Title:

Turning red to green repeatedly: ancient LWS duplications drive convergent visual tuning in teleost fishes.

Fabio Cortesi[1*], Daniel Escobar Camacho[2,3], Gina Maria Sommer[4], Martin Luehrmann[1], Flurin Hidber[5,6], Xavier Deupi[5,7,8], Karen Carleton[2], Zuzana Musilova[4*]

[1] The School of the Environment, The University of Queensland, Brisbane, QLD 4072, Australia

[2] Department of Biology, University of Maryland, College Park, MD 20742, USA

[3] Laboratorio de Ecología Acuática, Instituto BIÓSFERA, Universidad San Francisco de Quito, Quito 170150, Ecuador

[4] Department of Zoology, Faculty of Science, Charles University, Vinicna 7, 12844 Prague, Czech Republic

[5] Condensed Matter Theory Group, PSI Center for Scientific Computing, Theory, and Data, 5232 Villigen-PSI, Switzerland

[6] Department of Biology, ETH Zürich, Zürich, Switzerland

[7] Laboratory of Biomolecular Research, PSI Center for Life Sciences, 5232 Villigen-PSI, Switzerland

[8] Swiss Institute of Bioinformatics (SIB), 1015 Lausanne, Switzerland

* corresponding authors: f.cortesi@uq.edu.au, zuzana.musilova@natur.cuni.cz

Keywords: whole-genome duplication, vision, opsin, fish, gene duplications, parallel evolution, gene expression, retina

Abstract

Photopigments, formed by an opsin protein bound to a light-sensitive chromophore, underlie vertebrate vision. Long-wavelength–sensitive (LWS) opsins mediate red-light detection, and most teleosts retain a single functional LWS1. A shorter-shifted green-sensitive paralog (LWS2) is found only in a few lineages. By mining teleost genomes and sequencing retinal transcriptomes, we identify elopomorphs as an additional lineage retaining LWS2 (alongside characins and osteoglossomorphs), and we reveal a previously overlooked shorter-shifted paralog, LWS3, restricted to gobies (Percomorpha) and arising from an ancient duplication. Structural modeling of twelve LWS opsins uncovers convergent evolution at four key amino acid sites (214, 259, 261, 269) in the retinal-binding pocket, showing parallel substitutions in human MWS, teleost LWS2, and goby LWS3 relative to red-sensitive counterparts, consistent with repeated spectral shifts toward green wavelengths. In several lineages—including characins, mormyrids, gobies, and primates—these shorter-shifted LWS opsins have functionally replaced the canonical green-sensitive RH2 opsin. Retinal transcriptomes and in situ hybridization demonstrate variable *lws3* expression across gobies, with localization to a distinct double-cone member in *Amblygobius phalaena*, analogous to *rh2* expression in other fishes. Together, these results show that repeated convergent evolution toward green sensitivity over 500 million years involves coordinated changes at the molecular, regulatory, and functional levels, providing a striking example of multilevel sensory adaptation.

Significance statement:

Vertebrate vision depends on opsins, which detect specific wavelengths of light. Most teleosts retain a single long-wavelength–sensitive opsin (LWS1), while a green-sensitive paralog occurs only in a few lineages. We identify a previously overlooked green-shifted opsin, LWS3, restricted to gobies, and show that LWS2 and LWS3 have repeatedly evolved green sensitivity through convergent amino acid changes. These shorter-shifted opsins have functionally replaced the canonical green-sensitive RH2 opsin in multiple teleost lineages and even in primates. By integrating genomic, structural, and expression data, we reveal how multilevel convergent evolution—from molecular tuning to photoreceptor specialization—has repeatedly shaped green-light vision over 500 million years, illustrating the remarkable evolutionary flexibility of vertebrate visual systems.

Introduction

Animals rely on vision for a variety of fundamental tasks, including foraging, mating, avoiding predators, and navigating their environment. At the molecular level, the visual process is initiated when light is absorbed by photopigments, located in the outer segments of the retinal photoreceptors (Wald 1968; Hunt et al. 2014). Vertebrate photopigments consist of light-sensitive vitamin-A-derived chromophores, which are covalently bound to one of five different visual opsin protein types. These five types were already present in the vertebrate ancestor, and are thought to be the product of two rounds of whole-genome duplication (2R) ~ 500 million years ago (Mya) (Lamb, 2013). In the retina, they are expressed in an ancestral set of photoreceptor types (PR0 – PR4) (Baden et al. 2025). They are the dim-light active rod opsin (rhodopsin; RH1, expressed in PR0) and four cone-based photopigments [SWS1 (PR4), SWS2 (PR3), RH2 (PR2), and MWS/LWS (PR1)], which are active during bright light and used for colour and motion vision (Hunt et al. 2014). Their spectral sensitivity ranges are a remarkable example of niche partitioning, where each cone-opsin pigment occupies a constricted visual space, spanning from the ultraviolet (SWS1) to the red end of the light spectrum (LWS) (Marshall et al, 2024).

The evolution of vertebrate visual opsins is characterised by changes in opsin gene numbers, gene structure and gene expression (Hunt et al. 2014; Lamb 2019). Unlike in terrestrial vertebrates, where the basic opsin setup was either maintained [e.g., birds (Borges et al., 2015), reptiles (Katti et al., 2019)] or reduced [e.g., mammals (Jacobs, 2013)], opsin genes have continued to proliferate in teleost fishes (Musilova et al., 2021; Policarpo et al. 2025). With over 35,000 species, teleosts are not only the most species-rich clade of vertebrates (Fricke et al., 2021), they also show extreme ecological, physiology and morphological diversity as they have conquered almost every aquatic and even some terrestrial habitats (Cortesi et al., 2018). It is those differences in photic environment and life history traits that are the main drivers for opsin gene diversity and evolution in fishes (Carleton et al. 2020).

Opsin gene evolution affects teleost vision at all taxonomic levels from species over families to entire classes (Cortesi et al. 2015; Lin et al. 2017; Musilova et al. 2019; reviewed in Carleton et al. 2020 and Musilova et al., 2021). Novel opsins commonly arise through tandem duplication but, as for the ancestral five types, may also be the product of whole-genome duplications (Musilova et al., 2021; Policarpo et al. 2025) and in some rare cases may arise from retrotransposition (Mano et al. 1999; Ward et al. 2008). While most genes are lost shortly after duplication and disappear or become non-functional pseudogenes (Walsh, 2003), some genes persist by acquiring novel functions. This neofunctionalization of visual opsins

usually involves changes in 'key' amino acid sites in the retinal binding pocket, that shift the spectral sensitivity of the photopigment, which is often associated with the colonisation of novel light habitats or 'visual niches' (Yokoyama 2008; Carleton et al. 2020). Neofunctionalization can also occur due to differences in gene expression with ontogeny (Carleton et al. 2008; Cortesi et al. 2016), location in the retina or other tissues (Davies et al. 2015; Morrow et al. 2016), or by changing non-visual characteristics of the protein (e.g. Castiglione et al., 2018; Luk et al., 2016). Finally, teleost opsins are prone to gene conversion, a form of reticulate evolution whereby sequence information is unidirectionally exchanged between sister genes (Cortesi et al. 2015; Sandkam et al. 2017). Gene conversion has the effect of homogenising paralogs, which may repair and sometimes even resurrect gene function (Cortesi et al. 2015).

The genomes of extant teleosts contain a median of seven visual opsin genes: one rod opsin and six cone opsins (Musilova et al. 2019). Most of the cone opsin expansions can be attributed to more recent lineage or species-specific tandem duplications of the *rh2* gene, which is sensitive to the most common, blue to green light spectrum underwater (peak spectral sensitivity range, λ_{max} ~ 450 to 540 nm) (Carleton et al. 2020; Musilova & Cortesi, 2023). Several older tandem duplications have also persisted, such as those found in the neoteleost and percomorph violet to blue-sensitive *sws2* duplicates (Cortesi et al. 2015). Interestingly, at their origin ~ 320 Mya, teleosts underwent a third round of whole-genome duplication (3R) (Ravi and Venkatesh 2018), and two extant opsin gene duplicates, one *rh1* (Chen et al., 2018) and one *lws* (Liu et al. 2019), have been attributed to this event. The ancestral *lws*2 duplicate has long been overlooked but has recently been described from bony tongues (Osteglossiformes) and characins (Characiformes) (Liu et al. 2019; Escobar-Camacho et al. 2020).

Most teleosts possess a single LWS visual pigment (LWS1), which is sensitive from the yellow to red spectrum (λ_{max} ~ 560 to 620 nm) (Carleton et al. 2020; Schweikert et al., 2018), or they might have lost the gene altogether, as commonly seen in deep-sea lineages (Musilova et al. 2019; Lupše et al., 2021). The second copy, LWS2, in extant species, but likely already in the ancestor shortly after its duplication by 3R, has shifted its spectral sensitivity to shorter, greener wavelengths (λ_{max} ~ 530 nm) (Escobar-Camacho et al. 2020). This is similar to the MWS/LWS duplicate in primates, which re-enabled trichromatic vision after the loss of *rh2* in the mammalian ancestor (Carvalho et al., 2017). Concurrently, both bony tongues and characins have either lost or severely reduced the expression of *rh2*, suggesting that there is a trade-off between the overlap of pigment spectral sensitivities and the potential for evolutionary

retention (Liu et al. 2019; Escobar-Camacho et al. 2020). Interestingly, a shift towards shorter spectral sensitivities can also be observed in some more recent LWS1 duplicates in several fish species, as well as between different alleles of the same gene (Seehausen et al. 2008; Sandkam et al. 2018). Given the repeated pattern of *lws* duplication followed by occupation of the green-light niche, it is surprising that teleosts only possess a limited number of these genes. Moreover, the convergence of green-sensitive LWS visual pigments in distantly related vertebrate groups presents a unique opportunity to examine how evolution is repeated to achieve similar visual functions.

Against this background, we explore the evolutionary history of the *lws* opsins in teleost fishes from a phylogenetic representative sampling of more than 100 species. By combining next-generation genome and transcriptome sequencing with open-source data mining, we uncover a rich history of gene duplication, loss and gene conversion. Among these, we find two more teleost groups: tarpons (Elopomorpha), which also retain the *lws2* duplicate, and gobies (Percomorpha), which have a third *lws* copy (*lws3*) of ancestral origin. We then focus on the (re)evolution of green sensitivity and investigate the protein-level signatures of parallelism between the teleost green-sensitive LWS2 and LWS3 visual pigments and the human MWS, uncovering remarkable signatures of repeated amino acid substitutions and functional convergence in gene expression patterns across the vertebrate tree of life.

Results and Discussion

To reconstruct the evolutionary history of LWS opsins, we analyzed available whole-genome data and sequenced retina transcriptomes from 20 selected species carrying a second, short-wavelength–shifted LWS paralog, primarily gobies and osteoglossiforms (Table S1). Combined with structural protein modeling, these data allowed us to focus on the (re)evolution of green sensitivity and the convergent emergence of green–red LWS opsin pairs in teleost fishes and humans.

Lws gene duplications and synteny

Reconstructing a comprehensive *lws* gene tree based on 109 teleost genomes and adding non-teleost ray-finned (Actinopterygian), lobe-finned (Sarcopterygii), and several ancestrally derived fishes as outgroups, we confirm that most teleosts have one red-sensitive *lws* gene (*lws1*). In addition, we find that the tarpons (Elopomorpha) have retained the *lws2* copy, which was previously reported only from the bony tongues and characins, and emerged during 3R (Liu et al. 2019; Escobar-Camacho et al. 2020). Surprisingly, however, we uncover a third

ancestral *lws* (*lws3*) gene in gobies (Gobiiformes) (Figs. 1, 2, and S1). The teleost-specific *lws* duplication that gave rise to *lws2* is currently the earliest described cone opsin duplication (reviewed in Musilová et al., 2021); our results show that *lws3* is possibly of a similar or even more ancient origin. However, accurately dating the *lws3* duplication is more challenging than for *lws2*, mainly due to insufficient phylogenetic signal in the data. Three scenarios are likely:

(1) *Lws3* is an ancient duplicate predating the teleost *lws1–lws2* duplication, which is supported by the nucleotide tree (Fig. S1). However, this finding may be influenced by the saturation of the third-codon positions (Breinholt & Kawahara, 2013), making the phylogenetic signal questionable. Moreover, as gobies are a crown-percomorph group, such an ancestral origin would require at least 19–20 independent *lws3* losses across teleosts (Fig. 1).

(2) *Lws3* is homologous to *lws2* (i.e., derived from the same ancestral duplicate); however, there is no phylogenetic support for this scenario besides for exon 3, which groups the goby *lws3* together with the exon 3 of *lws2* (Fig. S4). This hypothesis also lacks any syntenic support (Fig. S3), and, similar to scenario 1, it implies the independent loss of *lws2/lws3* in all non-goby euteleosts. Nevertheless, five identical amino-acid pairs in the retinal-binding pocket of the LWS2 and LWS3 proteins (Fig. 2B) might suggest otherwise, albeit these might have arisen convergently.

(3) *Lws3* is a more recent duplication originating in the percomorph ancestor. Supporting this scenario, to recover the monophylum with goby *lws1* and *lws3* always involves other percomorphs in the trees based on amino-acid sequences or reduced nucleotide trees (excluding the third codon position), although these trees remain largely unresolved (Fig. S5). It would also suggest an unusually rapid sequence evolution in goby *lws3* and convergence with *lws2*, which is not found in extant goby species.

An additional and unexpected outcome of our phylogenetic analyses is that *lws* genes from some tetrapods (amphibians, snakes, lizards, and mammals) and the non-tetrapod lungfish cluster within the teleost *lws* clade, whereas *lws* genes from other tetrapods (birds and turtles) follow the species tree and branch outside the ray-finned fishes (Fig. S1). This pattern suggests the possibility of yet another additional ancestral *lws* duplication, potentially in the common ancestor of ray-finned and lobe-finned fishes. Under this scenario, birds and turtles would have retained a different *lws* paralog than other tetrapods. However, more extensive sampling across tetrapod lineages will be required to fully resolve this aspect of *lws* evolutionary history.

A detailed analysis of synteny (±10 flanking genes) around teleost *lws* loci revealed a highly conserved genomic context for *lws1*, which is consistently located between *hcfc1* or *sws2* upstream and *gnl3l* downstream (Fig. S3). In contrast, *lws2* shows reduced syntenic

199 conservation (Fig. 3) and occurs on distinct chromosomal regions in bony tongues and
200 characins. Notably, we detected interspecific synteny between two *lws2* genomic regions in
201 tarpon (Elopomorpha) and the *lws2* regions of *Paramormyrops* (Osteoglossomorpha) and
202 cavefish (Characiformes). One tarpon *lws2* region closely matches the *Paramormyrops* cluster,
203 whereas the second corresponds more closely to the characiform cluster (Figs. 3, S2, and S3).
204 This pattern implies that the teleost ancestor possessed two *lws2* paralogs located in distinct
205 genomic regions, both of which are retained in tarpons, while only one copy persists in
206 osteoglossomorphs and the other in ostariophysans.
207 In gobies, *lws3* is located in a distinct genomic region, flanked by *rab32* upstream and
208 *stxbp5* or *androglobin* downstream (Fig. S3). This region lacks any opsin gene in other teleosts
209 (Fig. S3), and no conserved synteny is detectable among the *lws1*, *lws2*, and *lws3* loci. One
210 possibility is that both *lws2* and *lws3* have undergone genomic relocation over evolutionary
211 time. Transposable elements are known to play a major role in opsin gene dynamics in teleosts,
212 as demonstrated in African cichlids (Carleton et al., 2020b), and both teleost *rh1* and one *lws1*
213 copy in guppies and other livebearers originated via retrotransposition (Mano et al., 1999;
214 Ward et al., 2008; Sandkam et al., 2017). Alternatively, ancestral synteny may have been
215 obscured by lineage-specific gene loss or extensive chromosomal rearrangements over deep
216 evolutionary time.
217 To assess regulatory divergence among the ancient LWS duplicates (*LWS1*, *LWS2*, and
218 *LWS3*), we examined conservation of cis-regulatory elements and found that canonical
219 vertebrate *LWS1* transcription factor binding sites (OTX, THR, RAX; Ghinia Tegla et al., 2020;
220 Emerson et al., 2013; Volkov et al., 2020; Mackin et al., 2019; Nelson et al., 2009; Furukawa
221 et al., 1997) are absent from the upstream regions of *LWS2* and *LWS3*, and that putative miRNA
222 signals detected upstream of one *LWS2* copy in *Megalops* and *LWS3* in *Neogobius* exhibit low
223 folding energies (miRNAfold; Tav et al., 2016) and limited similarity to vertebrate miR726,
224 indicating distinct regulatory architectures whose functional elements remain to be identified
225 (Table S2 and Fig. S7).
226
227 **LWS evolutionary history and dynamics of multiple opsin paralogs**
228 Overall, we found that most shallow marine and freshwater fish species have retained at least
229 one copy of the *lws1* gene (Fig. 1). On the contrary, many deeper-living species, including
230 members of the anglerfishes (Lophiiformes), cusk-eels (Ophidiiformes) and toad fishes
231 (Batrachoidiformes), have lost *lws* altogether. Similarly, the deep-water Sanzo's goby,
232 *Lesueurigobius sanzi*, has lost both *lws1* and *lws3* copies (Figs. 1 and 5). This might not be

surprising as longer wavelengths are quickly absorbed with depth (Jerlov 1976), which would make the use of *lws* obsolete (also see Musilova et al. 2019, Policarpo et al. 2025). Various groups also have more recent *lws1* duplicates (Fig. 1). These are the result of lineage-specific whole-genome duplications (e.g., Cypriniformes, Salmoniformes, Esociformes) and, in many cases, tandem gene duplications (e.g., in Atherinomorpha and Anabantiformes). *Lws1* duplications are common in species that live in murky, red-shifted waters, such as some characins (Escobar-Camacho et al. 2020) and pike (Esociformes). However, *lws1* also appears to proliferate in species that exhibit elaborate yellow, orange, and red body patterns, such as wrasses (Labriformes)(Phillips et al., 2016) and livebearers (Windsor & Owens, 2009). In the most extreme cases, brown trout, *Salmo trutta* (Salmoniformes), Siamese fighting fish, *Betta splendens* (Anabantiformes), the humphead or Napoleon wrasse, *Cheilinus undulatus*, and the slingjaw wrasse, *Epibulus insidiator* (both Labriformes), have acquired five *lws1* paralogs (Figs. 1 and 2). Gene duplications can also be found for both *lws2* and *lws3*. While bony tongues have a single copy of *lws2*, the Indo-Pacific tarpon, *Megalops cyprinoides* (Elopiformes), has two copies, and characins may have up to three copies of the gene (Fig. 1) (see Escobar-Camacho et al. 2020 for three *LWS2* copies in *Piabucina panamensis*).

Gene conversion is a common process between paralogous genes and has been documented among *lws1* copies as well as between *lws1* and *lws2* (Escobar-Camacho et al., 2020; Sandkam et al., 2017). We found clear signs of conversion between the LWS1 and LWS2 opsins in Megalops in exon 3 based on single-exon gene tree (Fig. S4). We further screened for recombination breakpoints between *LWS1* and *LWS3* in gobies using GARD (Kosakovsky Pond et al., 2006) and detected evidence of gene conversion in the peacock gudgeon *Tateurndina ocellicauda* (Fig. S6), resulting in two genes with highly similar functional properties (Fig. 5). However, patterns of gene conversion in gobies were less pronounced than those reported for other opsin duplicates, such as *SWS2* (Cortesi et al., 2015). This suggests that *lws3* diverged rapidly after duplication, with subsequent gene conversion acting as a partial homogenizing force—primarily affecting the 5′ portion of the coding sequence—while exons containing spectral tuning sites remained comparatively unaffected.

Convergent red to green functional shifts in LWS opsins

Shorter-wavelength (green)-shifted LWS opsins have evolved convergently in teleost fishes through both *lws2* and *lws3*, representing a reversal of the primate pattern, in which a red-sensitive opsin evolved from a green-sensitive ancestor following gene duplication (Carvalho

267 et al., 2017). Here, we compare structural features of teleost LWS1 (red-sensitive) and
268 LWS2/LWS3 (green-sensitive) opsins with those of the human MWS–LWS pair. In both
269 teleosts and primates, duplication of an ancestral LWS gene enabled the evolution of an
270 additional spectral sensitivity—green in teleosts and red in primates (Fig. 2A). More
271 specifically, LWS2 in characins and bony tongues, as well as several derived teleost LWS1
272 copies, have shifted to green sensitivity via convergent substitutions at similar amino acid sites
273 (Liu et al., 2019; Escobar-Camacho et al., 2020; Sandkam et al., 2018), and here we observe
274 an analogous convergent shift in goby *LWS3* (Table 1).

275 To further test for parallel molecular evolution, we generated structural models for 12
276 opsins, comprising one opsin pair from each of six species: human (*Homo sapiens*, MWS and
277 LWS), the osteoglossomorph aba (*Gymnarchus niloticus*, LWS1 and LWS2), the elopomorph
278 tarpon (*Megalops cyprinoides*, LWS1 and LWS2), the ostariophysan characin (*Cyphocharax*
279 *magdalenae*, LWS1 and LWS2), and two percomorph gobies (*Amblygobius phalaena* and
280 *Stiphodon ornatus*, both LWS1 and LWS3). The models focused on amino acid variation
281 within the retinal-binding pocket, encompassing 10 variable sites (Fig. 2B; Fig. S8; Table 2).
282 We identified three key positions—Y261F, T269A, and F259L (bovine rhodopsin
283 numbering)—that show clear signatures of convergence between human and teleost LWS
284 opsins; in addition, site 214 consistently distinguishes green- and red-sensitive opsins, albeit
285 via different amino acid substitutions.

286 Together, these four sites show parallel changes separating red and green opsins in
287 humans, teleosts with *LWS2*, and teleosts with *LWS3* (Fig. 2B). Several of these positions,
288 particularly 261 and 269, are well-established contributors to spectral tuning and form part of
289 the "five-sites rule" or the "OH-sites rule" (Yokoyama and Radlwimmer, 1998; Yokoyama,
290 2008; Sekharan et al., 2012; Chi et al., 2020). The five-sites rule—positions 180, 197, 277,
291 285, and 308 in human MWS/LWS (bovine rhodopsin equivalents 164, 181, 261, 269, and
292 292)—accounts for most major spectral shifts in vertebrate LWS photopigments, although not
293 all cases are fully explained (Yokoyama and Radlwimmer, 1998; Yokoyama et al., 2008; Chi
294 et al., 2020; Hagen et al., 2023). In addition, we identified five further amino acid substitutions
295 (L88V, A118S, A164A, Q201K, and A275S) shared between green-shifted LWS2 and LWS3
296 opsins, along with site 270, which exhibits elevated variability but involves different residues
297 (Fig. 2B). These shared substitutions further support convergent evolution between the two
298 teleost LWS paralogs, unless *LWS2* and *LWS3* ultimately derive from the same ancestral gene
299 copy (see scenarios above).

300 Ancestral-state reconstruction of the three most variable tuning sites (164, 261, and
301 269) further highlights functional divergence among the paralogs (Fig. 4). *LWS1* consistently
302 retains red sensitivity by preserving longer-wavelength–shifted residues at all three positions,
303 whereas *LWS2* predominantly carries green-shifted variants. *LWS3* exhibits an intermediate
304 pattern: site 261 is green-shifted in all gobies, site 164 is variably green-shifted among species,
305 and site 269 is green-shifted only in mudskippers, with other gobies retaining the longer-
306 wavelength variant (Fig. 4; Table 1).
307 We used the machine-learning approach OPTICS (Frazer and Oakley, 2025) to predict
308 the spectral sensitivities of all tested LWS opsins from 15 goby species, as well as from three
309 teleost species carrying *LWS2* (Table 1). Predicted λmax values for goby LWS1 ranged from
310 551.9 to 567.3 nm, whereas green-shifted LWS3 ranged from 523 to 557 nm, and *LWS2* from
311 other teleosts ranged from 517 to 529 nm. Overall, these predictions indicate that *LWS2* is
312 generally shifted toward shorter (greener) wavelengths than *LWS3*.
313 Predicted spectral sensitivities for goby LWS paralogs largely overlapped with
314 measured mid-wavelength–sensitive (MWS) and long-wavelength–sensitive photoreceptor
315 sensitivities reported for other goby species. However, several empirically measured MWS
316 values (λmax = 527–531 nm; Pierotti et al., 2020 and references therein) were slightly shorter
317 than many predicted values for LWS3 (λmax = 533–550 nm). This discrepancy likely reflects
318 the contribution of additional amino acid sites beyond the canonical five tuning positions to
319 LWS3 spectral sensitivity, consistent with the need for further experimental validation,
320 including in vitro mutagenesis, to fully resolve tuning mechanisms in cone opsins (Yokoyama
321 and Yokoyama, 2020). Alternatively, we considered whether MWS photoreceptors might
322 express a mixture of *rh2* and *lws3* rather than *lws3* alone, as coexpression of *rh2* with *lws2* has
323 been reported in some short-wavelength–shifted characin photoreceptors (Escobar-Camacho
324 et al., 2020). However, because gobies generally lack retinal expression of *rh2* (Fig. 5), this
325 explanation appears unlikely for LWS3.
326
327 **Neofunctionalization of LWS3 in gobies**
328 We confirmed the functional relevance of multiple *LWS* copies across teleost lineages
329 by examining their retinal expression patterns in light of the structural and spectral-tuning
330 differences described above. Both *LWS1* and *LWS2* are expressed in the retinas of tarpon
331 (*Megalops*; Elopomorpha), gymnarchids (Gymnarchidae) and elephantfishes (Mormyridae;
332 both Osteoglossomorpha), as well as in characins (Characiformes; Otomorpha) (Fig. 1),
333 consistent with the distinct spectral sensitivities inferred for these paralogs. Similarly, both

334 *LWS1* and *LWS3* are expressed in gobies (Gobiiformes), although expression patterns vary
335 markedly among species (Figs. 1 and 5).
336 Using a comparative transcriptomic approach, we sequenced retinal transcriptomes
337 from ten goby species representing nine genera and most major subfamilies (Fig. 5). Five
338 species expressed *LWS3* to varying degrees, whereas *LWS1* remained the dominant paralog in
339 most taxa. Notably, in the subfamily Sycidiinae, we observed a pronounced shift in paralog
340 usage: *LWS3* was the most highly expressed copy in the red lipstick goby (*Syciopus*
341 *rubicundus*), which expressed only this paralog, and predominated in the rainbow goby
342 (*Stiphodon ornatus*) (Fig. 5C).
343 All gobies expressed at least one copy of *SWS2*, albeit at low levels in Sycidiinae and
344 in the spinycheek sleeper (*Eleotris pisonis*). In contrast, the canonical teleost green-sensitive
345 opsin *RH2* was generally absent or expressed only at trace levels, with the exception of
346 Sycidiinae, where *RH2* expression was comparable to *SWS2* levels in other gobies (Fig. 5C).
347 Taken together with the predicted and modeled green sensitivity of LWS3, the reduced reliance
348 on *RH2* supports a functional replacement of RH2-mediated green vision by LWS3 in gobies,
349 providing strong evidence for neofunctionalization of this paralog.
350 Using fluorescent in situ hybridization in *Amblygobius phalaena*, we further showed
351 that the *SWS2* opsins (*SWS2B* and *SWS2Aβ*) are expressed exclusively in single cones (Fig. S9
352 A, B). Within single cones, *SWS2Aβ* was expressed in nearly all photoreceptors, whereas
353 *SWS2B* showed variable and often lower expression, resulting in partial coexpression within
354 the same single cones (Fig. S9 A–C). In contrast, *LWS1* and *LWS3* were expressed exclusively
355 in double cones (Fig. 5 F–H; Fig. S9 D–F), where they localized to opposite members of each
356 double cone pair (Fig. 5F). This cellular segregation demonstrates that *LWS3* has acquired an
357 independent photoreceptor role in *A. phalaena* and, together with *SWS2*- and *LWS1*-expressing
358 photoreceptors, likely underpins trichromatic vision in this species.
359
360 **Conclusion**:
361 By reconstructing the evolutionary history of LWS opsins across ray-finned fishes, we show
362 that repeated gene duplication, divergence, and differential retention have generated three
363 ancient LWS paralogs—*LWS1*, *LWS2*, and *LWS3*—that remain functional in extant lineages
364 despite extensive genomic rearrangement and regulatory divergence. Our discovery of *LWS3*
365 reveals an unrecognized layer of sensory diversity and demonstrates that ancient duplicates can
366 be repeatedly co-opted to expand visual function long after their origin. In gobies, *LWS3* has
367 undergone neofunctionalization to mediate green sensitivity, functionally replacing the

368 canonical green-sensitive opsin *RH2*. This role parallels that of *LWS2* in osteoglossiforms and
369 characins and mirrors primate color vision, in which duplication of an LWS gene enabled
370 green-sensitive cones following loss of *RH2* in the mammalian ancestor. Across these systems,
371 convergent amino acid substitutions at key spectral-tuning sites underpin parallel functional
372 shifts, highlighting common molecular solutions to similar selective pressures. Together, our
373 results show that opsin gene evolution in teleost fishes is exceptionally dynamic, with ancient
374 gene duplicates serving as a recurrent substrate for sensory innovation. This evolutionary
375 flexibility contrasts with the more constrained opsin repertoires of other vertebrates and
376 underscores the power of comparative genomic and functional approaches to reveal hidden
377 aspects of visual system evolution.

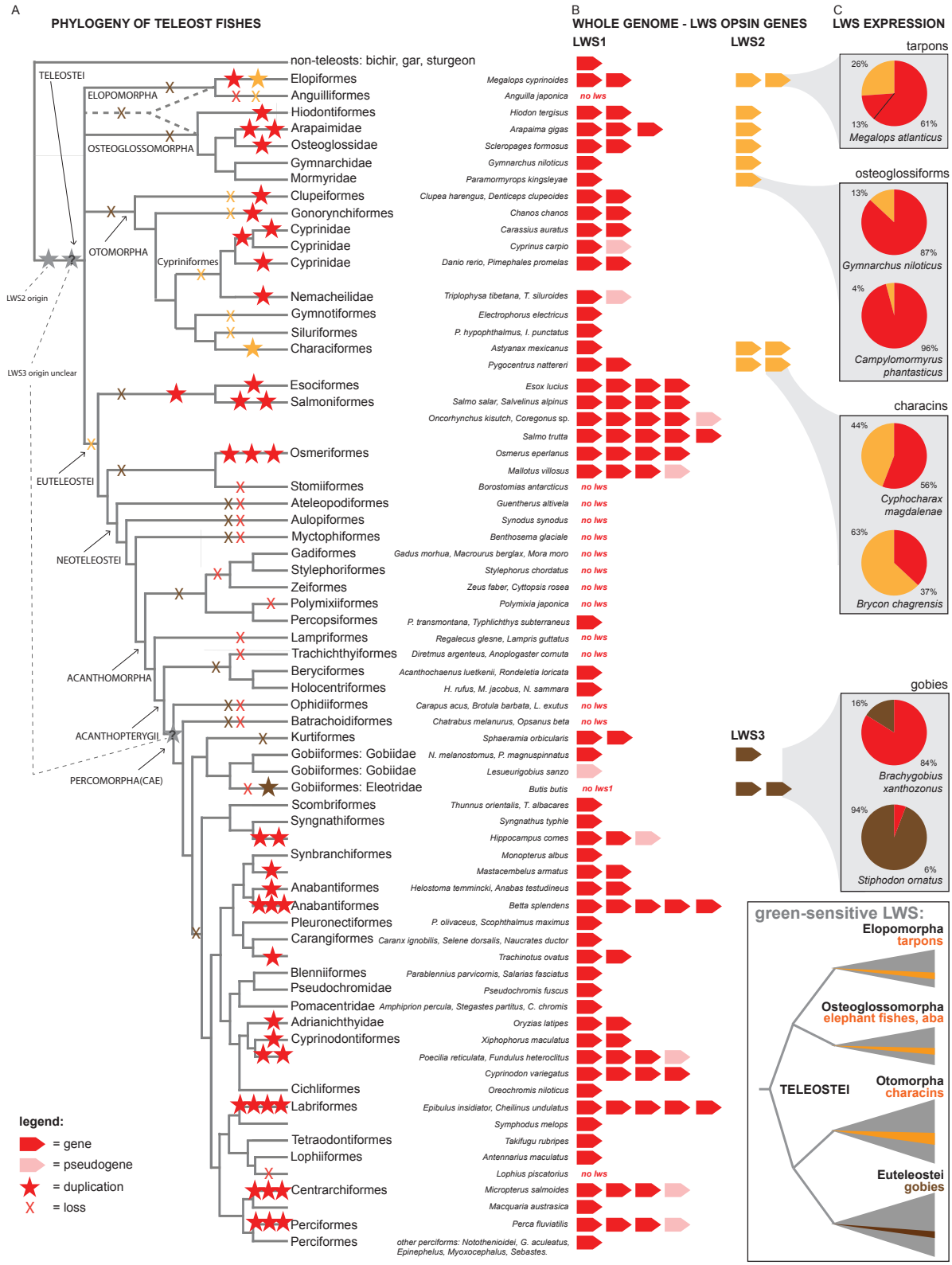
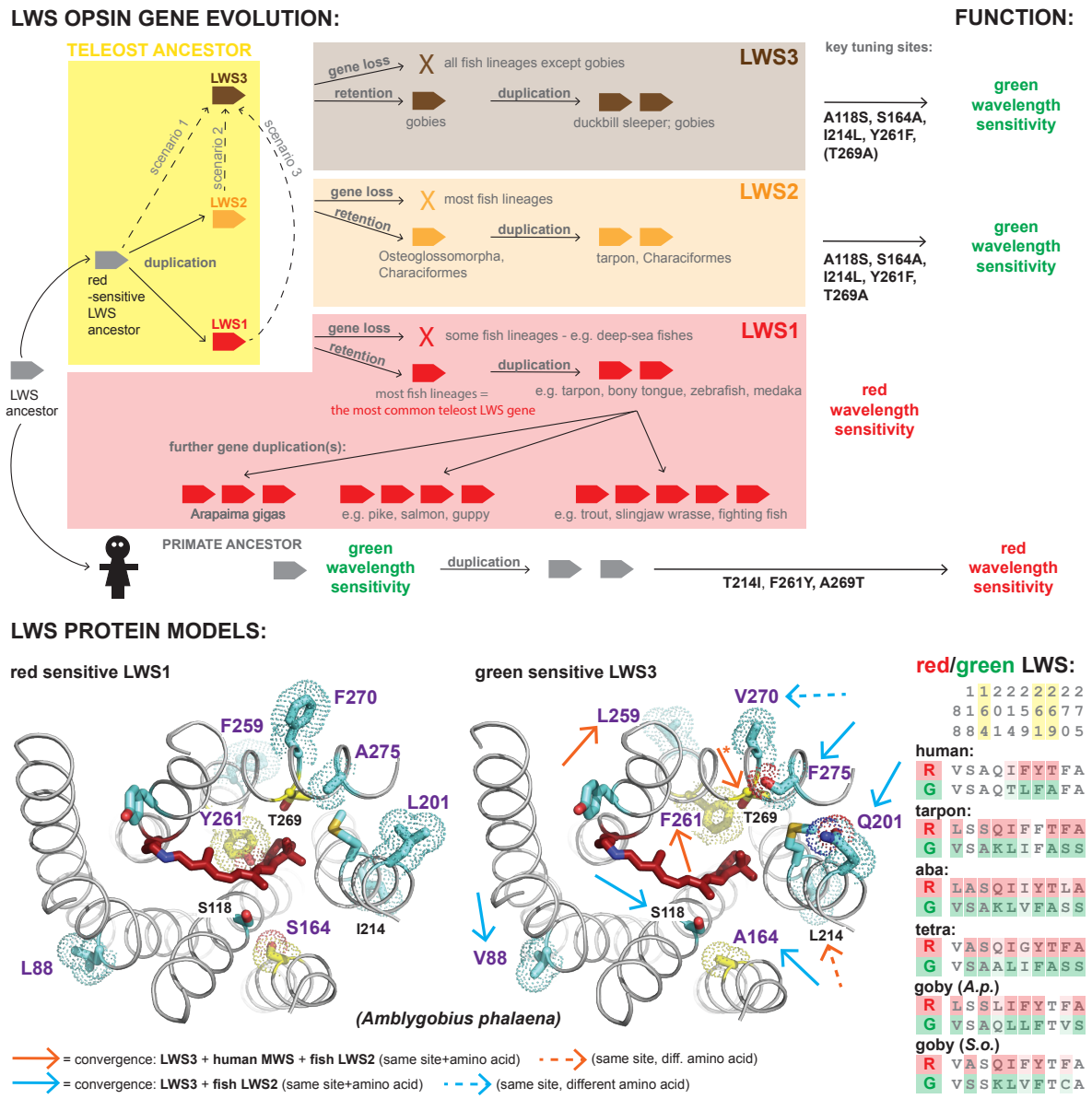


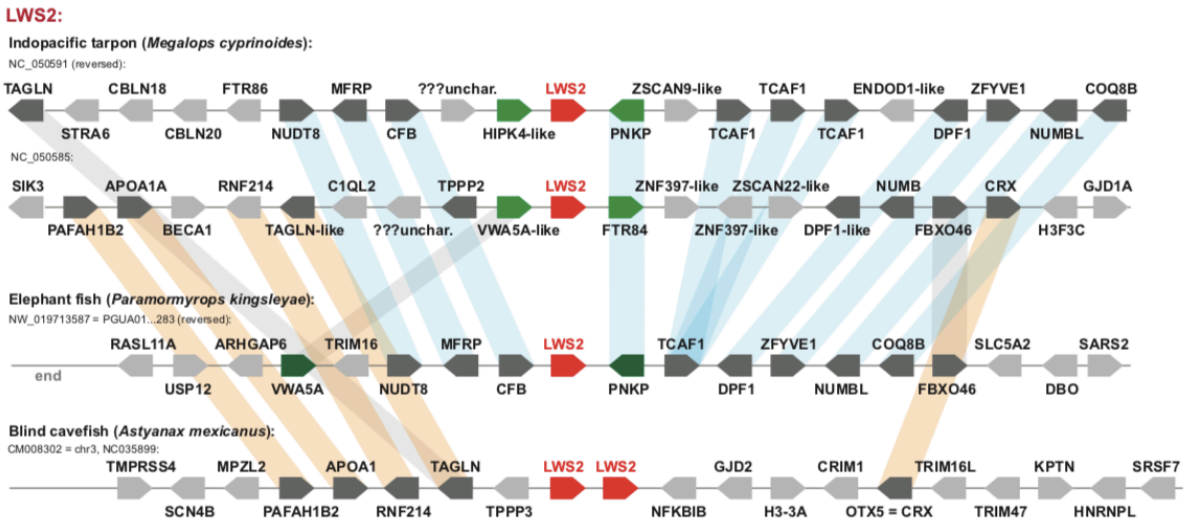
***Figure 1**: LWS gene repertoire in teleost fishes. A) Putative gene duplications (star symbols) and gene losses (X symbols) mapped on the teleost phylogenetic tree (simplified after Betancur et al., 2017). Most of the duplications are species/genus specific, except for the Esociformes+Salmoniformes and Cyprinodontiformes which are older. Two ancient gene*

383 *duplications are marked at the onset of teleosts (grey stars). B) Number and type of LWS opsin*
384 *genes found in the whole genomes of the teleost fishes. Red - LWS1 (red-sensitive), orange -*
385 *LWS2, brown - LWS3 (both shifted towards shorter wavelenghts ~ green). Note that only few*
386 *lineages of Elopomorpha, Osteoglossomorpha and Characiformes have the LWS2 gene, and*
387 *only gobies (Gobiiformes) have the LWS3 gene. C) Proportional gene expression pattern in*
388 *selected species with multiple LWS gene types: tarpon (Elopomorpha), elephant fishes*
389 *(Osteoglossiformes), characins (Ostariophysi) and gobies (Percomorpha). Position of these*
390 *groups within teleost phylogeny highlighted in the inset below the expression plots. For the*
391 *LWS gene tree, please see Figure S1, for the full list of the analysed genomes, please see Table*
392 *S1. For the details on the goby phylogeny (15 species), please refer to Figure 5.*
393



394

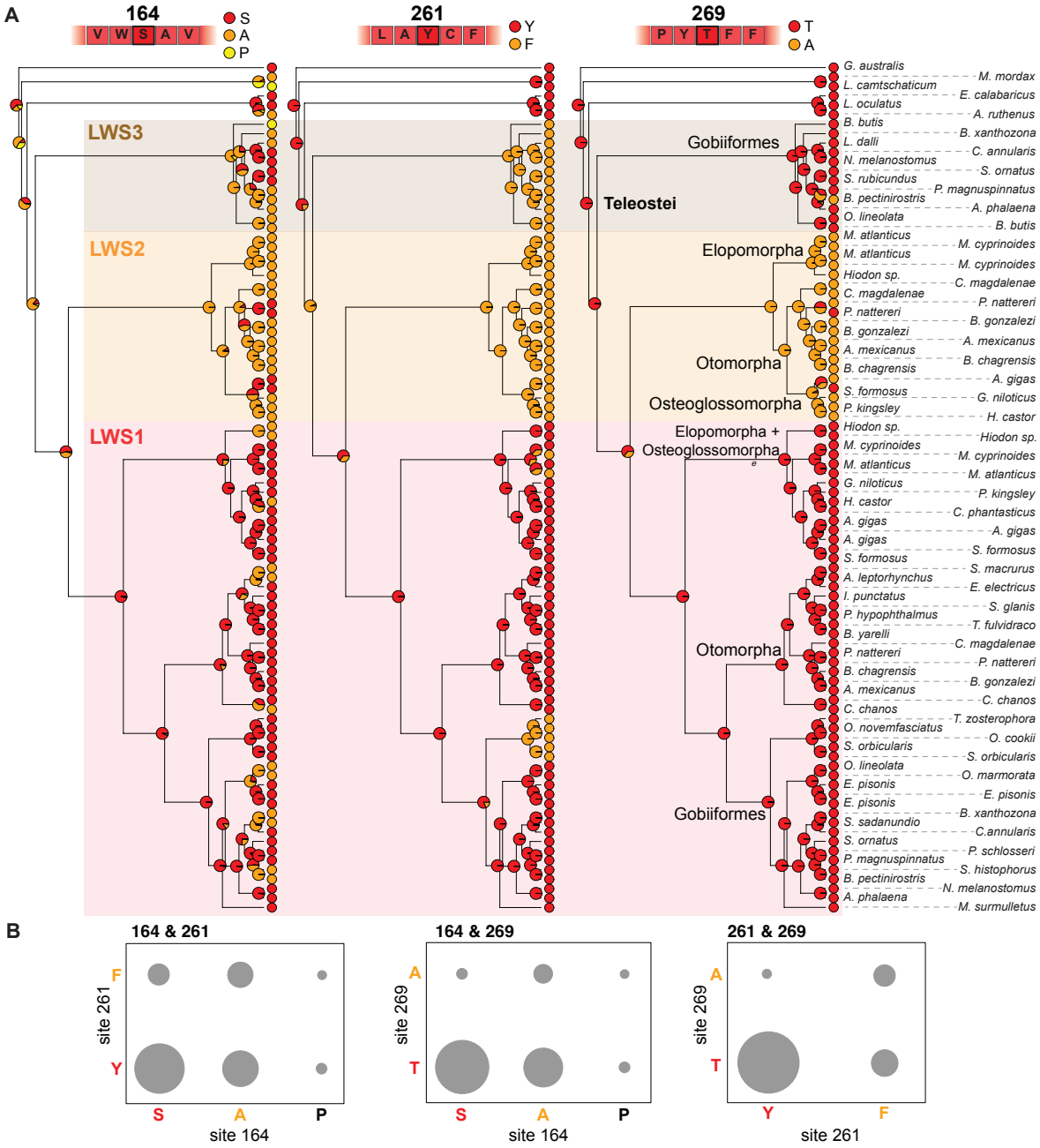
395 ***Figure 2*** *- Evolution of green-light sensitivity using LWS opsins. A) Potential models of fate*
396 *and dynamics of the three LWS teleost copies. The most prevalent gene copy is LWS1, which*
397 *is present in most fish lineages and functions mainly in the longer wavelengths (red). The LWS2*
398 *copy is retained in Osteoglossomorpha, Elopomorpha and Characiformes (and is not found in*
399 *any Euteleostei). The LWS3 has been detected only in gobies (Gobiiformes) so far and has been*
400 *lost in all other fish lineages. Both LWS2 and LWS3 are functionally shifted towards shorter*
401 *wavelengths. B) LWS opsin protein models by Boltz. Green sensitivity evolved from the red-*
402 *sensitive LWS opsin by amino acid mutations in the retinal pocket. Convergently, similar*
403 *residues have mutated in human and teleost green + red opsin pairs. Three amino acid sites*
404 *show the same mutations in human MWS (green) opsin and in both teleost LWS2 and LWS3*
405 *opsins (red arrows). Further sites have evolved convergently among teleosts, i.e. in LWS2 and*
406 *in LWS3 in parallel (blue arrows). * marks the mutation T269A, which is found in humans,*
407 *LWS2 gene in teleosts, and also in terrestrial gobies (mudskippers) but not in the two goby*
408 *species selected for the modelling. For all 12 structural models performed in this study, please*
409 *refer to Fig. S8.*
410



411

412 ***Figure 3*** *– Synteny of the LWS2 and upstream and downstream genes. Two clusters with the*
413 *LWS2 gene located on two different chromosomes are found in tarpon (Megalops), while only*
414 *one cluster is in elephant fish (Paramormyrops) and cavefish (Astyanax). One of the tarpon*
415 *clusters (NC_050591) shows more conserved synteny with the elephant fish cluster (same*
416 *genes connected by blue shading), while the other tarpon cluster (NC_050585) is more similar*
417 *to the cavefish (connected by orange bars). Note, that the LWS2 genes themselves do not carry*
418 *this phylogenetic signal anymore (see gene tree in Figure 4 and Figure S1). For the synteny*

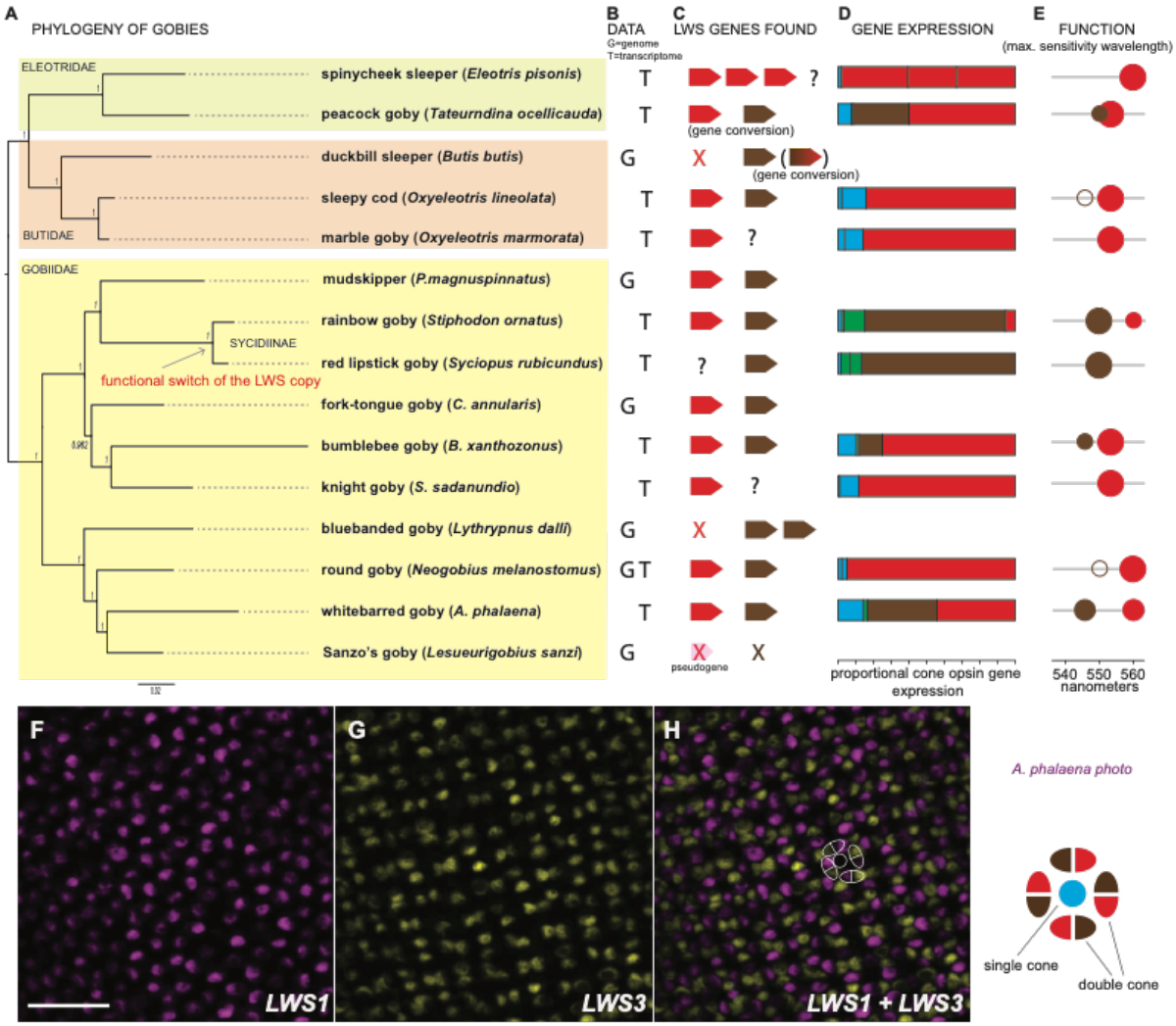
419 *evidence on a longer genomic fragment (up to 2 Mbp) refer to Figure S2 and for the synteny*
420 *of the LWS1 and LWS3 genes, please see Figure S3.*
421



422

423 ***Figure 4*** *- Ancestral state reconstruction of the three most variable key-tuning sites in selected*
424 *teleosts with multiple LWS gene types. A) Sites 164, 261 and 269 are known to shift the spectral*
425 *sensitivities. The shorter-shifting (orange circles) and longer-shifting (red circles) amino acids*
426 *have been found differentially evolving in groups with multiple LWS, i.e., in gobies (LWS1 and*
427 *LWS3) and elopomorphs (tarpons) + osteoglossomorphs (mormyrids) + otomorphs*
428 *(characins) (LWS1 and LWS2). While LWS1 has predominantly longer-shifting amino acid*

429 *variants in all three sites, the LWS2 gene demonstrates shorter-shifting amino acid variant in*
430 *all three positions, and, hence, the notable shift towards the green sensitivity. The LWS3 gene*
431 *shows a mixed signal with a uniform shorter-shifting variant (F) in site 261, a mostly longer-*
432 *shifting variant (T) in site 269 and a mix of both variants at site 164 (S and A). Pie charts on*
433 *the nodes indicate the scaled likelihoods of each variant, calculated using the ace function in*
434 *APE. B) Summary of the association of amino acid variants in the LWS opsin evolution*
435 *including all analysed species and all genes together (325 haplotypes). Most of the LWS genes*
436 *show similar signal of red shift (longer-shifting variant) in most sites (driven mostly by LWS1*
437 *sequences).*
438



439

440 ***Figure 5:*** *Evolution of the LWS opsin genes in gobies (Gobiiformes)**.*** *A) phylogenetic*
441 *relationship among the studied gobies based on the genomic and transcriptomic data set. Three*
442 *goby families are included in this study. B) Source of data: genomic (G), transcriptomic (T) or*
443 *both (GT) data available. C) List of the found LWS genes. Note that if only transcriptomic data*

444 *are used, we cannot make any conclusions on the gene loss (marked by ?). D) proportional*
445 *cone opsin gene expression in the gobies. Colour code in the horizontal bars stands for the*
446 *sensitivity of the opsin gene - blue = SWS2, green = RH2, red = LWS1, brown = LWS3. Note*
447 *that the LWS genes dominate the cone expression profiles in all studied species, and it is either*
448 *LWS1 (most gobies) or LWS3 (Sycidiinae). E) function of the two LWS genes in vision. Spectral*
449 *sensitivity (λmax) of the LWS1 (red) and LWS3 (brown) opsin gene has been predicted (see*
450 *Table 1) and is shown on the wavelength scale. Larger vs. smaller circles represent the*
451 *dominant vs. rare transcripts in the expression profile. Empty circle means no expression. Note*
452 *the functional proximity of the two LWS1 and LWS3 genes in the peacock goby (*Tateurndina
453 ocellicauda*) where gene conversion has been detected (see also Fig S6).* ***F-H)*** *Double-*
454 *labelling in-situ hybridisation of expressed opsin mRNAs in retinal double cone photoreceptors*
455 *in adult* Amblygobius phalaena*. LWS1 (magenta) and LWS3 (yellow) mRNAs were expressed*
456 *in opposite members of double cones. Scale bars: 10 µm. For in-situ of the single cones refer*
457 *to Fig S9.*
458
459

460 ***Table 1****: LWS opsins and predicted values of the maximum sensitivity (λmax) in gobies (LWS1,*
461 *LWS3) and other teleosts (groups with LWS2). For gobies, we applied the five-sites rule by*
462 *adding + 6 nm in 164, + 9 nm in 261, + 13 nm in 269 if polar residue is present (Sekharan et*
463 *al., 2012; Yokoyama 2008). For gobies we started with predicted values for the mudskipper*
464 *opsins (You et al., 2014) and the only predicted combination was S-H-F-T-A (‡). For the LWS2*
465 *species (Megalops, characin, Gymnarchus), we started with the value from bovine rhodopsin*
466 *~ 500 nm (+28 nm for E181H). We used OPTICS for the machine-learning predictions with*
467 *the whole-dataset (mine-and-match or heterologous) and considering amino-acid properties*
468 *(AA_prop parameter). LWS2 and LWS3 sensitivity is shifted towards green compared the*
469 *LWS1 (red). The rhodopsin sites 164, 181, 261, 269, and 292 correspond to the 180, 197, 277,*
470 *285, and 308 sites in the human LWS gene.*

species	gene	five-sites rule amino acids					λmax prediction - 5-sites rule	λmax prediction by OPTICS:			
		180	197	277	285	308		WDS-mnm; AA prop; median + SD		par: WDS; AA prop; median + SD	
bovine rhodopsin equivalent:		*164*	*181*	*261*	*269*	*292*					
bovine rhodopsin amino acid:		*A*	*E*	*F*	*A*	*A*					
GOBY SPECIES WITH BOTH TYPES OF LWS (LWS1, LWS3) FOUND IN THE TRANSCRIPTOME/GENOME:											
Amblygobius phalaena	LWS1	S	H	Y	T	A	560 nm	560.3 nm	4.7	563.4 nm	10.8
	LWS3	A	H	F	T	A	546 nm	546.6 nm	3.4	548 nm	7.6
Boleophthalmus pectinirostris	LWS1	A	H	Y	T	A	553 nm	531.4 nm	3.1	530.9 nm	5.8
	LWS3	A	H	F	A	A	531 nm	551.1 nm	4.3	551.5 nm	7
Brachygobius xanthozona	LWS1	A	H	Y	T	A	553 nm	549.9 nm	4	551.9 nm	7.3
	LWS3	A	H	F	T	A	546 nm	543.9 nm	2.9	544.1 nm	6
Chaenogobius annularis	LWS1	S	H	Y	T	A	560 nm	560.4 nm	5.1	565.1 nm	11.3
	LWS3	S	H	F	T	A	551 nm‡	549.1 nm	4.4	555.7 nm	7.6
Neogobius melanostomus	LWS1	S	H	Y	T	A	560 nm	559.9 nm	4.9	562.7 nm	10.7
	LWS3	S	H	F	T	A	551 nm‡	544.9 nm	4.4	550.6 nm	6.8
Oxyeleotris lineolata	LWS1	A	H	Y	T	A	553 nm	552.8 nm	3.9	556.2 nm	8.8
	LWS3	A	H	F	T	A	546 nm	544.7 nm	4.8	542.4 nm	10.4
Periophthalmus magnuspinnatus	LWS1	S	H	Y	T	A	560 nm	559.6 nm	4.3	562.7 nm	8
	LWS3	A	H	F	T	A	546 nm	543.7 nm	4	543.4 nm	6.8
Stiphodon ornatus	LWS1	S	H	Y	T	A	560 nm	552.6 nm	5.4	558.6 nm	9
	LWS3	S	H	F	T	A	551 nm‡	551.5 nm	3.4	555.8 nm	8.8
Tateurndina ocellicauda	LWS?	A	H	Y	T	A	553 nm	552.9 nm	3.7	554.1 nm	8.4
	LWS?	S	H	F	T	A	551 nm‡	548.2 nm	3.5	552.7 nm	7.2
GOBY SPECIES WITH ONLY ONE TYPE OF LWS FOUND IN THE TRANSCRIPTOME/GENOME:											
Butis butis (mislabelled as *O. marmorata* on GenBank*)*	LWS3-1	A	H	F	T	A	546 nm	n/a	n/a	n/a	n/a
	LWS3-2	P	H	F	T	A	546 nm	n/a	n/a	n/a	n/a
Eleotris pisonis	LWS1-1	S	H	Y	T	A	560 nm	552.4 nm	4	557.4 nm	7.4
	LWS1-2	S	H	Y	T	A	560 nm	552.5 nm	3.7	557.6 nm	7.1
	LWS1-3	S	H	Y	T	A	560 nm	551.9 nm	3.7	557.1 nm	7
Oxyeleotris marmorata	LWS1	A	H	Y	T	A	553 nm	547 nm	3.4	552.3 nm	6.1
Syciopus rubicundus	LWS3	S	H	F	T	A	551 nm‡	550.8 nm	3.6	556 nm	9.2
Stigmatogobius sadanundio	LWS1	A	H	Y	T	A	553 nm	552.5 nm	3.8	556.9 nm	7.8
Lythrypnus dalli	LWS3	A	H	F	T	A	546 nm	n/a			n/a
TELEOSTS WITH LWS2 (ELOPOMORPHA + OSTEOGLOSSOMORPHA + CHARACIFORMES) + HUMAN† :											
Megalops cyprinoides	LWS1a	S	H	F	T	A	544 nm	551 nm	1.6	551.6 nm	3.4
	LWS1b	S	H	Y	T	A	553 nm	561.1 nm	1.4	560.7 nm	3.4
	LWS2_715	A	H	F	A	A	528 nm	530.6 nm	3.4	530.3 nm	5.3
	LWS2_709	A	H	F	A	A	528 nm	514.4 nm	8.5	524.2 nm	8.1
Gymnarchus niloticus	LWS1	S	H	Y	T	A	553 nm	550.8 nm	3.4	554.7 nm	6.1
	LWS2	A	H	F	A	A	528 nm	529.3 nm	3.5	529.6 nm	6
Cyphocharax magdalenae	LWS1	S	H	Y	T	A	553 nm	553.3 nm	3.5	556.6 nm	7.5
	LWS2_1	A	H	F	A	A	528 nm	513.2 nm	8.7	524.9 nm	8.4
	LWS2_2	A	H	F	A	A	528 nm	528.6 nm	3.4	531.1 nm	3.9
Homo sapiens	LWS	A	H	Y	T	A	547 nm	557.9 nm	0.9	557 nm	1.1
	MWS	A	H	F	A	A	528 nm	529.7 nm	0.5	530.2 nm	0.5

471

472

Table 2. Variable position in the retinal binding pocket. Five-sites rule amino acid positions 164, 261 and 269 highlighted by yellow. Predictions for lambda max values using five-sites rule are provided.

bovine rhodopsin position	λ max /	88	118	164	201	208	214	259	261	269	270	275	293
human_NM_020061_LWS_OPN1LW	547	V	S	A	Q	M	I	F	Y	T	F	A	Y
human_NM_000513_MWS_OPN1MW	528	V	S	A	Q	M	T	L	F	A	F	A	F
Aphalaena_FZ1_LWSx1	560	L	S	S	L	M	I	F	Y	T	F	A	Y
Aphalaena_FZ3_LWSx2	546	V	S	A	Q	M	L	L	F	T	V	S	Y
Stiphodon_ornatus_LWS01	560	V	A	S	Q	M	I	F	Y	T	F	A	Y
Stiphodon_ornatus_LWS02	551	V	S	S	K	M	L	V	F	T	C	A	Y
Cmagdalenae_MT310731	553	V	A	S	Q	M	I	G	Y	T	F	A	Y
Cmagdalenae_MT310746	528	V	S	A	A	L	L	I	F	A	S	S	Y
Gymnarchus_niloticus_119C1_LWS1	553	L	A	S	Q	M	I	I	Y	T	L	A	Y
Gymnarchus_niloticus_119C1_LWS2	528	V	S	A	K	M	L	V	F	A	S	S	Y
Megalops_cyprinoides_CM023712_LWS01A	544	L	S	S	Q	M	I	F	F	T	F	A	Y
Megalops_cyprinoides_CM023709_LWS2	528	V	S	A	K	L	L	I	F	A	S	S	Y

Materials and Methods

Ethical approval and relevant permits

All experimental procedures involving live animals within Australia were approved by the University of Queensland Animal Ethics Committee (QBI/304/16). Fish were collected under the Great Barrier Reef Marine Parks Permit (G17/38160.1) and Queensland general fisheries permits (207976 and 199434). Collecting samples in the Czech Republic has been permitted by the Czech Fisheries Association permit. Species collected for the purpose of the retina transcriptome sequencing have been euthanized by MS222 following the regulation and by researchers with required animal research certificate.

Taxon sampling

Samples of *Amblygobius phalaena* and *Epibulus insidiator* for expression have been collected and previously published in Musilova et al., 2019 (see Table S1 for sample list and Accession numbers). *Oxyeleotris lineolata* (n = 4 sub-adults) were collected from the Dawson River at Taroom, Queensland, Australia in May 2020. *Amblygobius phalaena* specimens (n = 3 adults) for in-situ hybridisation were collected on snorkel from shallow reefs (1-2 m) surrounding

493 Lizard Island, Great Barrier Reef, Australia in March 2021 using an anaesthetic clove oil
494 solution (10% clove oil, 40% ethanol, and 50% seawater) and hand nets. Eyes were enucleated
495 and the retina was dissected out and either fixed in 4% PFA for in-situ hybridisation or stored
496 on RNAlater at -20 C until further processing for transcriptome sequencing. Samples of
497 *Neogobius melanostomus* were collected in the Elbe River (Děčín, Czech Republic) using the
498 electro fishing device. Samples of the remaining seven species, *Eleotris pisonis, Eleotris*
499 *marmorata, Tateurndina ocellicauda, Stiphodon ornatus, Syciopus rubicundus, Brachygobius*
500 *xanthozonus* and *Stigmatogobius sadanundio*, have been obtained from aquarium trade.
501
502 *Transcriptome sequencing*
503 Retinal RNA was extracted using a Qiagen mini kit column extraction following the
504 manufacturer's protocol and including a DNAse step. RNA extractions concentrations were
505 measured on a Fluorometer and integrity checked with a Pico chip on a 2100 Bioanalyzer from
506 Agilent. Library preparation (unstranded, 300 bp insert) was conducted using Illumina's
507 NEBNext Ultra II Directional RNA library preparation kit, NEBNext Multiplex Oligos and the
508 NEBNext Poly(A) mRNA Magnetic Isolation Module (New England Biolabs) for all
509 osteoglossiform, tarpon and goby species used for transcriptomics except for *Oxyleotris*
510 *lineolata, Neogobius melanostomus* and *Amblygobius phalaena* whose library preparations
511 have been performed on the commercial basis in Novogene, Singapore
512 (https://en.novogene.com/). Transcriptome sequencing (RNAseq HiSeq Paired-End PE150)
513 was outsourced to Novogene for all species.
514
515 *Assembly and gene extraction*
516 Transcriptome raw reads have been filtered and mapped with Medium sensitivity settings
517 against the general goby reference (based on the genome of *Neogobius melanostomus*; Adrian-
518 Kalchauser et al., 2019) in the Geneious 9.0 software (https://www.geneious.com). Species
519 specific gene sequences have been retrieved from the mapped alignments and used as a species-
520 specific reference for another round of mapping with the Lowest sensitivity settings. Number
521 of reads have been calculated to the FPKM (fragments per kilo base per million reads) and the
522 proportion of each opsin from the total cone opsin gene expression has been calculated. More
523 details on the protocols described in de Busserolles et al. 2017, Musilova et al., 2019, and
524 Tettamanti et al. 2019.
525
526 *Whole genome analysis*

527 High quality teleost whole genomes have been downloaded from GenBank (accession numbers
528 as per sample Table S1). Majority of the high-quality genomes has been sequenced by the
529 Sanger Institute (Rhie et al., 2021). We further complemented this data set by the genomic data
530 of lower quality (draft genomes as per Cortesi et al., 2015 and Musilova et al., 2019).
531 Assembled genomic scaffolds / chromosomes have been mapped against the general single
532 exon reference of the example LWS1 (*Oreochromis niloticus*), LWS2 (*Astyanax mexicanus*)
533 and LWS3 (*Neogobius melanostomus*) copy. Mapping has been performed in Geneious 9.1.8
534 (https://www.geneious.com) using Medium to Medium High sensitivity settings - this settings
535 has been tested to pick all scaffolds with any LWS opsin (higher sensitivity yielded to false
536 hits). Single LWS genes have been extracted from the scaffolds either following the annotation
537 (if available) or by mapping of single exons against the scaffold with the Highest sensitivity.
538
539 *Phylogenetic analysis*
540 Coding sequences of the single LWS genes have been aligned in Geneious 9.1.8 using the
541 MAFFT Alignment plugin (Katoh et al., 2019). Five independent runs (each with two parallel
542 runs) have been performed in MrBayes 3.1.2 (Ronquist and Huelsenbeck, 2003) software for
543 50 million generations. The output files have been manually inspected in the Tracer 1.5
544 (https://github.com/beast-dev/tracer/releases) for reaching the lnL plateau, the best -lnL score
545 and to identify the burnin break point. Three out of the ten independent output runs have been
546 used for the consensus tree construction with the burnin set to 50%.
547
548 *Ancestral state reconstruction*
549 In order to assess the evolution of the key-tuning amino acid sites, we focused on the three key
550 sites that are known to cause large shifts in λ_{max} (sites 164, 261 and 269 – 7,10 and 16 nm,
551 respectively) and reconstructed the ancestral states for each key site with the gene tree pruned
552 to the taxa with multiple LWS types, i.e., the Elopomorpha+Osteoglossomorpha+Otomorpha
553 with LWS1 and LWS2 and for GObiiformes with LWS1 and LWS3 gene types. Using the
554 occurrence of amino acids variants that shift λ_{max} as character states, and with the equal
555 transition rates model (ER), we performed the ancestral state analysis the ace function in the
556 APE package (Paradis & Schliep, 2019). The ace function employs a maximum likelihood
557 approach where the reconstructed ancestral states are given as a proportion of the total
558 likelihood for each state at each node. Trees were also analysed, visualized, and edited with the
559 phytools package.
560

561 *Gene conversion analysis*

562 GARD (Kosakovsky Pond et al. 2006) was run using general discrete site-to-site variation with
563 three rate classes on the Datamonkey platform (www.datamonkey.org) (Weaver et al. 2018)
564 on an alignment of *LWS1* and *LWS3* from goby species for which both copies were recovered
565 (Fig. 3). The sliding window analysis followed the methods described in Cortesi et al. 2015. In
566 short, the intraspecific dS (neutral process) along the coding sequences of *LWS1* and *LWS3*
567 was calculated using a window size of 30 and a step size of 1 in DNAsp v.6 (Rozas et al. 2017).
568

569 *Regulatory elements analysis*

570 The key regulatory elements controlling expression of the *LWS* opsin gene were examined
571 including the locus control region (LCR), which is shared across vertebrates (Smallwood et al
572 1992; Wakefield et al 2018), and the associated miRNA (miR-726) shared across fishes
573 (O'Quin et al 2011; Smallwood et al., 2002; Wakefield et al., 2008). In most fish, the miRNA
574 and LCR are within a 500 bp region. We aligned these regions from *Oryzias latipes,*
575 *Gasterosteus aculeatus, Oreochromis niloticus* and *Danio rerio,* to identify the core LCR and
576 miRNA regions. These were then blasted to the genomes of *Astyanax, Paramormyrops,*
577 *Meglalops,* and *Neogobius* to identify homologous regions upstream of any of the *LWS* genes.
578 Within the LCRs, we identified binding sites for key regulators of the LWS gene (OTX2, THR
579 or RAX; Ghinia Tegla et al., 2020; Emerson et al., 2013; Volkov et al., 2020; Mackin et al.,
580 2019; Nelson et al., 2009; Furukawa et al., 1997) using Jaspar 2020 (Fornes et al., 2019). If
581 there was homologous sequence to miR-726, we tested miRNA stability by calculating the
582 miRNA folding energy with miRNAfold (https://evryrna.ibisc.univ-evry.fr/miRNAFold). In
583 this method, a more negative energy signifies more stability and so a more likely functional
584 miRNA.
585

586 *Protein structure prediction and retinal binding pocket analysis*

587 Three-dimensional structures of LWS opsins were predicted using Boltz (Passaro et al.,
588 2025), a deep learning-based protein structure prediction method. Predicted structures were
589 converted from mmCIF to PDB format for downstream analysis. Bovine rhodopsin (PDB:
590 1GZM, chain A) and human OPN1LW (NM_020061) and OPN1MW (NM_000513) were
591 included as structural references.

592 The retinal chromophore binding pocket was characterized by identifying residues within 6 Å
593 of 11-cis-retinal and its covalently bound lysine. Residues were categorized into three distance

594 shells from the retinal-lysine complex: close (<4 Å), medium (4–5 Å), and far (5–6 Å).
595 Analysis was performed using PyMOL (Schrödinger, LLC) with custom Python scripts.
596 Opsin sequences were aligned to establish positional homology across species. Binding pocket
597 residues identified by distance criteria were mapped to alignment positions, enabling cross-
598 species comparison of equivalent sites. Residues at positions corresponding to the "five-sites
599 rule" for LWS/MWS spectral tuning (positions 180, 197, 277, 285, and 308 in human opsin
600 numbering) were extracted from the binding pocket analysis.
601
602 *In-situ hybridisation*
603 Dual-labelling RNA-FISH was performed on two wholemount retinas (left and right)
604 of one adult *A. phalaena* collected at Lizard Island on 16 March 2021 following standard
605 protocols (e.g., Barthel and Raymond 2000, Raymond and Barthel 2004, Allison et al. 2010,
606 Dalton et al. 2014, Dalton et al. 2015). Retinas were prepared as follows. After one-hour dark
607 adaptation, the eyes were enucleated, vitreous was enzymatically digested, and retinas were
608 extracted from the eyecups. The retinal pigment epithelium layer (RPE) was removed by PBS-
609 jetting using a single use syringe and needle (Terumo). Retinas were flattened and pinned,
610 photoreceptor side down, in Sylgard-filled petri dishes, and fixed in 4% PFA at 4°C overnight.
611 Fixed retinas were triple washed in PBS for 10 min, briefly rinsed in 70% methanol in PBS,
612 and then transferred to 100% methanol and stored at -20°C.
613 Previously extracted retinal mRNA was reverse transcribed using a High Capacity
614 RNA-to-cDNA Reverse Transcription Kit (Applied Biosystems). Riboprobe templates were
615 synthesized from cDNA via standard PCR using opsin specific primers (Table S3) and Q5
616 High Fidelity DNA polymerase (New England Biolabs). Primers were designed to bind to the
617 coding sequence of target opsins (*SWS2B*, *SWS2Aβ*, *LWS1*, *LWS3*) and to contain a T3 or T7
618 RNA polymerase promoter sequence at their 5'-ends (T3, reverse primer; T7, forward primer).
619 Amplicons were isolated via gel-electrophoresis and gel-extraction (Qiagen Gel Extraction
620 Kit), and subsequently used as template in enrichment PCR using the same primers. Anti-sense
621 riboprobes labelled with either digoxigenin-UTP (DIG) or fluorescein-UTP (FL) were
622 synthesized from cDNA templates via T3 RNA polymerase mediated strand-specific RNA
623 transcription using DIG/FL RNA labelling mix (Roche). Hybridised, labelled riboprobes were
624 detected using anti-digoxigenin or anti-fluorescein antibodies conjugated to horseradish
625 peroxidase (Roche). Fluorescent tagging was performed using Alexa Fluor 594 and 488 dyes
626 with streptavidin Tyramide Signal Amplification (TSA, Invitrogen). Finally, retinas were

627 mounted in 70% glycerol in PBS, photoreceptor side up, on microscopy slides and covered
628 with a coverslip.
629 Fluorescent tagged photoreceptor cells were imaged using a CFI Apo Lambda 60x/1.4
630 NA oil immersion objective on a spinning disk confocal microscope (Diskovery, Andor
631 Technologies, UK) built around a Nikon Ti-E body (Nikon Corporation, Japan) equipped with
632 two Zyla 4.2 sCMOS cameras (Andor Technology), and controlled by Nikon NIS Elements
633 software (Nikon Corporation, Japan). Exported images were further processed (max. intensity
634 z-stack projection, channel merging) with ImageJ v. 1.52p (National Institute of Health, USA).
635
636 **Data Availability:** All sequenced transcritpomic data have been submitted to GenBank and
637 are available under accession numbers listed in the Table S1**.**
638
639 **Acknowledgements**
640 We would like to thank Dr. Andrew Mather for help with sample collection. We thank
641 Veronika Truhlářová for lab management and logistical support. FC was supported by an
642 Australian Research Council DECRA Research Fellowship (ARC DE200100620) and by a
643 University of Queensland Development Fellowship, Australia. GMS has been funded by
644 GAUK (1524119; Charles University) and SVV 260-792/2025. ZM was supported by Swiss
645 National Science Foundation (SNF grant PROMYS, 166550), Charles University (Primus),
646 Czech Science Foundation (21-31712S) and ERC Consolidator (SensingDeep).
647
648 **Author contributions**: Z.M. and F.C. designed research and analyzed data; D.E.C. performed
649 the ancestral state reconstruction, C.K. performed the tests of regulatory regions and
650 contributed to the overall interpretations. G.M.S. performed the RNA extractions and
651 transcriptome sequencing. F.H. and X.D. performed the structural protein modelling. M.L.
652 performed the fluorescent in situ hybridization. Z.M. and F.C. wrote the initial version of the
653 manuscript. All authors have contributed to the advanced versions of the manuscript.
654
655 **Competing interests:** The authors declare no competing interest.
656
657 **References:**

658 Adrian-Kalchhauser, I., Blomberg, A., Larsson, T., Musilova, Z., Peart, C. R., Pippel, M., …
659 Winkler, S. (2020). The round goby genome provides insights into mechanisms that may

660 facilitate biological invasions. *BMC Biology*, *18*(1), 1–33. http://doi.org/10.1186/s12915-019-
661 0731-8

662 Baden, T., Angueyra, J. M., Bosten, J. M., Collin, S. P., Conway, B. R., Cortesi, F., ... &
663 Corbo, J. C. (2025). A standardized nomenclature for the rods and cones of the vertebrate
664 retina. *PLoS Biology*, *23*(5), e3003157.
665
666 Borges, R., Khan, I., Johnson, W. E., Gilbert, M. T. P., Zhang, G., Jarvis, E. D., ... &
667 Antunes, A. (2015). Gene loss, adaptive evolution and the co-evolution of plumage
668 coloration genes with opsins in birds. *BMC genomics,* 16(1), 1-14.
669
670 Breinholt, J. W., & Kawahara, A. Y. (2013). Phylotranscriptomics: saturated third codon
671 positions radically influence the estimation of trees based on next-gen data. Genome Biology
672 and Evolution, 5(11), 2082-2092.
673
674 Castiglione, G. M., Schott, R. K., Hauser, F. E., & Chang, B. S. (2018). Convergent selection
675 pressures drive the evolution of rhodopsin kinetics at high altitudes via nonparallel
676 mechanisms. Evolution, 72(1), 170-186.
677
678 Carleton, K.L., Conte, M.A., Malinsky, M., Nandamuri, S.P., Sandkam, B.A., Meier, J.I.,
679 Mwaiko, S., Seehausen, O. and Kocher, T.D., 2020. Movement of transposable elements
680 contributes to cichlid diversity. *Molecular Ecology*, *29*(24), pp.4956-4969.
681
682 Carleton, K.L., Escobar-Camacho, D., Stieb, S.M., Cortesi, F. and Marshall, N.J., 2020.
683 Seeing the rainbow: mechanisms underlying spectral sensitivity in teleost fishes. *Journal of*
684 *Experimental Biology*, *223*(8).
685
686 Carvalho, L. S., Pessoa, D., Mountford, J. K., Davies, W. I., & Hunt, D. M. (2017). The
687 genetic and evolutionary drives behind primate color vision. *Frontiers in Ecology and*
688 *Evolution,* 5, 34.
689
690 Chen J-N, Samadi S, Chen W-J. 2018. Rhodopsin gene evolution in early teleost fishes.
691 *PLOS ONE*. 13(11):e0206918
692
693 Chi, H., Cui, Y., Rossiter, S.J. and Liu, Y., 2020. Convergent spectral shifts to blue-green
694 vision in mammals extends the known sensitivity of vertebrate M/LWS pigments.
695 *Proceedings of the National Academy of Sciences*, *117*(15), pp.8303-8305.
696
697 Cortesi, F., Cheney, K. L., Cooke, G. M., & Ord, T. J. (2018). Opsin gene evolution in
698 amphibious and terrestrial combtooth blennies (Blenniidae). BioRxiv, 503516.
699
700 Cortesi, F., Musilová, Z., Stieb, S.M., Hart, N.S., Siebeck, U.E., Malmstrøm, M., Tørresen,
701 O.K., Jentoft, S., Cheney, K.L., Marshall, N.J. and Carleton, K.L., 2015. Ancestral
702 duplications and highly dynamic opsin gene evolution in percomorph fishes. *Proceedings of*
703 *the National Academy of Sciences*, *112*(5), pp.1493-1498.
704
705 Cortesi, F., Musilová, Z., Stieb, S.M., Hart, N.S., Siebeck, U.E., Cheney, K.L., Salzburger,
706 W. and Marshall, N.J., 2016. From crypsis to mimicry: changes in colour and the
707 configuration of the visual system during ontogenetic habitat transitions in a coral reef fish.
708 *Journal of Experimental Biology*, *219*(16), pp.2545-2558.

Davies, W. I., Tamai, T. K., Zheng, L., Fu, J. K., Rihel, J., Foster, R. G., ... & Hankins, M. W. (2015). An extended family of novel vertebrate photopigments is widely expressed and displays a diversity of function. *Genome research*, 25(11), 1666-1679.

de Busserolles, F., Cortesi, F., Helvik, J.V., Davies, W.I., Templin, R.M., Sullivan, R.K., Michell, C.T., Mountford, J.K., Collin, S.P., Irigoien, X. and Kaartvedt, S., 2017. Pushing the limits of photoreception in twilight conditions: the rod-like cone retina of the deep-sea pearlsides. *Science advances*, *3*(11), p.eaao4709.

Emerson, M. M., Surzenko, N., Goetz, J. J., Trimarchi, J., & Cepko, C. L. (2013). Otx2 and Onecut1 promote the fates of cone photoreceptors and horizontal cells and repress rod photoreceptors. *Developmental cell*, *26*(1), 59-72.

Escobar-Camacho, D., Carleton, K.L., Narain, D.W. and Pierotti, M.E., 2020. Visual pigment evolution in Characiformes: the dynamic interplay of teleost whole-genome duplication, surviving opsins and spectral tuning. *Molecular Ecology*, *29*(12), pp.2234-2253.

Fornes O, Castro-Mondragon JA, Khan A, et al. JASPAR 2020: update of the open-access database of transcription factor binding profiles. Nucleic Acids Res. 2019; doi: 10.1093/nar/gkz1001

Frazer, S. A., & Oakley, T. H. (2025). Accessible and Robust Machine Learning Approaches to Improve the Opsin Genotype-Phenotype Map. *bioRxiv*, 2025-08.

Fricke, R., Eschmeyer, W. N. & Van der Laan, R. (eds) 2021. ESCHMEYER'S CATALOG OF FISHES: GENERA, SPECIES, REFERENCES. (http://researcharchive.calacademy.org/research/ichthyology/catalog/fishcatmain.asp). Electronic version accessed 02/04/2021.

Furukawa, T., Kozak, C. A., & Cepko, C. L. (1997). Rax, a novel paired-type homeobox gene, shows expression in the anterior neural fold and developing retina. *Proceedings of the National Academy of Sciences*, *94*(7), 3088-3093.

Galbraith, J.D., Ludington, A.J., Suh, A., Sanders, K.L. and Adelson, D.L., 2020. New Environment, New Invaders—Repeated Horizontal Transfer of LINEs to Sea Snakes. *Genome Biology and Evolution*, *12*(12), pp.2370-2383.

Ghinia Tegla, M. G., Buenaventura, D. F., Kim, D. Y., Thakurdin, C., Gonzalez, K. C., & Emerson, M. M. (2020). OTX2 represses sister cell fate choices in the developing retina to promote photoreceptor specification. *elife*, *9*, e54279.

Hagen, J. F., Roberts, N. S., & Johnston Jr, R. J. (2023). The evolutionary history and spectral tuning of vertebrate visual opsins. Developmental Biology, 493, 40-66.

Hunt DM, Hankins MW, Collin SP, Marshall NJ. 2014. Evolution of Visual and Non-Visual Pigments. Boston: Springer

Jacobs, G. H. (2013). Losses of functional opsin genes, short-wavelength cone photopigments, and color vision--a significant trend in the evolution of mammalian vision. *Visual Neuroscience*, 30(1-2), 39.

Jerlov NG. 1976. Marine Optics. Amsterdam: Elsevier. 2nd ed.

Katoh, K., Rozewicki, J., & Yamada, K. D. (2019). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in bioinformatics*, *20*(4), 1160-1166.

Katti, C., Stacey-Solis, M., Coronel-Rojas, N. A., & Davies, W. I. L. (2019). The diversity and adaptive evolution of visual photopigments in reptiles. *Frontiers in Ecology and Evolution*, 7, 352.

Kosakovsky Pond, S.L., Posada, D., Gravenor, M.B., Woelk, C.H. and Frost, S.D., 2006. Automated phylogenetic detection of recombination using a genetic algorithm. *Molecular biology and evolution*, *23*(10), pp.1891-1901.

Lamb, T.D., 2013. Evolution of phototransduction, vertebrate photoreceptors and retina. *Progress in retinal and eye research*, *36*, pp.52-119.

Lamb, T.D., 2019. Evolution of the genes mediating phototransduction in rod and cone photoreceptors. *Progress in retinal and eye research*, p.100823.

Lin, J.J., Wang, F.Y., Li, W.H. and Wang, T.Y., 2017. The rises and falls of opsin genes in 59 ray-finned fish genomes and their implications for environmental adaptation. *Scientific reports*, *7*(1), pp.1-13.

Liu, D.W., Wang, F.Y., Lin, J.J., Thompson, A., Lu, Y., Vo, D., Yan, H.Y. and Zakon, H., 2019. The cone opsin repertoire of osteoglossomorph fishes: gene loss in mormyrid electric fish and a long wavelength-sensitive cone opsin that survived 3R. *Molecular biology and evolution*, *36*(3), pp.447-457.

Luk, H. L., Bhattacharyya, N., Montisci, F., Morrow, J. M., Melaccio, F., Wada, A., ... & Olivucci, M. (2016). Modulation of thermal noise and spectral sensitivity in Lake Baikal cottoid fish rhodopsins. Scientific Reports, 6(1), 38425.

Lupše, N., Cortesi, F., Freese, M., Marohn, L., Pohlmann, J. D., Wysujack, K., ... & Musilova, Z. (2021). Visual Gene Expression Reveals a cone-to-rod Developmental Progression in Deep-Sea Fishes. *Molecular Biology and Evolution*, 38(12), 5664-5677.

Lynch, M., & Conery, J. S. (2000). The evolutionary fate and consequences of duplicate genes. *Science*, 290(5494), 1151-1155.

Mackin, R. D., Frey, R. A., Gutierrez, C., Farre, A. A., Kawamura, S., Mitchell, D. M., & Stenkamp, D. L. (2019). Endocrine regulation of multichromatic color vision. Proceedings of the National Academy of Sciences of the United States of America, 116(34), 16882–16891. https://doi.org/10.1073/pnas.1904783116.

Mano H, Kojima D, Fukada Y. 1999. Exo-rhodopsin: a novel rhodopsin expressed in the zebrafish pineal gland. *Mol. Brain Res*. 73(1–2):110–18

Marshall, J., Cronin, T., Johnsen, S., Douglas, R., Hurlbert, A., Boddy, J., & Cortesi, F. (2024). Color in Nature. Princeton University Press. ISBN 978-0-691-25861-4

Morrow, J.M., Lazic, S., Fox, M.D., Kuo, C., Schott, R.K., Gutierrez, E.D.A., Santini, F., Tropepe, V. and Chang, B.S., 2017. A second visual rhodopsin gene, rh1-2, is expressed in zebrafish photoreceptors and found in other ray-finned fishes. *Journal of Experimental Biology*, *220*(2), pp.294-303.

Musilova, Z., Cortesi, F., Matschiner, M., Davies, W.I., Patel, J.S., Stieb, S.M., de Busserolles, F., Malmstrøm, M., Tørresen, O.K., Brown, C.J. and Mountford, J.K., 2019. Vision using multiple distinct rod opsins in deep-sea fishes. *Science*, *364*(6440), pp.588-592.

Musilova, Z., & Cortesi, F. (2023). The evolution of the green-light-sensitive visual opsin genes (RH2) in teleost fishes. Vision Research, 206, 108204.

Musilova Z., Salzburger, W., Cortesi, F. (2021). The Visual Opsin Gene Repertoires of Teleost Fishes: Evolution, Ecology and Function. Annual Review of Cell and Developmental Biology.

Nelson, S. M., Park, L., & Stenkamp, D. L. (2009). Retinal homeobox 1 is required for retinal neurogenesis and photoreceptor differentiation in embryonic zebrafish. *Developmental biology*, *328*(1), 24-39.

O'Quin, K. E., Smith, D., Naseer, Z., Schulte, J., Engel, S. D., Loh, Y. H. E., ... & Carleton, K. L. (2011). Divergence in cis-regulatory sequences surrounding the opsin gene arrays of African cichlid fishes. *BMC evolutionary biology*, *11*(1), 120.

Paradis, E., & Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, *35*(3), 526-528.

Passaro, S., Corso, G., Wohlwend, J., Reveiz, M., Thaler, S., Somnath, V. R., ... & Barzilay, R. (2025). Boltz-2: Towards accurate and efficient binding affinity prediction. bioRxiv 2025. Doi: 10.1101/2025.06.14.659707

Phillips, G. A., Carleton, K. L., & Marshall, N. J. (2016). Multiple genetic mechanisms contribute to visual sensitivity variation in the Labridae. *Molecular biology and evolution*, 33(1), 201-215.

Pierotti, M.E., Wandycz, A., Wandycz, P., Rebelein, A., Corredor, V.H., Tashiro, J.H., Castillo, A., Wcislo, W.T., McMillan, W.O. and Loew, E.R., 2020. Aggressive mimicry in a coral reef fish: The prey's view. *Ecology and Evolution*, *10*(23), pp.12990-13010.

Policarpo, M., Fogg, L. G., Cortesi, F., & Salzburger, W. (2025). Evolution of the non-visual and visual opsin gene repertoire in ray-finned fishes. *Genome Biology and Evolution*, evaf129.

855 Rabosky, D.L., Chang, J., Title, P.O., Cowman, P.F., Sallan, L., Friedman, M., Kaschner, K.,
856 Garilao, C., Near, T.J., Coll, M. and Alfaro, M.E., 2018. An inverse latitudinal gradient in
857 speciation rate for marine fishes. *Nature*, *559*(7714), pp.392-395.
858
859 Ravi, V. and Venkatesh, B., 2018. The divergent genomes of teleosts. *Annual review of*
860 *animal biosciences*, *6*, pp.47-68.
861
862 Rhie, A., McCarthy, S. A., Fedrigo, O., Damas, J., Formenti, G., Koren, S., ... & Jarvis, E. D.
863 (2021). Towards complete and error-free genome assemblies of all vertebrate
864 species. *Nature*, *592*(7856), 737-746.
865
866 Revell, L. J. (2012). phytools: an R package for phylogenetic comparative biology (and other
867 things). *Methods in ecology and evolution*, (2), 217-223.
868
869 Ronquist, F. and J. P. Huelsenbeck. 2003. MRBAYES 3: Bayesian phylogenetic inference
870 under mixed models. *Bioinformatics* 19:1572-1574.
871
872 Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J.C., Guirao-Rico, S., Librado, P., Ramos-
873 Onsins, S.E. and Sánchez-Gracia, A., 2017. DnaSP 6: DNA sequence polymorphism analysis
874 of large data sets. *Molecular biology and evolution*, *34*(12), pp.3299-3302.
875
876 Sandkam, B.A., Joy, J.B., Watson, C.T. and Breden, F., 2017. Genomic environment impacts
877 color vision evolution in a family with visually based sexual selection. *Genome biology and*
878 *evolution*, *9*(11), pp.3100-3107.
879
880 Sandkam, B., Dalton, B., Breden, F. and Carleton, K., 2018. Reviewing guppy color vision:
881 Integrating the molecular and physiological variation in visual tuning of a classic system for
882 sensory drive. *Current zoology*, *64*(4), pp.535-545.
883
884 Seehausen, O., Terai, Y., Magalhaes, I.S., Carleton, K.L., Mrosso, H.D., Miyagi, R., Van Der
885 Sluijs, I., Schneider, M.V., Maan, M.E., Tachida, H. and Imai, H., 2008. Speciation through
886 sensory drive in cichlid fish. *Nature*, *455*(7213), pp.620-626.
887
888 Sekharan, S., Katayama, K., Kandori, H., & Morokuma, K. (2012). Color vision:"OH-site"
889 rule for seeing red and green. Journal of the American Chemical Society, 134(25), 10706-
890 10712.
891
892 Smallwood, P. M., Wang, Y., & Nathans, J. (2002). Role of a locus control region in the
893 mutually exclusive expression of human red and green cone pigment genes. *Proceedings of the*
894 *National Academy of Sciences*, *99*(2), 1008-1011.
895
896 Tav, C., Tempel, S., Poligny, L., & Tahi, F. (2016). miRNAFold: a web server for fast miRNA
897 precursor prediction in genomes. *Nucleic acids research*, *44*(W1), W181-W184.
898
899 Tettamanti, V., de Busserolles, F., Lecchini, D., Marshall, N.J. and Cortesi, F., 2019. Visual
900 system development of the spotted unicornfish, Naso brevirostris (Acanthuridae). *Journal of*
901 *Experimental Biology*, *222*(24).
902

903 Volkov, L. I., Kim-Han, J. S., Saunders, L. M., Poria, D., Hughes, A. E., Kefalov, V. J., ... &
904 Corbo, J. C. (2020). Thyroid hormone receptors mediate two distinct mechanisms of long-
905 wavelength vision. *Proceedings of the National Academy of Sciences*, *117*(26), 15262-15269.
906
907 Wakefield, M. J., Anderson, M., Chang, E., Wei, K. J., Kaul, R., Graves, J. A. M., ... & Deeb,
908 S. S. (2008). Cone visual pigments of monotremes: filling the phylogenetic gap. *Visual*
909 *Neuroscience*, *25*(3), 257-264.
910
911 Wald, G., 1968. The molecular basis of visual excitation. *Nature*, *219*(5156), pp.800-807.
912
913 Walsh, B. (2003). Population-genetic models of the fates of duplicate genes. In *Origin and*
914 *Evolution of New Gene Functions (*pp. 279-294). Springer, Dordrecht.
915
916 Ward MN, Churcher AM, Dick KJ, Laver CR, Owens GL, et al. 2008. The molecular basis of
917 color vision in colorful fish: Four Long Wave-Sensitive (LWS) opsins in guppies (Poecilia
918 reticulata) are defined by amino acid substitutions at key functional sites. *BMC Evol. Biol*.
919 8(1):210
920
921 Weaver, S., Shank, S.D., Spielman, S.J., Li, M., Muse, S.V. and Kosakovsky Pond, S.L.,
922 2018. Datamonkey 2.0: a modern web application for characterizing selective and other
923 evolutionary processes. *Molecular biology and evolution*, *35*(3), pp.773-777.
924
925 Windsor, D. J., & Owens, G. L. (2009). The opsin repertoire of Jenynsia onca: a new
926 perspective on gene duplication and divergence in livebearers. BMC Research Notes, 2(1),
927 159.
928
929 Yokoyama, S., 2008. Evolution of dim-light and color vision pigments. *Annu. Rev. Genomics*
930 *Hum. Genet.*, *9*, pp.259-282.
931
932 Yokoyama, S. and Radlwimmer, F.B., 1998. The" five-sites" rule and the evolution of red
933 and green color vision in mammals. *Molecular biology and evolution*, *15*(5), pp.560-567.
934
935 You, X., Bian, C., Zan, Q., Xu, X., Liu, X., Chen, J., Wang, J., Qiu, Y., Li, W., Zhang, X.
936 and Sun, Y., 2014. Mudskipper genomes provide insights into the terrestrial adaptation of
937 amphibious fishes. *Nature communications*, *5*(1), pp.1-8.

Supplementary Material for:

Turning red to green repeatedly: ancient LWS duplications drive convergent visual tuning in teleost fishes.

Fabio Cortesi, Daniel Escobar Camacho, Gina Maria Sommer, Martin Luehrmann, Flurin Hidber, Xavier Deupi, Karen Carleton, Zuzana Musilova

Table S1. Samples and accession numbers used in this study.

(table provided in a separate file)

Table S2: *The LWS regulatory elements found in the LWS1 genes but not much in the LWS2 and LWS3 upstream regions of the same species of cavefish (A. mexicanus), elephant fish (P. kingsleyae), tarpon (M. cyprinoides) and goby (N. melanostomus). The miRNA folding energy determined with https://evryrna.ibisc.univ-evry.fr/miRNAFold. More negative energy means more stable and so more likely functional miRNA. The LCR binding sites (OTX2, THR or RAX) identified with Jaspar. Grey marks the non-tested combinations (gene not present). For more details on miRNA detection, please check Figure S5:*

	LWS1		LWS2		LWS2.II		LWS3	
	miR726	LCR	miR726	LCR	miR726	LCR	miR726	LCR
Astyanax	Yes E=-69 kcal/mol	Yes 2OTX 2THR			No	No		
Paramormyrops	Yes E=-67	Yes 2OTX, 1THR, 1RAX	No	No				
Megalops	Yes E=-61.3	Yes 2OTX, 1THR	No	No	Yes E=-32 kcal/mol	No		
Neogobius	Yes E=-59.2	Yes 2OTX					Yes E=-24 kcal/mol	No

Table S3. Primers used for probe template synthesis in *Amblygobius phalaena*. T7 (forward primer) and T3 (reverse primer) RNA polymerase promoter sequences are shown in small letters.

Target gene Probe length (bp) Probe similarity (gene, %)	Primer	Sequence
SWS2B 547 *SWS2Aβ*, 75.4%	SWS2B_F1 SWS2B_R2	5'-TAATACGACTCACTATAGGGACAGGCCTTTTCTACTCTTTGTCT-3' 5'-AATTAACCCTCACTAAAGGGGGACTCGTTGTTGTACTTGTTGTT-3'
SWS2Aβ 740 *SWS2B*, 75.4%	SWS2Aβ_F1 SWS2Aβ_R1	5'-TAATACGACTCACTATAGGGTGGGAGGTTTTATCAACGCTCTTA-3' 5'-AATTAACCCTCACTAAAGGGAAAGTTTCTCCACGGTTATTGACC-3'
LWS1 740 *LWS3*, 81.7%	LWS1_F1 LWS1_R1	5'-TAATACGACTCACTATAGGGGCTCTGGATGTGTTTTGTGGTTAT-3' 5'- AATTAACCCTCACTAAAGGGAAAGCAGGCAAAGAATGTGTAAGG-3'
LWS3 690 *LWS1*, 81.7%	LWS3_F2 LWS3_R2	5'-TAATACGACTCACTATAGGGATTTTCGAGGGCTACACTGTCTC-3' 5'-AATTAACCCTCACTAAAGGGTCGTCCACCTCTTTTCCAAACAG-3'

Figure S1 - Phylogenetic gene tree and key tuning amino acids of LWS. LWS2 and LWS3 genes are highlighted in orange and brown, respectively.

(figure provided as a separate pdf file)

Figure S2: Vista plot of the elephant fish (*P. kingsleyae*) and a cavefish (*A. mexicanus*) *LWS2* regions compared to two tarpon (*M. cyprinoides*) *LWS2* regions located on different chromosomes.

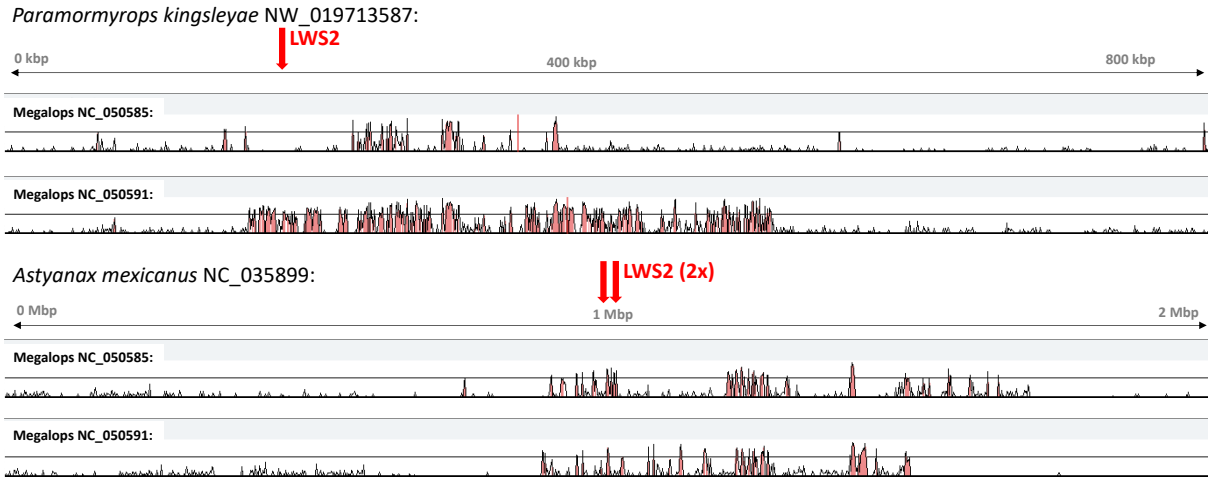


Figure S3 - Synteny of the *LWS1*, *LWS2* and *LWS3* genes and the flanking regions in the round goby (*N. melanostomus*), elephant fish (*P. kingsleyae*), blind cavefish (*A. mexicanus*) and tarpon (*Megalops cyprinoides*).

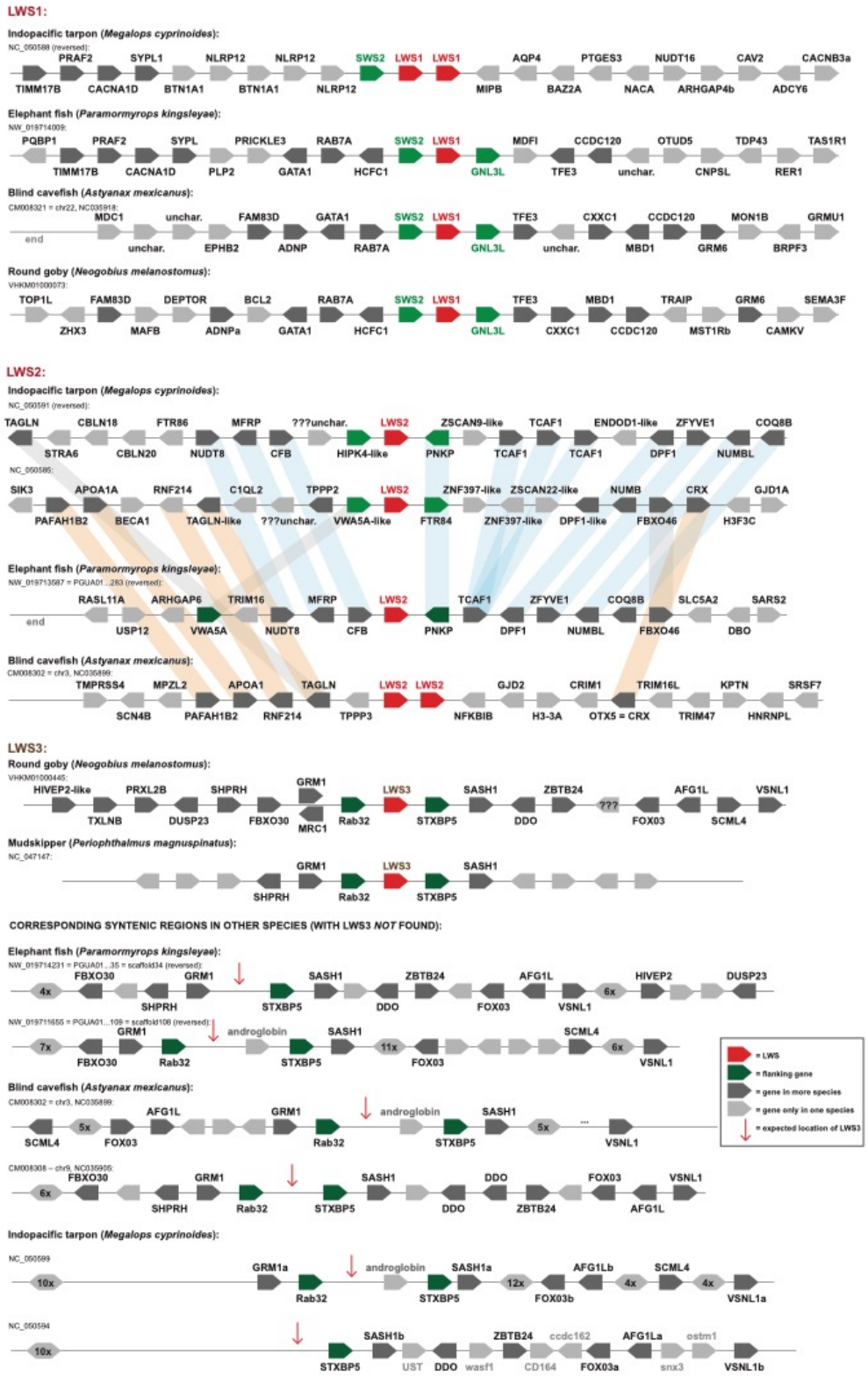


Figure S4: Single-exon phylogenies calcuated in MrBayes. Orange highlights the *lws2* of osteoglossiform, characin and elopomorph fishes, blue are goby *lws3* opsins. Note that in exon 3, *Megalops* LWS1 and LWS2 are similar, most likely due to gene conversion. In exon 4, the LWS3 and LWS2 form a grouping, althought not correctly corresponding to the species tree.

	exon2	exon3	exon4	exon5
nuc:	65-361	362-530	531-696	697-936
AA:	23-120	122-176	178-232	233-312

Figure S5: Phylogenetic analysis without the third codon position shows unresolved topology that involves percomorph fishes mixed with gobies.

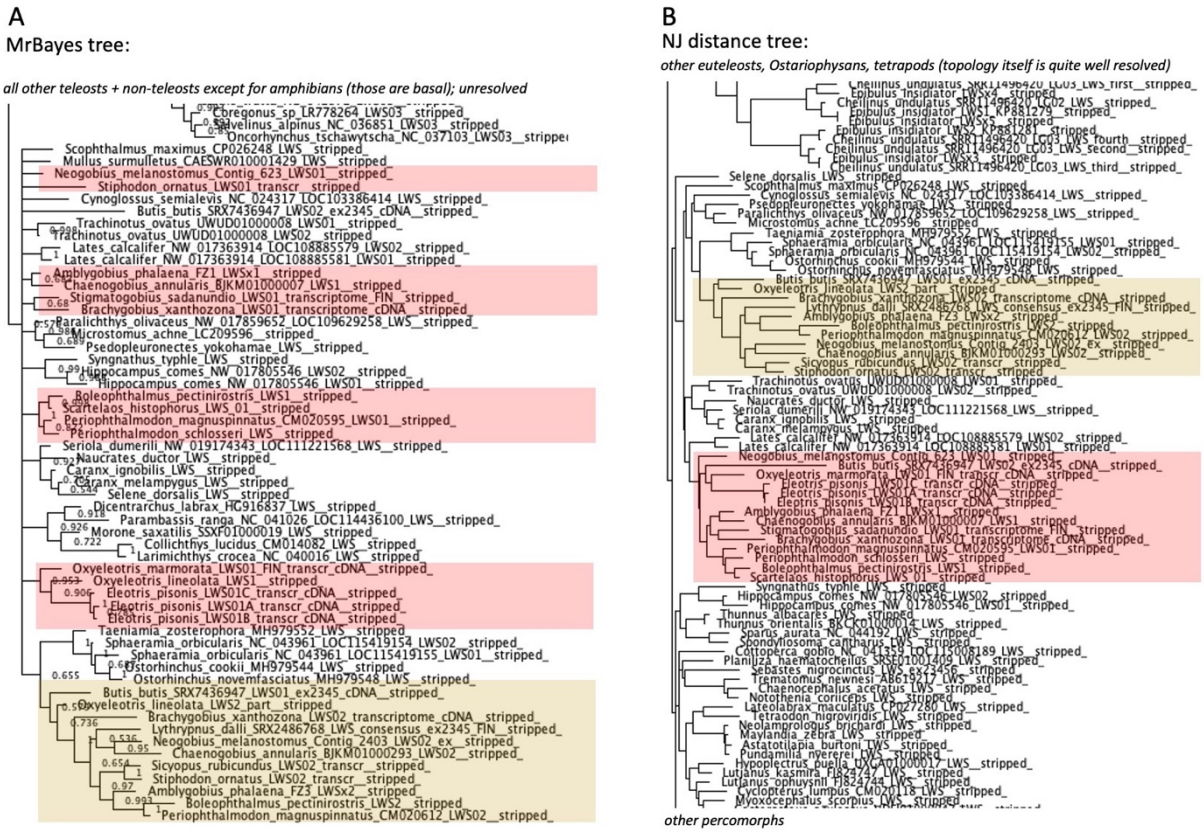


Figure S6: Gene conversion among the two *LWS* copies (*LWS1* and *LWS3*) of the goby, *Tateurndina ocellicauda*. Large-scale more ancient gene conversion impacted exons 2-3 and exons 5-6. These two sequences compromised the phylogenetic signal within the overall gene tree and had to be removed from the final analysis. A) Identity plot of the two nucleotide sequences. Green bars mark identical stretches of the coding region, white "gaps" indicate differences in sequence. Yellow boxes highlight the putative regions of gene conversion i.e., regions of high sequence similarity. B) GARD results for the comparison of all goby *LWS1* and *LWS3* sequences. The most likely breakpoints for *T. ocellicauda* are highlighted in yellow. C) Drops in the dS ratio along the gene correspond to the detected converted regions. D) A phylogenetic tree based on the longest converted region of the coding region i.e., exon5-6 (position 526 – 1074 of the genes) shows that the two *T. ocellicauda* sequences cluster together.

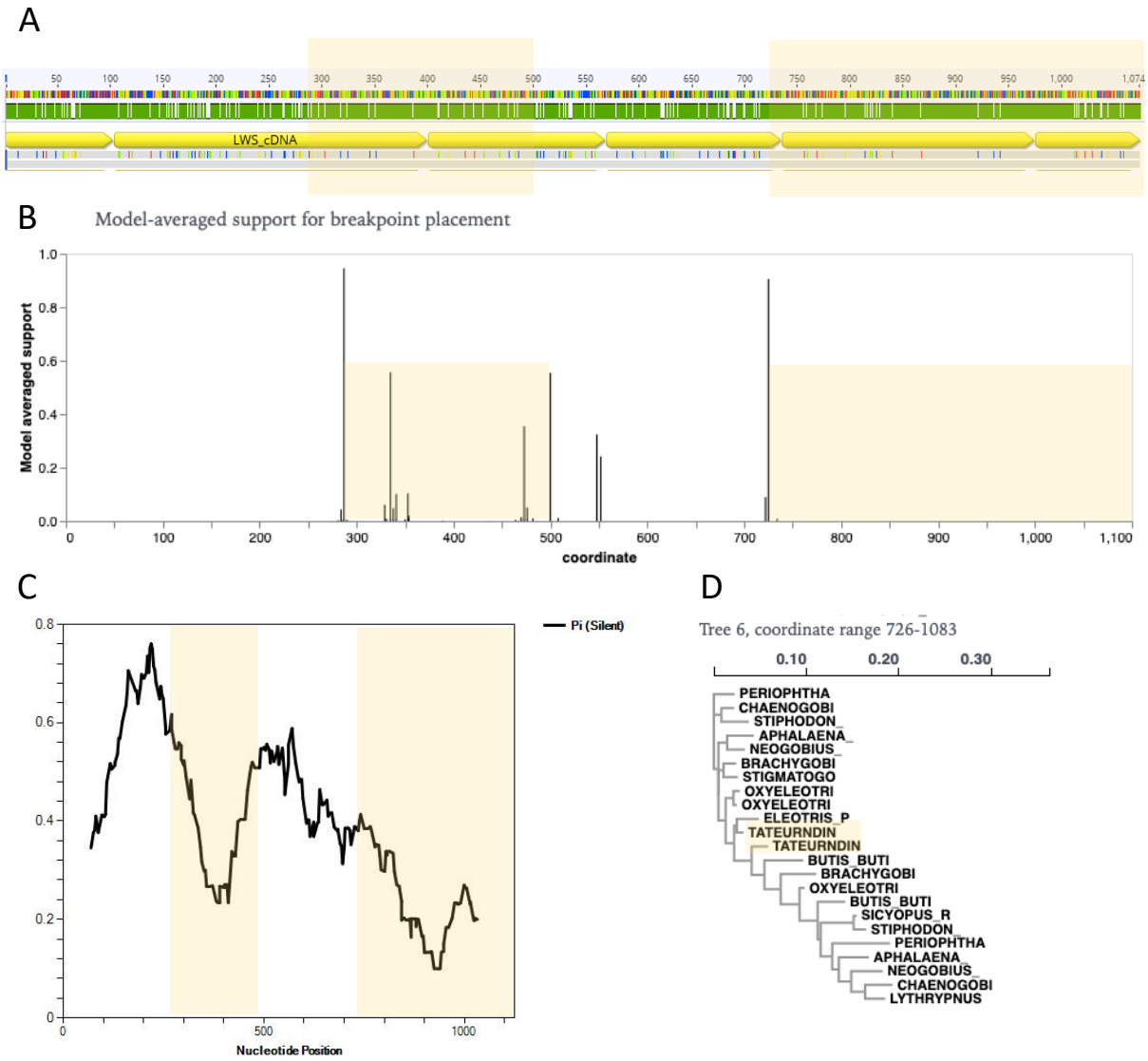


Figure S7: The miRNA detected in the upstream region of *LWS1* (all species), *LWS2* (*Megalops*, one copy) and *LWS3* (*Neogobius*).

LWS1:

Zebrafish -61.5 kcal/mole

```
>Zfish_LWS miR726 region
TAAACCGTGCTACTATGTTGAGCTGTCGTTGAGAATTCAGTTGTATATCCGAGTTCTGCTAGTAGTGAACTTTTGCCAATCACGAATAGTTCAGAACTAGCGGAATTCCAGGCATACGACTGACTGAGTGTTGAC

11 152 [-61.5] ((.(((.(((((.(((((.......(.(((((((((.((.(((((((((((((..((((((.((((.....)))).)).))))))..))))))))))))))..)).))))))
```

Paramormyrops -67

```
>Paramormyrops_LWS1 miR726 region
GGGCGCACATGGTCCGGCGGGGGCGCCCCATCAGCACCGGCCGCCGACGTGAGCCGGGTCCAAATCCAAGTTCTGCTAGTAGTGAACTCCTGACCAGCGGCGCCCAGTTCAAAGCTAGCGGAATTCCGGACGTGC

22 163 [-67] (((.......(((..(((((.....((..((..(((((((.((..(((.(((((((((((((..((((((.((((.....)))).....)))))))..))))))))))))..)))..)).)))))))..)
```

Astyanax -69.2

```
>Astyanax_LWS1 MiR726 region
GAAGATCAGAAAGAGTCATGAGCCTGAGCTTCCGCTGGCATTCAGGTGTAAATCCGAGTTCTGCTAGTAGTGAACTTTTACCAATCACGAATAGTTCAGAACTAGCGGAATTCCAGACTTACACCTGATAAAGAG

12 160 [-69.2] ((...((.(((((.(.((((..(((.....((((((((((.((.((((((((((((((..((((((.((.(((.........)).)))))))..))))))))))))))..)).))))))))))....
```

Megalops -61.3

```
>Megalops_LWS1 miR726 region
CTGACCCAGGCACATTCTCTCGCCCCTAGCTGGATTTCTCGTTGACAGCCGGGTGCAAGTCCGAGTTCTGCTAGTAGTGAACTTTTGCCAATCATGAATAGTTCAGAACTAGCGGAATTCCAGACGTACCCCCGA

16 161 [-61.3] (((.((((..(.((((..(((.(((.........((((((...((.((((((((((((((..((((((.((.(((.........)).)))))))..))))))))))))))..)).)))))))...
```

Neogobius -59.2

```
>Neogobius_LWS1 miR726 region
AACCGGACCTCCTGCTGGTGGCTTCGCTGGCAGTCGGG-GCAAATCCAAGTTCTGCTAGTAGTGAACTTTCAACAAACACAAGTAGCTCAGAACTAGCGGAATTCCAGATTTACCCTCGAGACGTCGGGGA

3 148 [-59.2] ((((.(.......(((((.(((.(((((..((((((.(((((..((((((((((((..(((.((.(.............).)).)))..))))))))))))...))))))..)))))))
```

LWS2:

Megalops -32.7

```
>Megalops_LWS2.II miR726 region      not great free energy
GAATGAACTAAAAAGAAGATGAGCCATCCTTTGCCGTCACTCCTTCTTGGAATGCTGTTACACAGGTCTGAACTGTTTCTAGCAGGATACTGGACCATTCTTGAAGAAGGGAAAGCCTCCAGTTCTTTGAG

46 157 [-32.7] ((((..((..(((((((...(((((.......(((((...(((..(((((((...(((((.....))))))).....)))))))..)))..))))))....)))))..)))))..))..))))
```

LWS3:

Neogobius -24

```
>Neobogius_LWS3   not good folding
CAAACTCCACTTACAGTTTAGGTGTGATTAAAATGATTATGTCTCAATTGAACTGGCAGATAGTTATTTCTAGGTTACTTTTTGCAGTATCACTTCAAGTAGAACTCAGACCAAACCCTAACTCCTGGAAA

16 128 [-24] ((((((.((...........((..((((.(.((((((.(.((((((.(((.((.....))).)))).))))))..).....))))))..).....))))..)).....)).))))))
```

Figure S8: Structural models of LWS opsins in teleost fishes and in human

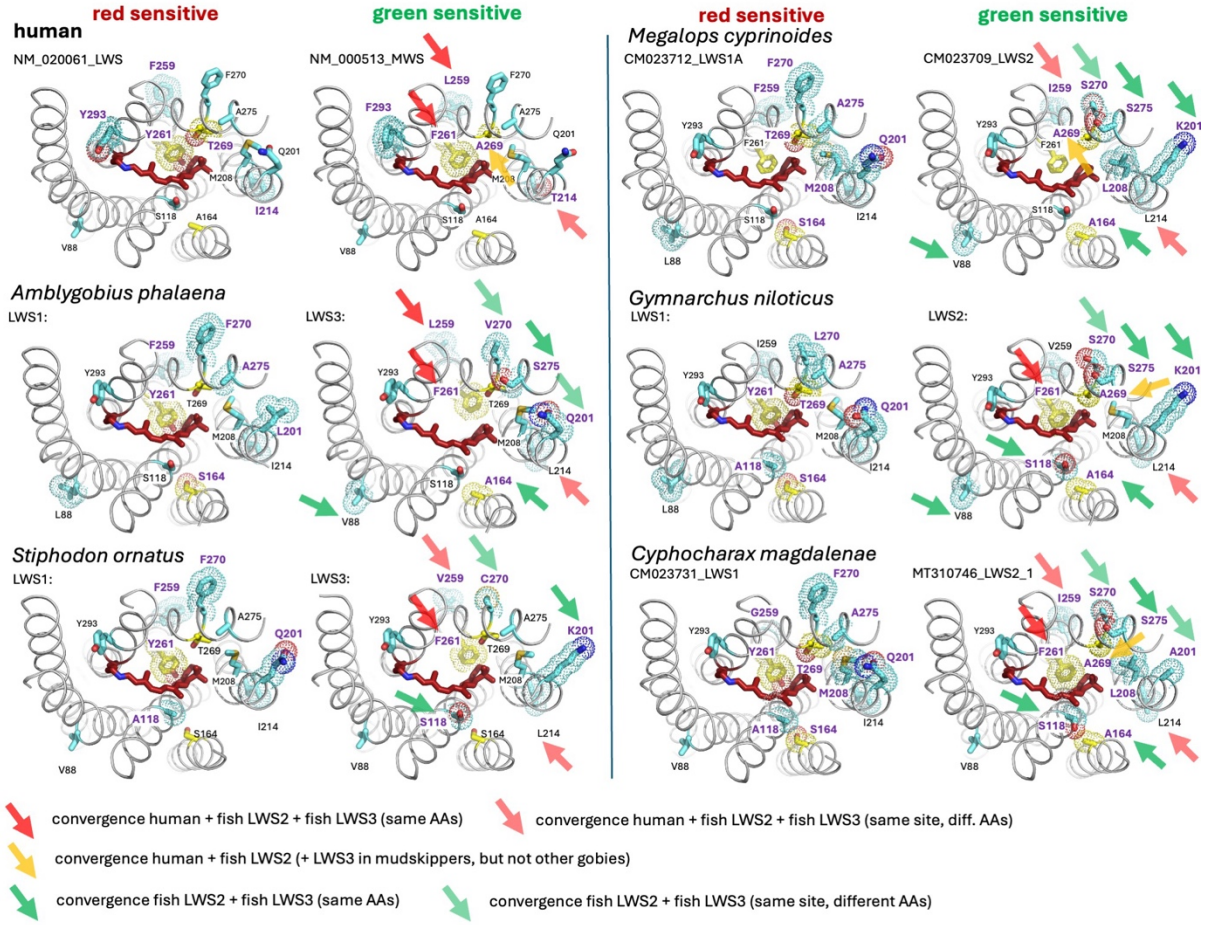
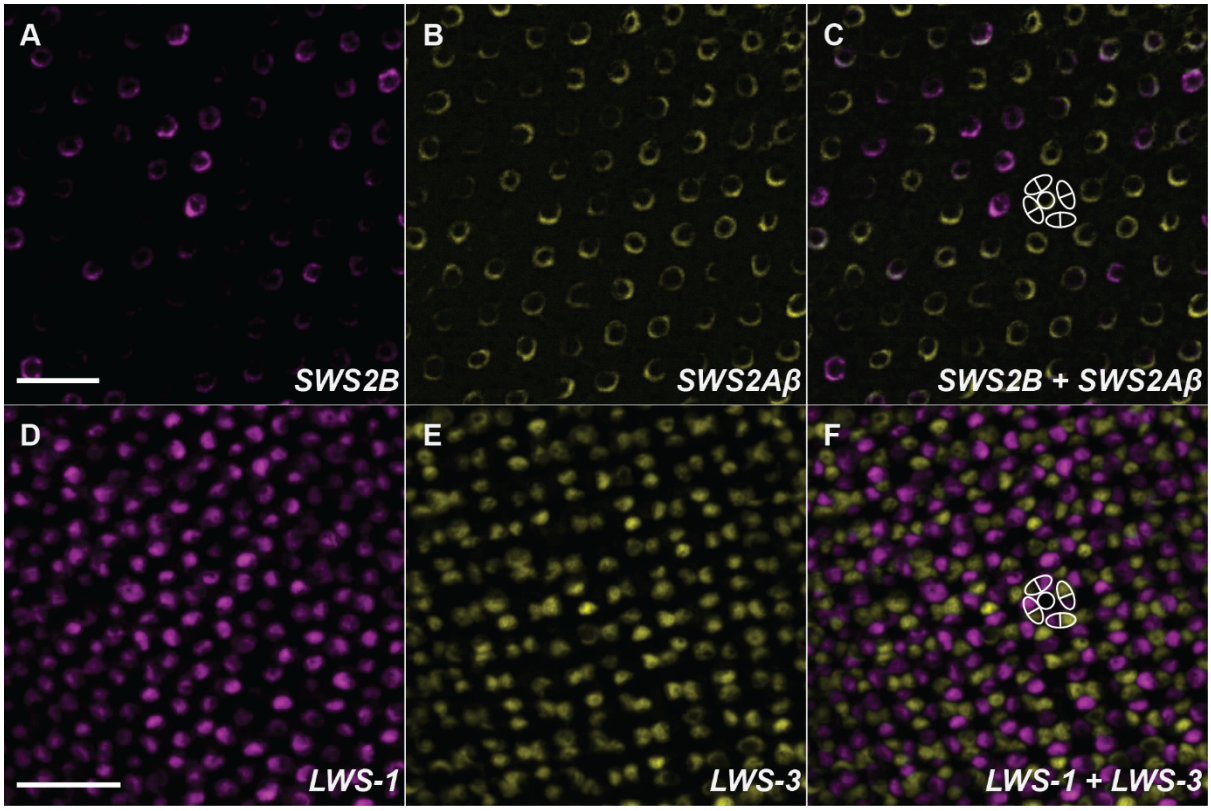



Figure S9: Double-labelling in-situ hybridisation of opsin mRNAs in retinal single cone (A-C) and double cone (D-F) photoreceptors of adult whitebarred goby, *Amblygobius phalaena*. **(A-C)** *SWS2B* (magenta) and *SWS2Aβ* (yellow) mRNAs were co-expressed in some single cones. **(D-F)** *LWS1* (magenta) and *LWS3* (yellow) mRNAs were expressed in opposite members of double cones. Scale bars: 10 µm.