspecific comparison of the remaining fugu *RH2-2* coding sequence paradoxically indicates that the gene was under strong purifying selection until the truncation occurred.

© 2005 Elsevier B.V. All rights reserved.

**Keywords:** Pseudogenes; Gene duplication; Transposable elements; *Takifugu rubripes*; Pufferfish

## 1. Introduction

Gene family evolution is a dynamic process, in which the birth of new member genes from tandem or whole-genome duplication is balanced by mutational decay of some duplicates into pseudogenes. Theory and empirical observation suggest that most gene duplicates that will be lost begin to decay soon after they originate, unless one copy assumes a new function or mutation subdivides the functions of the original copy among the pair (Ohno, 1970; Li, 1980; Force et al., 1999). The loss of anciently duplicated genes exhibiting evidence of functional divergence or subdivision is expected to be a rare occurrence under these models.

Here we report the recent convergent loss of an opsin gene duplicate exhibiting strong evidence of functional divergence that we estimate to be ~80 million years old. The opsins compose a small family of retinal transmembrane proteins that facilitate vertebrate color vision via sensitivity to different wavelengths of light. Opsins are members of the G-protein-coupled receptor (GPCR) family that are expressed in the membranes of retinal cone cells. Like rhodopsin, which is expressed in retinal rod cells, they bind with a retinylidene chromophore via a protonated Schiff base at Lys-296. Illumination isomerizes the chromophore and initiates the G-protein signal transduction cascade. Interactions between the opsin protein and the chromophore determine the maximal wavelength sensitivity ( $\lambda_{\max}$ ) of the chromophore. Amino acid substitutions in opsin proteins affecting these interactions shift the  $\lambda_{\max}$ , yielding visual pigments maximally sensitive to wavelengths of light in the ultra-violet (<380 nm) or “visual” (380–740 nm) range of the electromagnetic spectrum.

While characterizing the opsin gene families in the sequenced *Tetraodon nigroviridis* and *Takifugu rubripes* (fugu) pufferfish genomes, we discovered the convergent

**Abbreviations:** LTR, long terminal repeat; GPCR, G-protein-coupled receptor.

\* Corresponding author. The Broad Institute of MIT and Harvard, 320 Charles St., Cambridge, MA 02141, USA. Tel.: +1 617 324 5259; fax: +1 617 452 4588.

E-mail address: [neafsey@broad.mit.edu](mailto:neafsey@broad.mit.edu) (D.E. Neafsey).

loss of a medium-wavelength-sensitive (“green”) opsin paralog, designated *RH2-2*, from both the *Tetraodon* and fugu pufferfish genomes. Evidence from extensive structural, biochemical, and comparative evolutionary studies of opsins and the related rhodopsin protein (for review see Sakmar et al., 2002) strongly suggests that the spectral sensitivity of *RH2-2* is distinctly blue-shifted relative to its sister paralog *RH2-1*. Comparison to intact orthologs in medaka indicates that pufferfish *RH2-1* and *RH2-2* were, until recently, subject to similarly strong levels of purifying selection.

In this manuscript we describe our efforts to understand the evolutionary circumstances surrounding the loss of *RH2-2* in these two pufferfish lineages. The *Tetraodon* copy of *RH2-2* is highly degenerate, and may have become a pseudogene soon after the *Tetraodon* and fugu lineages diverged 18–30 million years ago (Crnogorac-Jurcevic et al., 1997). We focus our efforts on fugu, where *RH2-2* was lost much more recently as the result of a nearby non-Long Terminal Repeat (LTR) retroelement insertion. A deletion that likely accompanied the retroelement insertion caused the truncation of the N-terminal 115 codons of fugu *RH2-2* (Fig. 1), removing the first two of seven transmembrane domains. We checked for expression of the truncated gene in a fugu cDNA library to substantiate its nonfunctional status, because the truncation abuts a methionine codon (which could act as a novel start codon) and the remaining coding sequence exhibits no fixed or segregating indels or nonsense mutations. The presence of non-truncated, apparently functional copies of *RH2-2* in five other members of the genus *Takifugu* indicates that the truncation fixed in *T. rubripes* in the recent past, but we were unable to detect strong evidence of selection associated with the fixation event from polymorphism data. We conclude that the fixation event may have occurred neutrally by genetic drift, or alternatively, that if truncation was driven to fixation by

selection, the fixation may not have occurred sufficiently recently for a detectable signal of natural selection to persist.

## 2. Materials and methods

### 2.1. Identification of pufferfish opsin genes

Versions 3 and 6 of the fugu and *Tetraodon* genome assemblies, respectively, were queried with T-BLASTX using a zebrafish rhodopsin (RH1) coding sequence (accession no. NM131084). Highly similar nucleotide sequences were aligned to each other and to a selection of publicly available teleost opsin and rhodopsin sequences with Clustal X using the default parameter settings (see Appendix for accession numbers). Medaka *RH2-1* was assembled from two expressed sequence tags (ests; ID=BJ491781, BJ495952) downloaded from the MBase medaka est library Olestall0309 (<http://mbase.bioweb.ne.jp>). A neighbor-joining tree was constructed from the aligned sequences in Clustal X with a Kimura correction for multiple substitutions to identify fugu and *Tetraodon* orthologs belonging to the four major clades of the vertebrate opsin gene family: RH2 (“green-sensitive”), SWS1 (“UV-sensitive”), SWS2 (“blue-sensitive”), and LWS (“red-sensitive”). The tree was rooted with rhodopsin (RH1) sequences and nodal support was determined with 1000 bootstrap replicates. All teleost opsin sequences were tested for statistical evidence of gene conversion with Sawyer’s Runs Tests, v1.4.1 (Sawyer, 1989) using both polymorphic informative sites and silent informative sites.

Sequences belonging to the RH2 clade were further analyzed phylogenetically using a maximum likelihood approach to substantiate the timing of the duplication event leading to pufferfish *RH2-1* and *RH2-2*. Fugu rhodopsin (accession no. AF137214) was used as an outgroup. Hier-

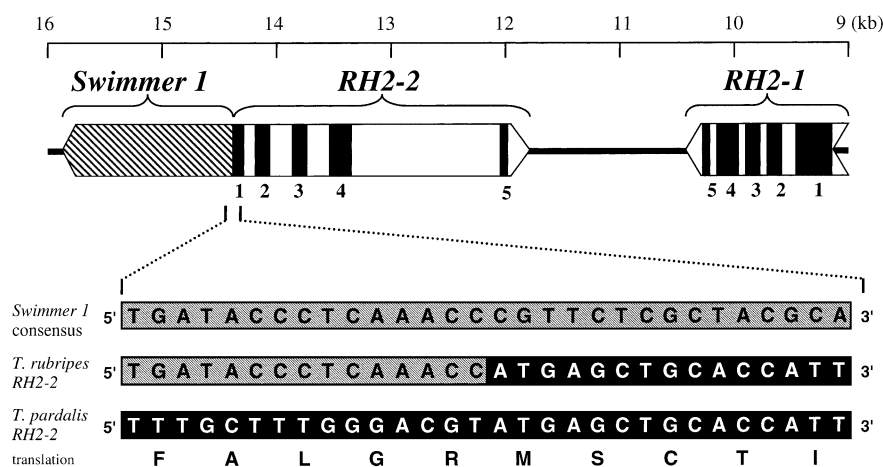


Fig. 1. Physical map of fugu *RH2-1* and *RH2-2* genes, showing truncation of *RH2-2* induced by *Swimmer 1* non-LTR element. Kilobase scale at top shows location of genes on scaffold 2650 from the JGI Fugu Genome Assembly v. 3.0. Open reading frames are left-to-right for *RH2-2* and right-to-left for *Swimmer 1* and *RH2-1*. Numbers below dark bars indicate exons. An expanded view of the junction between the *Swimmer 1* insertion and *RH2-2* is presented below the physical map, together with non-truncated *Swimmer 1* and *RH2-2* sequences.

archical model testing implemented in Modeltest 3.0 (Posada and Crandall, 1998) indicated that a general time-reversible (GTR) model allowing for invariant sites and gamma-distributed variation in rates among sites was most appropriate for the data. Maximum likelihood analysis was implemented in Paup v.4.02b using a heuristic search with random sequence addition and TBR branchswapping. Bootstrap support was determined with 500 replicates using a neighbor-joining starting tree.

Retroelement family affiliation of the partial reverse transcriptase (RT) sequence abutting the truncated fugu *RH2-2* opsin gene was determined by protein BLAST analysis using a translation of the RT sequence.

## 2.2. Functional analysis of teleost *RH2* opsins

Amino acid translations of *RH2* nucleotide sequences from zebrafish, medaka, and pufferfish were analyzed for amino acid identity at sites known to affect the spectral sensitivity of rhodopsin and/or opsin proteins (Sakmar et al., 1989; Lin et al., 1998; Yokoyama and Radlwimmer, 1998; Cowing et al., 2002). Site numbering followed bovine rhodopsin numbering conventions (Hargrave et al., 1993).

## 2.3. Tissue samples

All *T. rubripes* and *Takifugu pseudomus* samples were collected from the waters of Quingdao, China, except for *T. rubripes* SH5, which was acquired from a fish market in Shimonoseki, Japan. Samples of *Takifugu exascurus*, *Takifugu pardalis*, *Takifugu porphyreus*, *Takifugu snyderi*, and *Takifugu stictonotus* were collected from unknown locations. All samples were graciously donated by other investigators, who are listed in the Acknowledgements.

## 2.4. Expression analysis

We obtained a high-density nylon-gridded cDNA library prepared from Fugu eye tissue from MRC Geneservice London. The library was probed for evidence of expression of *RH2-2* using a DIG High Prime DNA Labeling and Detection Starter Kit II from Roche, according to the protocol included with the kit. A 512 base pair (bp) probe sequence containing exons 2, 3 and 4 of *RH2-2* was generated from fugu genomic DNA using primers GF (GCTGTCCCACCTCTTGTG) and G2R2 (AGACTCTGATTCCTGCTGCTG). Seven clones that appeared to positively hybridize with the probe were ordered from MRC Geneservice and sequenced with vector primers according to the protocol described below.

## 2.5. DNA extraction, amplification, and sequencing

Tissue samples preserved in ethanol were soaked in distilled water at 4 °C for 5 min prior to DNA extraction. All extractions were performed using NucleoSpin columns (BD

Biotech) according to the manufacturer's protocol. PCR was carried out in 20 µL volumes containing 1 µL of template DNA, 0.1 µL *taq* polymerase (Roche), 2 µL 10X PCR buffer (Roche) and final concentrations of 50 µM for each dNTP and 0.5 µM for each primer. Amplifications were performed as follows: initial denaturation at 94 °C for 90 s, followed by 35 cycles of denaturation at 94 °C for 30 s, annealing at 58 °C for 30 s, and extension at 72 °C for 1 min. After the final cycle, the temperature was held at 72 °C for an additional 5 min. Amplification products were analyzed on 1.5% agarose gels.

Successful amplifications were cleaned using a Qiaquick PCR Purification kit (Qiagen). Purified products were ligated into a pGEM T vector (Promega), which was then used to transform chemically competent XL-1 Blue *E. coli* (Stratagene). Colonies were PCR screened using vector primers. Positive transformations were sequenced in both directions using vector primers with BigDye3 (ABI) on an Applied Biosystems 3100 sequencer.

## 2.6. Intraspecific polymorphism analysis

A recent analysis of 1962 polymorphic RAPD markers failed to find evidence of genetic differentiation between *T. rubripes* and *T. pseudomus* (Song et al., 2001). Hence, they were considered a single species for this analysis. The truncated *RH2-2* opsin and a portion of the abutting retroelement reverse transcriptase (RT) sequence were amplified from each *T. rubripes* and *T. pseudomus* sample by PCR in five separate reactions spanning overlapping regions. An 835 bp fragment of the upstream RT sequence was amplified with the primer pair SWIMF (GAGCATGAAGGAGCGAAAAG) and SWIMR (CTCGTCATGGACAGCTTCAC). An 832 bp fragment beginning in the RT region and containing the complete sequences of exons 1, 2, 3, and a portion of exon 4 was amplified using primers SWF (GAAGCTGTCCATGACGAGTG) and G2R2. A 341 bp fragment containing the complete sequence of exon 4 was amplified with primers G4F (TCCCCTCATTTGATTGTCTCC) and G4R (AAAACTGTGTGGTTTCCTGTAG). A 977 bp fragment spanning most of the large intron between exons 4 and 5 was amplified with primers G4F2 (TTGCTGCCTGGATTTTCTT) and G5R2 (ACACCCTCGAATGAGTTGCT). A 501 bp fragment containing exon 5 and surrounding intron sequence was amplified using primers G5F (GTATCCAGCATCCCCAAATG) and G5R (TTTGAGCATGGAGACGTCAG). Each amplification product was cloned, and at least six colonies per amplification were sequenced. Singleton substitutions were verified through a second round of amplification and sequencing from genomic DNA to eliminate *taq* error from the alignment. The linkage phase of haplotypes from different amplification reactions was checked by performing PCR with the SWIMF+G5R primers on each genomic DNA sample and cloning the resultant 3.1 kilobase (kb)

product using a Topo XL kit (Invitrogen). PCR and sequencing were then carried out as necessary on the 3.1 kb template using internal primer pairs.

Sequences from separate reactions were concatenated and overlapping regions removed to create two 3121 bp haplotypes for each genomic DNA sample. Sequences were aligned by hand. The population genetics parameter  $\theta = 4N_e\mu$ , where  $N_e$  is the effective population size and  $\mu$  is the neutral mutation rate, was calculated for various subsets of the polymorphism data using the  $\pi$  and  $\theta_W$  estimators.  $\pi$  is calculated as the average number of pairwise differences between two random sequences (Tajima, 1989), and  $\theta_W$  is derived from the number of segregating sites. Under neutral equilibrium conditions the two estimators are expected to be equal. Tajima's  $D$  statistic (Tajima, 1989) measures the significance of deviations between the two estimators, and Fay and Wu's  $H$  test is designed to detect the signature of selective sweeps in the form of an excess of derived mutations segregating at high frequency (Fay and Wu, 2000). Fay and Wu's  $H$  Test was implemented using the web interface at <http://crimp.lbl.gov/hctest.html>. The ancestral state of polymorphic loci in the fugu *RH2-2* opsin was determined by sequencing those loci in congeners. The ancestral state of polymorphic loci in the adjacent *Swimmer* sequence was determined from a consensus sequence (accession no. AY598941) of recent *Swimmer* insertions in the JGI Fugu Genome Assembly v.3.0.

### 2.7. Interspecific comparisons

The absence of the transposition-induced *RH2-2* truncation in congeners of *T. rubripes* was confirmed through PCR using the degenerate primer G2F2 (SAA-CYTGGCNGTGGCTG) with primer G2R2. Primer G2F2 anneals to a region of exon 1 in *RH2-1* and *RH2-2* that has been truncated in *T. rubripes RH2-2*. Differences in intron size between the paralogs yielded a 736 bp product from *RH2-1* and an 889 bp product from the non-truncated version of *RH2-2*. Amplification products were cloned and sequenced as described above. Sequences were aligned by hand.

A 652 bp region of mitochondrial D-loop was PCR amplified from all sampled species of *Takifugu* using primers FDLF (CCATAAAGCAAGTACAGGAAGC) and FDLR (TGGTTCAAACTTAAAGGTCTCTCC). Amplification products were directly sequenced and sequences were aligned by hand. Genetic distances between sequences were computed with DNADIST v3.6 software (Felsenstein, 1993) using an F84 substitution model.

Relative levels of purifying selection acting on the *RH2-1* and *RH2-2* loci were estimated by computing orthologous pairwise estimates of the  $K_a/K_s$  ratio between species. We estimated  $K_a/K_s$  using a maximum likelihood method that accounts for bias in the transition/transversion ratio and codon usage, as implemented in PAML v.3.13 (Yang, 1997).

$K_a$  is the number of nonsynonymous substitutions per nonsynonymous site, and  $K_s$  is the number of synonymous substitutions per synonymous site. A  $K_a/K_s$  ratio close to 0 indicates strong purifying selection, whereas a ratio close to 1 signifies neutral evolution.

## 3. Results

### 3.1. Identification of pufferfish opsin genes and gene family history

Fig. 2 presents a neighbor-joining tree showing the placement of fugu and *Tetraodon* opsin genes relative to other teleost opsins. Teleost LWS sequences were used to root the tree. Bootstrap support for the LWS, SWS1, SWS2, and RH2 clades is 100%, allowing confident determination of pufferfish opsin orthologs. No evidence for gene conversion among the teleost opsin sequences was detected using Sawyer's Runs Tests (Sawyer, 1989). Sequences not previously described were submitted to Genbank (fugu *LWS*: AY598942; *Tetraodon LWS*: AY598943; *Tetraodon RH2-1*: AY598944; fugu *RH2-2*: AY598945; *Tetraodon RH2-2*: AY598946; fugu *SWS2*: AY598947; *Tetraodon SWS2*: AY598948). Note that membership in a particular clade does not guarantee that an individual opsin pigment has a  $\lambda_{\max}$  typical of that clade. For example, an opsin might be phylogenetically grouped with the LWS "red-sensitive" clade, and yet harbor substitutions that make it maximally sensitive to green wavelengths of light.

We failed to detect SWS1 ("UV-sensitive") orthologs in either the fugu or *Tetraodon* genomes, indicating that this gene was most likely lost in a common Tetraodontiform ancestor after the Percomorph radiation. Both pufferfish genome projects are over 90% complete, so the probability of both assemblies missing a gene present in both genomes is only 1%. In the SWS2 ("blue-sensitive") clade, tree topology and two copies of SWS2 opsin in Nile tilapia and killifish indicate a gene duplication event preceding the divergence of superorder Percomorpha (represented by Nile tilapia, medaka, killifish, flounder, and pufferfish). Only one SWS2 gene was recovered from each pufferfish genome, indicating the probable loss of the other paralog in the Tetraodontiformes lineage and perhaps one or more independent losses of a paralog in the medaka and flounder lineages. A duplication of the LWS ("red-sensitive") opsin occurred after divergence of the common ancestor of killifish, medaka, Nile tilapia, and flounder from the pufferfish lineage. A separate LWS opsin duplication occurred in the zebrafish lineage after divergence from goldfish (Chinen et al., 2003).

The RH2 ("green-sensitive") clade is the most evolutionary dynamic group among teleost opsins. This clade exhibits evidence of frequent gene duplications, particularly within order Cypriniformes (zebrafish, goldfish, and carp). The very high bootstrap values associated with the nodes in

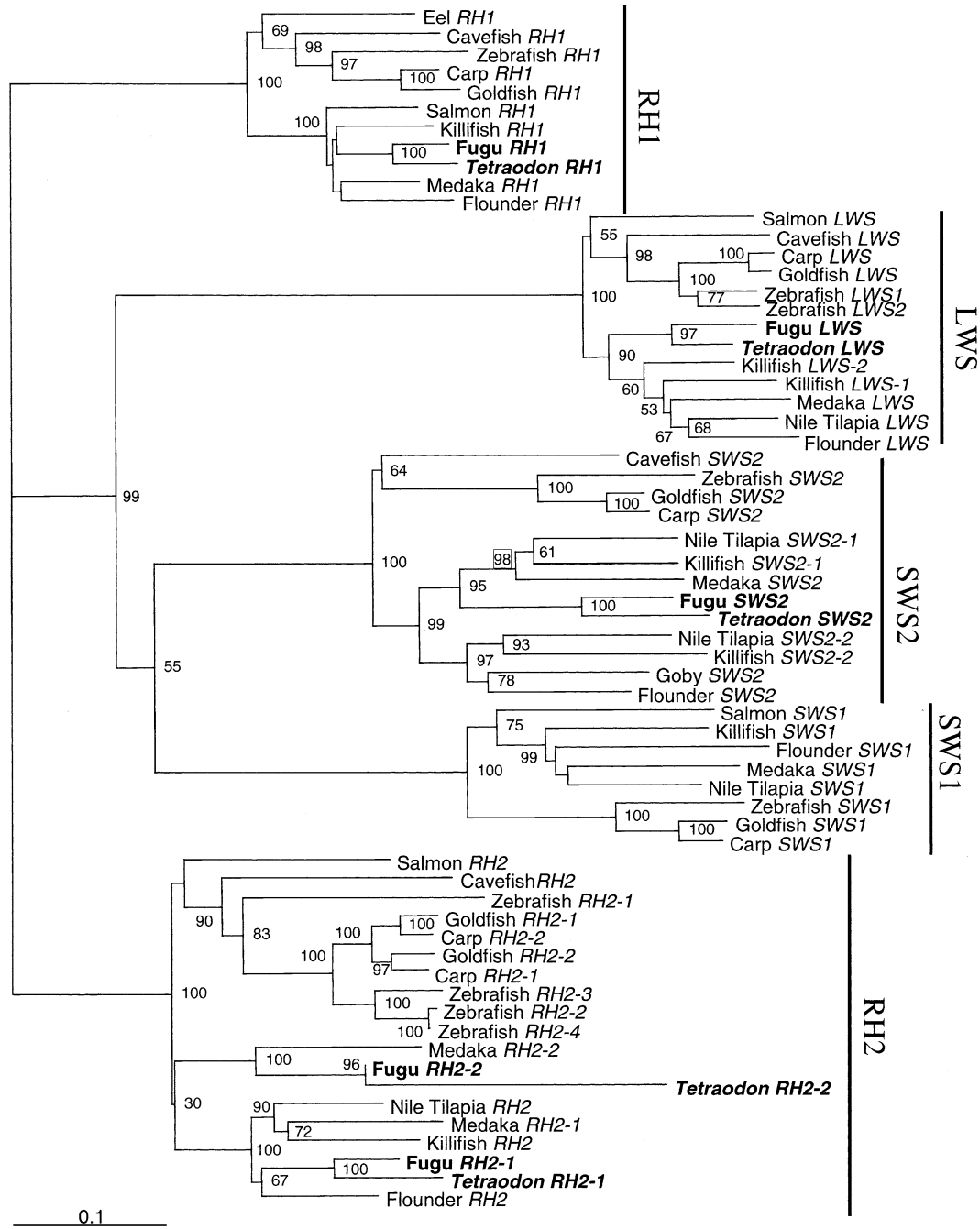


Fig. 2. Neighbor-joining phylogenetic tree of teleost LWS, SWS1, SWS2, and RH2 opsins, rooted with teleost LWS opsin sequences, with correction for multiple substitutions. Numbers at nodes indicate percentage bootstrap support of at least 50% from 1000 replicates. Pufferfish sequences are indicated in bold. Accession numbers for previously reported sequences are listed in the Appendix. Scale bar indicates 0.1 nucleotide substitutions/site.

this clade, together with the lack of statistical evidence for gene conversion, strongly suggest that the timing of duplication events relative to lineage divergences can be accurately determined. The neighbor-joining and maximum likelihood trees (Figs. 2 and 3) are identical in topology except in the placement of salmon RH2, which does not affect our conclusions. As has been previously observed, there is evidence of one RH2 duplication event preceding the divergence of zebrafish and goldfish, another duplica-

tion event preceding the divergence of carp and goldfish after their divergence from zebrafish, and two more duplication events in the zebrafish lineage after divergence from goldfish/carp (Chinen et al., 2003).

A single RH2 duplication event is observed outside Cypriniformes, preceding the divergence of orders Tetraodontiformes (pufferfish) and Beloniformes (medaka) 60–80 million years ago (Hedges and Kumar, 2002). Nodal support grouping medaka and pufferfish orthologs of RH2-1 and

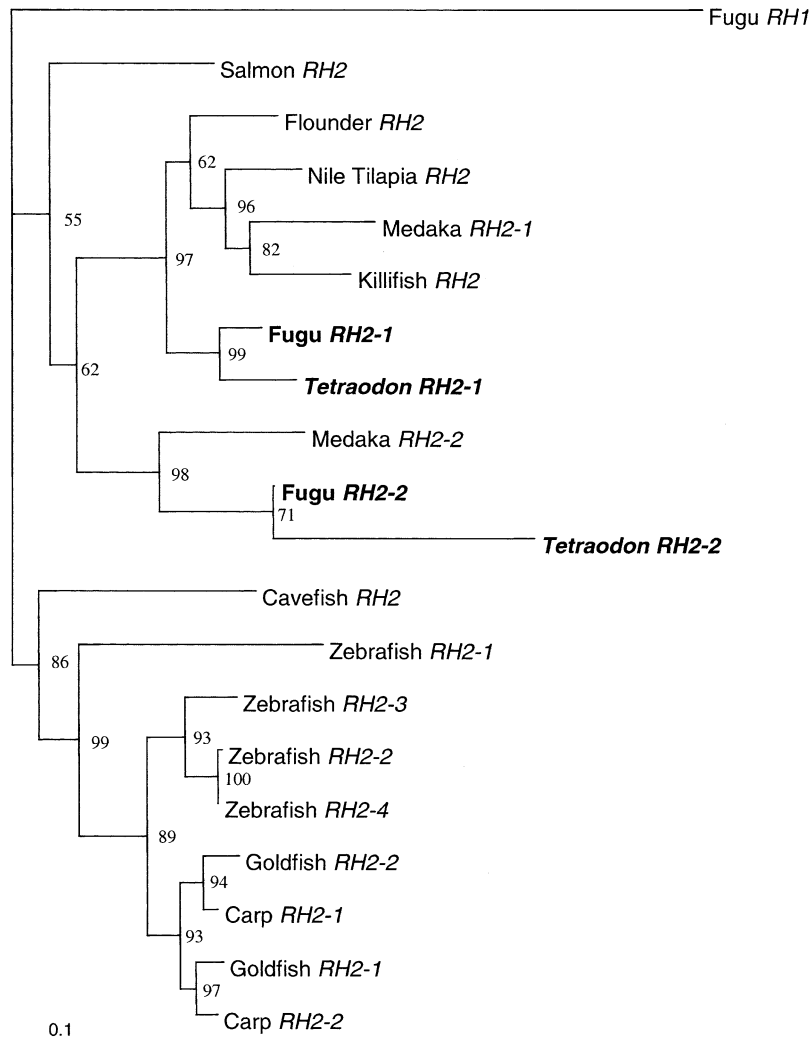


Fig. 3. Maximum likelihood phylogenetic tree of teleost RH2 (green-sensitive) opsins constructed using GTR+I+G substitution model, rooted with Fugu RH1. Numbers at nodes show bootstrap support from 500 replicates. Pufferfish sequences are indicated in bold. Scale bar indicates 0.1 nucleotide substitutions/site.

RH2-2 is high in both the neighbor joining and maximum likelihood trees, with respective bootstrap percentage values of 100/97 for RH2-1 and 100/98 for RH2-2. Only orthologs of RH2-1 are observed in the other Percomorph species, indicating RH2-2 has not been sampled or is lost in these lineages.

Both the *Tetraodon* and fugu RH2-2 sequences have been subject to dramatic mutations. Roughly two-thirds of the *Tetraodon* RH2-2 opsin is missing (Fig. 4). The remaining 342 bp of coding sequence contains two frame-shift-inducing deletions and a high proportion of non-synonymous substitutions, assuring its status as a pseudogene. The fugu RH2-2 opsin lacks the N-terminal 115 amino acids, corresponding to the first two transmembrane domains of the protein (Figs. 1 and 4). The apparent agent of this truncation event is a recently inserted non-LTR retroelement belonging to the *Swimmer 1* family, first characterized in medaka (Duvernell and Turner, 1998). This retroelement family was overlooked in the first annotation of the fugu genome (Aparicio et al., 2002), possibly because

BLAST analysis only indicates three other *Swimmer 1* insertions in the genome assembly. No indels or nonsense mutations were observed in the remaining fugu RH2-2 sequence. Because fugu RH2-2 has not been subject to further mutation and the truncation event abuts an in-frame methionine codon, we checked for expression to confirm that it was no longer a functional gene in its truncated form.

### 3.2. Expression analysis

Seven clones on the high-density gridded eye cDNA library appeared to hybridize with the RH2-2 probe, and were ordered from MRC GeneService for further analysis. Sequencing of the seven inserts yielded only one opsin gene, however. Clone G33-06 was found to contain a fragment of RH2-1 opsin sequence, which exhibits 80% nucleotide similarity to the RH2-2 opsin and therefore serves as a good positive control for our hybridization conditions. Three of the other clones (G33-05, G33-09, and G34-03) were in close physical proximity to G33-06 on the

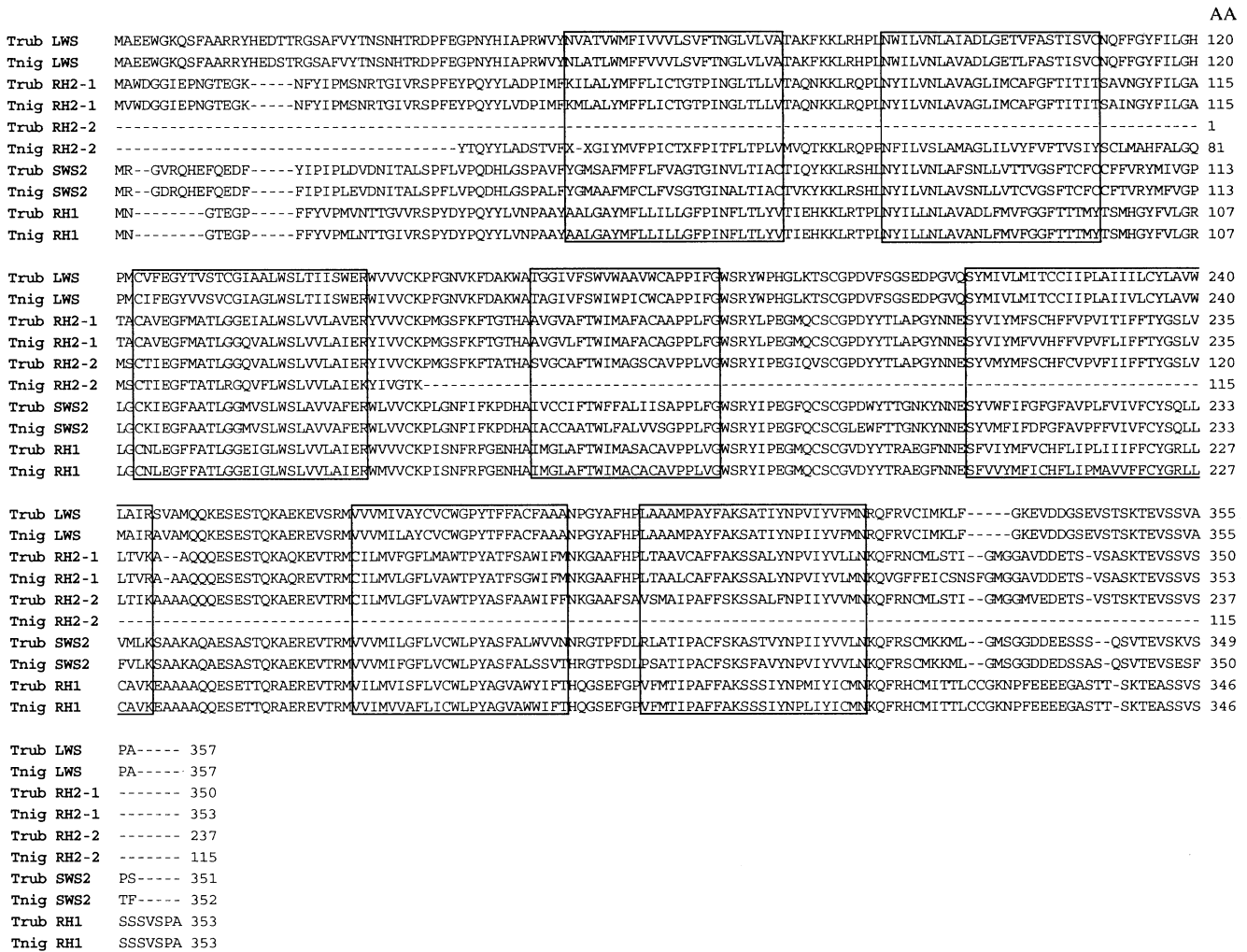


Fig. 4. Amino acid alignment of fugu and *Tetraodon* opsin sequences. Site numbering for rhodopsin (RH1) sequences corresponds to system established for bovine rhodopsin (Hargrave et al., 1993) and is used to identify tuning sites analyzed in Table 1. Predicted transmembrane domains are indicated by boxes.

filter but were found to contain no sequence similarity to the probe, suggesting physical cross-contamination during the gridding process. The last three clones (G26-C2, G41-P1, and G48-A3) contained sequence from unrelated open reading frames, and likely represent spurious hybridization. We conclude that it is unlikely that *RH2-2* is still expressed in the fugu retina. Although it is possible that fugu *RH2-2* is expressed in the retina only transiently or in non-retinal tissue, there are no GPCRs described in the literature or the GPCRDB database (<http://www.gpcr.org/7tm/>) that lack transmembrane domains I and II. Even if fugu *RH2-2* is expressed and translated, its product is therefore most likely nonfunctional.

### 3.3. Functional analysis of *RH2* teleost opsins

Significant evidence of functional divergence among zebrafish, medaka, and pufferfish *RH2-2* opsins was observed (Table 1). Zebrafish *RH2-1* and *RH2-2* did not exhibit amino acid differences at known functional tuning

sites, though in vitro measurement of the  $\lambda_{max}$  of these genes recovers a difference of about 9 nm ( $\lambda_{max}$  = 466.5 and 475.7 for *RH2-1* and *RH2-2*, respectively; Chinen et al., 2003). Zebrafish *RH2-3* and *RH2-4* ( $\lambda_{max}$  = 488.0 and 504.9, respectively; Chinen et al., 2003) show possible evidence of functional divergence at sites 122, 124, and 299.

Medaka and pufferfish *RH2-1* and *RH2-2*, resulting from an independent duplication, also show evidence of functional divergence, including some substitutions convergent with zebrafish at sites 122 and 124. Medaka and pufferfish both exhibit divergence at sites 122 and 295, indicating these changes took place soon after the initial tandem duplication event before the Percomorph radiation. Medaka has lineage-specific divergence at sites 124 and 292, while pufferfish shows lineage-specific divergence at sites 86 and 90.

Our contention that these amino acid substitutions confer functional differences to the medaka and pufferfish *RH2-1* and *RH2-2* paralogs is supported by coordination in the direction of the spectral shift among the tuning sites. All six

Table 1  
Amino acid identity in RH2 (green-sensitive) teleost opsins at known spectral tuning sites

	M86L <sup>a</sup>	G90S <sup>a</sup>	A117G <sup>a</sup>	T118A <sup>b</sup>	E122Q <sup>c</sup>	A124T <sup>a</sup>	S164A <sup>d</sup>	H181Y <sup>d</sup>	W265Y <sup>a</sup>	Y261F <sup>d</sup>	T269A <sup>d</sup>	A292S <sup>a</sup>	A295S <sup>a</sup>	A299C <sup>a</sup>
Zebrafish <i>RH2-1</i>	M	G	A	T	Q	A	A	E	W	F	A	A	S	A
Zebrafish <i>RH2-2</i>	M	G	A	T	Q	A	A	E	W	F	A	A	S	A
Zebrafish <i>RH2-3</i>	M	G	A	T	Q	S	A	E	W	F	A	A	S	S
Zebrafish <i>RH2-4</i>	M	G	A	T	E	A	A	E	W	F	A	A	S	A
Medaka <i>RH2-1</i>	M	G	A	T	E	A	A	E	W	F	A	A	A	A
Medaka <i>RH2-2</i>	M	G	A	T	Q	S	A	E	W	F	A	S	S	A
Puffer <i>RH2-1</i>	M	G	A	T	E	A	A	E	W	F	A	A	A	A
Puffer <i>RH2-2</i>	L <sup>c</sup>	V <sup>c</sup>	A	T	E/Q	A	A	E	W	F	A	A	S	A

<sup>a</sup> Lin et al. 1998 (mutations shift towards blue).

<sup>b</sup> Cowing et al. 2002 (mutations shift towards blue).

<sup>c</sup> Sakmar et al. 1989 (mutation shifts towards blue 20 nm).

<sup>d</sup> Yokoyama and Radlwimmer 1998 (mutations shift towards blue).

<sup>e</sup> Sites present in *Tetraodon* only.

of the substitutions observed in pufferfish and medaka RH2 proteins conspire to give *RH2-2* a shorter, blue-shifted  $\lambda_{\max}$  relative to *RH2-1*. The exact binomial probability of all six mutations occurring in this coordinated fashion is only 1.6%, suggesting the coordination is non-random.

Further, site-directed mutagenesis indicates that substituting glutamine for glutamic acid at site 122 (E122Q) may be a mutation of major effect, owing to interactions between this amino acid residue and the chromophore. Sakmar et al. (1989) found that this substitution alone can confer a 20 nm blue-shift in  $\lambda_{\max}$  in bovine rhodopsin. Yokoyama et al. (1999) measured a 25 nm blue-shift in the  $\lambda_{\max}$  of coelacanth rhodopsin attributable to changes E122Q and A292S. Spectrophotometry of zebrafish opsins reconstituted in vitro indicates that this effect likely applies to RH2 as well as rhodopsin. Zebrafish *RH2-1* and *RH2-2* differ from *RH2-4* only at site 122 among the tuning sites in Table 1, and yet exhibit  $\lambda_{\max}$  values shortened by 38.4 nm and 29.2 nm relative to *RH2-4*, respectively (Chinen et al., 2003). E122Q alone, then, may be sufficient to infer significant functional divergence between pufferfish/medaka *RH2-1* and *RH2-2*.

### 3.4. Intraspecific polymorphism at swimmer 1 and fugu *RH2-2*

Summary statistics for 835 bp of *Swimmer 1* RT sequence and 2285 bp of opsin *RH2-2* sequence are presented in Table 2. Individual segregating sites are illustrated in Table 3. Among 14 genomic isolates (accession nos. AY599599–AY599612), there were six segregating sites in the RT sequence and 13 segregating sites in the opsin *RH2-2* sequence. All polymorphisms were single-

nucleotide substitutions, except for a 1 bp deletion at site 506 in the opsin sequence. There were no nonsense mutations or indels found to be segregating in the coding sequence of either region, providing additional evidence that the retrotransposon inserted and fixed very recently. Also supporting this assumption is the observation that many of the segregating mutations have occurred at hyper-mutable CpG dinucleotides. When methylated, CpG dinucleotides experience C to T transitions through de-amination of the 5-methyl cytosine at a rate approximately an order of magnitude greater than the normal nucleotide mutation rate (Kondrashov, 2003). These transition mutations are therefore expected to be among the first mutations observed at coding sequence that has recently begun to evolve neutrally. Four of the six segregating mutations in *Swimmer 1* and three of the thirteen segregating mutations in *RH2-2* represent C to T transitions at CpG dinucleotides in either the sense or antisense strands.

Five of the six polymorphisms in the *Swimmer 1* sequence would cause amino acid substitutions in a translated product, reflecting a lack of purifying selection at this 5' truncated 'dead on arrival' retroelement insertion. Three of the six mutations segregating in the *RH2-2* coding region would cause amino acid substitutions, also suggesting the gene has entered a phase of neutral evolution.

Our calculated values of  $\pi$  and  $\theta_w$  were similar for all sequence subsets considered. Tajima's *D* statistic (Tajima, 1989) confirms that we are unable to reject a null model of neutral evolution for the *Swimmer 1* sequence ( $p < 0.549$ ), fugu *RH2-2* ( $p < 0.545$ ), or both sequences considered together ( $p < 0.453$ ). Fay and Wu's *H* test also fails to indicate recent selection associated with the *Swimmer 1*

Table 2  
Summary statistics for polymorphisms segregating in *Swimmer 1* and fugu *RH2-2*

Locus	Sites	S	$\pi$	$\theta$	# Haps	<i>D</i>	<i>H</i>
<i>RH2-2</i> (silent sites only)	2285 (1727)	13 (10)	0.0019 (0.0019)	0.0019 (0.0018)	9	−0.0972	−0.8352
<i>Swimmer 1</i>	835	6	0.0018	0.0023	6	−0.6626	0.6154
<i>RH2-2</i> + <i>Swimmer 1</i>	3121	19	0.0019	0.0020	11	−0.3017	−0.2198



All pairwise comparisons within *Takifugu* yielded divergence values of less than 10%. *T. pardalis* and *T. porphyreus* exhibited the least divergence from *T. rubripes*, at approximately 5%. [Shedlock et al. \(1992\)](#) have computed that mitochondrial D-loop evolves in salmonid fish at a rate of approximately 0.75% per million years, indicating that the *T. rubripes* lineage has been separately evolving for ~6.66 million years, or 66N generations.

*RH2-1* and *RH2-2* sequences from the sampled members of genus *Takifugu* are insufficiently divergent to compare levels of selective constraint between the two loci.  $K_a/K_s$  ratios for orthologous fugu-medaka comparisons at *RH2-1* and *RH2-2* were 0.1265 and 0.0469, respectively. This indicates that *RH2-2* was subject to an equal, if not greater, magnitude of purifying selection than *RH2-1* prior to truncation.

#### 4. Discussion

Opportunities to study the transformation of functional genes into pseudogenes are rare. Because such events are relatively infrequent, most sequences characterized as pseudogenes are mutation-ridden fossils that offer no hope of understanding why they were abandoned by purifying selection. Inference of causality in ancient pseudogene formation may only be possible under unique circumstances. For example, the convergent acquisition of trichromatic vision in apes and howler monkeys may be able to explain the greater proportion of pseudogenes in olfactory gene families in those primate lineages ([Gilad et al., 2004](#)).

There are two selective scenarios under which the loss of functionality of *RH2-2* in fugu and *Tetraodon* may have come about. First, the non-functionalizing mutations in *RH2-2* may have fixed because they were neutral or nearly neutral. Under this scenario, *RH2-2* would have ceased to perform a selectively valuable function for the organism, presumably because habitat shift, change at another locus, or some other factor altered the “adaptive landscape” that had maintained *RH2-2* under strong purifying selection for ~80 million years. Mutations creating less-functional or nonfunctional copies of *RH2-2* would drift to fixation over time, or be fixed by a selective sweep at a nearby linked gene.

The other selective scenario that may have precipitated the demise of *RH2-2* is positive selection. The blue-shifted spectral sensitivity of *RH2-2* may have recently become a fitness liability to fugu and *Tetraodon*, rather than just a neutral or redundant sensory trait. The first mutations to knock out these genes may have been swept quickly to fixation by positive selection. Our analysis of polymorphisms segregating at *RH2-2* in the fugu genome found a reduced level of polymorphism in the vicinity of *RH2-2* but failed to detect significant evidence of a sweep ([Table 2](#)). Our power to reject the neutral null hypothesis may have been limited, however.

Neither a neutral nor a positive selection model of the fixation of the truncated form of fugu *RH2-2* accounts for the extremely abrupt transition from purifying selection to neutral evolution at this locus. Interspecific comparison with medaka indicates the remaining *RH2-2* coding sequence shows no evidence of an elevated rate of amino acid substitution or change in nucleotide composition, suggesting that the truncation abruptly signaled the beginning of neutral evolution at this locus. This is not the expected model of pseudogene formation for coding sequence under relaxed purifying selection. Previous studies of genes released partially or fully from selective constraint have found a gradual loss of functionality. [Hendriks et al. \(1987\)](#) observed an elevated rate of non-synonymous substitution in the  $\alpha A$  crystallin gene expressed in the rudimentary lens of blind mole rats (*Spalax ehrenbergi*). [Yokoyama et al. \(1995\)](#) detected an elevated rate of nucleotide substitution and C to T transition at CpG dinucleotides in the opsins of blind Mexican cave fish (*Astyanax fasciatus*). In both of these studies, it appears that the functionality of the genes in question is gradually being eroded by mutation while expression continues.

Speculation about a potential selective motive for the loss of *RH2-2* in fugu and *Tetraodon* is difficult without microspectrophotometric measurements of the  $\lambda_{max}$  of pufferfish cone cells or reconstituted visual pigments. Though both fugu and *Tetraodon* are demersal and presumably use vision to help locate their favored arthropod and crustacean prey (FishBase: [www.fishbase.org](http://www.fishbase.org)), the two species live in markedly different habitats. Fugu inhabits temperate and sub-tropical seas, whereas *Tetraodon* lives in tropical freshwater or brackish streams. Freshwater environments often contain dissolved and suspended organic matter, which elevate short-wave light absorbance relative to marine environments ([Levine et al., 1979](#)). A common mechanism to explain the loss of *RH2-2* in fugu and *Tetraodon* would likely rely on more precise knowledge of the relative spectral sensitivities of pufferfish opsin genes and the photic microhabitats of these species. Fine-scale spatial and temporal expression analysis of *RH2-1* and the intact *RH2-2* opsin in other species of *Takifugu*, in particular, may be valuable in further explaining the loss of this anciently-duplicated, functionally divergent gene in the two genomic model pufferfish species.

#### Acknowledgements

We thank David Haig, John Wakeley, and Justin Blumenstiel for helpful comments on the manuscript. We thank Toshiaki Itami, Linsheng Song, Mari Yotsu-Yamashita, Yusuki Yamanoue, and Masaki Miya for generously donating tissue samples. This project was funded by a student dissertation grant to D.N. from the Department of Organismic and Evolutionary Biology at Harvard University and a National Science Foundation Graduate Fellowship.

## Appendix A

The following teleost sequences used in phylogenetic analyses were retrieved from Genbank: eel (*Anguilla anguilla*) *RH1* (accession no. L78007), Mexican cavefish (*Astyanax* spp.) *RH1* (U123281), *LWS* (AH003047), *SWS2* (AH007939), and *RH2* (S75251), zebrafish (*Danio rerio*) *RH1* (AB087811), *LWS-1* (AB087803), *RH1-2* (AB087804), *SWS2* (AB087809), *SWS1* (AB087810), *RH2-1* (AB087805), *RH2-2* (AB087806), *RH2-3* (AB087807), and *RH2-4* (AB087808), carp (*Cyprinus carpio*) *RH1* (U02475), *LWS* (AB055656), *SWS2* (AB113668), *SWS1* (AB113669), *RH2-1* (AB110602), and *RH2-2* (AB110603), goldfish (*Carassius auratus*) *RH1* (L11863), *LWS* (L11867), *SWS2* (L11864), *SWS1* (AB113669), *RH2-1* (L11866), and *RH2-2* (L11865), salmon (*Oncorhynchus* spp.) *RH1* (AY214146), *LWS* (AY214145), *SWS1* (AY214148), and *RH2* (AY214147), killifish (*Lucania goodei*) *RH1* (AY296738), *LWS-1* (AY296740), *RH1-2* (AY296741), *SWS2-1* (AY296736), *SWS2-2* (AY296737), *SWS1* (AY296735), and *RH2* (AY296739), fugu (*Takifugu rubripes*) *RH1* (AF201471) and *RH2-1* (AF226989), *Tetraodon nigroviridis* *RH1* (AJ293018), medaka (*Oryzias latipes*) *RH1* (AU244308), *LWS* (AB001604), *SWS2* (AB001602), *SWS1* (AB001605), and *RH2-2* (AB001603), flounder (*Hippoglossus hippoglossus*) *RH1* (AF156265), *LWS* (AF316498), *SWS2* (AF316497), *SWS1* (AF156264), and *RH2* (AF156263), Nile tilapia (*Oreochromis niloticus*) *LWS* (AF247124), *SWS2-1* (AF247116), *SWS2-2* (AF247120), *SWS1* (AF191221), and *RH2* (AF247124).

Medaka *RH2-1* was assembled from two expressed sequence tags (ests; ID=BJ491781, BJ495952) downloaded from the MBase medaka est library Olestall0309 (<http://mbase.bioweb.ne.jp>).

## References

- Aparicio, S., et al., 2002. Whole-genome shotgun assembly and analysis of the genome of Fugu rubripes. *Science* 297, 1301–1310.
- Chinen, A., Hamaoka, T., Yamada, Y., Kawamura, S., 2003. Gene duplication and spectral diversification of cone visual pigments of zebrafish. *Genetics* 163, 663–675.
- 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.
- Crnogorac-Jurcevic, T., Brown, J.R., Lehrach, H., Schalkwyk, L.C., 1997. *Tetraodon fluviatilis*, a new puffer fish model for genome studies. *Genomics* 41, 177–184.
- Duvernell, D.D., Turner, B.J., 1998. Swimmer 1, a new low-copy-number LINE family in teleost genomes with sequence similarity to mammalian L1. *Mol. Biol. Evol.* 15, 1791–1793.
- Fay, J.C., Wu, C.L., 2000. Hitchhiking under positive Darwinian selection. *Genetics* 155, 1405–1413.
- Felsenstein, J., 1993. PHYLIP (Phylogeny Inference Package) Version 3.6. Department of Genetics, University of Washington, Seattle.
- Force, A., Lynch, M., Pickett, F.B., Amores, A., Yan, Y.L., Postlethwait, J., 1999. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151, 1531–1545.
- Gilad, Y., Wiebe, V., Przeworski, M., Lancet, D., Paabo, S., 2004. Loss of olfactory receptor genes coincides with the acquisition of full trichromatic vision in primates. *PLoS Biol.* 2, E5.
- Hargrave, P.A., Hamm, H.E., Hofmann, K.P., 1993. Interaction of rhodopsin with the G-protein, transducin. *Bioessays* 15, 43–50.
- Hedges, S.B., Kumar, S., 2002. Genomics. Vertebrate genomes compared. *Science* 297, 1283–1285.
- Hendriks, W., Leunissen, J., Nevo, E., Bloemendal, H., de Jong, W.W., 1987. The lens protein alpha A-crystallin of the blind mole rat, *Spalax ehrenbergi*: evolutionary change and functional constraints. *Proc. Natl. Acad. Sci. U. S. A.* 84, 5320–5324.
- Hudson, R.R., Kaplan, N.L., 1985. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* 111, 147–164.
- Kondrashov, A.S., 2003. Direct estimates of human per nucleotide mutation rates at 20 loci causing Mendelian diseases. *Hum. Mutat.* 21, 12–27.
- Levine, J.S., MacNichol Jr., E.F., Kraft, T., Collins, B.A., 1979. Intraretinal distribution of cone pigments in certain teleost fishes. *Science* 204, 523–526.
- Li, W.H., 1980. Rate of gene silencing at duplicate loci: a theoretical study and interpretation of data from tetraploid fishes. *Genetics* 95, 237–258.
- Lin, S.W., Kochendoerfer, G.G., Carroll, K.S., Wang, D., Mathies, R.A., Sakmar, T.P., 1998. Mechanisms of spectral tuning in blue cone visual pigments. Visible and raman spectroscopy of blue-shifted rhodopsin mutants. *J. Biol. Chem.* 273, 24583–24591.
- Lynch, M., Conery, J.S., 2003. The origins of genome complexity. *Science* 302, 1401–1404.
- Ohno, S., 1970. *Evolution by Gene Duplication*. Springer-Verlag, New York.
- Posada, D., Crandall, K.A., 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14, 817–818.
- 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. U. S. A.* 86, 8309–8313.
- Sakmar, T.P., Menon, S.T., Marin, E.P., Awad, E.S., 2002. Rhodopsin: insights from recent structural studies. *Annu. Rev. Biophys. Biomol. Struct.* 31, 443–484.
- Sawyer, S., 1989. Statistical tests for detecting gene conversion. *Mol. Biol. Evol.* 6, 526–538.
- Shedlock, A.M., Parker, J.D., Crispin, D.A., Pietsch, T.W., Burner, G.C., 1992. Evolution of the salmonid mitochondrial control region. *Mol. Phylogenet. Evol.* 1, 179–192.
- Song, L., Liu, B., Xiang, J., Qian, P.Y., 2001. Molecular phylogeny and species identification of pufferfish of the genus *Takifugu* (Tetraodontiformes, Tetraodontidae). *Mar. Biotechnol. (N.Y.)* 3, 398–406.
- Tajima, F., 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123, 585–595.
- Yang, Z., 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* 13, 555–556.
- 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., Meany, A., Wilkens, H., Yokoyama, R., 1995. Initial mutational steps toward loss of opsin gene function in cavefish. *Mol. Biol. Evol.* 12, 527–532.
- Yokoyama, S., Zhang, H., Radlwimmer, F.B., Blow, N.S., 1999. Adaptive evolution of color vision of the *Comoran coelacanth* (*Latimeria chalumnae*). *Proc. Natl. Acad. Sci. U. S. A.* 96, 6279–6284.