Multiple ancestral duplications of the red-sensitive opsin gene (LWS) in teleost fishes and convergent spectral shifts to green vision in gobies

Fabio Cortesi[1], Daniel Escobar Camacho[2], Martin Luehrmann[1], Gina Maria Sommer[3], Zuzana Musilova[3]

[1] Queensland Brain Institute, The University of Queensland, Brisbane, QLD 4072, Australia
[2] Department of Biology, University of Maryland, College Park, MD 20742, USA
[3] Department of Zoology, Faculty of Science, Charles University, Vinicna 7, 12844 Prague, Czech Republic

Abstract

Photopigments, consisting of an opsin protein bound to a light-sensitive chromophore, are at the centre of vertebrate vision. The vertebrate ancestor already possessed four cone opsin classes involved in colour perception during bright-light conditions, which are sensitive from the ultraviolet to the red-wavelengths of light. Teleosts experienced an extra round of whole genome duplication (3R) at their origin, and while most teleosts maintained only one long-wavelength-sensitive opsin gene *(LWS1*), the second ancestral copy (*LWS2*) persisted in characins and osteoglossomorphs. Following 3R, teleost opsins have continued to expand and diversify, which is thought to be a consequence of the different light environment fishes inhabit, from clear streams to the relative darkness of the deep-sea. Although many recent and a few ancestral opsin duplicates can be found, none predating the 3R were thought to exist. In this study we report on a second, previously unnoticed ancestral duplication of the red-sensitive opsin (*LWS3*), which predates the teleost-specific genome duplication and only persists in gobiid fishes. This is surprising, since it implies that *LWS3* has been lost at least 19-20 times independently along the teleost phylogeny. Mining 109 teleost genomes we also uncover a third lineage, the elopomorphs, that maintained the *LWS2* copy. We identify convergent amino acid changes that green-shift ancestral and recent *LWS* copies, leading to adaptive differentiation and the functional replacement of the original green-sensitive *RH2* opsin. Retinal transcriptomes and in-situ hybridisation show that *LWS3* is expressed to various extents in gobies and in the case of the whitebarred goby, *Amblygobius phalaena*, it occurs in a separate photoreceptor to *LWS1*. Our study highlights the importance of comparative studies to comprehend evolution of gene function.

Introduction

Animals rely on vision for a variety of fundamental tasks such as foraging, mating, to avoid predators, and to navigate through 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). They are the dim-light active rod opsin (rhodopsin; RH1) which typically operates in the blue-green light spectrum [peak spectral sensitivity (λ_{max}) ~ 445 to 525nm], and four differently tuned cone opsins: short-wavelength-sensitive 1 and 2 (SWS1: λ_{max} ~ 345 to 440 nm; SWS2: λ_{max} ~ 395 to 490 nm), rhodopsin-like 2 (RH2: λ_{max} ~ 450 to 540 nm), and mid/long-wavelength-sensitive (MWS/LWS: λ_{max} ~ 490 to 575 nm) (Yokoyama 2008; Carleton et al. 2020). The different cone-based photopigments are active during bright-light conditions where they are often involved in colour vision (Hunt et al. 2014).

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 (reviewed in Carleton et al. 2020 and Musilova et al., 2021) – with nearly 35,000 species, the most species-rich clade of vertebrates (Fricke et al., 2021). Teleosts are extremely diverse in ecology, physiology and morphology as they have conquered almost every aquatic habitat (few species have even transitioned to land) from clear mountain streams to the relative darkness of the deep-sea. 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 (reviewed in Musilova et al., 2021) 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 of the retinal binding pocket, that shift the spectral sensitivity of the photopigment (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), or location in the retina or other tissues (Davies et al. 2015; Morrow et al. 2016). Teleost opsins are also 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 (Carleton et al. 2020). Several older tandem duplications have also persisted, such as 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 Vankatesh 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* copy (*LWS1*) which is sensitive from the yellow to red spectrum (Carleton et al. 2020), or they might have lost the gene altogether (Musilova et al. 2019). LWS2 in extant species, but likely already in the ancestor shortly after its duplication, has shifted its spectral sensitivity to shorter, green-wavelengths as compared to LWS1 (Escobar-Camacho et al. 2020). This is similar to the green shifted 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 have severely reduced the expression of the green-sensitive *RH2*, suggesting that there is a trade-off between the overlap of opsin 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 found in some more recent LWS1 duplicates but also 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 such a limited number of these genes. It is also intriguing, considering that the predicted duplication rate of genes is 100 My (Lynch and Conery, 2000), that no opsin gene duplicates from the time between 2R and 3R (~ 200 My) have been found so far.

In this study, we explored the evolutionary history of the *LWS* opsins in teleost fishes from a phylogenetic representative sample of more than 100 species. Combining next generation genome and transcriptome sequencing with open-source data mining, we uncover a rich history of gene duplication, loss and gene conversion. This includes several recent *LWS1* duplications and an extra order, which has retained the *LWS2* copy. Surprisingly, we also discovered another major *LWS* duplication that predates the teleost-specific genome duplication and has only been retained in gobiid fishes. In a remarkable example of convergent evolution, this ancestral *LWS3* copy has also shifted its spectral sensitivity to shorter wavelengths. Like for the species that have retained *LWS2*, retinal transcriptomes in various goby species revealed a reduction or a loss in expression of the green sensitive *RH2* opsin. Using in-situ hybridisation in a species that expresses *LWS1* and *LWS3*, the whitebarred goby, *Amblygobius phalaena*, we show that each gene is expressed in a separate photoreceptor. *A. phalaena* also expresses two *SWS2* opsins located in a third photoreceptor, thus this goby is likely to have a trichromatic visual system sensitive to blue, green and red wavelengths.

Results and Discussion

LWS duplication and gene synteny

Reconstructing a comprehensive *LWS* phylogeny based on 109 teleost genomes and including non-teleost ray-finned (Actinopterygian), lobe-finned (Sarcopterygii), and several ancestrally derived fishes as outgroups, we first uncover an ancestral *LWS3* duplication in gobies (Gobiiformes), which seems to predate the teleost-specific genome duplication (Fig. 1, and Fig. S1). We also find that Elopiformes (tarpons, ladyfish and relatives) in addition to the bony tongues and characins have retained the *LWS2* copy, which emerged during 3R (Liu et al. 2019; Escobar-Camacho et al. 2020). Since *LWS* in non-teleost ray-finned fishes (bichir, sturgeon, gar and bowfin) is ancestral to the *LWS1-3* paralogs found in modern teleosts, it is likely that all three duplicates originated in the teleost ancestor (Fig. 2 and Fig. S1).

Gobies are part of the crown euteleosts and diverged from their closest sister lineage, cardinalfishes and nurseryfishes (both Kurtiformes), around 100 Mya (Rabosky et al. 2018). Considering the most direct phylogenetic route from the teleost ancestor to the goby ancestor and ignoring any possibly extinct lineages, *LWS3* must have been lost at least 19-20 times independently in the teleost radiation; a truly amazing feat considering how numerous and

diverse teleost fishes are (Fig. 1). An alternative explanation for how gobies might have acquired *LWS3* is through horizontal gene transfer. While arguably unlikely, horizontal gene transfer has been shown to contribute to genetic diversity in sea-snakes which have several transposable elements that are most like the ones found in teleosts and corals (Galbraith et al. 2020). However, based on the vertebrate genomes we surveyed; it is currently not possible to discern a potential hypothetical donor for *LWS3*.

Another astounding finding from our phylogenetic reconstruction is that the *LWS* genes of some tetrapods (amphibians, snakes, lizards, mammals) and the non-tetrapod lungfish form a cluster within the teleosts, while the *LWS* genes in other tetrapods (birds, turtles) follow the species tree with a topology outside the ray-finned fishes (Fig. S1). This pattern is puzzling, but it does indicate that there has been even more ancestral *LWS* duplications, possibly in the common ancestor of ray-finned and lobe-finned fishes. More exhaustive sampling of the tetrapod lineage is needed to fully resolve the *LWS* evolutionary history in this case.

A closer inspection of the syntenic region (10 genes upstream and downstream) surrounding the different teleost *LWS* opsins, revealed a highly conserved structure for *LWS1*, which is located between *HCFC1* or *SWS2* upstream and *GNL3L* downstream. The synteny for *LWS2* is less conserved as it is located on different chromosomal stretches in bony tongues, and characins. Lastly, *LWS3* in gobies is located between *RAB32* upstream and *STXBP5* or androglobin downstream, a different chromosomal region compared to *LWS1*. In other teleosts including the cardinalfishes, this region contains no traces of an opsin gene (Supp Fig S2). It is possible that both *LWS2* and *LWS3* have relocated over evolutionary time or that *LWS3* emerged from transposition in the first place. Indeed, a recent study in African cichlids has revealed that transposable elements are a key factor in shaping the evolutionary dynamics of opsin genes in fishes (Carleton et al. 2020). This is also highlighted by the fact that both, the teleost *RH1* and one of the *LWS1* copies in guppies and other livebearers (Atherinomorpha), originated from retrotransposition (Mano et al. 1999, Sandkam et al. 2017, Ward et al. 2008). Alternatively, the synteny might no longer be discernible as the duplicated genes in the original syntenic region were lost at different rates between fish lineages or the chromosomes were rearranged.

LWS evolutionary history

Overall, we found that most shallow marine and freshwater fishes have kept at least one copy of the longer wavelength sensitive *LWS1* gene. On the contrary, many deeper-living species, including members of the anglerfishes (Lophiiformes), cusk-eels (Ophidiiformes) and toad

fishes (Batrachoidiformes), have lost *LWS* altogether. Similarly, the deep-water Sanzo's goby, *Lesueurigobius sanzi*, has lost both *LWS1* and *LWS3* copies (Fig. 1 and Fig. 3). 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). 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 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 seems to proliferate in species that show elaborate yellow, orange and red body patterns such as wrasses (Labriformes) and livebearers (Phillips et al., 2016). In the most extreme cases, brown trout, *Salmo trutta* (Salmoniformes), 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). Tandem 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*). In the cases where *LWS2* has duplicated, it always co-occurs with *LWS1* (Fig. 1; Escobar Camacho et al. 2020). However, this is not the case for the only *LWS3* duplication we found, in which case *LWS1* is missing from the genome of the bluebanded goby, *Lythrypnus dalli* (Fig. 4).

Gene conversion between *LWS1* paralogs but also between *LWS1* and *LWS2* (Escobar-Camacho et al. 2020) is quite common (e.g., Sandkam et al. 2017; Escobar-Camacho et al. 2020). In characins, gene conversion seems to affect mostly the first couple of exons, while the regions that carry the key tuning sites on exons 3 – 5, remain unaffected (Escobar-Camacho et al. 2020). This is not the case in poeciliids, where gene conversion between *LWS1-1* and *LWS1-3* seems ubiquitous, albeit in the guppies (*Poecilia*) which rely more heavily on red-colouration for mate choice, gene conversion seems less prevalent (Sandkam et al. 2017). Using GARD (Kosakovsky Pond et al. 2006) to detect recombination points between *LWS1* and *LWS3*, we found two weakly supported break points in exons 2 (position 287 bp; 42% support) and 5 (903 bp; 33% support). A subsequent sliding window analysis measuring the neutral divergence along duplicates [rate of synonymous substitutions per synonymous sites (dS)] supported partially by the single-exon phylogenies, confirmed that gene conversion has most likely affected the region surrounding at least the first break point. However, the strength of the signal

differed markedly between species ranging from a rather equal dS ratio in the round goby, *Neogobius melanostomus*, to clearly discernible drops in dS in the duckbill sleeper, *Butis butis*, and the peacock gudgeon/goby, *Tateurndina ocellicauda*. It appears that in general, gene conversion between *LWS1* and *LWS3* occurred early on during the diversification of the paralogs, with the GARD support values being low and the drop in dS being rather modest compared to what is found for example between *SWS2* duplicates (Cortesi et al. 2015). It is possible that following duplication, *LWS3* diverged rapidly, and gene conversion acted as a homogeniser between paralogs to stabilise at least the first part of the coding region, with the exons carrying the tuning sites seemingly less affected.

Convergent shifts to green vision in *LWS* duplicates

Changes in five key tuning sites, the so called 'five-sites rule' (human MWS/LWS, amino acid position: 180, 197, 277, 285 and 308; bovine rhodopsin equivalent: 164, 181, 261, 269 and 292), can explain most, albeit not all (e.g., Chi et al. 2020), major spectral sensitivity shifts of the *LWS* photopigments in vertebrates (Yokoyama and Radlwimmer 1998; Yokoyama et al. 2008). As for the primate MWS/LWS case (Carvalho et al., 2017), LWS2 in characins and bony tongues but also several recent teleost LWS1 copies have functionally shifted to green sensitivity via convergent changes in these five sites (Liu et al. 2019; Escobar-Camacho et al. 2020; Sandkam et al. 2018). Comparing these key tuning sites in LWS3, we also find a convergent shift to shorter wavelength sensitivity in gobies (Table 1). Like the substitutions in LWS2, changes in LWS3 mostly affected sites 180 and 277, where shorter shifts are caused by substitutions from serine to alanine and tyrosine to phenylalanine, respectively. However, as opposed to LWS2 where changes in 285 were also common, we only found one such change from the longer shifted threonine to the shorter shifted alanine in the LWS3 of the great blue-spotted mudskipper, *Boleophthalmus pectinirostris* (Table 1) (also see You et al. 2014). Spectral sensitivity predictions for the goby LWS paralogs largely overlapped with measured mid-wavelength-sensitive (MWS) and long-wavelength-sensitive photoreceptor sensitivities from a variety of other goby species. However, several of the measured MWS values (λ_{max} = 527 – 531 nm) were slightly shorter than many of the predictions for LWS3 (λ_{max} = 533 – 550 nm) (Pierotti et al. 2020 and references therein). This could be explained if amino acids at sites other than the five commonly acknowledged ones are also tuning LWS3. Alternatively, it may be that the goby MWS photoreceptors express the green-sensitive RH2 opsin or a mixture of RH2 and LWS3 rather than pure LWS3. Coexpression of RH2 with LWS2 has been reported for some shorter shifted characin photoreceptors (Escobar-Camacho et al. 2020).

239

240 **Neofunctionalization of LWS3 in gobies**

241 To test which visual opsins gobies use and what photoreceptors they might be expressed in, we
242 sequenced the retinal transcriptomes of ten species from nine different genera and representing
243 most goby subfamilies (Fig. 3). Five out of the ten species were found to express *LWS3* to
244 various degrees. In eight species *LWS1* was the dominant paralog. However, *LWS3* was the
245 highest expressed copy in the sub-family Sycidiinae, with the red lipstick goby, *Syciopus*
246 *rubicundus*, only expressing this copy and the rainbow goby, *Stiphodon ornatus*, showing very
247 low levels of *LWS1* expression (Fig. 4C). All gobies also expressed at least one copy of *SWS2*,
248 albeit at very low levels in the Sycidiinae and the spinycheek sleeper, *Eleotris pisonis*. *RH2*
249 on the other hand, was either not found to be expressed or only very marginally expressed in
250 most species, except for the Sycidiinae where *RH2* was expressed at levels similar to the *SWS2*
251 expression in other species.

252 Using fluorescent in-situ hybridisation in *A. phalaena*, revealed that the *SWS2* opsins,
253 *SWS2B* and *SWS2Aβ*, were expressed exclusively in the single cones (Fig. 4 A, B). Among
254 single cones, *SWS2B* and *SWS2Aβ* showed a varying degree of coexpression in the same single
255 cone photoreceptors, with *SWS2Aβ* expressed in virtually every single cone, and *SWS2B* only
256 expressed in some but not (or perhaps only lowly expressed) in others (Fig. 4 A-C). *LWS1* and
257 *LWS3*, were exclusively expressed in double cones, two single cones that are fused together
258 and may also be optically coupled (Fig 4 D-F). Among these, *LWS1* was expressed in one
259 double cone member and *LWS2* was expressed in the other double cone member (Fig. x F).
260 This shows that for *A. phalaena* at least, *LWS3* has acquired an independent function which
261 together with the *SWS2* and *LWS1* expressing photoreceptors, is likely to provide the species
262 with trichromatic vision.

263

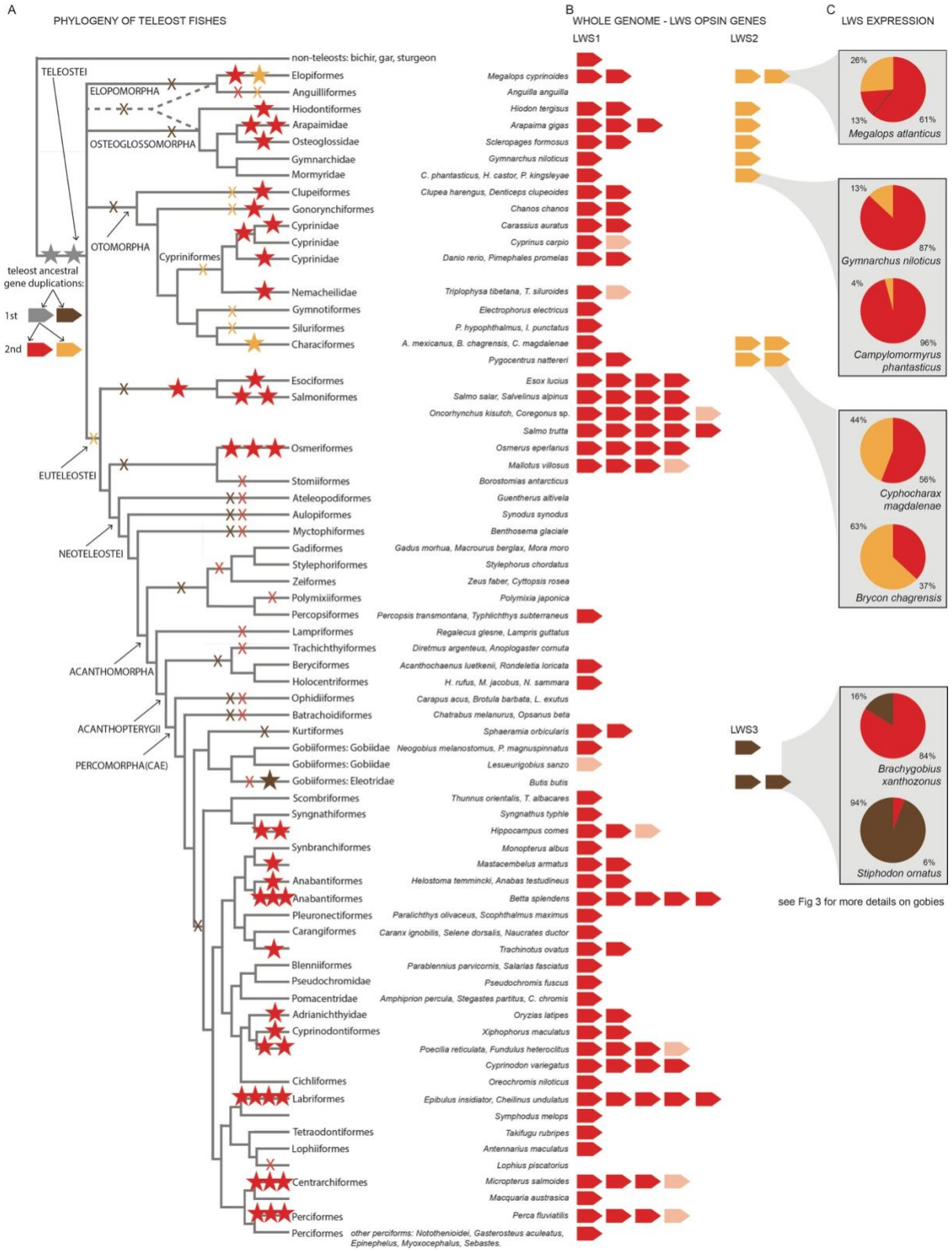

***Figure 1**: The scenario of the LWS evolution in teleost fishes. A) The reconstructed gene duplications and gene losses are mapped on the phylogenetic tree (after Betancur et al., 2017). Most of the duplications are species/genus specific, except for the Esociformes+Salmoniformes and Cyprinodontiformes which are older. B) Number and type of LWS opsin genes found in*

the whole genomes of the teleost fishes. Red - LWS1, orange - LWS2, brown - LWS3. Note that only lineages of Elopomorpha, Osteoglossomorpha and Characiformes have the LWS2 gene and only gobies (Gobiiformes) have the LWS3 gene. C) Gene expression pattern in selected species with multiple LWS gene types: tarpon, elephant fishes, characins and gobies.

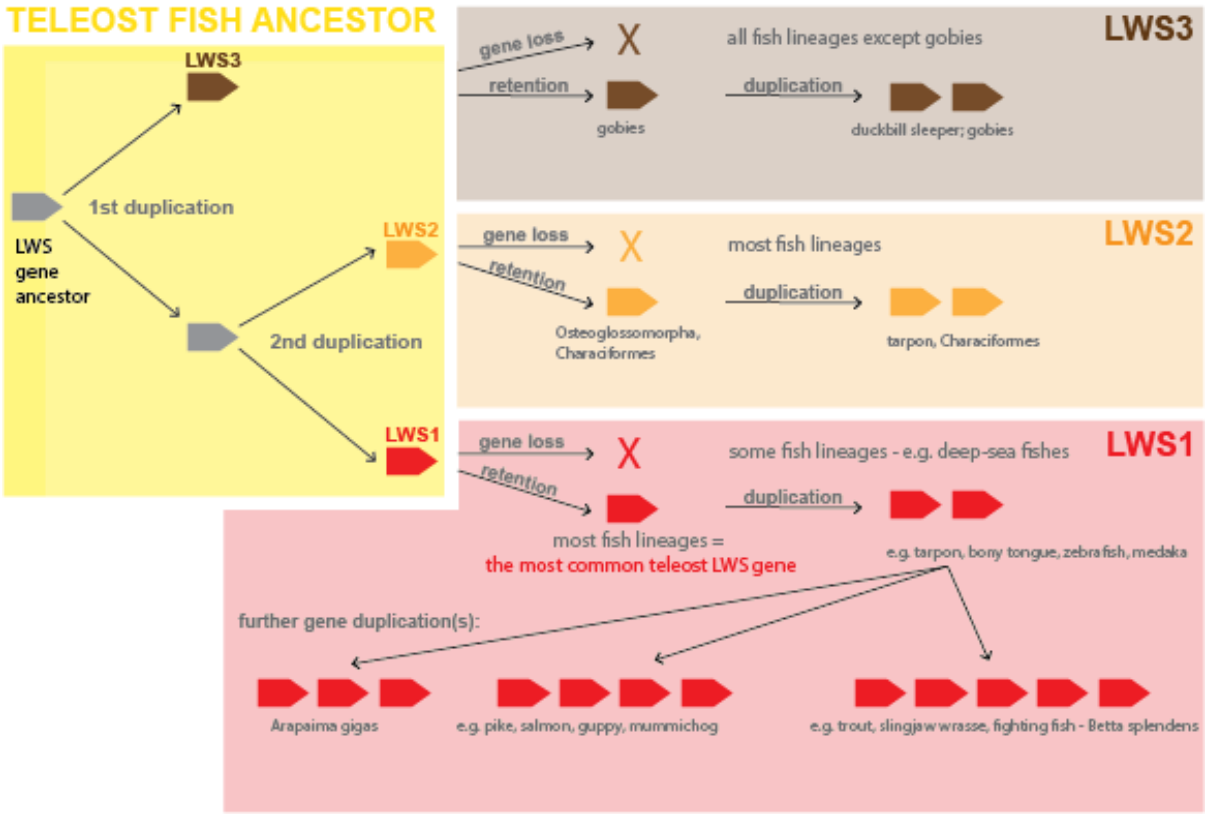

***Figure 2** - The fate and dynamics of the three ancestral teleost copies of LWS in teleost fishes. The most prevalent gene copy is LWS1, which is present in the vast majority of fish lineages. The LWS2 copy is retained in Osteoglossomorpha, Elopomorpha and Characiformes (and is not found in any Euteleostei). The LWS3 has been detected only in gobies (Gobiiformes) so far and has been lost in all other fish lineages.*

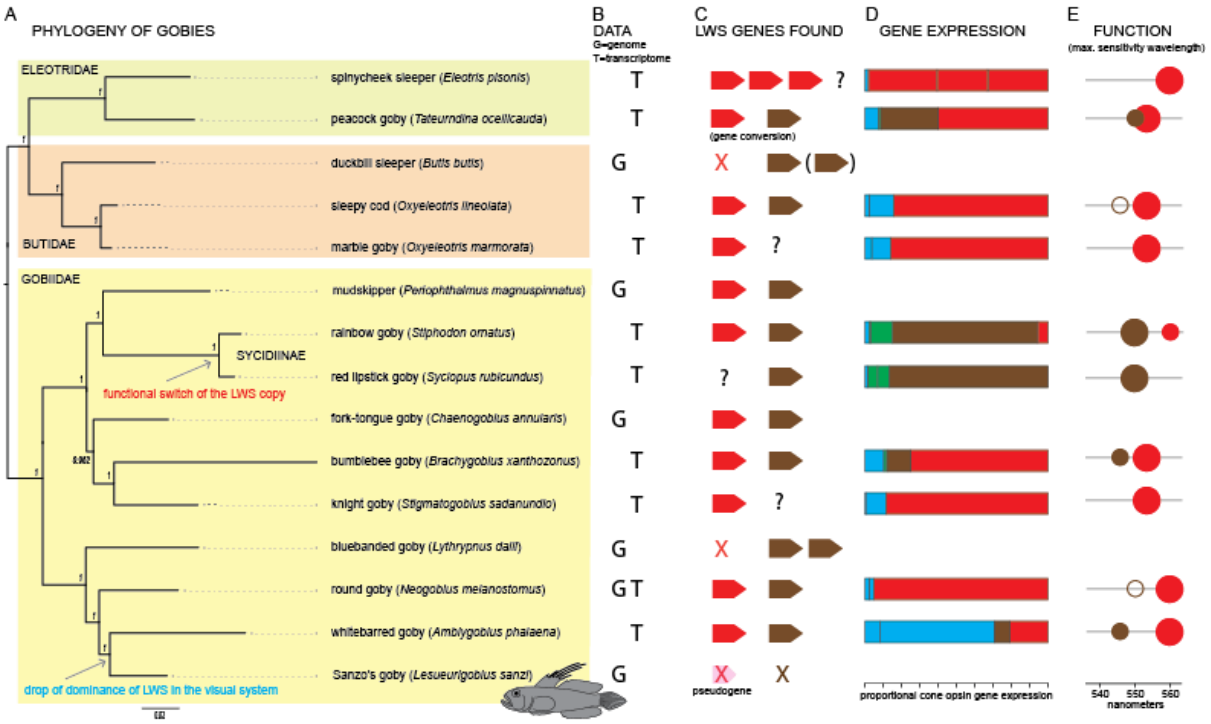

Figure 3: *Evolution of the LWS opsin genes in gobies (Gobiiformes)***.** *A) phylogenetic relationship among the studied gobies based on the genomic and transcriptomic data set. Three goby families are included in this study. B) Source of data: genomic (G), trancriptomic (T) or both (GT) data available. C) List of the found LWS gene.s Note that if only transcriptomic data are used, we cannot make any conclusions on the gene loss. C) proportional cone opsin gene expression in the gobies. Colour code in the horizontal bars stands for the sensitivity of the opsin gene - blue = SWS2, green = RH2, red = LWS1, brown = LWS3. Note that the LWS genes dominate the cone expression profiles in all but one (whitebarred goby,* A. phalaena*) studied species. D) function of the two LWS genes in vision. Spectral sensitivity (lmax) of the LWS1 (red) and LWS3 (brown) opsin gene has been predicted (Table 1) and is shown on the wavelength scale. Larger vs. smaller circles represent the dominant vs. rare transcripts in the expression profile. Note the functional proximity of the two LWS1 and LWS3 genes in the peacock goby (*Tateurndina ocellicauda*) where gene conversion has been detected.*

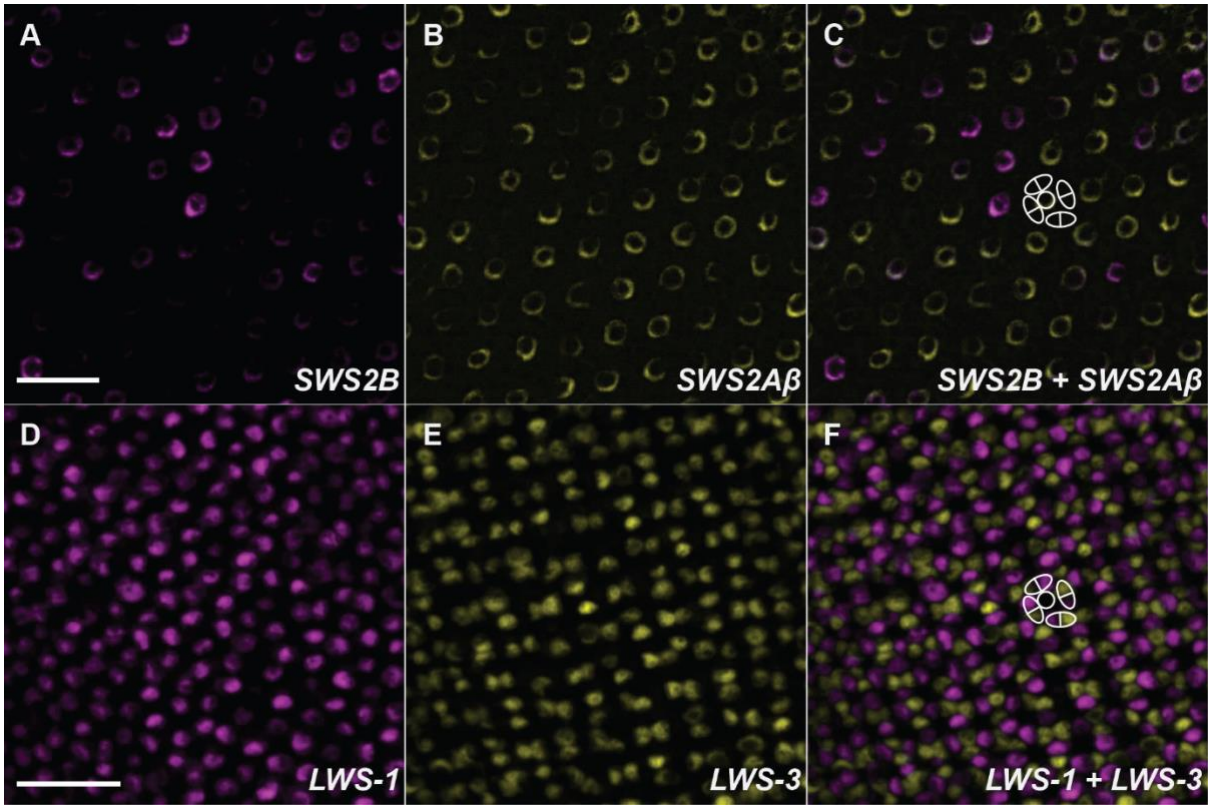

***Figure 4**: Double-labelling in-situ hybridisation of expressed opsin mRNAs in retinal single cone (A-C) and double cone (D-F) photoreceptors in adult* Amblygobius phalaena. ***(A-C)** SWS2B (magenta) and SWS2Aβ (yellow) mRNAs were co-expressed in some single cones. **(D-F)** LWSx1 (magenta) and LWSx2 (yellow) mRNAs were expressed in opposite members of double cones. Scale bars: 10 µm.*

species	ecology	gene	key tuning amino acid site					max. spectral sensitivity	reference	note
			180	197	277	285	308			
bovine rhodopsin equivalent:			*164*	*181*	*261*	*269*	*292*			
bovine rhodopsin amino acid:			*A*	*E*	*F*	*A*	*A*			
SPECIES WITH BOTH TYPES OF LWS (LWS1, LWS3) FOUND IN THE TRANSCRIPTOME/GENOME:										
Boleophthalmus pectinirostris	terrestrial	LWS1	A	H	Y	T	A	553 nm	You et al., 2014	only genome
	mudskipper	LWS3	A	H	F	A	A	531 nm	You et al., 2014	
Periophthalmus magnuspinnatus	terrestrial	LWS1	S	H	Y	T	A	560 nm	You et al., 2014	only genome
	mudskipper	LWS3	A	H	F	T	A	546 nm	You et al., 2014	
Amblygobius phalaena	coral reef goby	LWS1	S	H	Y	T	A	560 nm	this study/Musilova et al., 2019	
		LWS3	A	H	F	T	A	546 nm	this study/Musilova et al., 2019	
Neogobius melanostomus	freshwater temperate	LWS1	S	H	Y	T	A	560 nm	Adrian-Kalchhauser et al., 2020	
		LWS3	S	H	F	T	A	550 nm*	Adrian-Kalchhauser et al., 2020	
Brachygobius xanthozona		LWS1	A	H	Y	T	A	553 nm	this study	
		LWS3	A	H	F	T	A	546 nm	this study	
Stiphodon ornatus		LWS1	S	H	Y	T	A	560 nm	this study	
		LWS3	S	H	F	T	A	550 nm*	this study	
Tateurndina ocellicauda		LWS?	A	H	Y	T	A	553 nm	this study	gene conversion
		LWS?	S	H	F	T	A	550 nm*	this study	gene conversion
Oxyeleotris lineolata		LWS1	A	H	Y	T	A	553 nm	this study	
		LWS3	A	H	F	T	A	546 nm	this study	
Chaenogobius annularis		LWS1	S	H	Y	T	A	560 nm	this study	only genome
		LWS3	S	H	F	T	A	550 nm*	this study	
SPECIES WITH ONLY ONE TYPE OF LWS FOUND IN THE TRANSCRIPTOME/GENOME:										
Butis butis (misidentified as *O. marmorata* on GenBank)	sleeper	LWS3-1	A	H	F	T	A	546 nm	this study	only genome
		LWS3-2	P	H	F	T	A	???	this study	
Eleotris pisonis		LWS1-1	S	H	Y	T	A	560 nm	this study	
		LWS1-2	S	H	Y	T	A	560 nm	this study	
		LWS1-3	S	H	Y	T	A	560 nm	this study	
Oxyeleotris marmorata		LWS1	A	H	Y	T	A	553 nm	this study	
Sycíopus rubicundus		LWS3	S	H	F	T	A	550 nm*	this study	
Stigmatogobius sadanundio		LWS1	A	H	Y	T	A	553 nm	this study	
Lythrypnus dalli		LWS3	A	H	F	T	A	546 nm	this study	only genome

* = predicted by the key tuning sites, and Y261F shift of 10 nm; Yokoyama 2008 (review)

Table 1*: Predicted values of the maximal sensitivity (lmax) of the LWS genes in gobies. We applied the five amino acid rule, i.e. similar to application in the mudskippers (You et al., 2014). Note that the LWS3 is sensitive to the shorter wavelengths than the LWS1.*
The rhodopsin sites 164, 181, 261, 269, and 292 correspond to the 180, 197, 277, 285, and 308 sites in the LWS gene.

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.

Taxon sampling

Samples of *Amblygobius phalaena* and *Epibulus insidiator* for expression have been collected and previously published in Musilova et al., 2019 (see Table 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 Lizard Island, Great Barrier Reef, Australia in March 2021 using an anaesthetic clove oil solution (10% clove oil, 40% ethanol, and 50% seawater) and hand nets. Eyes were enucleated and the retina was dissected out and either fixed in 4% PFA for in-situ hybridisation or stored on RNAlater at -20 C until further processing for transcriptome sequencing. Samples of *Neogobius melanostomus* were collected in the Elbe River (Děčín, Czech Republic) using the electro fishing device. Samples of the remaining seven species, *Eleotris pisonis, Eleotris marmorata, Tateurndina ocellicauda, Stiphodon ornatus, Syciopus rubicundus, Brachygobius xanthozonus* and *Stigmatogobius sadanundio*, have been obtained from aquarium trade.

Transcriptome sequencing

Retinal RNA was extracted using a Qiagen mini kit column extraction following the manufacturer's protocol and including a DNAse step. Library preparation (unstranded, 300 bp insert) and transcriptome sequencing (RNAseq HiSeq Paired-End PE150) was outsourced to Novogene (https://en.novogene.com/).

Assembly and gene extraction

Transcriptome raw reads have been filtered and mapped with Medium sensitivity settings against the general goby reference (based on the genome of *Neogobius melanostomus*; Adrian-Kalchauser et al., 2019) in the Geneious 9.0 software (https://www.geneious.com). Species specific gene sequences have been retrieved from the mapped alignments and used as a species-specific reference for another round of mapping with the Lowest sensitivity settings. Number of reads have been calculated to the FPKM (fragments per kilobase and million reads) and the proportion of each opsin from the total cone opsin gene expression has been calculated. More details on the protocols described in de Busserolles et al. 2017 and Tettamanti et al. 2019.

Whole genome analysis

High quality teleost whole genomes have been downloaded from GenBank (accession numbers as per sample Table). Majority of the high-quality genomes has been sequenced by the Sanger Institute (Rhie et al., 2021). We further complemented this data set by the genomic data of lower quality (draft genomes as per Cortesi et al., 2015 and Musilova et al., 2019). Assembled genomic scaffolds / chromosomes have been mapped against the general single exon reference of the example LWS1 (*Oreochromis niloticus*), LWS2 (*Astyanax mexicanus*) and LWS3 (*Neogobius melanostomus*) copy. Mapping has been performed in Geneious 9.1.8

(https://www.geneious.com) using Medium to Medium High sensitivity settings - this settings has been tested to pick all scaffolds with any LWS opsin (higher sensitivity yielded to false hits). Single LWS genes have been extracted from the scaffolds either following the annotation (if available) or by mapping of single exons against the scaffold with the Highest sensitivity.

Phylogenetic analysis

Coding sequences of the single LWS genes have been aligned in Geneious 9.1.8 using the MAFFT Alignment plugin (Katoh et al., 2019). Five independent runs (each with two parallel runs) have been performed in MrBayes 3.1.2 (Ronquist and Huelsenbeck, 2003) software for 50 million generations. The output files have been manually inspected in the Tracer 1.5 (https://github.com/beast-dev/tracer/releases) for reaching the lnL plateau, the best -lnL score and to identify the burnin break point. Three out of the ten independent output runs have been used for the consensus tree construction with the burnin set to 50%.

Gene conversion analysis

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

In-situ hybridisation

Dual-labelling RNA-FISH was performed on two wholemount retinas (left and right) of one adult *A. phalaena* collected at Lizard Island on 16 March 2021 following standard protocols (e.g., Barthel and Raymond 2000, Raymond and Barthel 2004, Allison et al. 2010, Dalton et al. 2014, Dalton et al. 2015). Retinas were prepared as follows. After one-hour dark adaptation, the eyes were enucleated, vitreous was enzymatically digested, and retinas were extracted from the eyecups. The retinal pigment epithelium layer (RPE) was removed by PBS-jetting using a single use syringe and needle (Terumo). Retinas were flattened and pinned, photoreceptor side down, in Sylgard-filled petri dishes, and fixed in 4% PFA at 4°C overnight. Fixed retinas were triple washed in PBS for 10 min, briefly rinsed in 70% methanol in PBS, and then transferred to 100% methanol and stored at -20°C.

Previously extracted retinal mRNA was reverse transcribed using a High Capacity RNA-to-cDNA Reverse Transcription Kit (Applied Biosystems). Riboprobe templates were synthesized from cDNA via standard PCR using opsin specific primers (Table S1) and Q5 High Fidelity DNA polymerase (New England Biolabs). Primers were designed to bind to the coding sequence of target opsins (*SWS2B*, *SWS2Aβ*, *LWS1*, *LWS3*) and to contain a T3 or T7 RNA polymerase promoter sequence at their 5'-ends (T3, reverse primer; T7, forward primer). Amplicons were isolated via gel-electrophoresis and gel-extraction (Qiagen Gel Extraction Kit), and subsequently used as template in enrichment PCR using the same primers. Anti-sense riboprobes labelled with either digoxigenin-UTP (DIG) or fluorescein-UTP (FL) were synthesized from cDNA templates via T3 RNA polymerase mediated strand-specific RNA transcription using DIG/FL RNA labelling mix (Roche). Hybridised, labelled riboprobes were detected using anti-digoxigenin or anti-fluorescein antibodies conjugated to horseradish peroxidase (Roche). Fluorescent tagging was performed using Alexa Fluor 594 and 488 dyes with streptavidin Tyramide Signal Amplification (TSA, Invitrogen). Finally, retinas were mounted in 70% glycerol in PBS, photoreceptor side up, on microscopy slides and covered with a coverslip.

Fluorescent tagged photoreceptor cells were imaged using a CFI Apo Lambda 60x/1.4 NA oil immersion objective on a spinning disk confocal microscope (Diskovery, Andor Technologies, UK) built around a Nikon Ti-E body (Nikon Corporation, Japan) equipped with two Zyla 4.2 sCMOS cameras (Andor Technology), and controlled by Nikon NIS Elements software (Nikon Corporation, Japan). Exported images were further processed (max. intensity z-stack projection, channel merging) with ImageJ v. 1.52p (National Institute of Health, USA).

Acknowledgements

We would like to thank Dr. Andrew Mather for help with sample collection. FC was supported by an Australian Research Council (ARC) DECRA Fellowship DE200100620, and by a University of Queensland Development Fellowship, Australia. We thank Veronika Truhlářová for lab management and logistical support. FC was supported by an Australian Research Council DECRA Research Fellowship (ARC DE200100620). GMS has been funded by GAUK (1524119; Charles University). ZM was supported by Swiss National Science Foundation (SNF grant PROMYS, 166550), Charles University (Primus), and Czech Science Foundation (21-31712S).

Data Availability: All sequenced data have been submitted to GenBank and are available under accession numbers as per sample Table**.**

References:

Adrian-Kalchhauser, I., Blomberg, A., Larsson, T., Musilova, Z., Peart, C. R., Pippel, M., … Winkler, S. (2020). The round goby genome provides insights into mechanisms that may facilitate biological invasions. *BMC Biology*, *18*(1), 1–33. http://doi.org/10.1186/s12915-019-0731-8

Borges, R., Khan, I., Johnson, W. E., Gilbert, M. T. P., Zhang, G., Jarvis, E. D., ... & Antunes, A. (2015). Gene loss, adaptive evolution and the co-evolution of plumage coloration genes with opsins in birds. *BMC genomics*, 16(1), 1-14.

Carleton, K.L., Conte, M.A., Malinsky, M., Nandamuri, S.P., Sandkam, B.A., Meier, J.I., Mwaiko, S., Seehausen, O. and Kocher, T.D., 2020. Movement of transposable elements contributes to cichlid diversity. *Molecular Ecology*, *29*(24), pp.4956-4969.

Carleton, K.L., Escobar-Camacho, D., Stieb, S.M., Cortesi, F. and Marshall, N.J., 2020. Seeing the rainbow: mechanisms underlying spectral sensitivity in teleost fishes. *Journal of Experimental Biology*, *223*(8).

Carvalho, L. S., Pessoa, D., Mountford, J. K., Davies, W. I., & Hunt, D. M. (2017). The genetic and evolutionary drives behind primate color vision. *Frontiers in Ecology and Evolution*, 5, 34.

Chen J-N, Samadi S, Chen W-J. 2018. Rhodopsin gene evolution in early teleost fishes. *PLOS ONE*. 13(11):e0206918

Chi, H., Cui, Y., Rossiter, S.J. and Liu, Y., 2020. Convergent spectral shifts to blue-green vision in mammals extends the known sensitivity of vertebrate M/LWS pigments. *Proceedings of the National Academy of Sciences*, *117*(15), pp.8303-8305.

Cortesi, F., Musilová, Z., Stieb, S.M., Hart, N.S., Siebeck, U.E., Malmstrøm, M., Tørresen, O.K., Jentoft, S., Cheney, K.L., Marshall, N.J. and Carleton, K.L., 2015. Ancestral duplications and highly dynamic opsin gene evolution in percomorph fishes. *Proceedings of the National Academy of Sciences*, *112*(5), pp.1493-1498.

Cortesi, F., Musilová, Z., Stieb, S.M., Hart, N.S., Siebeck, U.E., Cheney, K.L., Salzburger, W. and Marshall, N.J., 2016. From crypsis to mimicry: changes in colour and the configuration of the visual system during ontogenetic habitat transitions in a coral reef fish. *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.

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.

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.

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.

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.

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

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

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

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.

Rabosky, D.L., Chang, J., Title, P.O., Cowman, P.F., Sallan, L., Friedman, M., Kaschner, K., Garilao, C., Near, T.J., Coll, M. and Alfaro, M.E., 2018. An inverse latitudinal gradient in speciation rate for marine fishes. *Nature*, *559*(7714), pp.392-395.

Ravi, V. and Venkatesh, B., 2018. The divergent genomes of teleosts. *Annual review of animal biosciences*, *6*, pp.47-68.

Rhie, A., McCarthy, S. A., Fedrigo, O., Damas, J., Formenti, G., Koren, S., ... & Jarvis, E. D. (2021). Towards complete and error-free genome assemblies of all vertebrate species. *Nature*, *592*(7856), 737-746.

Ronquist, F. and J. P. Huelsenbeck. 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572-1574.

Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J.C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S.E. and Sánchez-Gracia, A., 2017. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular biology and evolution*, *34*(12), pp.3299-3302.

Sandkam, B.A., Joy, J.B., Watson, C.T. and Breden, F., 2017. Genomic environment impacts color vision evolution in a family with visually based sexual selection. *Genome biology and evolution*, *9*(11), pp.3100-3107.

Sandkam, B., Dalton, B., Breden, F. and Carleton, K., 2018. Reviewing guppy color vision: Integrating the molecular and physiological variation in visual tuning of a classic system for sensory drive. *Current zoology*, *64*(4), pp.535-545.

Seehausen, O., Terai, Y., Magalhaes, I.S., Carleton, K.L., Mrosso, H.D., Miyagi, R., Van Der Sluijs, I., Schneider, M.V., Maan, M.E., Tachida, H. and Imai, H., 2008. Speciation through sensory drive in cichlid fish. *Nature*, *455*(7213), pp.620-626.

Tettamanti, V., de Busserolles, F., Lecchini, D., Marshall, N.J. and Cortesi, F., 2019. Visual system development of the spotted unicornfish, Naso brevirostris (Acanthuridae). *Journal of Experimental Biology*, *222*(24).

Wald, G., 1968. The molecular basis of visual excitation. *Nature*, *219*(5156), pp.800-807.

Walsh, B. (2003). Population-genetic models of the fates of duplicate genes. In *Origin and Evolution of New Gene Functions (*pp. 279-294). Springer, Dordrecht.

Ward MN, Churcher AM, Dick KJ, Laver CR, Owens GL, et al. 2008. The molecular basis of color vision in colorful fish: Four Long Wave-Sensitive (LWS) opsins in guppies (Poecilia reticulata) are defined by amino acid substitutions at key functional sites. *BMC Evol. Biol.* 8(1):210

Weaver, S., Shank, S.D., Spielman, S.J., Li, M., Muse, S.V. and Kosakovsky Pond, S.L., 2018. Datamonkey 2.0: a modern web application for characterizing selective and other evolutionary processes. *Molecular biology and evolution*, *35*(3), pp.773-777.

Yokoyama, S., 2008. Evolution of dim-light and color vision pigments. *Annu. Rev. Genomics Hum. Genet.*, *9*, pp.259-282.

Yokoyama, S. and Radlwimmer, F.B., 1998. The" five-sites" rule and the evolution of red and green color vision in mammals. *Molecular biology and evolution*, *15*(5), pp.560-567.

You, X., Bian, C., Zan, Q., Xu, X., Liu, X., Chen, J., Wang, J., Qiu, Y., Li, W., Zhang, X. and Sun, Y., 2014. Mudskipper genomes provide insights into the terrestrial adaptation of amphibious fishes. *Nature communications*, *5*(1), pp.1-8.