Opsin Gene Duplication in Lepidoptera: Retrotransposition, Sex Linkage, and Gene Expression

Peter O. Mulhair (iD),[1,*] Liam Crowley,[1] Douglas H. Boyes,[2] Owen T. Lewis,[1] and Peter W.H. Holland[1]

[1]Department of Biology, University of Oxford, Oxford OX1 3SZ, UK
[2]UK Centre for Ecology & Hydrology, Wallingford OX10 8BB, UK

*Corresponding author: E-mail: peter.mulhair@biology.ox.ac.uk.
Associate editor: Belinda Chang

Abstract

Color vision in insects is determined by signaling cascades, central to which are opsin proteins, resulting in sensitivity to light at different wavelengths. In certain insect groups, lineage-specific evolution of opsin genes, in terms of copy number, shifts in expression patterns, and functional amino acid substitutions, has resulted in changes in color vision with subsequent behavioral and niche adaptations. Lepidoptera are a fascinating model to address whether evolutionary change in opsin content and sequence evolution are associated with changes in vision phenotype. Until recently, the lack of high-quality genome data representing broad sampling across the lepidopteran phylogeny has greatly limited our ability to accurately address this question. Here, we annotate opsin genes in 219 lepidopteran genomes representing 33 families, reconstruct their evolutionary history, and analyze shifts in selective pressures and expression between genes and species. We discover 44 duplication events in opsin genes across ∼300 million years of lepidopteran evolution. While many duplication events are species or family specific, we find retention of an ancient long-wavelength-sensitive (LW) opsin duplication derived by retrotransposition within the speciose superfamily Noctuoidea (in the families Nolidae, Erebidae, and Noctuidae). This conserved LW retrogene shows life stage–specific expression suggesting visual sensitivities or other sensory functions specific to the early larval stage. This study provides a comprehensive order-wide view of opsin evolution across Lepidoptera, showcasing high rates of opsin duplications and changes in expression patterns.

Key words: opsins, color vision, Lepidoptera, gene duplication, retrotransposition, gene expression.

Introduction

Opsin genes belong to the G protein–coupled receptor multigene family; the encoded opsin proteins form a photosensitive complex through covalent binding to a retinal-based chromophore (Palczewski et al. 2000; Briscoe and Chittka 2003; Nilsson 2009). The spectral sensitivities of an organism are directly linked to its opsin gene content and expression; thus, this gene family provides a useful framework to relate changes in the genotype to the evolution of visual adaptations and behaviors (Arikawa 2017; Hauser and Chang 2017; van der Kooi et al. 2021; Van Nynatten et al. 2021). Opsin gene duplication and divergence, and gene loss, are 2 mechanisms by which insects have evolved altered visual sensitivities to adapt to niche-specific light environments and external cues (Spaethe and Briscoe 2004; Frentiu et al. 2007; Briscoe 2008; Thoen et al. 2014; Futahashi et al. 2015; Chen et al. 2016; Feuda et al. 2016; Sharkey et al. 2017; Armisén et al. 2018; Almudi et al. 2020; Feuda et al. 2021; Sondhi et al. 2021; Guignard et al. 2022; McCulloch, Macias-Muñoz, Mortazavi, et al. 2022). However, other mechanisms are known to increase the diversity of opsin specificities resulting in variation in light-sensing abilities, such as

amino acid substitutions without gene duplication causing shifts in absorption spectra (Shichida and Matsuyama 2009; Wakakuwa et al. 2010; Hauser et al. 2014; Liénard et al. 2021; Sharkey et al. 2023), coexpression of certain opsin proteins in a single photoreceptor cell (Wakakuwa et al. 2004; Perry et al. 2016; Satoh et al. 2017; Macias-Muñoz et al. 2019; Ilić et al. 2022), differences in ommatidia structure of the compound eye (Lau and Gross 2007; Lau and Meyer-Rochow 2007), and paralog-specific opsin gene expression (Arikawa 2003; Perry et al. 2016; Finkbeiner and Briscoe 2021; Roberts et al. 2023).

In Lepidoptera (butterflies and moths), color vision is determined by rhabdomeric type r-opsins encoded by 3 genes, with each opsin having a different peak wavelength sensitivity (λ_{max}): long-wavelength-sensitive opsins (LW opsins; λ_{max} 500 to 600 nm) that can respond to green light, short-wavelength-sensitive opsins (blue or SWS opsins; 400 to 500 nm) sensitive to blue light, and ultraviolet opsins (UV opsins; 300 to 400 nm) that respond to ultraviolet light (Briscoe and Chittka 2003; Stavenga and Arikawa 2006; Henze and Oakley 2015; Feuda et al. 2016; van der Kooi et al. 2021). In addition to this core set of opsin genes, Lepidoptera also possess Rh7 and c-opsin genes,

Open Access

thought not to be linked to vision but whose full function and phylogenetic distribution are unknown (Feuda et al. 2016; Ni et al. 2017; Sakai et al. 2017). Several duplications of opsin genes have been observed in lepidopteran species. For example, many butterflies in the genus *Heliconius* can see from the ultraviolet to the red extremes of the light spectrum, owing to duplication of the UV opsin gene followed by amino acid changes driven by positive selection in the ancestor of the clade (Briscoe et al. 2010). This also resulted in sexually dimorphic UV color vision, with females of certain species able to distinguish between different UV wavelengths (Finkbeiner and Briscoe 2021; Chakraborty et al. 2023). This specialization in vision is thought to have coevolved with wing coloration (Finkbeiner et al. 2017). Similarly, a blue opsin gene duplication in the family Lycaenidae allowed evolution of a green-shifted blue opsin paralog which, in combination with a red-shifted LW opsin gene, may allow finer wavelength discrimination, again coinciding with wing coloration evolution (Bernard and Remington 1991; Sison-Mangus et al. 2006, 2008; Liénard et al. 2021). Such diversity in the mechanisms of opsin evolution has resulted in massive variation in color vision, habitat adaptation, and feeding habits in this diverse order of insects (van der Kooi et al. 2021). Additionally, the number of transitions between diurnal and nocturnal behavior within this order (Kawahara et al. 2018) means that this group is an ideal system to relate changes in opsin content and evolution with transitions in lifestyle and behavior.

Despite these interesting case studies, much remains unknown concerning opsin diversity across Lepidoptera. The paucity of species sampling is also confounded by the fact that, until recently, high-quality genome assemblies for this order have been limited in number, compromising efforts to annotate opsin genes, assign orthology, and confirm true gene losses (Feuda et al. 2016; Sondhi et al. 2021). Here, we examine newly generated, chromosome-level assemblies for 219 lepidopteran species, representing the largest, most phylogenetically representative data set used to date. Chromosome-level genome assemblies publicly released by the Darwin Tree of Life Project (The Darwin Tree of Life Project Consortium 2022) allow for accurate identification and analysis of opsin genes in all species. We assessed the rate of opsin gene duplication and loss across the tree, revealing dynamic changes in opsin copy number. Although there is conservation of 5 core lepidopteran opsin genes, we also identify 44 distinct duplication events, many of which occurred in the stem lineage of certain lepidopteran clades, including Tortricidae, Micropterigidae, Lycaenidae, and Noctuoidea. Sequencing and analysis of transcriptomic data from members of the Noctuoidea superfamily revealed life stage–specific expression of LW paralogs, suggesting subfunctionalization of expression domains. Finally, we also tested whether transitions from nocturnality to diurnality are associated with shifts in evolutionary rate within the visual opsin genes.

Results

From a data set of 219 lepidopteran genome sequences representing 33 families, we constructed a species tree and annotated opsin genes (Fig. 1a, supplementary fig. S1, Supplementary Material online). The dense sampling of species represents much of the lepidopteran diversity and includes species from the early diverging Micropterigidae family. The data set includes 152 nocturnal, 56 diurnal species, and 11 species with both night- and day-flying behaviors (supplementary fig. S1, Supplementary Material online). These recently generated chromosome-level genome assemblies, most generated by the Darwin Tree of Life Project (The Darwin Tree of Life Project Consortium 2022), allowed for more accurate and comprehensive identification of opsin gene open reading frames (ORF) than in earlier studies. We used a combination of BLAST (Altschul et al. 1990) and Exonerate (Slater and Birney 2005) in an iterative approach to annotate opsin genes in all sampled genomes (see Materials and Methods). We identified 1,279 opsin genes across 219 lepidopteran species (Fig. 1a). Phylogenetic reconstruction of opsin genes using a maximum likelihood approach recovers the monophyly of all main opsin groups: 4 r-opsins (UV, blue, LW, and Rh7) and 1 c-opsin (Fig. 1b, supplementary fig. S2, Supplementary Material online). The branching patterns within the gene tree, in combination with gene structure and surrounding gene synteny where required, were used to infer the phylogenetic node of origin for each opsin gene duplication event (Fig. 1c).

Blue and LW Opsin Duplications Were Prevalent during Lepidopteran Evolution

We find variation in opsin gene copy number within different lineages, caused by 34 independent duplications of the LW opsin across 10 families, 6 duplication events of the blue opsin in 3 different butterfly families, and 1 species of moth, 1 Rh7 duplication, and 3 c-opsin duplications (Fig. 1c). These include some previously described duplication events: the duplication of the blue opsin in the butterfly family Lycaenidae (Bernard and Remington 1991; Sison-Mangus et al. 2006, 2008; Liénard et al. 2021); a blue opsin duplication within the Pieridae (Awata et al. 2009; Wakakuwa et al. 2010), which we find to be missing from the wood white butterfly (*Leptidea sinapis*) and underwent an additional tandem duplication in the clouded yellow butterfly (*Colias croceus*); and 2 LW duplications in the swallowtail butterfly *Papilio machaon* (Arikawa 2003; Saito et al. 2019). Increased taxon sampling in certain lineages refines the point of origin in lepidopteran evolution for certain duplication events. For example, the blue duplication in Lycaenidae is present in all species in our data set confirming that this event occurred in an ancestor to this family. Additionally, the blue duplication in the Pieridae family is now placed on the branch after divergence from *L. sinapis* (Fig. 1c). We find another duplication of the blue opsin within Hesperiidae, where a duplicate copy is shared between *Thymelicus sylvestris*,
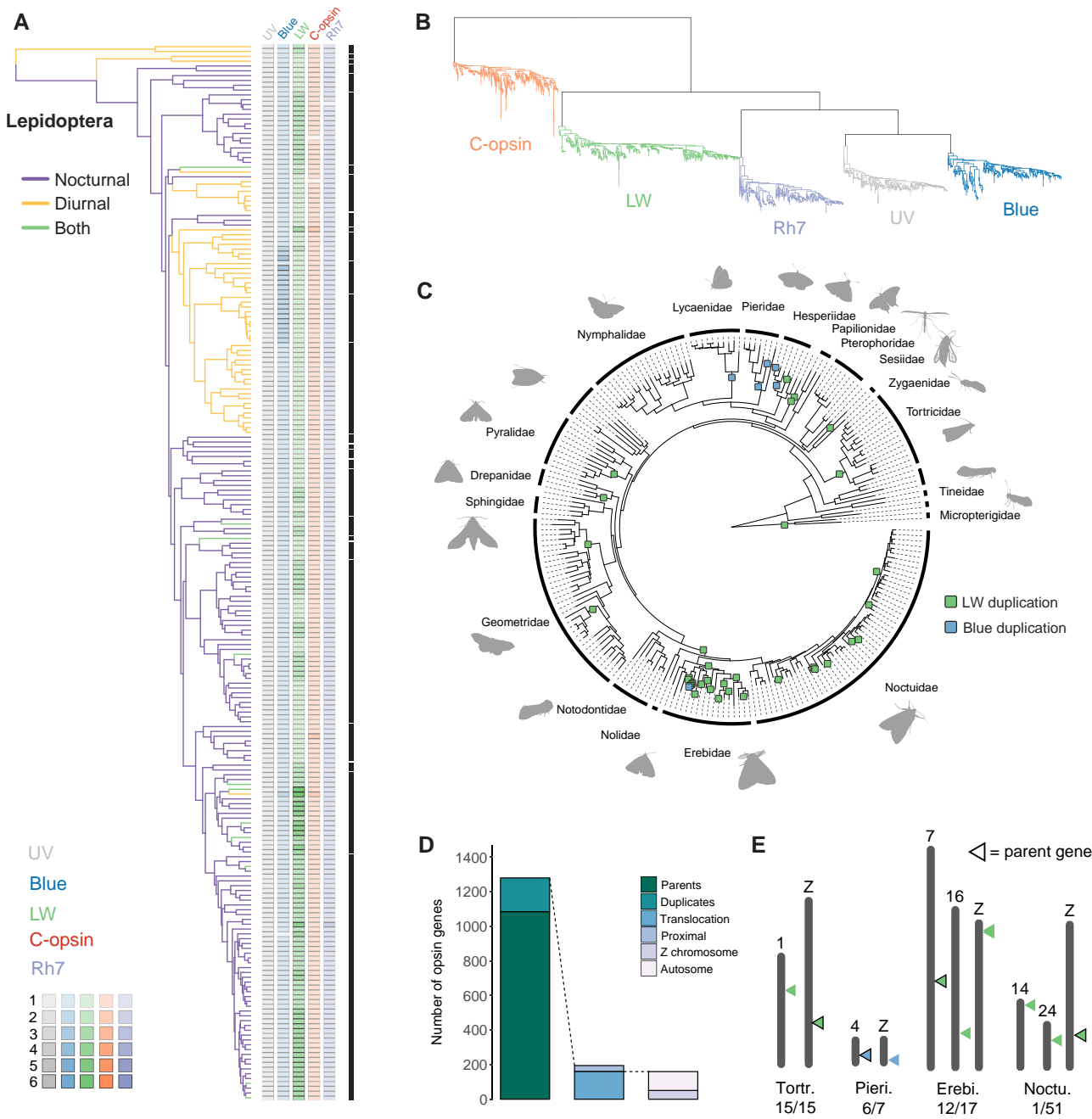
Fig. 1. Opsin diversity and copy number across Lepidoptera. a) Species phylogeny (left) with branches representing day/night flight activity of adults, along with copy number of each opsin gene type in each species (right). Each opsin gene has a distinct color, with the strength of the color representing copy number (ranges from 0 to 6 gene copies). Black bars represent lepidopteran families. Species names are given in supplementary fig. S1, Supplementary Material online. b) Gene tree of opsin genes present in Lepidoptera (amino acid sequences; maximum likelihood); each opsin gene groups into its own monophyletic clade. c) Species tree with families labeled with black bars, and blue and LW opsin gene duplication events mapped to their minimal branch of origin. d) Bar chart representing the proportion of opsin genes that (i) underwent duplication in Lepidoptera, (ii) duplicated near the parent gene or were translocated to a new genomic location, and (iii) the number of translocation events resulting in duplicate copy placed on another autosomal chromosome versus sex chromosome (Z or W). e) Opsin genes that translocated from an autosome to a sex chromosome mapped to chromosome graphs for each family where this occurred (Tortricidae, Pieridae, Erebidae, and Noctuidae). Numbers show the number of species within the family that share this sex-linked opsin gene.

Hesperia comma, and *Ochlodes sylvanus*. We find it originated from a tandem duplication event in the ancestor of these species in Hesperiidae, and that there was also a further tandem duplication of the blue opsin in *O. sylvanus* (Fig. 1c, supplementary fig. S3, Supplementary Material

online). We also uncover a novel duplication of the blue opsin in the day-flying Mother Shipton moth, *Euclidia mi* (Noctuoidea: Erebidae), which was the only example of duplication of this gene outside of the butterflies. This additional blue opsin originated by tandem duplication from

the parent opsin gene and has transcriptional orientation in the opposite direction (supplementary fig. S3, Supplementary Material online).

Duplication of the blue opsin in butterflies has previously been shown to lead to shifts in wavelength sensitivities in the paralogous copies resulting in finer wavelength discrimination (Wakakuwa et al. 2010; Liénard et al. 2021). In Lycaenidae, duplication of the blue opsin resulted in a typical blue opsin with λ_{max} of 435 to 440 nm and a green-shifted blue opsin with λ_{max} 495 to 500 nm (Liénard et al. 2021). In Pieridae, the duplication resulted in 1 normal blue opsin and 1 violet-shifted blue opsin with λ_{max} of 420 nm (Wakakuwa et al. 2010). Spectral tuning sites responsible for these shifts in wavelength peak absorbance have been functionally characterized (Wakakuwa et al. 2010; Liénard et al. 2021), with a substitution from serine to alanine at position 116 responsible for a 5- to 13-nm shift in peak absorbance in both Eumaeus atala (Lycaenidae) and Pieris rapae (Pieridae). Two further substitutions (G175S and Y177F) were combined to give a 73-nm bathochromic shift in E. atala (Liénard et al. 2021). We find the S116A substitution present in all paralogous blue copies within Lycaenidae, implying this spectral shifting substitution occurred once following the duplication event (supplementary fig. S4, Supplementary Material online). Within Pieridae, the S116A substitution is present in the closely related P. rapae, Pieris napi, and Anthocharis cardamines duplicated copies; however, it has reverted to alanine in Pieris brassicae. Within Hesperiidae, all blue opsin copies possess glycine at this site, while both blue copies in E. mi have the ancestral serine (supplementary fig. S4, Supplementary Material online).

We discovered a large number of gene duplication events involving the LW opsin, amounting to 34 duplications across the species tree (Fig. 1c). These include a gene duplication shared between both Micropterigidae species; a duplication shared between all 15 species analyzed in the family Tortricidae; 2 LW copies in the day-flying 6-spot burnet moth (Zygaena filipendulae) suggesting a recent duplication event; a gene duplication in the small skipper butterfly T. sylvestris; an LW duplication on the branch leading to the Pyralidae, with a duplicate subsequently lost in 2 species in this clade; a shared duplicate in the 2 Drepanidae species; 2 independent duplications within the Geometridae; a shared duplicate within the Noctuoidea superfamily (Nolidae, Erebidae, and Noctuidae families); and 23 subsequent duplication events of this paralog (see next section; Fig. 1c). Duplications of the c-opsin gene were also found in the European swallowtail butterfly (P. machaon), the Mother Shipton moth (E. mi), and the buff-tip moth (Phalera bucephala; Fig. 1a).

Of the opsin genes present in greater than 1 copy (199 genes in 130 species), 163 of the duplicate genes (82%) are present on a different chromosome to their parent gene of origin (Fig. 1d). This suggests a high rate of translocation following opsin duplication in Lepidoptera. There are 36 opsin genes that are in multicopy and present on the same chromosome, either in tandem or proximal locations to the parent opsin gene. Many of these pertain to evolutionarily recent duplication events, such as the c-opsin duplicate in P. machaon and LW duplicates in Spodoptera frugiperda, Laspeyria flexula, and Spilosoma lubricipeda. Some opsin paralogs, present in all species in a family, such as the LW duplication in Micropterigidae and the blue duplication in Lycaenidae (Fig. 1c), are likely much older yet have been retained in close genomic association following radiation of the clade. In the case of the 2 blue paralogs in all Lycaenidae species (10 species in our data set representing 2 out of 7 subfamilies), these genes are tightly linked with an average of ~6 kb of intergenic sequence separating the 2 paralogs (10 species, range ~3 to ~13 kb). In comparison, when we assess the intergenic distance between all paralogous gene pairs that are found on the same chromosome, in each of the 10 lycaenid species (see Materials and Methods), we find a median intergenic distance of ~16 kb (mean ~459 kb). This suggests a selective pressure to retain close linkage in the blue opsin paralogs in Lycaenidae.

Considering all opsin genes that underwent translocation following duplication, we find 51 opsin paralogs (from 36 species, representing 4 different taxonomic families) are located on the Z chromosome (Fig. 1d and e); we do not find any opsin genes located on the female-specific W chromosome in this data set (in Lepidoptera, males have ZZ sex chromosomes and females have ZW or occasionally Z0). This contrasts with a W-linked UV opsin gene in some Heliconius species giving female-specific expression in the heterogametic (ZW) sex (Chakraborty et al. 2023). We note, however, that only 59/219 of the genome assemblies in our data set were constructed from female individuals, so there could be hidden diversity of opsin genes on the W chromosome in Lepidoptera. Nonetheless, the fact that many duplicated opsin genes are located on the homogametic sex chromosome (Z in Lepidoptera) gives potential for sex-specific regulation.

We find that translocation to the Z chromosome following duplication may occur in 2 ways. The first is a consequence of an autosome-Z chromosome fusion event, which results in the opsin duplicate copy located on the neo-Z chromosome. This is the case for ancestral, parent LW opsin within the family Tortricidae and the parent LW opsin in the Noctuid species, Anorthoa munda. In both of these cases, an ancestral autosome containing an LW opsin copy was relocated to the Z chromosome as a result of a single chromosome fusion event (supplementary fig. S5, Supplementary Material online; Wright et al. 2023). The second case of Z-linked opsin genes occurs as a result of translocation of the individual opsin gene. For example, the blue opsin duplicate present in 6 out of the 7 species in the family Pieridae in our data set jumped to the Z chromosome, without any evidence for autosome-Z fusion (Wright et al. 2023). Interestingly, this translocation to the Z chromosome resulted in the duplicate blue opsin being located downstream and in the opposite orientation as a gene homologous to paraplegin (part of the AAA family

proteins) and in the same orientation as a gene that contains a 3′5′-cyclic nucleotide phosphodiesterase, catalytic domain, in all species except for *Aporia crataegi*. A direct role of cyclic nucleotide phosphodiesterases in phototransduction remains uncertain; however, they have been shown to localize in photoreceptor cells in the fly *Calliphora erythrocephala* (Schraermeyer et al. 1993). Another example of this type of translocation to the Z chromosome is seen in the LW opsin duplicate within the family Erebidae, present in 12/17 of the species in this family. This opsin paralog translocation was also likely due to an ancestral gene translocation event to the sex chromosome on the branch following the split from the subfamily Lymantriinae.

An Ancient, Conserved LW Retrocopy within the Noctuoidea Superfamily Shows Life Stage–Specific Expression

One of the most ancient cases of opsin duplication within the Lepidoptera occurred within the Noctuoidea superfamily (at the base of the Nolidae, Erebidae, and Noctuidae families), where the LW opsin underwent duplication (Fig. 1). Previously described in an erebid moth (Feuda et al. 2016) and a noctuid moth (Xu et al. 2016), we now suggest that this gene likely duplicated once, shared between 3 families within the superfamily Noctuoidea, by assessing the LW gene tree (Figs. 1b and 2a; supplementary fig. S6, Supplementary Material online), which shows 2 distinct monophyletic LW opsin groups, here named LWS1 and LWS2 (Fig. 2a). Given that this superfamily diverged ~80 million years ago (Kawahara et al. 2019), it is particularly striking that the 2 copies are retained in every species in our data set, consistent with functional divergence and selective retention. Further evidence for the potential adaptive benefit of this ancient LW duplication is the finding that further duplication events of this LWS2 gene occurred on 19 separate branches within Erebidae and Noctuidae, 13 of which are recent events (species specific in our data set; Fig. 1c). The LWS2 is intronless in all species (Feuda et al. 2016; Xu et al. 2016), compared to LWS1 that contains 7 introns, indicating that LWS2 is derived from LWS1 by retrotransposition (Betrán et al. 2002; Booth and Holland 2004; Kaessmann 2010). Further support for retrotransposition as the mechanism of origin is the fact that LWS2 is always located on a separate chromosome from LWS1.

Examining branch lengths in the gene tree of Noctuoidea LW opsins, it is clear that the protein encoded by intronless LWS2 genes is evolving at a faster rate than the protein encoded by the parent LWS1 gene (Fig. 2a). Assessing dN/dS values, we find increased global dN/dS ratio in the LWS2 gene relative to its parent gene LWS1, suggesting relaxation of purifying selection following the duplication event (Fig. 2b). This relaxation in selection was confirmed using a robust model-based approach, applying RELAX implemented in HyPhy (Wertheim et al. 2015; Kosakovsky Pond et al. 2020), which measures shifts in the stringency of selection acting on a gene; where a

value of $k > 1$ indicates intensified strength of selection, and $k < 1$ indicates relaxation of selection strength. Measuring shifts (i) on the branch leading to the LWS2 clade ($k = 0.63$, $P = 0.02$) and (ii) on all branches within the LWS2 clade of the gene tree ($k = 0.334$, $P = 0$), we find significant relaxation in the intensity of selection acting on the duplicate LWS2 gene relative to the background rate of the LWS1 clade (Fig. 2a). We next tested for positive selection in both LW copies in Noctuoidea using the branch-site model aBSREL implemented in HyPhy (Smith et al. 2015). Testing the branch leading to the LWS2 clade, and separately all branches within the LWS2 clade, we cannot identify amino acid sites with significant evidence for positive selection, relative to the background of LWS1. This suggests that, while there is relaxation of selection on the LWS2 copy, there is still overall selective pressure maintaining the function of this gene in all species.

The genomic location of the intronless LWS2 duplicate copy differs between taxonomic families within Noctuoidea (Nolidae, Erebidae, and Noctuidae; Fig. 2c, supplementary fig. S7, Supplementary Material online). In the 2 Nolidae species (*Meganola albula* and *Nycteola revayana*), the LWS2 retrogene is present in conserved syntenic regions, located downstream and in the same orientation as a leucine-rich repeat domain-containing gene (LRR-8), homologous to *Connectin* in *Drosophila melanogaster* (Fig. 2c, supplementary fig. S7, Supplementary Material online). In the 17 Erebidae species, the LWS2 gene is present in a different genomic location, within the intron of another gene (a Rab GTPase gene) in all species (Fig. 2c and d right; supplementary figs. S7 and S8, Supplementary Material online). As noted above, this LWS2 locus underwent an additional duplication and subsequent translocation in some Erebidae species (Fig. 1c), with translocation resulting in LWS2 copies on the Z chromosome of 12 species in our data set. Of these, 8 have multiple copies of LWS2 on the Z chromosome suggesting further tandem duplication following the translocation event. In the 51 Noctuidae species, the LWS2 copy is always present in the same syntenic region in all species in the family, downstream and in the same orientation of an ATP bind1 domain-containing gene (Fig. 2c, supplementary fig. S7, Supplementary Material online).

It was previously observed that the intronless LWS2 duplicate gene was more highly expressed in the first instar larval stage of the noctuid moth, *Helicoverpa armigera*, while the parental LWS1 gene has higher expression in the adult (Xu et al. 2016). To assess the consistency of this pattern, we measured the expression level of all opsin genes in larval and adult stages in an Erebidae species, *Orgyia antiqua* (vapourer moth). RNA extraction was carried out on the first instar larvae of *O. antiqua*, as well as heads of adult male and adult female *O. antiqua* (see Materials and Methods). We found that the LWS1 gene is highly expressed in the adult stage while the duplicate LWS2 gene is most highly expressed in the first instar larval
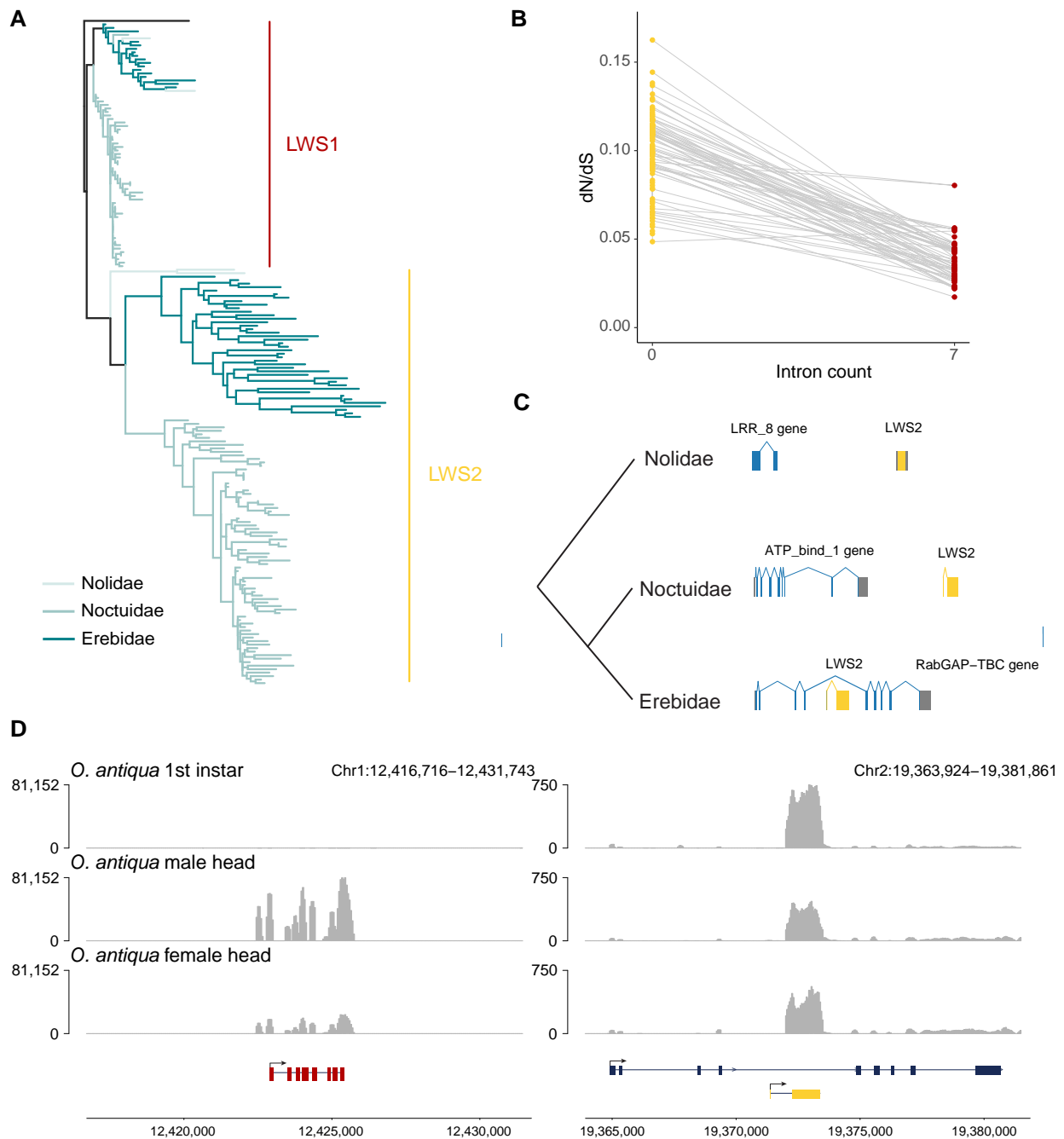
FIG. 2. Evolution and expression of LW duplicate within Noctuoidea. a) Gene tree of LW opsin genes in Noctuoidea families Nolidae, Erebidae, and Noctuidae, with branches colored according to family. LWS1 represents the ancestral parent LW opsin gene; LWS2 represents LW paralog duplicated at the base of the Noctuoidea superfamily. b) dN/dS values of both parent and duplicate LW copies in Noctuoidea, and intronless copy represents the LW paralog while the parent LW opsin contains the normal 7 introns. Gray lines linking points (LW genes) represent parent–duplicate copies in a single species. c) Syntenic locations of the LWS2 genes in Nolidae, Noctuidae, and Erebidae. LWS2 gene is shown in yellow, and the closest upstream gene in the same orientation is shown in blue. Blocks in genes represent exons, while lines joining them represent introns. d) Expression levels of parent (left) and duplicated (right) LW copy in *O. antiqua* (Erebidae). Tracks represent depth of mapped RNA reads on the genome. The top track shows expression levels for the first instar larva whole body, the second track shows the male head, and the third track shows the female head. Chromosome locations are given for both copies, and the gene structures (including orientation and exons/intron) are shown below.

stage (Fig. 2d, supplementary fig. S9, Supplementary Material online). This conservation of expression domains, combined with the selective pressure analyses, suggests functional importance of the LWS2 gene in early larval stages of erebid and noctuid moths.

Evidence of Divergent Molecular Evolution in Opsins of Day-Flying Species

To investigate the patterns of molecular evolution of opsin genes relative to the nocturnal or diurnal behavior of the lepidopteran species, we carried out tests to measure the
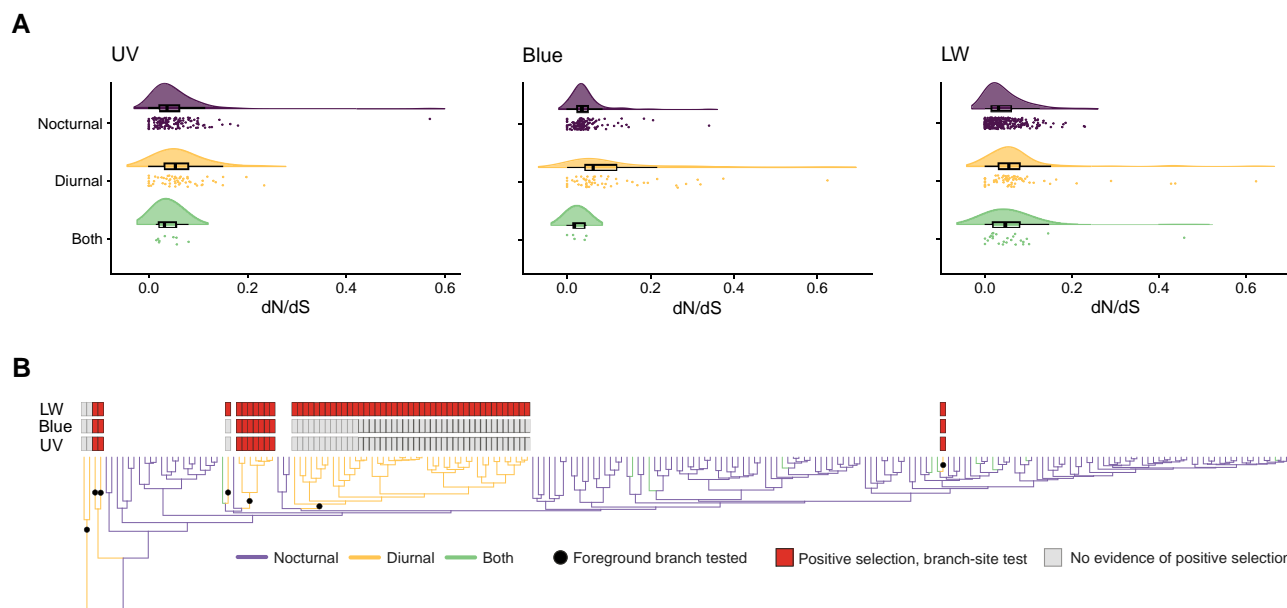
Fig. 3. Levels of selection acting on opsin genes in species with different photic niches. a) Level of selective pressure (dN/dS) acting on opsin genes in species with nocturnal activity, diurnal activity, or evidence for both. Raincloud plots show dN/dS values for all opsin genes in each species in the data set. b) Results from branch-site test on the 3 opsins. Species tree below has branches colored by photic niche of each species, and black circles represent foreground branches tested for evidence of positive selection. Gray bars indicate no positive selection found in tested foreground branches, while red bars show opsin genes with some evidence of positive selection in the foreground branches.

strength and form of selection acting on each gene. First, we classified each species in our data set based on whether the adults were nocturnal (152 species), diurnal (56), or have evidence for being both day and night flying (11; Fig. 1a, supplementary fig. S1, Supplementary Material online). For each of the opsin genes with functions directly related to vision (UV, blue, and LW), we measured rates of synonymous and nonsynonymous substitutions, as well as their ratio (omega = dN/dS), first using a codon model where omega does not vary across sites or branches in the tree, as implemented in HyPhy (Muse and Gaut 1994; Kosakovsky Pond et al. 2020). Comparing these values between the 3 lifestyle classes, we find that diurnal species have significantly higher dN/dS values in each visual opsin gene compared to the other 2 lifestyle categories (UV $P = 0.04146$; blue $P = 3.591e-06$; LW $P = 0.0001309$; Fig. 3a), suggesting different selective pressures acting on the opsin genes in diurnal lepidopteran species.

To investigate further, we then tested for positive selection acting on all branches within the clades containing day-flying species in the orthologous opsin genes using the BUSTED-PH model in HyPhy (Kosakovsky Pond et al. 2020). This tests for episodic diversifying selection on foreground branches with the same trait where there is no evidence of positive selection on the background branches. For the blue and LW opsins, we find evidence that selective pressures are significantly different between the background (nocturnal) and foreground (diurnal) branches; however, there is evidence for selection on both sets of branches. The UV opsin showed no evidence of episodic diversifying selection in the tested branches. Finally, to

test for positive selection acting on specific sites, we applied both the random site model and the branch-site model implemented in PAML (Yang 2007) to the UV, blue, and LW genes, with transitional branches to diurnal species labeled as foreground. We tested 7 independent transitions from nocturnal to diurnal species (Fig. 3b, supplementary fig. S1, Supplementary Material online). The branch-site model revealed positive selection in all 3 opsin genes, with the LW opsin in particular having the highest number of sites with evidence for change due to positive selection, with 31 unique sites in total under positive selection in all branches tested (Fig. 3b, supplementary fig. S10, Supplementary Material online). In the case of the LW and blue opsins, only one of the positively selected sites overlaps with known retinal-binding sites and may directly affect spectral tuning (supplementary fig. S10, Supplementary Material online). There are positively selected sites in all 3 of the tested opsins that map to the transmembrane regions near the retinal-binding sites (supplementary fig. S10, Supplementary Material online), which may have indirect impact on opsin spectral sensitivity. Further experimental work will be required to test whether these impact opsin spectral tuning.

While we find evidence that rates of sequence evolution are different in day-flying species, a phylogenetic correlation test did not find any evidence that gene duplication of the visual opsin genes was associated with diel niche (Pagel's λ, $P = 0.66$). This suggests that opsin duplication alone does not predict photic niche in a species but may have a wider range of implications, such as subfunctionalization of expression or changes in sensory roles outside of vision.

Discussion

The Lepidoptera provide a particularly interesting group to study the evolution of vision-related genes given the changes in behavior through the lifecycle and the number of independent transitions from nocturnal to diurnal activity of adults. Color discrimination is also known to differ between families and, in some cases, even between closely related species or sexes of the same species. Shifts in sensitivities to different wavelengths can occur through a range of different mechanisms, including structural changes to ommatidia, changes to opsin expression in specific photoreceptor cells, and processes of molecular evolution such as spectral-shifting amino acid substitutions in opsin genes. The role of opsin gene duplication and loss is also relevant and has been studied across arthropods, revealing some dramatic cases of gene turnover (Futahashi et al. 2015; Sharkey et al. 2017; Almudi et al. 2020; McCulloch, Macias-Muñoz, and Briscoe 2022). Recent advances in characterizing the spectral sensitivity of opsin proteins in vitro provides a promising platform to pinpoint the functional role of opsin duplication and loss in driving variation in visual sensitivities (Liénard et al. 2021, 2022).

A major question in the field is whether there is a relation between diel activity and opsin gene evolution. Here, we exploited newly generated high-quality genome assemblies to compare the opsin genes between 219 species of Lepidoptera, representing a broad phylogenetic span. Overall, we find a conserved opsin gene complement across Lepidoptera, with only 1 putative loss in the visual opsin genes: loss of the blue opsin in the twin-spotted Quaker moth (*A. munda*). We uncover 44 cases of duplication in the opsin genes, 40 of which affect the blue (short wave sensitive) and LW (long wave sensitive) visual opsin genes but no cases of duplication of the UV opsin. Among the duplicated genes, we find 4 cases of translocation of an opsin paralog to the Z chromosome following duplication. These events have the potential to lead to sex-linked expression differences related to dosage effects between homogametic males (ZZ) and heterogametic females (ZW). Whether such potential is realized may be dependent on the mechanism of transposition: either by chromosome fusion or movement of a small genomic region. Following autosome fusion to an ancestral Z chromosome, the distinct regions of the composite lepidopteran sex chromosome retain their respective patterns of dosage compensation, with the (now transferred) neo-Z region having a 2-fold increase in transcription in ZZ females (Gu et al. 2019). We suggest this may be the case for the Tortricidae LW gene duplicate, which we find was translocated to the Z chromosome through autosomal fusion in the ancestor of this family. In contrast, the translocation of the blue opsin duplicate to the Z chromosome in Pieridae was not a result of chromosomal fusion but rather a translocation of a DNA locus; we suggest this newly Z-linked opsin is likely to have dosage balance between sexes. It is also notable that this single gene translocation resulted in the blue paralog located downstream of a 3′5′-cyclic nucleotide phosphodiesterase gene. The orthologous gene in the fly *C. erythrocephala* is expressed in photoreceptor cells (Schraermeyer et al. 1993) and in *D. melanogaster*, it is expressed in the sensory system (manually curated expression from FlyBase FB2023_05 release [Gramates et al. 2022]), and thus, this locus may share regulatory sequences with the translocated opsin gene.

Of the 44 opsin duplication events identified, 16 occurred within lineages containing day-flying species. The majority of Lepidoptera are nocturnal with day-flying being a derived trait in most, if not all, cases. Opsin duplication in day-flying species shows some interesting lineage-specific patterns, such as an LW duplication shared between the Micropterigidae species, LW duplication in the 6-spot burnet moth (*Z. filipendulae*), blue opsin duplication in the Mother Shipton moth (*E. mi*), and 5 independent duplications of the blue opsin within the butterflies. These duplications could provide a novel source of variation to generate shifts in visual spectra for these species (Wakakuwa et al. 2010; Liénard et al. 2021). However, while these are interesting individual candidates for opsin duplication related to diurnality, we did not find a significant correlation between opsin duplication and diel activity across our species phylogeny (Pagel's λ, $P = 0.66$), suggesting that opsin duplicates may have been coopted for a range of possible roles relating to vision and other sensory functions (Feuda et al. 2022). One such case of opsin duplication not directly related to diurnal behavior is an ancient LW duplication, which likely occurred at the base of 3 families (Nolidae, Erebidae, and Noctuidae) within the superfamily Noctuoidea. This event, which occurred via retrotransposition (all copies possess no introns), generated a new LW opsin paralog, LWS2, which was then retained in every species in our data set. Differential expression of the 2 LW copies between life stages has now been shown in 2 species from different families (Noctuidae and Erebidae) and likely represents the ancestral state, although more data for species in the Nolidae family are required to confirm this. This route to subfunctionalization of opsin paralogs through changes in temporal or spatial expression has been found in other insects, for example in dragonflies, damselflies, mayflies, mosquitos, and other lepidopterans (Futahashi et al. 2015; Giraldo-Calderón et al. 2017; Almudi et al. 2020; Kuwalekar et al. 2022; Roberts et al. 2023). In many of these other cases, the larval and adult stages occupy very different niches and light environments (larvae are aquatic while adults are terrestrial). This is not the case for moth species within the Noctuoidea superfamily (most larvae and adults are nocturnal). Interestingly, we note that early instar larvae of Noctuidae and Erebidae are highly active and have rapid dispersal while some other larvae show more sedentary behavior in early instars. Whether the function of the LWS2 retrocopy is linked to this behavior is unknown.

While the exact function of LWS2 requires further analyses, it highlights the role of retrotransposition in generating new gene copies, with the potential for different functions (Kaessmann 2010; Carelli et al. 2016).

Retrotransposition of opsin genes has also been noted in Diptera, where the *Rh3* gene (UV-sensitive opsin) and *Rh6* gene (LW sensitive) originated via this mechanism in the *Drosophila* genus and different mosquito lineages, respectively (Giraldo-Calderón et al. 2017; Feuda et al. 2021). Retrogenes are easy to distinguish from their parent copy due to the missing introns but, importantly, from an evolutionary perspective, will not inherit promoters or most regulatory elements (Kaessmann 2010). This suggests that shifts in expression patterns are likely and could be driven by genomic features around the new insertion site in the genome (Carelli et al. 2016). In the case of the LWS2 retrogene, it is unclear where the original insertion site was because the retrogene has a different genomic location in the 3 extant families examined. In Erebidae, LWS2 is located within the intron of another gene (Fig. 2c, supplementary fig. S8, Supplementary Material online), a RabGAP-TBC domain-containing gene that we propose is homologous to human TBC1 domain family member 20 (TBC1D20). Interestingly, TBC1D20, along with 5 other TBC domain proteins, are crucial for trafficking G protein–coupled receptors (GPCRs), to which opsin genes belong, from the endoplasmic reticulum, through the Golgi apparatus, to their final location in the plasma membrane (Wei et al. 2019). Integration into the intron of this host gene may have provided crucial transcriptional as well as functional integrity to the duplicate LW copy within Erebidae. We note, however, that while LWS2 is expressed strongly in early instar larvae of the vapourer moth *O. antiqua*, we find relatively low expression of the TBC1D20 gene (Fig. 2d).

Conclusion

We uncover extensive and previously underappreciated opsin gene duplication and evolutionary change in the Lepidoptera. Due to broad sampling of high-quality genomes, we could confirm gene losses and the timing of duplication events using gene structure and genome synteny. We find different modes of gene duplication, including retrotransposition and tandem duplication, both sometimes followed by translocation or chromosomal rearrangement. These events have provided opportunities for substitution accumulation and sequence divergence, which have likely increased the transcriptional and functional diversity of opsin genes in this group (Lynch and Conery 2000; Kaessmann 2010). For example, we find evidence for transcriptional divergence between LW opsin genes in different lifecycle stages of some moths and evidence for differential selective pressures acting on the opsin genes of day-flying moths and butterflies.

Materials and Methods

Data Acquisition

The majority of the genomes used in this analysis were produced by the Darwin Tree of Life Project (The Darwin Tree of Life Project Consortium 2022), which can be found on the Darwin Tree of Life (DToL) portal page (https://portal.darwintreeoflife.org) or under accession number PRJEB40665 in the European Nucleotide Archive (ENA; https://www.ebi.ac.uk/ena/browser/home). The genomes from the remaining additional lepidopteran species were obtained from NCBI. A list of all species and their associated genomes and sources can be found in supplementary table S1, Supplementary Material online.

Species Tree Reconstruction

In order to map the presence, absence, and copy number of the opsin genes to the species, we required a species tree. Species tree reconstruction was carried out using a data set of 1,465 genes annotated with BUSCO v5.1.2 (Manni et al. 2021) using the Lepidoptera gene set. Each gene annotated this way contained all species sampled and was aligned with MAFFT v7.4 (Katoh et al. 2005) and trimmed with trimAl (Capella-Gutiérrez et al. 2009) before concatenation into a supermatrix using the create_concatenation_matrix option in PhyKIT (Steenwyk et al. 2021). The species tree was inferred from this supermatrix using IQ-TREE v2.0 (Minh et al. 2020) and the LG model with a gamma distribution with 4 categories. Tree visualization was carried out using Toytree (Eaton 2020) and ggtree (Yu et al. 2017).

Opsin Gene Annotation

A total of 615 protein sequences obtained from 2 studies on opsin evolution in Diptera (Feuda et al. 2021) and Lepidoptera (Sondhi et al. 2021) were used as seeds in an initial tBLASTn search of the lepidopteran genomes used in this analysis. ORF of opsins discovered by the BLAST search were constructed using Exonerate v2.4 (Slater and Birney 2005). Next, to ensure accurate annotation of the full ORF for each opsin gene, we reran the tBLASTn and exonerate pipeline this time using the newly found opsin genes that had a start codon and appropriate protein length. Opsin genes were assigned to a specific type (UV, blue, LWS, c-opsin, or RH7) based on our initial BLAST and exonerate annotation, as well as a subsequent BLASTp search against the nr BLAST database. Finally, accurate classification of the opsin genes was confirmed by building a gene tree of all opsins (see next section) and ensuring each annotated opsin grouped in the correct clade in the tree.

Within-Species Paralog Gene Pair Linkage

In order to test the significance of the close linkage between the blue opsin paralog gene pairs in the Lycaenidae species (mean ~6 kb intergenic distance between gene pairs across 10 species), we carried out a systematic search of all paralogous gene pairs in the 10 species. First, to find paralogous gene pairs for a given species, we carried out a BLASTp search of the species proteome against itself. The top, nonself hit for each gene was extracted, and these were filtered to obtain a set of paralogous gene pairs for each species. We filtered potential

paralogous gene pairs by (i) ensuring a potential paralogous gene pair each had a best reciprocal hit (i.e. the paralog reciprocally matched the same gene that had it as a top hit), (ii) paralogous gene pairs were located on the same chromosome, (iii) query of subject coverage was greater than or equal to 70%, and (iv) sequence percent identity was greater than or equal to 65% (which was the average percent identity between the blue opsin paralog pairs in each lycaenid species). This strict filtering ensured we were comparing within-species paralogous gene pairs that were most similar to the blue opsin gene pairs. This resulted in varying numbers of paralogous gene pairs per species, ranging from 501 in *Celastrina argiolus* to 1,963 gene pairs within *Glaucopsyche alexis*. These paralogous gene pairs will be a mixture of old duplication events as well as recent, species-specific events; however, ensuring similar coverage and sequence identity as is found in the blue opsin paralogs will ensure these will serve as a good proxy for measuring the rate of gene linkage. Intergenic distance was measured by taking the distance between the end of the first paralog gene and the start of the other, and this was averaged across all species analyzed.

Opsin Gene Tree Inference

Inference of the opsin gene tree was important for a number of parts of this study. All inferred opsin genes were aligned using MAFFT v7.4 (Katoh et al. 2005). To identify the relationships between the opsin genes, we performed tree inference using maximum likelihood with IQ-Tree (Minh et al. 2020) applying ModelFinder to find the model of best fit (Kalyaanamoorthy et al. 2017). The gene trees were used to confirm correct annotation of the opsin genes by the sequence homology pipeline described above. Additionally, the patterns in the complete opsin gene tree were used as 1 line of evidence to infer when duplication events occurred along the species tree, by carrying out a manual inspection of the gene tree to reconcile the gene tree with the species tree. Finally, individual gene trees were also constructed for each of the opsin types using the same approach as above, and these were used when carrying out the tests for selective pressure (see next section).

Selective Pressure Analyses

Tests for selection were carried out on the opsin genes in a number of different analyses. First, regarding the LW duplication event in Noctuoidea, we calculated rates of synonymous substitutions (dS) and nonsynonymous substitution (dN) per site, as well as their ratio (dN/dS), for both copies of this gene (i.e. LWS1 and LWS2 as described in the Results section). First, protein sequences for all copies of LWS1 and LWS2 in Noctuoidea species were aligned using MAFFT v7.4 (Katoh et al. 2005). Next, codon alignments were created using the protein alignments and corresponding nucleotide sequences as input to PAL2NAL (Suyama et al. 2006). dN/dS was calculated for all copies of the Noctuoidea LW genes using the

Muse–Gaut (MG94) model (Muse and Gaut 1994) (–type local) implemented in HyPhy (Kosakovsky Pond et al. 2020) and the Noctuoidea LWS gene tree (Fig. 2a). We also carried out a test for signatures of relaxation or intensification of selection, with the branches leading to and within the LWS2 clade set as the foreground and the LWS1 clade set as the background. To test this, we employed the RELAX model (Wertheim et al. 2015) in HyPhy (Kosakovsky Pond et al. 2020). Finally, the aBSREL model in HyPhy (Kosakovsky Pond et al. 2020) was employed to test for evidence of positive selection on any branches in the Noctuoidea LW gene tree. This model, which does not require a priori partitioning or selection of branches on the phylogeny, estimates dN/dS on all branches of the tree (Kosakovsky Pond et al. 2011).

We used several tests to measure selection within the day-flying lineages relative to nocturnal species in our data set. First, we calculated the dN/dS for each opsin gene in each species using the Muse–Gaut (MG94) model in HyPhy (Kosakovsky Pond et al. 2020) (–type local). Each dN/dS value for every species was used to summarize the general patterns of selection between day-flying species, night-flying species, and species with evidence of both. These summarized group values were visualized using Raincloud plots in R (Allen et al. 2019). Next, we used the BUSTED-PH test within HyPhy (Kosakovsky Pond et al. 2020) with all day-flying branches in the tree labeled as foreground test branches. Finally, we used codon based models employed in codeml (Yang and Nielsen 2002) using the pipeline Vespasian (zenodo.org/record/5779869; github.com/bede/vespasian), to test for evidence of episodic events of positive and divergent selection on selected branches leading to day-flying lineages in the lepidopteran phylogeny. We employed both site and branch-site models, comparing standard nested models using likelihood ratio tests, as implemented in Vespasian. Sites found to be under positive selection were mapped to protein models by predicting transmembrane helices for all 3 opsins using Phobius (Käll et al. 2007) through the webserver Protter (Omasits et al. 2014). Retinal-binding sites were inferred by including data from the jumping spider rhodopsin-1 into the model (Varma et al. 2019).

RNA Sequencing and Opsin Expression Quantification

A captive-reared female *O. antiqua* was mated with a wild caught male; resultant fertilized eggs were maintained over winter at ambient temperature (UK). On the day of larval hatching, 30 first instar larvae were homogenized using a sterile needle in RNAProtect; total RNA was purified several days later using a QIAGEN RNAeasy Plus Micro kit. Eggs from the same mating were reared until adult emergence the same year; 1 male and 1 female head were processed for RNA extraction directly using an QIAGEN RNAeasy Plus Micro kit, after removal of antennae. Paired-end 150 bp Illumina RNA-seq was performed commercially by Novogene (www.novogene.com) using

poly-A selection, random hexamer priming, and amplification giving 5.5 to 7.9 Gb of sequence per sample (NCBI SRA accessions SRX20677428, SRX20677423, and SRX20677422).

RNA-seq data from the 3 samples (first instar larvae, male head, and female head) were trimmed for quality using Trimmomatic v0.39 (Bolger et al. 2014). Next, transcriptome assembly was performed for each sample using Trinity v2.8.5 (Grabherr et al. 2011), and transcript abundance was calculated using kallisto v0.44 (Bray et al. 2016). Each opsin gene was identified by performing a reciprocal BLAST search, and gene expression abundance was measured using transcripts per million (TPM). Additionally, processed RNA reads were mapped to the *O. antiqua* genome using bowtie2 (Langmead and Salzberg 2012) in order to visualize RNA read depth and expression of the LW opsin genes. Read coverage was quantified using bedtools v2.25.0 (Quinlan and Hall 2010), and gene tracks and RNA coverage were visualized using trackplot (github.com/PoisonAlien/trackplot; Pohl and Beato 2014).

Supplementary Material

Supplementary material is available at *Molecular Biology and Evolution* online.

Acknowledgments

Funding

Data Availability

All data and code required to reproduce analyses and figures can be found in the Supplementary Material online and at GitHub (github.com/PeterMulhair/Lepidoptera_opsins).

References

Allen M, Poggiali D, Whitaker K, Marshall TR, Kievit RA. Raincloud plots: a multi-platform tool for robust data visualization. *Wellcome Open Res.* 2019:**4**:63. https://doi.org/10.12688/wellcomeopenres.15191.1.

Almudi I, Vizueta J, Wyatt CDR, de Mendoza A, Marlétaz F, Firbas PN, Feuda R, Masiero G, Medina P, Alcaina-Caro A, et al. Genomic adaptations to aquatic and aerial life in mayflies and the origin of insect wings. *Nat Commun.* 2020:**11**(1):2631. https://doi.org/10.1038/s41467-020-16284-8.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990:**215**(3):403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.

Arikawa K. Spectral organization of the eye of a butterfly, *Papilio. J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol.* 2003:**189**(11):791–800. https://doi.org/10.1007/s00359-003-0454-7.

Arikawa K. The eyes and vision of butterflies. *J. Physiol.* 2017:**595**(16):5457–5464. https://doi.org/10.1113/JP273917.

Armisén D, Rajakumar R, Friedrich M, Benoit JB, Robertson HM, Panfilio KA, Ahn S-J, Poelchau MF, Chao H, Dinh H, et al. The genome of the water strider Gerris buenoi reveals expansions of gene repertoires associated with adaptations to life on the water. *BMC Genomics* 2018:**19**(1):832. https://doi.org/10.1186/s12864-018-5163-2.

Awata H, Wakakuwa M, Arikawa K. Evolution of color vision in pierid butterflies: blue opsin duplication, ommatidial heterogeneity and eye regionalization in *Colias erate. J Comp Physiol A Neuroethol Sens Neural Behav Physiol.* 2009:**195**(4):401–408. https://doi.org/10.1007/s00359-009-0418-7.

Bernard GD, Remington CL. Color vision in Lycaena butterflies: spectral tuning of receptor arrays in relation to behavioral ecology. *Proc Natl Acad Sci U S A.* 1991:**88**(7):2783–2787. https://doi.org/10.1073/pnas.88.7.2783.

Betrán E, Thornton K, Long M. Retroposed new genes out of the X in Drosophila. *Genome Res.* 2002:**12**(12):1854–1859. https://doi.org/10.1101/gr.604902.

Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014:**30**(15):2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

Booth HAF, Holland PWH. Eleven daughters of NANOG. *Genomics* 2004:**84**(2):229–238. https://doi.org/10.1016/j.ygeno.2004.02.014.

Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol.* 2016:**34**(5):525–527. https://doi.org/10.1038/nbt.3519.

Briscoe AD. Reconstructing the ancestral butterfly eye: focus on the opsins. *J Exp Biol.* 2008:**211**(11):1805–1813. https://doi.org/10.1242/jeb.013045.

Briscoe AD, Bybee SM, Bernard GD, Yuan F, Sison-Mangus MP, Reed RD, Warren AD, Llorente-Bousquets J, Chiao C-C. Positive selection of a duplicated UV-sensitive visual pigment coincides with wing pigment evolution in *Heliconius* butterflies. *Proc Natl Acad Sci U S A.* 2010:**107**(8):3628–3633. https://doi.org/10.1073/pnas.0910085107.

Briscoe AD, Chittka L. The evolution of color vision in insects 2003. Available from: https://www.annualreviews.org/doi/abs/10.1146/annurev.ento.46.1.471.

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. Trimal: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 2009:**25**(15):1972–1973. https://doi.org/10.1093/bioinformatics/btp348.

Carelli FN, Hayakawa T, Go Y, Imai H, Warnefors M, Kaessmann H. The life history of retrocopies illuminates the evolution of new mammalian genes. *Genome Res.* 2016:**26**(3):301–314. https://doi.org/10.1101/gr.198473.115.

Chakraborty M, Lara AG, Dang A, McCulloch KJ, Rainbow D, Carter D, Ngo LT, Solares E, Said I, Corbett-Detig RB, et al. Sex-linked gene traffic underlies the acquisition of sexually dimorphic UV color vision in *Heliconius* butterflies. *Proc Natl Acad Sci U S A.* 2023:**120**(33):e2301411120. https://doi.org/10.1073/pnas.2301411120.

Chen P-J, Awata H, Matsushita A, Yang E-C, Arikawa K. Extreme spectral richness in the eye of the common bluebottle butterfly, *Graphium sarpedon. Front Ecol Evol.* 2016:**4**:18. https://doi.org/10.3389/fevo.2016.00018.

Eaton DAR. Toytree: a minimalist tree visualization and manipulation library for Python. *Methods Ecol. Evol.* 2020:**11**(1):187–191. https://doi.org/10.1111/2041-210X.13313.

Feuda R, Goulty M, Zadra N, Gasparetti T, Rosato E, Pisani D, Rizzoli A, Segata N, Ometto L, Stabelli OR. Phylogenomics of opsin genes in Diptera reveals lineage-specific events and contrasting evolutionary dynamics in *Anopheles* and *Drosophila*. *Genome Biol Evol.* 2021:**13**(8):evab170. https://doi.org/10.1093/gbe/evab170.

Feuda R, Marlétaz F, Bentley MA, Holland PWH. Conservation, duplication, and divergence of five opsin genes in insect evolution. *Genome Biol Evol.* 2016:**8**(3):579–587. https://doi.org/10.1093/gbe/evw015.

Feuda R, Menon AK, Göpfert MC. Rethinking opsins. *Mol Biol Evol.* 2022:**39**(3):msac033. https://doi.org/10.1093/molbev/msac033.

Finkbeiner SD, Briscoe AD. True UV color vision in a female butterfly with two UV opsins. *J Exp Biol.* 2021:**224**(18):jeb242802. https://doi.org/10.1242/jeb.242802.

Finkbeiner SD, Fishman DA, Osorio D, Briscoe AD. Ultraviolet and yellow reflectance but not fluorescence is important for visual discrimination of conspecifics by *Heliconius erato*. *J Exp Biol.* 2017:**220**(Pt\ 7):1267–1276. https://doi.org/10.1242/jeb.153593.

Frentiu FD, Bernard GD, Sison-Mangus MP, Brower AVZ, Briscoe AD. Gene duplication is an evolutionary mechanism for expanding spectral diversity in the long-wavelength photopigments of butterflies. *Mol Biol Evol.* 2007:**24**(9):2016–2028. https://doi.org/10.1093/molbev/msm132.

Futahashi R, Kawahara-Miki R, Kinoshita M, Yoshitake K, Yajima S, Arikawa K, Fukatsu T. 2015. Extraordinary diversity of visual opsin genes in dragonflies. *Proc Natl Acad Sci U S A.* **112**(11):E1247–E1256. https://doi.org/10.1073/pnas.1424670112.

Giraldo-Calderón GI, Zanis MJ, Hill CA. Retention of duplicated long-wavelength opsins in mosquito lineages by positive selection and differential expression. *BMC Evol Biol.* 2017:**17**(1):84. https://doi.org/10.1186/s12862-017-0910-6.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol.* 2011:**29**(7):644–652. https://doi.org/10.1038/nbt.1883.

Gramates LS, Agapite J, Attrill H, Calvi BR, Crosby MA, Dos Santos G, Goodman JL, Goutte-Gattat D, Jenkins VK, Kaufman T, et al. Flybase: a guided tour of highlighted features. *Genetics* 2022:**220**(4):iyac035. https://doi.org/10.1093/genetics/iyac035.

Gu L, Reilly PF, Lewis JJ, Reed RD, Andolfatto P, Walters JR. Dichotomy of dosage compensation along the neo Z chromosome of the monarch butterfly. *Curr Biol.* 2019:**29**(23):4071–4077.e3. https://doi.org/10.1016/j.cub.2019.09.056.

Guignard Q, Allison JD, Slippers B. The evolution of insect visual opsin genes with specific consideration of the influence of ocelli and life history traits. *BMC Ecol Evol.* 2022:**22**(1):2. https://doi.org/10.1186/s12862-022-01960-8.

Hauser FE, Chang BS. Insights into visual pigment adaptation and diversity from model ecological and evolutionary systems. *Curr Opin Genet Dev.* 2017:**47**:110–120. https://doi.org/10.1016/j.gde.2017.09.005.

Hauser FE, van Hazel I, Chang BSW. Spectral tuning in vertebrate short wavelength-sensitive 1 (SWS1) visual pigments: can wavelength sensitivity be inferred from sequence data? *J. Exp. Zool. B Mol. Dev. Evol.* 2014:**322**(7):529–539. https://doi.org/10.1002/jez.b.22576.

Henze MJ, Oakley TH. The dynamic evolutionary history of pancrustacean eyes and opsins. *Integr Comp Biol.* 2015:**55**(5):830–842. https://doi.org/10.1093/icb/icv100.

Ilić M, Chen P-J, Pirih P, Meglič A, Prevc J, Yago M, Belušič G, Arikawa K. Simple and complex, sexually dimorphic retinal mosaic of fritillary butterflies. *Philos Trans R Soc Lond B Biol Sci.* 2022:**377**(1862):20210276. https://doi.org/10.1098/rstb.2021.0276.

Kaessmann H. Origins, evolution, and phenotypic impact of new genes. *Genome Res.* 2010:**20**(10):1313–1326. https://doi.org/10.1101/gr.101386.109.

Käll L, Krogh A, Sonnhammer ELL. Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. *Nucleic Acids Res.* 2007:**35**(Web Server):W429–W432. https://doi.org/10.1093/nar/gkm256.

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. Modelfinder: fast model selection for accurate phylogenetic estimates. *Nat Methods.* 2017:**14**(6):587–589. https://doi.org/10.1038/nmeth.4285.

Katoh K, Kuma K-I, Toh H, Miyata T. MAFFT Version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* 2005:**33**(2):511–518. https://doi.org/10.1093/nar/gki198.

Kawahara AY, Plotkin D, Espeland M, Meusemann K, Toussaint EFA, Donath A, Gimnich F, Frandsen PB, Zwick A, Dos Reis M, et al. Phylogenomics reveals the evolutionary timing and pattern of butterflies and moths. *Proc Natl Acad Sci U S A.* 2019:**116**(45):22657–22663. https://doi.org/10.1073/pnas.1907847116.

Kawahara AY, Plotkin D, Hamilton CA, Gough H, St Laurent R, Owens HL, Homziak NT, Barber JR. Diel behavior in moths and butterflies: a synthesis of data illuminates the evolution of temporal activity. *Org Divers Evol.* 2018:**18**(1):13–27. https://doi.org/10.1007/s13127-017-0350-6.

Kosakovsky Pond SL, Murrell B, Fourment M, Frost SDW, Delport W, Scheffler K. A random effects branch-site model for detecting episodic diversifying selection. *Mol Biol Evol.* 2011:**28**(11):3033–3043. https://doi.org/10.1093/molbev/msr125.

Kosakovsky Pond SL, Poon AFY, Velazquez R, Weaver S, Hepler NL, Murrell B, Shank SD, Magalis BR, Bouvier D, Nekrutenko A, et al. Hyphy 2.5—a customizable platform for evolutionary hypothesis testing using phylogenies. *Mol Biol Evol.* 2020:**37**(1):295–299. https://doi.org/10.1093/molbev/msz197.

Kuwalekar M, Deshmukh R, Baral S, Padvi A, Kunte K. Duplication and sub-functionalisation characterise diversification of opsin genes in the Lepidoptera. *bioRxiv* [Internet]:2022.10.31.514481 2022. Available from: https://www.biorxiv.org/content/10.1101/2022.10.31.514481v1.abstract.

Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 2012:**9**(4):357–359. https://doi.org/10.1038/nmeth.1923.

Lau LTS, Gross E. Sexual dimorphism and light/dark adaptation in the compound eyes of male and female *Acentria ephemerella* (Lepidoptera: Pyraloidea: Crambidae). *Behav Processes.* 2007:**104**(2):247. https://doi.org/10.14411/eje.2007.066.

Lau TF, Meyer-Rochow VB. The compound eye of *Orgyia antiqua* (Lepidoptera: Lymantriidae): sexual dimorphism and light/dark adaptational changes. *Eur J Entomol.* 2007:**104**(2):247. https://doi.org/10.14411/eje.2007.039.

Liénard MA, Bernard GD, Allen A, Lassance J-M, Song S, Childers RR, Yu N, Ye D, Stephenson A, Valencia-Montoya WA, et al. The evolution of red color vision is linked to coordinated rhodopsin tuning in lycaenid butterflies. *Proc Natl Acad Sci U S A.* 2021:**118**(6):e2008986118. https://doi.org/10.1073/pnas.2008986118.

Liénard MA, Valencia-Montoya WA, Pierce NE. Molecular advances to study the function, evolution and spectral tuning of arthropod visual opsins. *Philos Trans R Soc Lond B Biol Sci.* 2022:**377**(1862):20210279. https://doi.org/10.1098/rstb.2021.0279.

Lynch M, Conery JS. The evolutionary fate and consequences of duplicate genes. *Science* 2000:**290**(5494):1151–1155. https://doi.org/10.1126/science.290.5494.1151.

Macias-Muñoz A, Rangel Olguin AG, Briscoe AD. Evolution of phototransduction genes in Lepidoptera. *Genome Biol Evol.* 2019:**11**(8):2107–2124. https://doi.org/10.1093/gbe/evz150.

Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. BUSCO update: novel and streamlined workflows along with broader

and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol.* 2021:**38**(10): 4647–4654. https://doi.org/10.1093/molbev/msab199.

McCulloch KJ, Macias-Muñoz A, Briscoe AD. Insect opsins and evodevo: what have we learned in 25 years? *Philos Trans R Soc Lond B Biol Sci.* 2022:**377**(1862):20210288. https://doi.org/10.1098/rstb.2021.0288.

McCulloch KJ, Macias-Muñoz A, Mortazavi A, Briscoe AD. Multiple mechanisms of photoreceptor spectral tuning in *Heliconius* butterflies. *Mol Biol Evol.* 2022:**39**(4):msac067. https://doi.org/10.1093/molbev/msac067.

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol.* 2020:**37**(5):1530–1534. https://doi.org/10.1093/molbev/msaa015.

Muse SV, Gaut BS. A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. *Mol Biol Evol.* 1994:**11**(5):715–724. https://doi.org/10.1093/oxfordjournals.molbev.a040152.

Ni JD, Baik LS, Holmes TC, Montell C. A rhodopsin in the brain functions in circadian photoentrainment in *Drosophila*. *Nature* 2017:**545**(7654):340–344. https://doi.org/10.1038/nature22325.

Nilsson D-E. The evolution of eyes and visually guided behaviour. *Philos Trans R Soc Lond B Biol Sci.* 2009:**364**(1531):2833–2847. https://doi.org/10.1098/rstb.2009.0083.

Omasits U, Ahrens CH, Müller S, Wollscheid B. Protter: interactive protein feature visualization and integration with experimental proteomic data. *Bioinformatics* 2014:**30**(6):884–886. https://doi.org/10.1093/bioinformatics/btt607.

Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, Le Trong I, Teller DC, Okada T, Stenkamp RE, *et al.* Crystal structure of rhodopsin: a G protein-coupled receptor. *Science* 2000:**289**(5480):739–745. https://doi.org/10.1126/science.289.5480.739.

Perry M, Kinoshita M, Saldi G, Huo L, Arikawa K, Desplan C. Molecular logic behind the three-way stochastic choices that expand butterfly colour vision. *Nature* 2016:**535**(7611):280–284. https://doi.org/10.1038/nature18616.

Pohl A, Beato M. Bwtool: a tool for bigWig files. *Bioinformatics* 2014:**30**(11):1618–1619. https://doi.org/10.1093/bioinformatics/btu056.

Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 2010:**26**(6):841–842. https://doi.org/10.1093/bioinformatics/btq033.

Roberts NS, Svensson EI, Liénard MA. Opsin gene expression plasticity and spectral sensitivity as mechanisms for search image formation in mate-searching male damselflies. *bioRxiv* [Internet]: 2023.07.31.551331. 2023. Available from: https://www.biorxiv.org/content/10.1101/2023.07.31.551331v1.abstract.

Saito T, Koyanagi M, Sugihara T, Nagata T, Arikawa K, Terakita A. Spectral tuning mediated by helix III in butterfly long wavelength-sensitive visual opsins revealed by heterologous action spectroscopy. *Zoological Lett.* 2019:**5**(1):35. https://doi.org/10.1186/s40851-019-0150-2.

Sakai T, Tsutsui K, Yamashita T, Iwabe N, Takahashi K, Wada A, Shichida Y. *Drosophila melanogaster* rhodopsin Rh7 is a UV-to-visible light sensor with an extraordinarily broad absorption spectrum. *Sci Rep.* 2017:**7**(1):7349. https://doi.org/10.1038/s41598-017-07461-9.

Satoh A, Stewart FJ, Koshitaka H, Akashi HD, Pirih P, Sato Y, Arikawa K. Red-shift of spectral sensitivity due to screening pigment migration in the eyes of a moth, *Adoxophyes orana*. *Zoological Lett.* 2017:**3**(1):14. https://doi.org/10.1186/s40851-017-0075-6.

Schraermeyer U, Stieve H, Rack M. Cyclic 3',5'-nucleotide phosphodiesterase: cytochemical localization in photoreceptor cells of the fly *Calliphora erythrocephala*. *J Neurocytol.* 1993:**22**(10): 845–853. https://doi.org/10.1007/BF01186356.

Sharkey CR, Blanco J, Lord NP, Wardill TJ. Jewel beetle opsin duplication and divergence is the mechanism for diverse spectral sensitivities. *Mol Biol Evol.* 2023:**40**(2):msad023. https://doi.org/10.1093/molbev/msad023.

Sharkey CR, Fujimoto MS, Lord NP, Shin S, McKenna DD, Suvorov A, Martin GJ, Bybee SM. Overcoming the loss of blue sensitivity through opsin duplication in the largest animal group, beetles. *Sci Rep.* 2017:**7**(1):8. https://doi.org/10.1038/s41598-017-00061-7.

Shichida Y, Matsuyama T. Evolution of opsins and phototransduction. *Philos Trans R Soc Lond B Biol Sci.* 2009:**364**(1531): 2881–2895. https://doi.org/10.1098/rstb.2009.0051.

Sison-Mangus MP, Bernard GD, Lampel J, Briscoe AD. Beauty in the eye of the beholder: the two blue opsins of lycaenid butterflies and the opsin gene-driven evolution of sexually dimorphic eyes. *J Exp Biol.* 2006:**209**(16):3079–3090. https://doi.org/10.1242/jeb.02360.

Sison-Mangus MP, Briscoe AD, Zaccardi G, Knüttel H, Kelber A. The lycaenid butterfly *Polyommatus icarus* uses a duplicated blue opsin to see green. *J Exp Biol.* 2008:**211**(3):361–369. https://doi.org/10.1242/jeb.012617.

Slater GSC, Birney E. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* 2005:**6**(1):31. https://doi.org/10.1186/1471-2105-6-31.

Smith MD, Wertheim JO, Weaver S, Murrell B, Scheffler K, Kosakovsky Pond SL. Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Mol Biol Evol.* 2015:**32**(5):1342–1353. https://doi.org/10.1093/molbev/msv022.

Sondhi Y, Ellis EA, Bybee SM, Theobald JC, Kawahara AY. Light environment drives evolution of color vision genes in butterflies and moths. *Commun Biol.* 2021:**4**(1):177. https://doi.org/10.1038/s42003-021-01688-z.

Spaethe J, Briscoe AD. Early duplication and functional diversification of the opsin gene family in insects. *Mol Biol Evol.* 2004:**21**(8):1583–1594. https://doi.org/10.1093/molbev/msh162.

Stavenga DG, Arikawa K. Evolution of color and vision of butterflies. *Arthropod Struct Dev.* 2006:**35**(4):307–318. https://doi.org/10.1016/j.asd.2006.08.011.

Steenwyk JL, Buida TJ, Labella AL, Li Y, Shen X-X, Rokas A. PhyKIT: a broadly applicable UNIX shell toolkit for processing and analyzing phylogenomic data. *Bioinformatics* 2021:**37**(16):2325–2331. https://doi.org/10.1093/bioinformatics/btab096.

Suyama M, Torrents D, Bork P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* 2006:**34**(Web Server):W609–W612. https://doi.org/10.1093/nar/gkl315.

The Darwin Tree of Life Project Consortium. Sequence locally, think globally: the Darwin Tree of Life Project. *PNAS.* 2022:**119**(4):e2115642118. https://doi.org/10.1073/pnas.2115642118.

Thoen HH, How MJ, Chiou T-H, Marshall J. A different form of color vision in mantis shrimp. *Science* 2014:**343**(6169):411–413. https://doi.org/10.1126/science.1245824.

van der Kooi CJ, Stavenga DG, Arikawa K, Belušič G, Kelber A. Evolution of insect color vision: from spectral sensitivity to visual ecology. *Annu Rev Entomol.* 2021:**66**(1):435–461. https://doi.org/10.1146/annurev-ento-061720-071644.

Van Nynatten A, Castiglione GM, de A Gutierrez E, Lovejoy NR, Chang BSW. Recreated ancestral opsin associated with marine to freshwater croaker invasion reveals kinetic and spectral adaptation. *Mol Biol Evol.* 2021:**38**(5):2076–2087. https://doi.org/10.1093/molbev/msab008.

Varma N, Mutt E, Mühle J, Panneels V, Terakita A, Deupi X, Nogly P, Schertler GFX, Lesca E. Crystal structure of jumping spider rhodopsin-1 as a light sensitive GPCR. *Proc Natl Acad Sci U S A.* 2019:**116**(29):14547–14556. https://doi.org/10.1073/pnas.1902192116.

Wakakuwa M, Stavenga DG, Kurasawa M, Arikawa K. A unique visual pigment expressed in green, red and deep-red receptors in

the eye of the small white butterfly, *Pieris rapae crucivora*. *J Exp Biol*. 2004:**207**(16):2803–2810. https://doi.org/10.1242/jeb.01078.

Wakakuwa M, Terakita A, Koyanagi M, Stavenga DG, Shichida Y, Arikawa K. Evolution and mechanism of spectral tuning of blue-absorbing visual pigments in butterflies. *PLoS One* 2010:**5**(11): e15015. https://doi.org/10.1371/journal.pone.0015015.

Wei Z, Zhang M, Li C, Huang W, Fan Y, Guo J, Khater M, Fukuda M, Dong Z, Hu G, *et al.* Specific TBC domain-containing proteins control the ER-Golgi-plasma membrane trafficking of GPCRs. *Cell Rep*. 2019:**28**(2):554–566.e4. https://doi.org/10.1016/j.celrep.2019.05.033.

Wertheim JO, Murrell B, Smith MD, Kosakovsky Pond SL, Scheffler K. RELAX: detecting relaxed selection in a phylogenetic framework. *Mol Biol Evol*. 2015:**32**(3):820–832. https://doi.org/10.1093/molbev/msu400.

Wright CJ, Stevens L, Mackintosh A, Lawniczak M, Blaxter M. Chromosome evolution in Lepidoptera. *bioRxiv* [Internet]: 2023.05.12.540473. 2023. Available from: https://www.biorxiv.org/content/10.1101/2023.05.12.540473v1.

Xu P, Feuda R, Lu B, Xiao H, Graham RI, Wu K. Functional opsin retrogene in nocturnal moth. *Mob DNA*. 2016:**7**(1):18. https://doi.org/10.1186/s13100-016-0074-8.

Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 2007:**24**(8):1586–1591. https://doi.org/10.1093/molbev/msm088.

Yang Z, Nielsen R. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol Biol Evol*. 2002:**19**(6):908–917. https://doi.org/10.1093/oxfordjournals.molbev.a004148.

Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y. Ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol*. 2017:**8**(1):28–36. https://doi.org/10.1111/2041-210X.12628.