Evolutionary Dynamics of Rhodopsin Type 2 Opsins in Vertebrates

Shozo Yokoyama* and Takashi Tada

Department of Biology, Rollins Research Center, Emory University

*Corresponding author: E-mail: syokoya@emory.edu

Associate editor: Takashi Gojobori

Abstract

Among the five groups of visual pigments in vertebrates, the rhodopsin type 2 (RH2) group shows the largest number of gene duplication events. We have isolated three intact and one nonfunctional RH2 opsin genes each from Northern lampfish (Stenobrachius leucopsar) and scabbardfish (Lepidopus fitchi). Using the deduced amino acid sequences of these and other representative RH2 opsin genes in vertebrates, we have estimated the divergence times and evolutionary rates of amino acid substitution at various stages of RH2 opsin evolution. The results show that the duplications of the lampfish and scabbardfish RH2 opsins have occurred ~60 and ~30 million years ago (Ma), respectively. The evolutionary rates of RH2 opsins in the early vertebrate ancestors were ~0.25 × 10⁻⁹/site/year, which increased to ~1 × 10⁻⁹ to 3 × 10⁻⁹/site/year in euteleost lineages and to ~0.3 × 10⁻⁹ to 5 × 10⁻⁹/site/year in coelacanth and tetrapods.

Key words: opsin, visual pigment, gene duplication, evolutionary rate, vertebrates.

Introduction

The fundamental mechanism of evolution of multigene families is gene duplication. Tandemly duplicated genes facilitate a further increase in the number of genes by unequal crossover. During the process of gene duplication, some genes may acquire new functions (Ohno 1970), but many others are also expected to become nonfunctional (Nei and Rooney 2005). Indeed, duplicated globin, sensory receptor, and many other genes reveal the dynamic processes of birth and death of duplicated genes in genomes (Nei 1987; Li 1997; Nei et al. 2008). The rates of sequence evolution can also vary considerably among duplicated genes (Li and Gojobori 1983; Li et al. 1985; Zhang 2006).

Visual pigments mediate the first step of vision, and each pigment consists of a chromophore (either 11-cis-retinal or 11-cis-3,4-dehydroretinal) and an opsin encoded by a specific opsin gene. Many vertebrate species have five major groups of visual pigments that have arisen through four gene duplication events: rhodopsin type 1 (RH1, rhodopsins), rhodopsin type 2 (RH2, RH1-like), short wavelength–sensitive type 1 (SWS1), short wavelength–sensitive type 2 (SWS2), and middle and long wavelength–sensitive (M/LWS) groups, having the evolutionary relationship of (((RH1, RH2), SWS2), SWS1), M/LWS (Yokoyama S and Yokoyama R 1996; Yokoyama et al. 2008). The vertebrate ancestor already possessed all five types of opsins (Yokoyama S and Yokoyama R 1996). During vertebrate evolution, additional gene duplications have occurred in all five groups (Spady et al. 2006; Yokoyama et al. 2008), among which the RH2 group shows the largest number of gene duplication events (Spady et al. 2006).

The major function of a visual pigment can be characterized by its wavelength of maximal absorbance (λmax). The λmax of the currently known RH2 pigment ranges from 467 nm in gecko (Kojima et al. 1992) and zebrafish (Chinen et al. 2005) to 535 nm in a cichlid (Carleton et al. 2005). The adaptive λmax shift of the coelacanth RH2 pigment to the depth of 200 m was caused by E122Q and M207L (Yokoyama et al. 1999; Yokoyama et al. 2008). The blue-shifted λmax of the RH2 pigment in the nocturnal gecko was caused mostly by specific amino acid substitutions at four sites (Takenaka and Yokoyama 2007). Because the amount of green–red light is reduced significantly at twilight (McFarland and Munz 1975), the blue-shifted λmax of the gecko pigment seems to be adaptive. The variable λmaxs (467–505 nm) of the zebrafish RH2 pigments are based largely on the amino acids (E or Q) at site 122 (Chinen et al. 2005); however, the biological significance of these variable λmaxs is not immediately clear. If these λmax shifts were needed for organisms to adapt to their ecological environments, then they had to wait for the necessary mutations to accumulate. Consequently, the speed of mutation accumulation is expected to significantly affect the efficiency of such phenotypic adaptations.

Here we have isolated four RH2 opsin genes, each from Northern lampfish (Stenobrachius leucopsar) and scabbardfish (Lepidopus fitchi); in each species, one RH2 gene has become nonfunctional. Considering these and other representative RH2 opsin genes in a wide range of vertebrate species, we shall evaluate two fundamental evolutionary statistics: the divergence times at all nodes and the rates of sequence evolution at all branches in the composite phylogenetic tree of RH2 opsins in vertebrates.

Materials and Methods

The DNA Samples and Molecular Clonings of RH2 Genes

High molecular weight DNAs of Northern lampfish (S. leucopsar) and scabbardfish (L. fitchi) have been isolated...
from their body tissues using a standard phenol–chloroform extraction procedure (Yokoyama et al. 2008).

The nucleotide sequences between exons 3 and 4 of lampfish RH2 genes were first cloned by the polymerase chain reaction (PCR) method using degenerate forward primer (5’-TBTGYAARCMATGGGNAGYTYYAATT-3’) and reverse primer (5’-GCTTTCRGRGTRGAHKCWGADTCYTGCTG-3’). The PCR products in the pBluescript SK(−) vector were sequenced by using an LI-COR 4300 automated DNA sequencer (LI-COR, Lincoln, NE). To clone the remaining coding regions of the lampfish opsin genes, inverse PCRs were performed using an identical set of primers: 5’-CTGTCATGATCCAYTTAGGCCAGC-3’ (forward) and 5’-CCACCTGAAASCAACTCRGCTCC-3’ (reverse). The PCR products were cloned into the pBluescript SK(−) vector and sequenced.

To clone the RH2 genes of the scabbardfish, a genomic DNA library was constructed with XhoI-digested 2FIXII vector and Sau3AI partially digested scabbardfish genomic DNA (10–20 kb). Screening ~1 × 10^6 recombinant plaques using the RH1 opsin gene of bovine (*Bos taurus*, GenBank accession number M21606), we obtained two clones that contained RH2 genes. One of them contained the entire coding region of an RH2 gene (referred to as scabbard 2A). Using the segment between the exons 1 and 5, including four introns, of this gene as the probe, we rescreened the same genomic DNA library and obtained additional six clones. From their restriction maps and DNA sequences, these eight clones were classified into four RH2 genes. All coding regions and four intron regions of the RH2 genes were then sequenced.

Data Analyses

We considered 25 representative RH2 opsins from a wide range of species: lamprey 2 (*Petromyzon marinus*: GenBank accession number AY366494), lampfish 2A (*S. leucopsarus*: GQ414753), lampfish 2B (GQ414754), lampfish 2D (GQ414756), scabbard 2A (*L. fitchi*: GQ414752), scabbard 2B (GQ421593), scabbard 2C (GQ421594), scabbard 2D (GQ421595), goldfish 2-1 (*Carassius auratus*: L11865), goldfish 2-2 (*C. auratus*: L11866), zebrafish 2-1 (*Danio rerio*: AB087805), zebrafish 2-2 (AB087806), zebrafish 2-3 (AB087807), zebrafish 2-4 (AB087808), medaka 2A (*Oryzias latipes*: AB223053), medaka 2B (AB223054), medaka 2C (AB223055), tilapia 2Aα (*Oreochromis niloticus*: DQ235683), tilapia 2Aβ (DQ235682), tilapia 2B (DQ235681), coelacanth 2 (*Latimeria chalumnae*: AF131258), chameleon 2 (*Anolis carolinensis*: AF134189), gecko 2 (*Gekko gekko*: M92035), pigeon 2 (*Columba livia*: AF149322), and chicken 2 (*Gallus gallus*: M92038). In the data analyses, we also considered six RH1 opsins, which is most closely related to the RH2 group (Yokoyama 1997): zebrafish 1 (AF109368), coelacanth 1 (AF131253), frog 1 (*Xenopus laevis*: U23463), chameleon 1 (L31503), chicken 1 (D00702), and bovine 1 (M21606).

To construct a rooted phylogenetic tree of the RH2 opsins, we applied the Neighbor-Joining (NJ) method (Saitou and Nei 1987) to their nucleotide sequences at a total of 330 codons (positions 1–324 and 327–332) and deduced amino acid sequences of the total of 31 RH1 and RH2 genes. The NJ trees were further modified by incorporating more widely accepted species relationships based on other molecular data (Hedges and Kumar 2009) as well as morphological data (Carroll 1988; Nelson 1994), and the composite tree topology was obtained. The divergence times and the evolutionary rates of amino acid substitution were inferred by using a new model (see the section “A Model for the Duplication of RH2 Opsins”). The numbers of nucleotide and site-specific amino acid substitutions at all branches in the composite tree were also inferred by using PAML (Yang 2007).

![Fig. 1. Phylogenetic tree of pigments A, B, X, and Y with two nodes M and N. Divergence times T1 and T2 are known, whereas T needs to be estimated and α, β; γ, and δ indicate the number of amino acid substitutions per site per year.](https://academic.oup.com/mbe/article-abstract/27/1/133/1130246)
To evaluate standard errors of these parameters, let $a, b, c, d,$ and $e$ be the numbers of amino acid substitutions per site between $N$ and $A$, between $N$ and $B$, between $M$ and $N$, between $M$ and $X$, and between $M$ and $Y$, respectively. Then, $a = (d_{AB} - d_{AX} + d_{BX})/2, b = (d_{AB} + d_{AX} - d_{BX})/2, c = -(d_{AB} + d_{AX} + d_{BY} + d_{BX})/4, d = (d_{AX} - d_{AY} + d_{BY} - d_{BX} - 2d_{XY})/4, and e = -(d_{AX} + d_{AY} - d_{AX} + d_{BY} + 2d_{XY})/4. Expressing the parameters in equation (2) in terms of $a, b, c, d,$ and $e$, we can obtain the standard errors of $T, \alpha, \beta, \gamma,$ and $\delta$ (Takezaki et al. 1995; Nei and Kumar 2000).

Previously, using various molecular data, the critical divergence times of different vertebrate species have been obtained: 1) lamprey and other vertebrates—608 million years ago (Ma) (Hedges 2009); 2) fish and tetrapods—455 Ma (Hedges 2009); 3) zebrafish/goldfish and other euteleosts—130 Ma (Wittbrodt et al. 2002); 4) coelacanth and tetrapods—430 Ma (Hedges 2009); 5) reptiles and leosts—130 Ma (Wittbrodt et al. 2002); 6) chameleon and tetrapods—130 Ma (Wittbrodt et al. 2002); 7) birds—275 Ma (Shedlock and Edwards 2009); 8) lampfish, scabbardfish, lamprey, medaka, tilapia, zebrafish, chicken—200 Ma (Hedges and Vidal 2009); and 9) gecko—200 Ma (Hedges and Vidal 2009). These NJ trees consistently show that the RH2 opsins in lampfish, medaka, tilapia, and scabbardfish (scabbard 2A, 2B, 2C, and 2D in figure 3) are most closely related among the four scabbard 2A, 2B, 2C, and 2D genes. Among these, lampfish 2C is a pseudogene, in which codon 175 has a stop codon (TGA) and the GT/AG splicing signal of its intron 2 has been changed to GA/AG. For this gene, we could not locate the exon 1 even after sequencing 853 nucleotides of its intron 1. Compared with this, the lengths of intron 1 of the other three lampfish RH2 genes are only 109–364 nucleotides. The other three lampfish RH2 genes consist of 341–348 codons (fig. 2). Nonfunctional lampfish 2C and functional lampfish 2D diverged ~7.5 Ma (see the Discussion), and they should not be allelic and must be located at different loci. Moreover, the closely related lampfish 2A and lampfish 2B (see the section “Molecular Phylogeny of Vertebrate RH2 Opsins”) consist of different number of amino acids (fig. 2). Hence, it is highly likely that lampfish has also four distinct genes.

**Molecular Phylogeny of Vertebrate RH2 Opsins**

Phylogenetic trees of the 25 RH2 opsins from 12 species (lampfish, scabbardfish, lamprey, medaka, tilapia, zebrafish, goldfish, coelacanth, chameleon, gecko, pigeon, and chicken) have been obtained by applying the NJ method to their nucleotide and deduced amino acid sequences. These NJ trees consistently show that the RH2 opsins in tetrapods are most closely related to those in coelacanth, euteleosts, and lamprey, in that order (fig. 3A), which is also consistent with the species tree based on morphological characters (Nelson 1994).

Figure 3A includes the following topologies as well: 1) scabbard 2A is most distantly related among the four scabbardfish RH2 opsins; 2) zebrafish 2-4 is more closely related to the two goldfish RH2 opsins than to zebrafish 2-3; and 3) chameleon 2 is more closely related to pigeon 2/chicken 2 than to gecko 2. First of all, the introns 1 (6 indels), 2 (5 indels), 3 (4 indels), and 4 (one indel) of the four scabbardfish RH2 genes clearly show that scabbard 2A and 2B (scabbard 2A/2B) and scabbard 2C/2D form two distinct clusters (supplementary fig. S1, Supplementary Material online), despite the highly supported clustering of scabbard 2B, 2C, and 2D in figure 3A. Second, when the corresponding nucleotide sequences are compared, the zebrafish 2-3/2-4 grouping is supported by a bootstrap value of 0.99, and it clusters with goldfish 2-1/2-2 group with a bootstrap value of 0.99. Hence, it is more likely that the zebrafish 2-3/2-4 is more closely related to the goldfish 2-1/2-2 than to the zebrafish 2-1/2-2, as previously suggested by several
Incorporating these three features into the tree topology in figure 3, we obtained the composite tree topology of RH2 opsins (fig. 3B). Applying PAML analysis to this predetermined tree topology, we evaluated the numbers of nucleotide substitutions per site. Compared with those in the amino acid tree (fig. 3A), the branches leading to the RH2 opsins genes in euteleosts and gecko are more pronounced in the nucleotide tree (fig. 3B).

In constructing the rooted phylogenetic trees of the RH2 opsins, we have compared the numbers (d) of amino acid substitutions per site within and between pairs of RH1 and RH2 opsins. The d values for the pairs of RH1 opsins (group 1), for those of the euteleost RH2 opsins (group 2), and for those of the coelacanth and tetrapod RH2 opsins (group 3) are as follows:

<table>
<thead>
<tr>
<th>Group</th>
<th>d Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>RH1</td>
<td></td>
</tr>
<tr>
<td>Euteleosts</td>
<td></td>
</tr>
<tr>
<td>Gecko</td>
<td></td>
</tr>
<tr>
<td>Coelacanth</td>
<td></td>
</tr>
<tr>
<td>Tetrapod</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 2. Alignment of the deduced amino acid sequences of scabbard 2A–2D and lampfish 2A–2D. The original nucleotide sequences of scabbard 2A (GenBank accession number GQ414752), 2B (GQ421593), 2C (GQ4321594), 2D (GQ421595), lampfish 2A (GQ414753), 2B (GQ414754), 2C (GQ414755), and 2D (GQ414756) have been deposited in GenBank. Dots and question marks indicate the identity of the amino acids with those of scabbard 2A and unknown amino acids, respectively, whereas dashes indicate indels.
ranged 0.168–0.321, 0.004–0.427, and 0.085–0.299, respectively, whereas the \( d \) values for the pairs between groups 1 and 2, between groups 1 and 3, and between groups 2 and 3 were 0.288–0.462, 0.277–0.401, and 0.311–0.419, respectively. Hence, some orthologous RH2 opsin pairs have larger distances than the two paralogous opsin pairs, showing that molecular clock does not hold for RH2 opsins, which can also be seen in figures 3A and 3B.

**Sequence X**

To satisfy the assumption of \( \gamma_1 = \gamma_2 \) in our estimation procedure (see Materials and Methods), an appropriate sequence X must be selected. The variable branch lengths in figure 3A indicate that our choice of sequence X is rather limited, and the method is highly data dependent. Hence, we inferred the parameters (\( \alpha, \beta, \gamma, \delta, \) and \( T \)) using various RH2 opsins as sequence X: 1) zebrafish 2-1 for the lampfish/medaka/tetrapod opsins; 2) scabbard 2A for the zebrafish/goldfish opsins; 3) zebrafish 2-4 for coelacanth/tetrapod opsins; 4) chicken 2, chameleon 2, and zebrafish 2-4 for tetrapod opsins; and 5) lamprey 2 for the euteleost/tetrapod opsins (supplementary table S1, Supplementary Material online).

To check the validity of using these specific sequences as sequence X, we can first compare the \( \gamma \) values obtained by considering closely related sequence pairs. Because of their rapid evolution, the comparison of \( \gamma \) values for the pairs of euteleost opsins is informative. The \( \gamma \) values evaluated for

---

**Fig. 3.** Rooted phylogenetic tree for the RH2 opsins. (A) The NJ tree, in which the numbers next to the different nodes indicate clustering percent supports generated by 1000 bootstrap analysis. (B) The maximum likelihood tree, in which the branch lengths were evaluated by nucleotide substitutions using the predetermined composed tree.
the pairs between lampfish 2A/2B/2D, between lampfish 2A/2B/2D and medaka 2A/tilapia 2B, between medaka 2B/2C and tilapia 2A, and between scabbard 2A/2B and scabbard 2C/2D are 1.36–1.38, 1.07–1.11, 1.26–1.30, and 1.26 to 1.29/site/year, respectively. Hence, the values within each group are very similar.

Second, by checking the equality of \( \gamma_1 \) and \( \gamma_2 \) directly. Assuming that the lampfish/medaka/tilapia/scabbardfish and zebrafish/goldfish RH2 opsin groups diverged 136 Ma (Wittbrodt et al. 2002) and using zebrafish 2-1 as sequence X, we considered six pairs of sequences A and B: 1) lampfish 2A and 2B; 2) medaka 2A and tilapia 2B; 3) medaka 2B and 2C; 4) tilapia 2A and 2B; 5) scabbard 2A and 2B; and 6) scabbard 2C and 2D. Then, using the evolutionary rates (or branch lengths) and appropriate \( T \) values given in figure 4, the six \( \gamma_1 \) and one \( \gamma_2 \) values were obtained (table 1). The results show that the \( \gamma_1 \) values vary between 1.28 x 10^{-9} to 1.84 x 10^{-9}/site/year, which differ significantly neither from each other nor from the \( \gamma_2 \) value of 1.59 x 10^{-9}/site/year, again justifying the assumption of \( \gamma_1 = \gamma_2 \).

**Divergence Times and Evolutionary Rates**

Using equation (2), we estimated \( T, \alpha, \) and \( \beta \) for all 48 branches that are specified in figure 3B (supplementary table S1, Supplementary Material online). Before studying these results, two comments may be in order. First, we evaluated the divergence times between euteleost and coelacanth/tetrapod (438 ± 36 Ma), between coelacanth and tetrapod (438 ± 34 Ma), between lampfish/medaka/tilapia/scabbardfish and zebrafish/goldfish (136 ± 15 Ma), between reptile and bird (253 ± 20 Ma), between chameleon and gecko (183 ± 15 Ma), and between pigeon and chicken.
Table 1. The Divergence Time (×10⁹) and Evolutionary Rate (×10⁻⁹) for an Ancestral Branch (M-N).

<table>
<thead>
<tr>
<th>Pigment A</th>
<th>Pigment B</th>
<th>T</th>
<th>γ₁</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lampfish 2A</td>
<td>Lampfish 2B</td>
<td>33</td>
<td>1.83 ± 0.24</td>
</tr>
<tr>
<td>Medaka 2A</td>
<td>Tilapia 2B</td>
<td>61</td>
<td>1.28 ± 0.21</td>
</tr>
<tr>
<td>Medaka 2B</td>
<td>Medaka 2C</td>
<td>17</td>
<td>1.81 ± 0.23</td>
</tr>
<tr>
<td>Tilapia 2A</td>
<td>Tilapia 2Aβ</td>
<td>27</td>
<td>1.84 ± 0.24</td>
</tr>
<tr>
<td>Scabbard 2A</td>
<td>Scabbard 2B</td>
<td>24</td>
<td>1.55 ± 0.22</td>
</tr>
<tr>
<td>Scabbard 2B</td>
<td>Scabbard 2D</td>
<td>28</td>
<td>1.56 ± 0.22</td>
</tr>
</tbody>
</table>

Note.—The γ value for a branch (M-X) is 1.59 ± 0.20 × 10⁻⁹.

(106 ± 11 Ma) RH2 opsins (fig. 4). These estimates are reasonably close to the times of the corresponding speciation events of 455, 430, 130, 275, 200, and 105 Ma (Wittbrodt et al. 2002; Hedges 2009; Hedges and Vidal 2009; Shedlock and Edwards 2009; van Tuinen 2009).

Second, using these and other divergence times obtained (fig. 4), we can also obtain the second set of evolutionary rates by inferring the actual numbers of site-specific amino acid substitutions at all branches using PAML and by dividing each by an appropriate evolutionary time span. The two sets of evolutionary rates are very similar at 38 out of a total of 48 branches (79%), but those at the remaining 10 branches differ significantly (supplementary table S2, Supplementary Material online). Among the latter group, the unreasonably high evolutionary rate (50.5 ± 16.2 × 10⁻⁹) for the common ancestor of coelacanth/tetrapod RH2 opsins (branch 40) seems to have occurred because of the short time span estimated (0.6 × 10⁶). If we adjust the split between euteleosts and tetrapods and that between coelacanth and tetrapods to 455 and 430 Ma, respectively (Hedges 2009), then the evolutionary rate becomes 1.21 ± 0.39 × 10⁻⁹/site/year and does not differ significantly from the estimate inferred from equation (2) (0.67 × 10⁻⁹/site/year, fig. 4).

The results in figure 4 show that the evolutionary rate of the RH2 opsin in the common ancestor of euteleosts and tetrapods was 0.26 × 10⁻⁹/site/year. This value is virtually identical to the evolutionary rates of lamprey 2, RH2 opsins in the euteleost ancestor, and coelacanth 2. Moreover, these rates do not differ significantly from that of the RH2 opsin in the common ancestor of coelalanth and tetrapods (0.67 × 10⁻⁹/site/year).

The RH2 opsin in the lampfish/medaka/tilapia/scabbardfish ancestor (2.8 × 10⁻⁹/site/year) evolved with a significantly faster rate than the zebrafish/goldfish ancestor (0.7 × 10⁻⁹/site/year) (P < 0.01). During the last 90–100 million years (My), the lineages leading to lampfish 2D and medaka 2B/2C have maintained high average rates of evolutionary changes at ~1 × 10⁻⁹ to 3 × 10⁻⁹/site/year, explaining their long branch lengths in figure 3A. On the other hand, during the last 30 My, the scabbardfish RH2 opsins, including nonfunctional scabbard 2D, reduced their evolutionary rates significantly (~1 × 10⁻⁹/site/year).

For 80–90 My after the separation from the lampfish/medaka/tilapia/scabbardfish opsin group, the zebrafish/goldfish RH2 opsins evolved with the rates of 0.7 × 10⁻⁹ to 1.1 × 10⁻⁹/site/year, causing the relatively shorter branch lengths of the zebrafish and goldfish RH2 opsins (fig. 3A).

Among the zebrafish and goldfish opsins, zebrafish 2-1 has the longest branch length (fig. 3A), which is explained by the accelerated mutant substitution during the last ~45 My.

The evolutionary rate of the RH2 opsin in the common ancestor of coelacanth and tetrapods is estimated to be 0.67 × 10⁻⁹/site/year, which do not differ significantly from that of coelacanth 2 (0.31 × 10⁻⁹/site/year). During the last ~180 My, gecko 2 has evolved with a much faster rate than chameleon 2, making an unusually long branch length among the tetrapod RH2 opsins as suggested by figure 3B. Even then, its evolutionary rate is much lower than those of many RH2 opsins in euteleosts. Hence, with the exception of gecko 2, the evolutionary rates of tetrapod RH2 opsins are ~0.2 × 10⁻⁹ to 0.5 × 10⁻⁹/site/year (fig. 4).

Discussion

We have seen that the evolutionary rates of amino acid substitution in the RH2 opsins in vertebrate ancestors were ~0.25 × 10⁻⁹/site/year, which is much lower than those of such well-known proteins as hemoglobin α and β chains, ~1.2 × 10⁻⁹/site/year (Nei 1987). The RH2 opsins in the euteleost ancestor evolved equally slowly, but during the 136 My of euteleost evolution, the evolutionary rates of RH2 opsins have been accelerated roughly by 5–6 folds, whereas those of the tetrapod RH2 opsins have increased only by at most 2-folds.

One curious aspect of RH2 opsin evolution is that some orthologous RH2 opsin pairs can be more distantly related than between paralogous RH1 and RH2 opsin pairs, showing that RH1 opsins evolved with much slower rates than some RH2 opsins. To evaluate the evolutionary rates for the RH1 opsins, we considered five representative RH1 opsins (zebrafish1, coelacanth 1, chameleon 1, chicken 1, and bovine 1) and used lamprey 2 as the outgroup. The evolutionary rates of the RH1 opsins obtained vary between 0.1 × 10⁻⁹ to 1.0 × 10⁻⁹/site/year (fig. 5). As suspected (Yokoyama 2000a), chameleon 1 has evolved more quickly than others, but its evolutionary rate (0.65 × 10⁻⁹/site/year) is similar to that of gecko 2 and is still significantly lower than those of many RH2 opsins in euteleosts. Figure 5 also shows that the RH1 opsin of the tetrapod ancestor had a relatively high evolutionary rate of 0.97 × 10⁻⁹/site/year. However, because this rate is associated with a large standard error, it seems reasonable to consider that the RH1 opsins have evolved with the rates of ~0.1 × 10⁻⁹ to 0.4 × 10⁻⁹/site/year, which are similar to those of RH2 opsins in the early vertebrate ancestors and tetrapods.

Because of its incomplete sequence, lampfish 2C has been excluded from the analyses. To study the evolutionary processes of this pseudogene, we also compared the 209 deduced amino acids at positions 122–324 and 327–332 of lampfish 2C, 25 RH2, and 6 RH1 opsins (see Materials and Methods). In this region, no amino acid difference can be found between lampfish 2A and lampfish 2B.
The NJ tree of the 32 sequences shows that lampfish 2A/2B and lampfish 2C/2D both cluster with bootstrap values of 1.0. When we compare the deduced amino acid sequences of lampfish 2C (sequence A), lampfish 2D (sequence B), zebrafish 2-1 (sequence X), and chameleon 1 (sequence Y), T = 7.6 ± 2.2 × 10^9, α = 7.33 ± 2.23 × 10^9/site/year, and β = 2.97 ± 1.41 × 10^9/site/year. Hence, the evolutionary rate of lampfish 2C is significantly higher than that of scabbard D (0.39 ± 10^9/site/year; P < 0.01).

Using zebrafish 2-1 (sequence X) and chameleon 1 (sequence Y), we can also show that lampfish 2A and lampfish 2C have evolved with the rates of 0.81 × 10^9 and 1.48 × 10^9/site/year, respectively, and lampfish 2A and lampfish 2D with the rates of 0.92 × 10^9 and 1.03 × 10^9/site/year, respectively. From these estimates, the common ancestor of the four lampfish RH2 opsins is traced back to 56 ± 6.9 Ma, and lampfish 2A (or lampfish 2B) and the lampfish 2C/2D ancestor have evolved at the rates of 0.87 ± 0.27 × 10^9 and 0.67 ± 0.26 × 10^9/site/year, respectively. The weighted average between the evolutionary rates of the lampfish 2C/2D ancestor and lampfish 2D is given by 0.98 ± 0.30 × 10^9/site/year and is very similar to the corresponding rate, 1.25 ± 0.26 × 10^9/site/year, in figure 4.

The variable evolutionary rates of mutation accumulation in the two pseudogenes (lampfish 2C and scabbard 2D) are comparable to the dramatically different evolutionary rates of the RH2 pseudogenes in fugu and Tetradon pufferfish (Neasey and Hartl 2005). Moreover, the SWS1 pseudogenes in two coelacanth species have evolved with an order of magnitude lower than those of hemoglobin α and β chains (Yokoyama and Tada 2000). Hence, scabbard 2D also shows that pseudogenes do not always evolve with a faster rate than the functional genes (see also fig. 3B), and their evolutionary rates can be much lower than a typical evolutionary rate of pseudogenes, −5 × 10^9/site/year (Li et al. 1985).

If we assume that lampfish and scabbardfish have four duplicate RH2 genes, we can identify 10 independent gene duplication events in figure 4. Curiously, these gene duplications can be found only in euteleosts. The significance of this observation is not immediately clear, but it may be related to the functional expansion of duplicated RH2 opsins in diverse water environments. For example, zebrafish 2-1, 2-2, 2-3, and 2-4 pigments have λ_maxs of 467, 478, 488, and 505 nm, respectively (Chinen et al. 2003), whereas the engineered ancestral pigments of zebrafish 2-1/2-2/2-3/2-4, of zebrafish 2-1/2-2, and of zebrafish 2-3/2-4 have had λ_maxs of 506, 474, and 506 nm, respectively (Chinen et al. 2005). These results suggest that zebrafish 2-4 has retained the ancestral phenotype (λ_max ≈ 505 nm), and the others have decreased their λ_max. Although they are not statistically significant, the evolutionary rates of zebrafish 2-1, 2-2, and 2-3 tend to be higher than that of zebrafish 2-4 (fig. 4). As noted in the beginning, the blue-shifted λ_max of the rapidly evolving gecko 2 is likely to be adaptive. These limited data suggest that the accelerated rate of sequence evolution and the gain of new function are positively correlated. The duplication of RH2 opsins seems to be playing a similar role.

To understand the biological significance of gene duplications, we would like to know 1) how gene duplications occur and 2) how these duplicate genes affect the fitness of organisms. The four scabbard RH2 genes may provide a rare opportunity in exploring the first question. As supplementary figure S1 (Supplementary Material online) suggests, the sequences of the four scabbardfish RH2 genes can be aligned easily, and the positions of indels and nucleotide substitutions can be relatively easily identified throughout the coding and noncoding segments. By obtaining the entire DNA sequence information in the intergenic regions of the RH2 genes, we may be able to reconstruct the detailed evolutionary history of the gene duplication events. The analyses may also shed light on the structural characteristics of why the RH2 genes have gained multiple copies so many times during euteleost evolution.

Although it is almost impossible to evaluate the fitness differences caused by various duplicated visual pigments (Lewontin 1979), we can still explore the mechanisms of molecular adaptation by studying the λ_max of RH2 pigments living in different water environments (Yokoyama et al. 2008). This is because not only are visual pigments associated strongly with organisms’ ecological environments (Lythgoe 1979; Yokoyama S and Yokoyama R 1996; Yokoyama 1997) but also the light (or color)
environments in the water can be reasonably well quantified (Jerlov 1976). Indeed, comparisons of various RH2 pigments in different fish species are expected to provide an excellent opportunity to study the adaptive significance of gene duplications.

**Supplementary Material**

Supplementary figure S1 and tables S1 and S2 are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournal.org/).

**Acknowledgments**

This work was supported by the National Institutes of Health (ROIY016400) and Emory University.

**References**


