

Origin and adaptation of green-sensitive (RH2) pigments in vertebrates

Shozo Yokoyama^{1,2}  and Huiyong Jia¹

¹ Department of Biology, Emory University, Atlanta, GA, USA

² Willamette View, Portland, OR, USA

Keywords

adaptive evolution; ancestral pigment reconstruction; gene duplication; RH2 opsin; spectral tuning

Correspondence

S. Yokoyama, Willamette View, 12705 SE River Rd, Portland, OR 97222, USA

Tel: +1 503-652-6757

E-mail: syokoya@emory.edu

(Received 12 January 2020, revised 20 February 2020, accepted 16 March 2020)

doi:10.1002/2211-5463.12843

One of the critical times for the survival of animals is twilight where the most abundant visible lights are between 400 and 550 nanometres (nm). Green-sensitive RH2 pigments help nonmammalian vertebrate species to better discriminate wavelengths in this blue-green region. Here, evaluation of the wavelengths of maximal absorption (λ_{\max} s) of genetically engineered RH2 pigments representing 13 critical stages of vertebrate evolution revealed that the RH2 pigment of the most recent common ancestor of vertebrates had a λ_{\max} of 503 nm, while the 12 ancestral pigments exhibited an expanded range in λ_{\max} s between 474 and 524 nm, and present-day RH2 pigments have further expanded the range to ~ 450–530 nm. During vertebrate evolution, eight out of the 16 significant λ_{\max} shifts (or $|\Delta\lambda_{\max}| \geq 10$ nm) of RH2 pigments identified were fully explained by the repeated mutations E122Q (twice), Q122E (thrice) and M207L (twice), and A292S (once). Our data indicated that the highly variable λ_{\max} s of teleost RH2 pigments arose from gene duplications followed by accelerated amino acid substitution.

Vision is initiated by the absorption of lights by visual pigments, each of which consists of an opsin, encoded by an opsin gene, and the retinal (11-*cis*-retinal or 11-*cis*-3, 4-dehydroretinal) derived from diet [1–4]. Prior to the appearance of the vertebrate ancestor, five groups of visual pigments already existed: rhodopsin (or RH1), rhodopsin-like (or RH2), shortwave length-sensitive type 1 (SWS1), SWS type 2 (SWS2) and middle- and long-wavelength-sensitive (M/LWS) pigments that absorb light maximally (λ_{\max}) at 480–510, 450–530, 360–440, 400–450 and 510–570 nm, respectively [5–7].

One of the critical times for the survival of organisms is twilight [8] where prevalent wavelengths are between 400 and 550 nanometres (nm) [9,10]. As lights pass through to deeper depth in the clear oceans and

lakes, the amount of light is reduced and short and long wavelengths are absorbed by water and the remaining lights also become limited to about 480 nm [11]. Visual acuity of many species in these dim-light environments has been improved by gene duplications within RH1 [12], RH2 [13–17] and SWS2 pigment lineages [14–15,18–26].

The molecular bases of the λ_{\max} shift (or spectral tuning) in visual pigments have been studied mostly by introducing various mutations into various present-day pigments. This approach is based on two implicit assumptions: (a) identical amino acid changes in different pigments shift the λ_{\max} by the same magnitude and (b) forward and reverse mutations shift the λ_{\max} in opposite directions by the same magnitude. These assumptions often fail because of the interactions

Abbreviations

λ_{\max} , the wavelength of maximal absorption; C terminus, carboxyl-terminus; JTT, Jones–Taylor–Thornton; M/LWS, middle- and long-wavelength-sensitive; N terminus, amino-terminus; PAML, phylogenetic analyses by maximum likelihood; RH1, rhodopsin; RH2, rhodopsin-like; SWS1, shortwave length-sensitive type 1; SWS2, SWS type 2; WAG, Whelan and Goldman.

among different amino acids, and conclusions derived from these mutagenesis experiments can be erroneous [27–30]. To elucidate the correct mechanism of spectral tuning, it is imperative to genetically engineer and manipulate ancestral pigments by following the actual evolutionary processes in forward directions [27–30]. This was done for RH1 [30], SWS1 [31], and M/LWS [32] pigments, but the similar molecular analyses have not been applied to the RH2 and SWS2 pigments.

Here, we inferred and genetically engineered the ancestral RH2 pigments at 13 critical stages of vertebrate evolution and determined their λ_{\max} s. Then, by introducing additional mutations, the molecular mechanisms of λ_{\max} shifts at eight evolutionary steps of RH2 pigments have been established.

Materials and methods

Inference of ancestral sequences

The amino acid sequences of ancestral RH2 pigments were inferred by applying the phylogenetic analyses by maximum likelihood (PAML) with Jones–Taylor–Thornton (JTT), Whelan and Goldman (WAG), and Dayhoff amino acid substitution models [33] to the amino acid sequences of the 37 vertebrate RH2 pigments and those of the more restricted 24 pigments with known λ_{\max} s (Fig. 1), together with the RH1 pigment of bovine (*Bos taurus*) and the RH1, SWS1 and SWS2 pigments of lamprey (*Geotria australis*) as the outgroup (Table S1). Among the four paralogous groups of pigments, RH2 pigments are most closely related to RH1, SWS2 and SWS1 pigments, in that order [6,34]. The divergence times of nonduplicated RH2 pigments were estimated from the TimeTree of Life Web server (www.timetree.org). The actual branch lengths of the composite phylogenetic tree were also determined by applying PAML to the user tree based on the amino acid sequences.

To infer the ancestral RH2 pigments, the amino acids between sites 31 and 311 were considered (Fig. S1A). In engineering ancestral pigments, we used an expression vector pMT5 that contained the amino (N) and carboxyl (C) termini (amino acids between positions 1–30 and 312–354, respectively) of the chameleon (P495; red letters in Fig. S1B) with the proper internal segment of a RH2 pigment. The N and C termini of chameleon (P495) do not affect the λ_{\max} of a RH2 pigment significantly. For example, zebrafish4 (P505) with its own and the chameleon (P495) termini with the proper internal segment of a RH2 pigment have λ_{\max} s of 505 nm [16] and 507 nm, respectively, and gecko (P467) with its own and the chameleon termini have λ_{\max} s of 467 nm [35] and 466 nm, respectively. Similarly, different amino acids at the N and C termini do not modify the λ_{\max} s of SWS1 pigments [31].

Thirteen ancestral pigments, inferred from JTT model based on the 24 sequence data with more evenly distributed λ_{\max} s, were engineered: those of all vertebrates (AncAgnatha, node 1), jawed fishes (AncJawedFish, node 2), Teleostei (AncTeleost, node 3), Clupeocephala (AncClupeo, node 4), Cypriniformes (AncCyprini1–2, nodes 5 and 6), Euteleosts (AncEuteleost1–4, nodes 7–10), Tetrapods (AncTetrapod, node 11), Sauropsids (AncSauropsid, node 12) and Squamates (AncSquamata, node 13; Fig. 1). In this process, we engineered AncCyprini2 and AncSquamata first by introducing six and 12 mutations into zebrafish4 (P505) and chameleon (P495), respectively, and the remaining 11 ancestral pigments were engineered by sequentially introducing additional 128 mutations into various ancestral pigments.

All mutant opsins were generated by using QuikChange Site-Directed Mutagenesis Kits (Stratagene, La Jolla, CA, USA). To rule out spurious mutations, the DNA fragments were sequenced by cycle sequencing reactions using the Sequitherm Excel II long-read kits (Epicentre Technologies, Madison, WI, USA) with dye-labelled M13 forward and reverse primers. Reactions were run on a LI-COR (Lincoln, NE, USA) 4300LD automated DNA sequencer.

The *in vitro* assay

Ancestral and other mutant opsins were expressed in COS1 cells by transient transfection [36]. The contiguous RH2 opsins between sites 31 and 311 were cloned into the *EcoRI* and *SalI* restriction sites of the expression vector pMT5, which contained the N terminus (amino acids between positions 1 and 30) and C terminus (amino acids between positions 312 and 354) of the chameleon (P495; red letters in Fig. S1B). The visual pigments were regenerated by incubating the opsins with 11-*cis*-retinal (provided by R. K. Crouch at Storm Eye Institute, Medical University of South Carolina and the National Eye Institutes) and were purified using immobilized 1D4 (The Culture Center, Minneapolis, MN, USA) in buffer W1 (50 mM *N*-(2-hydroxyethyl) piperazine-*N'*-2-ethanesulfonic acid (HEPES; pH 6.6), 140 mM NaCl, 3 mM MgCl₂, 20% (w/v) glycerol and 0.1% dodecyl maltoside). UV-visible spectra were recorded at 20 °C using a Hitachi U-3000 dual beam spectrophotometer (LI-COR Biosciences, Lincoln, NE, USA). Visual pigments were bleached for 3 min using a 60 W standard light bulb equipped with a Kodak Wratten #3 filter at a distance of 20 cm. Data were analysed using SIGMAPLOT software (Jandel Scientific, San Rafael, CA, USA).

Ethics

Research was carried out under approval of Emory University according to the university's animal ethics guidelines.

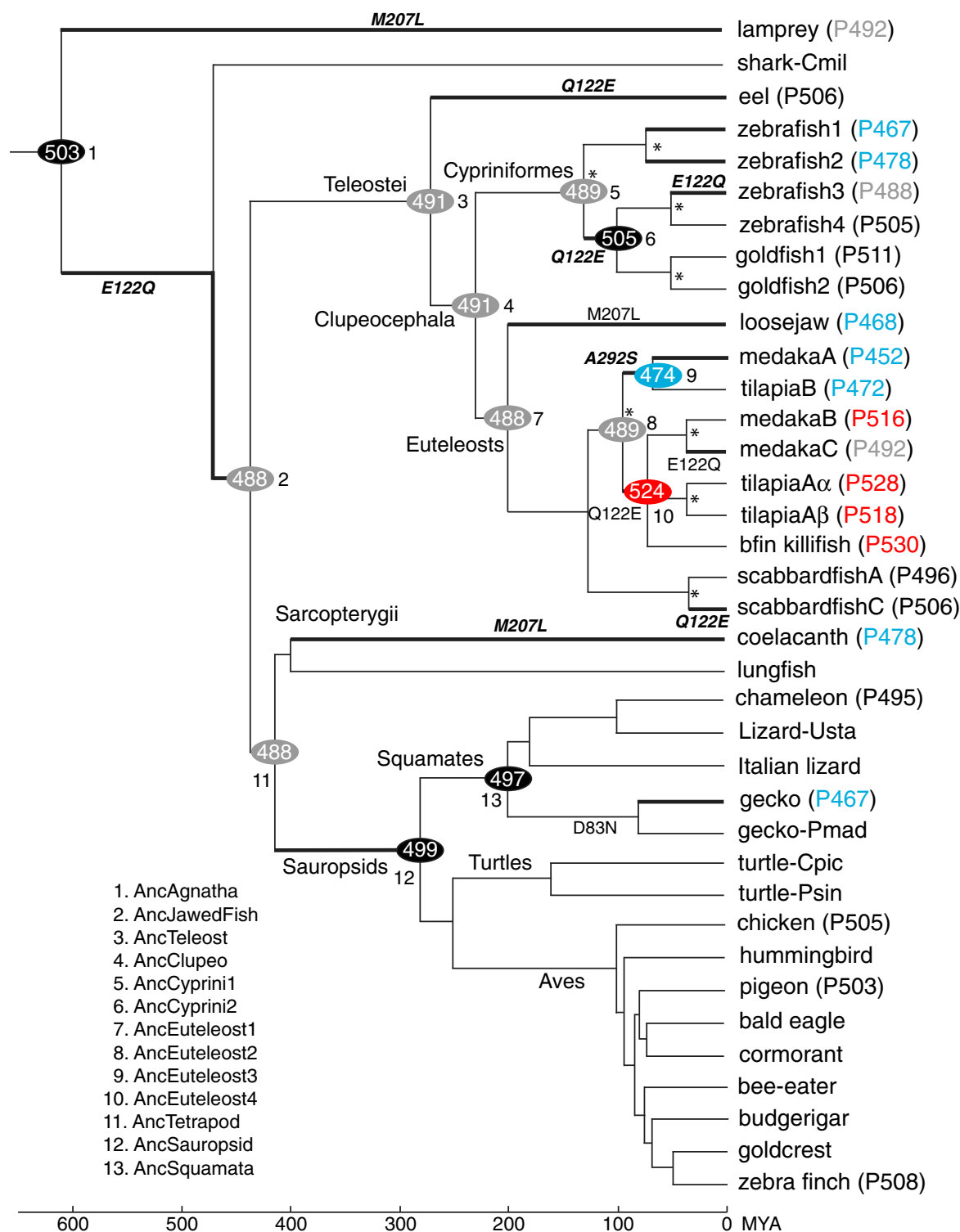


Fig. 1. A composite tree topology of 37 representative RH2 pigments in vertebrates. The numbers in ovals and after P in parentheses show λ_{\max} s of the ancestral and present-day pigments, respectively. Blue, grey, black and red indicate the λ_{\max} s of 452–478, 488–492, 495–511 and 516–530 nm, respectively. The critical amino acid substitutions are shown along branches. Asterisks (*) indicate the gene duplication events. Branches with significant λ_{\max} shift ($|\lambda_{\max}| \geq 10$ nm) are indicated by thick lines, and the amino acid changes that fully explain the λ_{\max} shift are indicated by bold italics. The divergence times at the bottom are taken from the TimeTree of Life Web server (www.timetree.org).

Results

The amino acid sequences of ancestral pigments

The ancestral amino acid sequences inferred are highly consistent. For example, when the earliest vertebrate ancestor AncAgnatha (node 1, Fig. 1) is considered, the amino acid sequences inferred by JTT and WAG models show that 241 (86%) out of a total of the 281 comparable sites are identical with Bayesian posterior probabilities (PPs) ≥ 0.95 , 22 out of the remaining 40 sites also have identical amino acids with a $0.70 \leq \text{PP} < 0.95$, and different amino acids are predicted only at eight highly variable sites. The ancestral sequences inferred using the 24 and 37 sequences are also very similar. For AncAgnatha, JTT model shows that amino acids at 263 out of 281 sites (94%) are identical and only those at the other 18 sites differ (indicated by * in Table S2). At these 18 sites, the present-day pigments have variable amino acids, but their λ_{max} s are similar, which implies that these amino acids are not critical in determining the λ_{max} s of visual pigments [34,37].

At potentially critical sites 122, 207 and 292 [5], the 24 and 37 data sets predict identical amino acids throughout vertebrate evolution (Table S3). Specifically, E122Q, Q122E and M207L occurred three, four and two times, respectively, while A292S occurred once during the vertebrate evolution (Fig. 1).

One of these results disagrees with the published result. Previously, using PAML with JTT model, it was suggested that AncCyprin1 (node 5, Fig. 1; or Ancestor1 in Fig. S2A) had E122 with a PP of 0.77 [38], but our result using the identical method shows that AncCyprin1 (Fig. S2B) had Q122 with PPs of 0.89 and 0.97 using the 24 and 37 sequences, respectively (Table S3). Comparing the two data sets, we can find the cause of the discrepancy. That is, Chinen *et al.* used the teleost data consisting of 10 pigments with E122 and four pigments with Q122, which were heavily biased towards E122. On the other hand, our teleost data consist of eight pigments with E122 and nine pigments with Q122 (Fig. S2A,B, respectively). Because of the less-biased data set and significantly higher PPs, it is most likely that AncCyprin1 had Q122.

Functional differentiation

We applied the *in vitro* assay to the 13 ancestral pigments engineered. The results show that (a) the amino acid sequence of AncJawedFish differs from that of AncTetrapod at multiple sites (Fig. S1B), but these

two pigments have identical λ_{max} values at 488 nm, and (b) the ratio of absorbance at ~ 280 nm to that at ~ 500 nm of the *in vitro* assay ranged from 2.5 (AncEuteleost) to 4.8 (AncSquamata; Fig. S3). In particular, AncAgnatha had a λ_{max} of 503 nm and its closest descendant, AncJawedFish (node 2, Fig. 1), decreased its λ_{max} to 488 nm, which had been maintained by six out of the remaining 11 ancestral pigments (nodes 3–5, 7, 8 and 11; grey ovals, Fig. 1; Fig. S3). However, the five others expanded their λ_{max} s from 474 nm (node 9) to 524 nm (node 10) and the λ_{max} s of present-day pigments have been expanded further between ~ 450 and 530 nm.

We can find a total of 16 significant λ_{max} shifts ($\Delta\lambda_{\text{max}} \geq 10$ nm), and three branches stand out with much larger λ_{max} shifts ($|\Delta\lambda_{\text{max}}| \geq 30$ nm): (a) AncEuteleost2 (P489) (node 8) \rightarrow AncEuteleost4 (P524) (node 10); (b) AncEuteleost4 (P524) (node 10) \rightarrow medakaC (P492); and (c) AncSquamata (P497) (node 13) \rightarrow gecko (P467) (represented by thick branches; Fig. 1).

The λ_{max} shifts of RH2 pigments reveal three characteristics (Fig. 1). First, despite their similarities, the λ_{max} s of zebrafish3 (P488) and medakaC (P492) did not evolve directly from that of AncJawedFish (P488), but these pigments reversed their λ_{max} s to over 500 nm before reaching about 490 nm. These changes can be found only by reconstructing ancestral pigments at intermediate evolutionary steps. Second, with the exceptions of eel (P506), loosejaw (P468), coelacanth (P478) and gecko (P467), significant λ_{max} shifts have been generated by gene duplications followed by critical amino acid substitutions. Third, the decreased λ_{max} s of loosejaw (P468) and coelacanth (P478) are closely connected with their unique adaptations to their highly species-specific light environments ([39–41]) and that of gecko (P467) indicates the adaptation to its nocturnal environment.

Our mutagenesis results show that E122Q, Q122E, M207L and A292S shifted the λ_{max} s of ancestral pigments by 7–20 nm (Table 1). These mutations explain eight evolutionary transitions: (a, b) AncAgnatha (P503) (node 1) \rightarrow AncJawedFish (P488) (node 2) and AncCyprin2 (P505) (node 6) \rightarrow zebrafish3 (P488) by E122Q; (c–e) AncTeleost (P491) (node 3) \rightarrow eel (P506), AncCyprin1 (P489) (node 5) \rightarrow AncCyprin2 (P505) (node 6), and AncEuteleost1 (P488) (node 7) \rightarrow scabbardfishC (P506) by Q122E; (f, g) AncAgnatha (P503) (node 1) \rightarrow lamprey (P492) and AncTetrapod (P488) (node 11) \rightarrow coelacanth (P478) by M207L; and (h) AncEuteleost2 (P489) (node 8) \rightarrow AncEuteleost3 (P478) (node 9) by A292S. However, the λ_{max} shifts of AncEuteleost1 (P488) (node

7) → Loosejaw (P468) and AncEuteleost4 (P524) (node 10) → medakaC (P492) cannot be explained by M207L and E122Q alone, respectively. D83N can also shift the λ_{\max} s of other visual pigments by ~3–5 nm [5], but it does not shift the λ_{\max} of AncSquamata (P497) (node 13; Table 1).

We also searched for additional critical mutations along two branches: AncEuteleost2 (P489) (node 8) → AncEuteleost4 (P524) (node 10; $\Delta\lambda_{\max} = 35$ nm) and AncTetrapod (P488) (node 11) → AncSauropsid (P499) (node 12; $\Delta\lambda_{\max} = 11$ nm; Fig. 1). For the first branch, the present-day pigments descended from

Table 1. Effects of various mutations on the λ_{\max} shift. The numbers after P show λ_{\max} values. When the λ_{\max} values of the mutant and its descendant pigments are similar ($|\Delta\lambda_{\max}| \leq 4$ nm), the results are shown by bold letters.

Ancestral pigment {node}	Mutation	λ_{\max} ($\Delta\lambda_{\max}$) (nm)	Descendant {node}
AncAgnatha (P503) {1}	E122Q	486 (−17)	AncJawedFish (P488) {2}
AncCyprini2 (P505) {6}	E122Q	489 (−16)	zebrafish3 (P488)
AncTeleost (P491) {3}	Q122E	508 (+17)	eel (P506)
AncCyprini1 (P489) {5}	Q122E	504 (+15)	AncCyprini2 (P505) {6}
AncEuteleost1 (P488) {7}	Q122E	503 (+15)	scabbardfishC (P506)
AncAgnatha (P503) {1}	M207L	493 (−10)	lamprey (P492)
AncTetrapod (P488) {11}	M207L	481 (−7)	coelacanth (P478)
AncEuteleost2 (P489) {8}	A292S	478 (−11)	AncEuteleost3 (P474) {9}
AncEuteleost1 (P488) {7}	M207L	480 (−8)	Loosejaw (P468)
AncEuteleost4 (P524) {10}	E122Q	504 (−20)	medakaC (P492)
AncSquamata (P497) {13}	D83N	496 (−1)	gecko (P467)
AncEuteleost2 (P489) {8}	Q122E	504 (+15)	AncEuteleost4 (P524) {10}
AncEuteleost2 (P489) {8}	Y96T/Q122E/C213F	509 (+20)	AncEuteleost4 (P524) {10}
AncEuteleost2 (P489) {8}	V60F/F74Y/Y96T/Q122E/T209V/C213F/I255V/L259M/A273G	510 (+21)	AncEuteleost4 (P524) {10}
AncTetrapod (P488) {11}	M44I	491 (+3)	AncSauropsid (P499) {12}
AncTetrapod (P488) {11}	L40V/A42C/M44I/I50T/V87A/A166S/I205L/L214I/L309M	491 (+3)	AncSauropsid (P499) {12}

AncEuteleost2 (P489) can be distinguished into two groups: (a) medakaB (P516), tilapiaA α (P528), tilapiaA β (P518) and bfin killifish (P530) with $\lambda_{\max} \geq 516$ nm; and (b) medakaA (P452), tilapiaB (P472) and medakaC (P492) with $\lambda_{\max} \leq 492$ nm (Fig. 1). Since nine changes (V60F, F74Y, Y96T, Q122E, T209V, C213F, I255V, L259M and A273G) are shared only by the pigments in the first group, we hypothesized that some of these mutations were responsible in increasing the λ_{\max} of AncEuteleost4 (P524). But, Q122E, Y96T/Q122E/C213F and V60F/F74Y/Y96T/Q122E/T209V/C213F/I255V/L259M/A273G explain only 43% (15 nm), 57% (20 nm) and 60% (21 nm) of the observed λ_{\max} shift (35 nm), respectively (Table 1). Hence, the interactions between E122Q and newly found critical changes Y96T and C213F do not improve the result significantly.

For the second branch, nine changes (L40V, A42C, M44I, I50T, V87A, A166S, I205L, L214I and L309M) are found in pigeon (P503), chicken (P505) and zebra finch (P508) with $\lambda_{\max} \geq 503$ nm, but are missing from coelacanth (P478) and gecko (P467) with $\lambda_{\max} \leq 478$ nm (Fig. 1). We found a new critical mutation (M44I), which increased the λ_{\max} by 3 nm, but neither M44I nor all nine changes (L40V, A42C, M44I, I50T, V87A, A166S, I205L, L214I and L309M) explain the λ_{\max} shift attained by AncSauropsid (P499) (Table 1).

These two examples reveal the complex nature of the spectral tunings in RH2 pigments. Chinen *et al.* [38] encountered the same problem in explaining the significant λ_{\max} shift from Ancestor1 (P506) to Ancestor2 (P474; supplementary information, Fig. S2A; for more details, see Discussion section).

Evolutionary rates of amino acid substitution

To evaluate the effects of gene duplication on the induction of the highly variable λ_{\max} s in teleost RH2 pigments, we considered representative 22 sequences and evaluated the numbers of amino acid substitutions per site per year at three lineages: (a) the vertebrate pigment lineage (lineage a), which excludes all Clupeocephala pigments (lineage c); (b) the Tetrapod pigment lineage (lineage b); and (c) the Clupeocephala pigment lineage where gene duplication events are prevalent (shown by a rectangle, lineage c; Fig. 2a). The branch lengths from these nodes to their descendant pigments were determined by applying PAML to the composite evolutionary tree of the amino acid sequences. The evolutionary rates were evaluated by taking the averages between bifurcated (or trifurcated) branches sequentially and assuming that lineages a, b and c

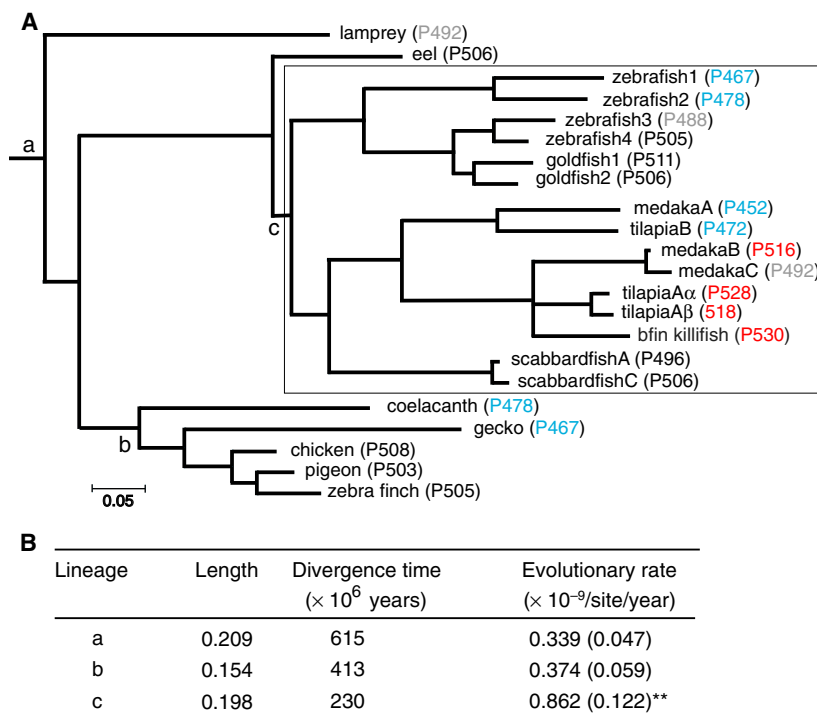


Fig. 2. (A) The numbers of amino acid substitutions per site for a composite tree of 22 RH2 pigments were determined by applying PAML [33] to the amino acids between sites 31 and 311. Blue, grey, black and red indicate the λ_{\max} s of 452–478, 488–492, 495–511 nm and 516–530 nm, respectively. (B) The evolutionary rates of amino acid substitution along the branches including gene duplications (c lineage) and without gene duplications (a and b lineages). ** $P < 0.01$ for the comparison of the branches between c-present and a-present and between those between c-present and b-present. The numbers in the parentheses indicate SE.

originated 615, 230 and 413 million years ago (MYA), respectively (www.timetree.org). The results show that the evolutionary rate for lineage c ($0.9 \pm 0.12 \times 10^{-9}$) is significantly higher than that for lineage a ($0.3 \pm 0.05 \times 10^{-9}$; $Z = 4.0$) and lineage b ($0.4 \pm 0.06 \times 10^{-9}$, $Z = 3.6$) at 1% level (Fig. 2b). Hence, the duplications of RH2 opsin genes were followed by the significantly accelerated amino acid substitutions, which led Clupeocephala RH2 pigments to expand the range of their λ_{\max} s.

Discussion

Depending on whether AncCyprini1 (node 5, Fig. 1) had Q122 or E122, the evolutionary processes of zebrafish RH2 pigments can be interpreted very differently. If AncCyprini1 had E122 and a λ_{\max} of 506 nm, the λ_{\max} of zebrafish3 (P488) decreased by E122Q and Ancestor2 (P474) by E122Q and additional unidentified mutations [38] (Fig. S2A). On the other hand, if AncCyprini1 had Q122 and a λ_{\max} of 489 nm and if the λ_{\max} of Ancestor2 (P474) still holds, the λ_{\max} shifts of AncCyprini2 (P505) and zebrafish3 (P488) are explained fully by Q122E and E122Q, respectively, but the critical mutations that caused the λ_{\max} shift of Ancestor2 (P474) remain to be discovered (Fig. S2B).

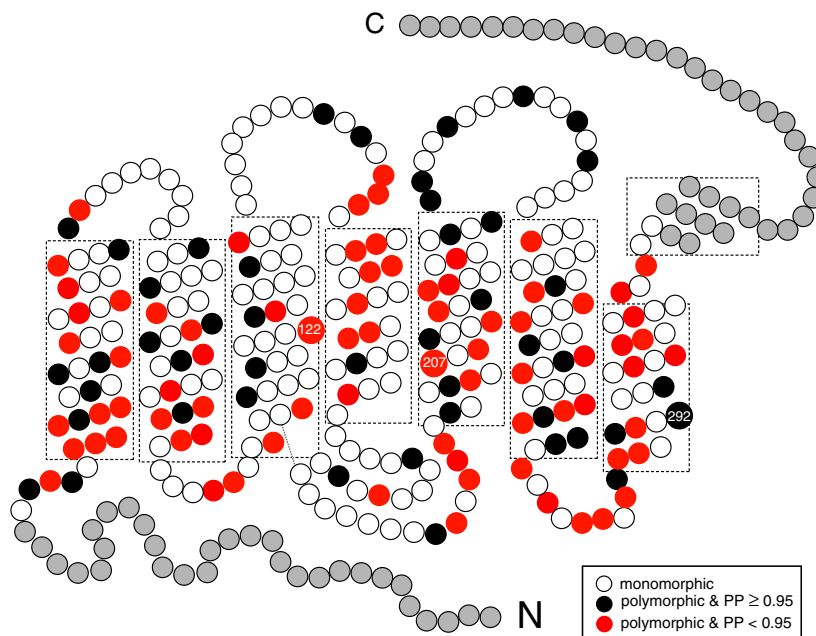
This example shows that the mechanism of phenotypic adaptation can be understood by inferring the

ancestral sequences correctly. For that, we need to use unbiased sequence data which, among other things, should consist of roughly equal numbers of molecules with different functions. Furthermore, to identify mutations that generated highly variable λ_{\max} s of ancestral and present-day RH2 pigments, it would be necessary to construct various sets of chimeric pigments between a pair of pigments with different λ_{\max} s and then perform extensive mutagenesis experiments [42–45].

The functional differentiation of euteleost RH2 pigments has been accelerated significantly by gene duplications. Evolution by gene duplication may be classified into four categories [46]: (a) gene duplication itself does not cause any selection, (b) the duplication itself renders selective advantage, (c) duplication occurs in a gene for which genetic variation already existed, and (d) duplications occur by whole-genome duplication or large segmental duplication. The first category includes Ohno's neofunctionalization model [47] and various subfunctionalization models [48–50].

Among these possibilities, neither selection caused by gene duplication itself (category 2) nor functionally meaningful genetic variation (category 3) have been established in the preduplication phase of RH2 pigments. During fish evolution, Euteleost and Cyprini-form ancestors appeared roughly 240 and 100 MYA, respectively [51]. Hence, RH2 gene duplications in

Fig. 3. The neutral and adaptive mutations in ancestral pigments. Polymorphic sites with $PP \geq 0.95$ and $PP \geq 0.70$ are indicated by black and red circles, respectively, while white circles show monomorphic sites. Grey circles in the N and C termini indicate those of the anole RH2 pigment (see Materials and Methods). Amino acid changes at sites 122, 207 and 292 cause significant λ_{\max} shifts. The 2-D model is after Palczewski [60].



fishes occurred much later than the fish-specific whole-genome duplication (WGD), which occurred about 350 MYA [52], showing that neither WGD nor large segmental duplication (category 4) [16] seem to be involved. The differential ontogenetic RH2 gene expression was described as ‘subfunctionalization’ [15], but actual dual functions of the ancestral RH2 gene have not been established. On the other hand, coelacanths (*Latimeria chalumnae*) seem to have moved into the depth at 200 m about 200 MYA [53] and the λ_{\max} s of the duplicated RH1 and RH2 pigments were decreased to 482 and 478 nm, respectively, and started to distinguish the narrow range of wavelengths available in its habitat [30,54]. Therefore, the λ_{\max} shifts of RH1 and RH2 pigments and newer duplicate RH2 pigments can be described best by the neofunctionalization model.

The accelerated evolutionary rates of the duplicate RH2 genes (lineage c, Fig. 2b) agree with the prediction [47,55] and observations [56–58] of the evolution of duplicate genes, and not the decelerated evolutionary rates that were caused presumably by the more critical biological functions of duplicate genes, which cause purifying selection [59]. In general, the latter evolutionary patterns of duplicate genes do not apply to opsin genes because in order to move into different ecological environments, organisms need to readjust the λ_{\max} s of their paralogous visual pigments.

Finally, the λ_{\max} s of 10 out of the 13 ancestral pigments and lamprey (P492), eel (P506), zebrafish3 (P488), scabbardfishC (P506) and coelacanth (P478) can be explained fully by the amino acid changes

E122Q, Q122E, M207L and A292S. When the amino acids at the comparable 281 sites (positions 31–311) of these 15 pigments are compared, 123 sites are polymorphic (red and black circles, respectively, Fig. 3). Hence, for these pigments, the amino acid changes only at sites 122, 207 and 292 (2.4%) out of the 123 sites cause λ_{\max} shifts and the remaining changes can be regarded as ‘selectively neutral’. Slightly higher magnitudes (~5%) of adaptive sites have been observed for RH1 [30], SWS1 [31] and M/LWS [32] pigments.

Conclusions

By inferring and engineering the RH2 pigments at 13 critical stages of vertebrate evolution, we have shown that the green-sensitive pigments of the vertebrate ancestor had a λ_{\max} of 503 nm, from which the 12 ancestral pigments changed their λ_{\max} s between 474 and 524 nm and the present-day RH2 pigments have further expanded the range to ~450–530 nm. Eight out of the 16 significant λ_{\max} shifts of RH2 pigments can be explained by the mutations E122Q (twice), Q122E (thrice), M207L (twice) and A292S (once). The highly variable λ_{\max} s of teleost RH2 pigments have been achieved by gene duplications followed by accelerated amino acid substitution.

Acknowledgements

SY was supported partially by National Institutes of Health (EY016400) and Emory University. We thank

Dr. Rosalie K. Crouch (Storm Eye Institute, Medical University of South Carolina), National Eye Institute, for supplying 11-*cis*-retinal to us, and Dr. Ruth Yokoyama and anonymous reviewers for helpful comments.

Conflict of interest

The authors declare no conflict of interest.

Author contributions

SY designed research, performed phylogenetic analyses and wrote the paper. HJ performed experiments and commented on the manuscript.

References

- Wald G (1935) Carotenoids and the visual cycle. *J Gen Physiol* **19**, 351–371.
- Wald G (1955) The photoreceptor process in vision. *Am J Ophthalmol* **40**, 18–41.
- Walls GL (1942) *The Vertebrate Eye and its Adaptive Radiation*. Cranbrook Institute of Science, Bloomfield Hills, MI.
- Lythgoe JN (1979) *The Ecology of Vision*. Clarendon, Oxford.
- Yokoyama S (2008) Evolution of dim-light and color visual pigments. *Annu Rev Genom Human Genet* **9**, 259–282.
- Yokoyama S (2000) Molecular evolution of vertebrate visual pigments. *Prog Retin Eye Res* **19**, 385–419.
- Ebrey T and Koutalos Y (2001) Vertebrate photoreceptors. *Prog Retin Eye Res* **20**, 49–94.
- Munz FW and McFarland WN (1977) Evolutionary adaptations of fishes to the photic environment. In *Handbook of Sensory Physiology* (Crescitelli F, ed), pp. 193–274. Springer-Verlag, Berlin.
- McFarland WN and Munz FW (1975) Part II: the photic environment of clear tropical seas during the day. *Vision Res* **15**, 1063–1070.
- McFarland WN and Munz FW (1975) Part III: the evolution of photopic visual pigments in fishes. *Vision Res* **15**, 1071–1080.
- Jerlov NG (1976) *Marine Optics*. Elsevier, Amsterdam.
- Musilova Z, Cortesi F, Matschiner M, Davies WIL, Suresh Patel JS, Stieb SM, de Busserolles F, Malmström M, Tørresen OK, Brown CJ *et al.* (2019) Vision using multiple distinct rod opsins in deep-sea fishes. *Science* **364**, 588–592.
- Johnson RL, Grant KB, Zankel TC, Boehm MF, Merbs SL, Nathans J and Nakanishi K (1993) Cloning and expression of goldfish opsin sequences. *Biochemistry* **32**, 208–14.
- Matsumoto Y, Fukamachi S, Mitani H and Kawamura S (2006) Functional characterization of visual opsin repertoire in Medaka (*Oryzias latipes*). *Gene* **371**, 268–278.
- Spady TC, Parry JW, Robinson PR, Hunt DM, Bowmaker JK and Carleton KL (2006) Evolution of the cichlid visual palette through ontogenetic subfunctionalization of the opsin gene arrays. *Mol Biol Evol* **23**, 1538–1547.
- Chinen A, Hamaoka T, Yamada Y and Kawamura S (2003) Gene duplication and spectral diversification of cone visual pigments of zebrafish. *Genetics* **163**, 663–675.
- Yokoyama S and Tada T (2010) Evolutionary dynamics of rhodopsin type 2 opsins in vertebrates. *Mol Biol Evol* **27**, 133–141.
- Yokoyama S, Takenaka N and Blow N (2007) A novel spectral tuning in the short wavelength-sensitive (SWS1 and SWS2) pigments of bluefin killifish (*Lucania goodei*). *Gene* **396**, 196–202.
- Korner KE, Schlupp I, Plath M and Loew ER (2006) Spectral sensitivity of mollies: comparing surface- and cave-dwelling Atlantic mollies, *Poecilia mexicana*. *J Fish Biol* **69**, 54–65.
- Archer SN and Lythgoe JN (1990) The visual pigment basis for cone polymorphism in the guppy, *Poecilia reticulata*. *Vision Res* **30**, 225–33.
- Watson CT, Gray SM, Hoffmann M, Lubieniecki KP, Joy JB, Sandkam BA, Weigel D, Loew E, Dreyer C, Davidson WS *et al.* (2011) Gene duplication and divergence of long wavelength-sensitive opsin genes in the guppy, *Poecilia reticulata*. *J Mol Evol* **72**, 240–252.
- Kawamura S, Kasagi S, Kasai D, Tezuka A, Shoji A, Takahashi A, Imai H and 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.
- Jordan R, Kellogg K, Howe D, Juanes F, Stauffer J and Loew E (2006) Photopigment spectral absorbance of Lake Malawi cichlids. *J Fish Biol* **68**, 1291–1299.
- Carleton KL, Parry JW, Bowmaker JK, Hunt DM and Seehausen O (2005) Colour vision and speciation in Lake Victoria cichlids of the genus *Pundamilia*. *Mol Ecol* **14**, 4341–4353.
- Parry JW, Carleton KL, Spady T, Carboo A, Hunt DM and Bowmaker JK (2005) Mix and match color vision: tuning spectral sensitivity by differential opsin gene expression in Lake Malawi cichlids. *Curr Biol* **15**, 1734–1739.
- Kasagi S, Mizusawa K, Murakami N, Andoh T, Furufuji S, Kawamura S and Takahashi A (2015) Molecular and functional characterization of opsins in barfin flounder (*Verasper moseri*). *Gene* **556**, 182–191.
- Yokoyama S, Tada T, Liu Y, Faggionato D and Altun A (2016) A simple method for studying the molecular

- mechanisms of ultraviolet and violet reception in vertebrates. *BMC Evol Biol* **16**, 64.
- 28 Yokoyama S (2012) Synthesis of experimental molecular biology and evolutionary biology: an example from the world of vision. *Bioscience* **62**, 939–948.
 - 29 Yokoyama S (2013) Synthetic biology of phenotypic adaptation in vertebrates: the next frontier. *Mol Biol Evol* **30**, 1495–1499.
 - 30 Yokoyama S, Tada T, Zhang H and Britt L (2008) Elucidation of phenotypic adaptations: molecular analyses of dim-light vision proteins in vertebrates. *Proc Natl Acad Sci USA* **105**, 13480–13485.
 - 31 Shi Y and Yokoyama S (2003) Molecular analysis of the evolutionary significance of ultraviolet vision in vertebrates. *Proc Natl Acad Sci USA* **100**, 8308–8313.
 - 32 Yokoyama S and Radlwimmer FB (2001) The molecular genetics and evolution of red and green color vision in vertebrates. *Genetics* **158**, 1697–1710.
 - 33 Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**, 1586–1591.
 - 34 Yokoyama S and Yokoyama R (1996) Adaptive evolution of photoreceptors and visual pigments in vertebrates. *Annu Rev Ecol Syst* **27**, 543–567.
 - 35 Takenaka N and Yokoyama S (2007) Mechanisms of spectral tuning in the RH2 pigments of Tokay gecko and American chameleon. *Gene* **399**, 26–32.
 - 36 Yokoyama S (2000) Phylogenetic analysis and experimental approaches to study color vision in vertebrates. In *Methods Enzymol* (Palczewski K, ed), pp. 312–25. Academic Press, San Diego, CA.
 - 37 Yokoyama S (1997) Molecular genetic basis of adaptive selection: examples from color vision in vertebrates. *Annu Rev Genet* **31**, 315–336.
 - 38 Chinen A, Matsumoto Y and Kawamura S (2005) Reconstitution of ancestral green visual pigments of zebrafish and molecular mechanism of their spectral differentiation. *Mol Biol Evol* **22**, 1001–1010.
 - 39 Schliewen U, Fricke H, Shartl M, Epplen JT and Paabo S (1993) Which home for coelacanth? *Nature* **363**, 405.
 - 40 Fricke H, Hissmann K, Schauer J and Plante R (1995) Yet more danger for coelacanths. *Nature* **374**, 314.
 - 41 Herring PJ and Cope C (2005) Red bioluminescence in fishes: on the suborbital photophores of *Malacosteus*, *Pachystomias* and *Aristostomias*. *Mar Biol* **148**, 383–394.
 - 42 Takahashi Y and Yokoyama S (2005) Genetic basis of spectral tuning in the violet-sensitive visual pigment of African clawed frog, *Xenopus laevis*. *Genetics* **171**, 1153–1160.
 - 43 Yokoyama S, Altun A, Jia H, Yang H, Koyama T, Faggionato D, Liu Y and Starmer WT (2015) Adaptive evolutionary paths from UV reception to sensing violet light by epistatic interactions. *Sci Adv* **1**, e1500162.
 - 44 Shi Y, Radlwimmer FB and Yokoyama S (2001) Molecular genetics and the evolution of ultraviolet vision in vertebrates. *Proc Natl Acad Sci USA* **98**, 11731–11736.
 - 45 Sun H, Macke JP and Nathans J (1997) Mechanisms of spectral tuning in the mouse green cone pigment. *Proc Natl Acad Sci USA* **94**, 8860–8865.
 - 46 Innan H and Kondrashov F (2010) The evolution of gene duplications: classifying and distinguishing between models. *Nature Rev Genet* **11**, 97–108.
 - 47 Ohno S (1970) *Evolution by Gene Duplication*. Springer-Verlag, New York, NY.
 - 48 Lynch M and Force A (2000) The probability of duplicate gene preservation by subfunctionalization. *Genetics* **154**, 459–473.
 - 49 Hughes AL (1994) The evolution of functionally novel proteins after gene duplication. *Proc R Soc Lond B* **256**, 119–124.
 - 50 Näsval J, Sun L, Roth JR and Andersson DI (2012) Real-time evolution of new genes by innovation, amplification, and divergence. *Science* **338**, 384–387.
 - 51 Near TJ, Eytan RI, Dornburg A, Kuhn KL, Moore JA, Davis MP, Wainwright PC, Friedman M and Smith WL (2012) Resolution of ray-finned fish phylogeny and timing of diversification. *Proc Natl Acad Sci USA* **109**, 13698–13703.
 - 52 Meyer A and Peer Y (2005) From 2R to 3R: evidence for a fish-specific genome duplication (FDGD). *BioEssays* **27**, 937–945.
 - 53 Yokoyama S and Tada T (2000) Adaptive evolution of the African and Indonesian coelacanths to deep-sea environments. *Gene* **261**, 35–42.
 - 54 Yokoyama S, Zhang H, Radlwimmer FB and Blow NS (1999) Adaptive evolution of color vision of the Comoran coelacanth (*Latimeria chalumnae*). *Proc Natl Acad Sci USA* **96**, 6279–6284.
 - 55 Ohno S (1973) Ancient linkage groups and frozen accidents. *Nature* **244**, 259–262.
 - 56 Kondrashov FA, Rogozin IB, Wolf YI and Koonin EV (2002) Selection in the evolution of gene duplications. *Genome Biol* **3**, RESEARCH0008.
 - 57 Lynch M and Conery JS (2000) The evolutionary fate and consequences of duplicate genes. *Science* **290**, 1151–1155.
 - 58 Li C, Li M, Dunwell JM and Zhang Y-M (2012) Gene duplication and an accelerated evolutionary rate in 11S globulin genes are associated with higher protein synthesis in dicots as compared to monocots. *MBC Evol Biol* **12**, 15.
 - 59 Jordan IK, Wolf YI and Koonin EV (2004) Duplicated genes evolve slower than singletons despite the initial rate increase. *MBC Evol Biol* **4**, 22.
 - 60 Palczewski K (2006) G protein-coupled receptor rhodopsin. *Annu Rev Biochem* **75**, 743–767.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Fig. S1. Aligned amino acid sequences of RH2 pigments. **(A)** Present-day pigments, where Bkillifish, scabbard, and Ital lizard are bluefin killifish, scabbardfish, and Italian lizard, respectively. The numbers after P in parentheses show λ_{\max} s. Amino acids sites 122, 207, and 292 are indicated by stars (*). **(B)** 13 ancestral pigments inferred by applying the PAML with JTT model to the 24 sequence data, where the ancestral amino acids with < PP of 95% or less are indicated by bold italic letters. The amino acids in red letters are those of chameleon (P495). Following the tradition in vision science, the amino acid site numbers are those of bovine RH1 (GenBank accession no. M21606).

Fig. S2. Two different inferences of the RH2 pigment evolution in Cypriniformes. The AncCyprinil was inferred to have either E122 **(A)**, Chinen et al. [38]) or

Q122 **(B)**, present analysis). The λ_{\max} s of Ancestors 1–3 are taken from (Chinen et al. [38]). The numbers in ovals and after P in rectangles show λ_{\max} s of the ancestral and present-day pigments, respectively. The amino acids at site 122 are given at the right column. E122Q decreases the λ_{\max} , whereas Q122E increases λ_{\max} . E122Q* explains about 47% of the λ_{\max} shift of Ancestor2. Blue, grey, black, and red indicates the λ_{\max} s of 452–478, 488–492, 495–511, and 516–530 nm, respectively.

Fig. S3. Absorption spectra of ancestral RH2 pigments. The λ_{\max} values of AncJawedFish and AncTetrapod are identical at 488 nm but their absorbance levels at ~ 280 nm are 1.1 and 1.6, respectively.

Table S1. The source of RH2 pigment sequences.

Table S2. Amino acids of AncAgnatha with PP < 0.95 (in parentheses) inferred using PAML with JTT model.

Table S3. Amino acids of ancestral pigments at three critical sites with PP (in parentheses) inferred using PAML with JTT substitution model.